# Refocusing the lens through which we view affect dynamics: The Skills, Difficulty, Value, Efficacy and Time Model

Jaclyn Ocumpaugh
University of Pennsylvania
Philadelphia, PA, USA
jlocumpaugh@gmail.com

Nidhi Nasiar
University of Pennsylvania
Philadelphia, PA, USA
nidhinasiar@gmail.com

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

Alex Goslen
North Carolina State University
Philadelphia, PA, USA
amgoslen@ncsu.edu

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

Jordan Esiason
North Carolina State University
Raleigh, NC, USA
jesiaso@ncsu.edu

Jonathan Rowe
North Carolina State University
Raleigh, NC, USA
jprowe@ncsu.edu

Stephen Hutt
University of Denver
Denver, CO, USA
stephen.hutt@du.edu

## Abstract

For more than a decade, a handful of theoretical models have shaped a substantial amount of the research related to students' emotional experiences during learning. This research has been productive, but articulating the underlying implicit assumptions in existing theories and their implications in our empirical interpretations can help to better investigate the reciprocal relationships between learning and emotion, and subsequently, to develop better interventions. This paper expands upon the existing theoretical frameworks, increasing the types of questions we ask about affect dynamics. We do so within the context of Crystal Island, a virtual world that allows middle school students to investigate microbiology questions. Specifically, we use this data to examine and revise the assumptions that are implicit in these models and the methods we use to investigate them.

## CCS Concepts

• **Human-centered computing**;; • **Social and professional topics**;; • **Applied computing**; • **Education; Interactive learning environments**;;

## Keywords

Affect Dynamics, Control Value Theory, Self-Efficacy, Game-based learning

**ACM Reference Format:**
Jaclyn Ocumpaugh, Nidhi Nasiar, Andres Felipe Zambrano, Alex Goslen, Jessica Vandenberg, Jordan Esiason, Jonathan Rowe, and Stephen Hutt. 2025. Refocusing the lens through which we view affect dynamics: The Skills, Difficulty, Value, Efficacy and Time Model. In *LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 03–07, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3706468.3706495



*LAK 2025, March 03–07, 2025, Dublin, Ireland*

ACM ISBN 979-8-4007-0701-8/25/03
https://doi.org/10.1145/3706468.3706495

## 1 Introduction

Research in learning analytics has been influenced by several theories that seek to explain how students' emotional experiences interact with their learning processes, including those related to flow [12] and metacognition [14]. For more than a decade, the field has been strongly influenced by theoretical models that build on Pekrun's Control Value Theory (CVT; [31]), with many papers examining the empirical evidence for D'Mello & Graesser's [13] theory of affect dynamics. As we discuss later, this theoretical model has significantly advanced the research on affect dynamics, but empirical evidence for this remains limited [21]), highlighting a need for the existing theory to adapt.

In part, the contradictory results may stem from aggregating specific constructs across the entire dataset, without considering potential variations across time. None of the studies in Karumbaiah et al.'s [22] review of affect dynamics partitioned student affective experiences across minutes, hours, days, weeks, or months of use. However, one could imagine that some students are more tolerant of challenges early in the learning session than after hours of struggling, or that some manage task difficulty better at the start of the school year than in the week before a major holiday.

Other contradictions may result from researchers aggregating data across student-level traits. For example, 29 of 39 studies examined in Karumbaiah et al. [22] only investigated affect dynamics in terms of learning gains (i.e., with pre and posttests of knowledge) without considering any trait-level factors. In contrast, other studies show these factors have influence both learning outcomes and affect dynamics. For example, Andres et al. [1] show that trait-level anxiety correlates with higher frustration, but only during the second half of each class period. Moreover, CVT hypothesizes that students' outcome expectations and task values may influence outcomes [31], suggesting that constructs like self-efficacy, prior knowledge (or skill level), intrinsic interest, and task utility could help us to better understand the relationship between learning and affect. Self-efficacy has been demonstrated to mediate affective states in learners [7; 30]. Specifically, students with higher self-efficacy are more likely to experience frustration when faced with obstacles, whereas those with moderate self-efficacy may first encounter confusion [9]. This differentiation shows how emotional responses may vary when engaging with material that challenges their current skill or knowledge levels [25].

Aggregating across student-level characteristics may also explain other contradictions in the literature, where there are sometimes inconsistent suggestions about how to label these experiences. For example, Baker et al., [2] have argued that we should collapse confusion and frustration into a single construct they call *confrustion.* Both appear to signal that an impasse has been detected [13], which likely explains why moderate levels of each are necessary for learning [26]. However, other research suggests that each may instead be multiple experiences—or at least they are experiences that may be interpreted differently. For example, Gee [18] argues for the construct of pleasurable frustration—an energizing experience that is distinct from frustration's typical manifestations—and Cloude et al. [11] find evidence that there may be multiple types of confusion and multiple types of frustration. Additionally, evidence suggests that frustration may be experienced and interpreted differently by students with high anxiety (e.g., [1]). It seems likely that the first group of researchers is picking up on the necessary conditions for learning (introducing material a student has not mastered), but that motivational constructs like self-efficacy might still shape the ways students respond to being challenged.

This study seeks to re-examine our theoretical models by reviewing the learning analytics literature on student affect, including assumptions from both the theoretical literature [12; 13; 31] and implicit techniques we use for empirical investigation (e.g., [13; 20]. We examine these assumptions using data from a classroom study of Crystal Island, a game-based learning system for middle school science. Specifically, we employ an epistemic network analysis (ENA) of BROMP-based [29] automated detectors of student affect. We follow past work of affect dynamics using ENA [20; 28] in terms of the ENA window size, but we also label larger temporal segments (20-minute intervals) to explore how students' affective cycles differ from the start of a two-day implementation to the end. Our results demonstrate significant changes in affective dynamics over time and suggest that trait-level student characteristics, including self-efficacy, may also influence these patterns. We use these results to motivate a new model for understanding students' affective experiences.

## 2 Prior Theoretical Models

### 2.1 Csikszentmihalyi's Flow Theory

One of the earliest models of emotion to influence learning research was Csikszentmihalyi's [12] Flow Theory, which describes the positive psychological state of flow, which occurs when a person's skill is aligned with the task's difficulty. Flow is associated with strong feelings of autonomy and a level of focus that becomes so intense that people no longer notice things like the passage of time [12]. It is sometimes described as the space between where a task is too difficult (where anxiety occurs) and the space where it is too easy (where boredom occurs). Flow theory conceptualizes this as a two-dimensional space related to person-external challenges (e.g. problem-level difficulty) and person-internal skills.

### 2.2 Pekrun's Control Value Theory

Like Csikszentmihalyi's [12] Flow Theory, Pekrun's [31] Control Value Theory (CVT) of Achievement Emotions emphasizes feelings of autonomy (e.g., control appraisals). Plausibly, this incorporates the two dimensions identified by Csikszentmihalyi; that is, feelings of autonomy are likely related to the matching of internal skills with external task requirements, but the two theories diverge elsewhere, with Pekrun et al. [32] providing evidence that boredom often occurs among low-achieving students—as a result of being over-challenged—which is not accounted for in Flow Theory.

Importantly, CVT includes another dimension found in many motivational theories, value [31], helps to determine whether the flow experience occurs. In Pekrun's [31] formulation of activity emotions, he suggests that flow (or at least the enjoyable part of it) is only experienced if the student also values the task; when students do not value the task, they are predicted to be angry. In this discussion, value also helps to characterize why students experience frustration and boredom. Students who assign low value to the task are predicted to be bored, regardless of whether or not their skill level is above or below the level required for the task. However, boredom and frustration sometimes overlap in this model, as students are predicted to experience frustration when their skill level is below the level required of the task, either because they do not value the task or because they do. This overlap likely suggests one of two things. Either we are using the same label for more than one emotional experience, or there are additional dimensions that we should consider. Figure 1 illustrates this conceptualization in terms of Flow Theory, illustrating both types of boredom (from being either overtaxed or undertaxed) and the places where boredom and frustration are theorized to overlap.

Figure 1, which is distinct from Pekrun's usual formulation of CVT, shows the alignment of skill and difficulty, as a highlighted plane. In this plane, anger is likely to occur when value is low, while Pekrun's flow (an enjoyable experience) only occurs if value is high. This dimension occurs regardless of difficulty level—but only if difficulty is matched to the student's skill. Notably, this figure does not distinguish between extrinsic and intrinsic value, as is typical in Pekrun's theory [31], and includes only *activity* emotions. Pekrun [31] argues that these interact with *outcome* emotions, such as hopelessness or hopefulness, which are not shown in this interpretation, but which likely also influence feelings of autonomy or control and therefore the state-level emotions of boredom, frustration, anger, and flow. Confusion, which has not always been considered a standard emotion (see [34]), is not included here.

However, this figure does illustrate how we might interpret certain affect transitions. For example, it suggests that sustained concentration (often the operationalized measure of flow) is indicative not just of the match between skill and difficulty, but also of value. Meanwhile, frustration (the nearest category to anger) could occur even when skill and difficulty are matched. It also suggests that cycles between boredom and frustration are most likely when the difficulty exceeds the skill level, although we should probably

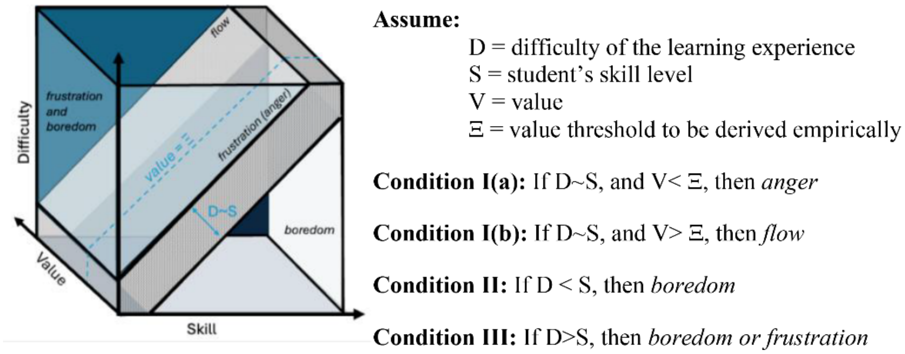



**Figure 1: Illustration of Pekrun's [31] critique of Flow Theory, which suggests value must also be high to achieve flow**

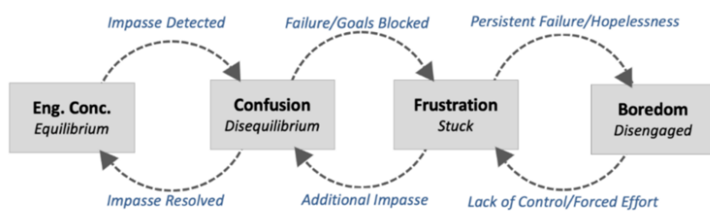



**Figure 2: Affective Dynamics Theory (D'Mello & Greasser [13])**

expect sustained frustration or sustained boredom in this space as well. Likewise, it suggests that we get boredom from under stimulation, which is likely uncontroversial but not well articulated in other models used in the learning analytics community (e.g., [13]).

## 2.3 D'Mello & Graesser's Affect Dynamics

D'Mello and Graesser's [13] Theory of Affect Dynamics (Figure 2) has been highly influential within the learning analytics community. Rather than considering the dimensionality of affective states, it considers the temporal shifts of students from one educationally relevant state to another. Influenced by CVT, this model suggests that learning occurs when an impasse is detected and resolved, but when left unresolved, students transition to frustration and boredom. The elegance of this model inspired many empirical studies (see review in [22]), but as discussed below (section 2.4), these have offered incomplete evidence for this conceptualization.

In part, this may be because the L-statistic suggested by [13] focuses the analyses on the time units being used to measure epistemic emotions. That is, if labels are applied at 20-second intervals, as is in many BROMP-based studies [6], then the L-statistic examines time scales that may represent only a minute (or less) of data. It seems likely that some students may take more than 60 seconds to transition from an impasse being detected to feelings of hopelessness.

However, another reason that empirical results do not reflect our theory is that the L-statistic deliberately flattens the data by calculating the likelihood of transitions according to the base rate of each affective state (though see Karumbaiah et al.'s [19] discussion on the assumptions involved in its initial instantiation). This normalization may amplify transitions that are actually artifacts of affective states with a small base rate, making rare events appear more frequent than they are. Although there are benefits to using L-statistic, it may obscure the true likelihood of transitions involving dominant affective states and frequent transitions. Combined with the abstractions involved in D'Mello & Graesser's [13] model, this technique may have obscured some of the nuance needed to effectively interpret this data, which likely requires examinations of the data at multiple levels of granularity.

## 2.4 Empirical Evidence on Affect and Cognition

*2.4.1 Affect Dynamics using Correlations or D'Mello's L..* Two primary methods have been used to investigate affect dynamics: correlations and D'Mello's L-statistic, which calculates the likelihood of a transition occurring relative to the base rate of the individual emotions [13]. Empirical research on affective experiences in learning has yielded important findings. Famously, we know that it is better to be frustrated than bored [3], as both frustration and confusion must be present for learning [26]. However, a recent reanalysis of data using these studies finds limited evidence for D'Mello and Graesser's [13] theoretical models [21].

The heatmap in Table 1 summarizes the number of studies where these transitions occured above chance, below chance, or at rates that were not significant. As it shows, theorized transitions are somewhat more likely to emerge than non-theorized transitions, but so were self-transitions, which are not explicitly included in theoretical model. Moreover, only one theorized transition is significantly above chance in more than half of the studies (engaged concentration to confusion, predicted to occur when a student detects an impasse). These results suggest that if the transitions hypothesized

**Table 1: Summary of Karumbaiah et al.'s [21] meta-analysis.**

| Framework | Transitions | Total Studies | Above Chance | Below Chance | Not Significant |
|---|---|---|---|---|---|
| D&G: Impasse Detected | Eng→Con | 18 | 9 | 3 | 6 |
| D&G: Impasse Resolved | Conf→Eng | 18 | 7 | 4 | 7 |
| D&G: Failure/Goals Blocked | Conf→Fru | 18 | 6 | 2 | 10 |
| D&G: Additional Impasse | Fru→Con | 16 | 3 | 0 | 13 |
| D&G: Persistent Failure/Hopelessness | Fru→Bor | 15 | 3 | 2 | 10 |
| D&G: Lack of Control/Forced Effort | Bor-Fru | 16 | 6 | 2 | 8 |
| Not theorized | Bor→Conf | 15 | 1 | 5 | 9 |
| Not theorized | Bor→Eng | 16 | 2 | 4 | 10 |
| Not theorized | Conf→Bor | 15 | 0 | 4 | 11 |
| Not theorized | Eng→Bor | 15 | 2 | 6 | 7 |
| Not theorized | Eng→Fru | 15 | 1 | 3 | 11 |
| Not theorized | Fru→Eng | 14 | 2 | 4 | 8 |
| Self-Transition | Bor→Bor | 13 | 10 | 1 | 2 |
| Self-Transition | Conf→Conf | 13 | 6 | 1 | 6 |
| Self-Transition | Eng→Eng | 13 | 8 | 1 | 4 |
| Self-Transition | Fru→Fru | 12 | 4 | 1 | 7 |

[a] Dark gray highlights findings at or above 50% of total studies, light gray highlights findings at or above 25%.

in D'Mello & Graesser's [13] model are the most important to understanding learning, their frequency is quite low when measured by the L-statistic. However, these results are calculated for entire data sets without respect to student-level characteristics like skill or task value and without respect to problem difficulty.

*2.4.2 Affect Dynamics Research Using Epistemic Network Analysis.* More recently, some researchers have shifted from investigating affect dynamics with D'Mello & Graesser's [12] L-statistic to using epistemic network analysis (ENA) and its related method, ordered network analysis (ONA; [36]). Unlike the L-statistic, which normalizes data to account for the base rate of each affective state, ENA applies cosine normalization to each student's data, considering the total count of affect transitions [35] and self-transitions (in the case of ONA). This approach highlights transitions between affective states that are more prevalent in students' affect dynamics, rather than underscoring transitions that might be more sporadic but appear significant due to base rate normalization. Accordingly, when Karumbaiah and Baker [20] compared the two methods, they found that the L-statistic highlighted a strong connection for only one of the theorized pathways (CONF→ENG), while their adapted ENA (which labeled receiving states differently to analyze ordering effects) showed strong connections for two pathways (impasse detected and impasse resolved). The analysis also showed that non-theorized pathways were common: BOR→ENG, ENG→BOR, and ENG→FRUS were identified by both techniques, and ENG→FRUS was identified by the ENA technique.

Nasiar et al. [28] took a related approach, using ONA to investigate two sets of detectors that were applied to a science-based learning game: one trained on self-report data and one trained on BROMP-based observation data. This research showed that self-transitions (not considered in [20]) were most common and heavily influenced by the base rate of each affective state. However, the strength of each connection also differed depending on the learning gains of the student, and these findings were strongest among the self-report trained detectors, where self-transitions for boredom, confusion, and frustration were more common among students with low learning gains, while self-transitions for engaged concentration and delight were more common among students with high learning gains.

## 3 METHODS

This study proposes an affect dynamics model that increases both the temporal dimensions considered in analyses as well as the number of learning and motivation constructs affecting these patterns.

### 3.1 Study Context & Learners

This study investigates the dimensionality of student affect within a science learning system aligned to the U.S. state standards for 8th grade microbiology. Specifically, it uses data from an open-world, game-based learning environment, called Crystal Island, where students assume the role of a scientist asked to diagnose an illness that has spread on a remote island. They must identify the associated pathogen of this disease and its transmission, which they investigate by exploring the island, interacting with non-player characters, reading educational materials, and conducting virtual tests for viruses and bacteria. Students are provided with virtual materials to track their hypotheses and virtual test results, allowing them to organize their findings and conclusions.

This study examines data from 120 middle schoolers who interacted with Crystal Island during two days of their regular science class at a school in the southeastern U.S. Students were well-balanced for gender (53 male, 66 female, and 5 who preferred to describe themselves) and included a strong representation from historically marginalized communities (46% Black, 16% Hispanic, 5% Asian, 5% Multiracial, 1% Native American, and only 27% White). Before playing the game, students completed a series of pre-surveys:

a demographic questionnaire, a science content pre-test, a situational interest scale [24], and a self-efficacy scale [8]. Afterwards, they completed post-surveys (i.e., a science content post-test and several more motivational measures). While interacting with the game, students' affective states were labeled with BROMP observations, which were used to develop the cross-validated detectors used in this analysis.

### 3.2 Automated Detectors of Student Affect

This study uses previously published [37], student-level cross-validated, interaction-based ML detectors of epistemic affective states to explore how time in game and student self-efficacy impact affect dynamics. Interaction-based detectors model what students do within a system when they are experiencing an epistemic emotion, allowing the system to infer these emotions in real-time (see review of this research across a wide range of learning systems [5]). Specifically, Zambrano et al. [37] developed detectors to predict BROMP-based observations of five epistemic emotions (i.e., boredom, confusion, frustration, engaged concentration, and delight) that are commonly included in affect research [29]. These detectors employed features such as the duration of conversations with NPCs, the time inside specific locations during the last minute, or the number of readings [37]. In keeping with previous standards [5], and to maintain consistency with the granularity of previous studies [6], these interaction-based detectors were applied using 20-second clips of students' log files. To control for any biases that might emerge during the detector-building process (i.e., resampling of less-common classes), detector output was normalized by mathematically adjusting each detector's output to align the distribution of predicted affective states with the proportions of each affective state found in the training data.

### 3.3 Ordered Network Analysis (ONA)

This study uses Ordered Network Analysis, a type of Epistemic Network Analysis [35], which, unlike traditional ENA, considers the temporal order of associations between constructs—here, affective states—while also accounting for self-transitions [36]. To analyze the evolution of students' affect dynamics over time, we segmented the prediction of their affective states throughout the gameplay session into 20-minute intervals. For each interval, we generated an ONA model to capture the strength of the ordered associations between pairs of affective states. As in previous research [20; 28], our ONA models used a moving stanza window of size 2, analyzing only two consecutive affective states at a time. The resulting model is visualized as a directed network graph, where nodes represent affective states and edges indicate the strength of each transition.

Furthermore, we use ONA to examine the role of self-efficacy in the evolution of affect dynamics. Although self-efficacy has not been explicitly incorporated into the previously described theoretical models, research indicates that students with varying levels of self-efficacy exhibit different affect dynamics when facing obstacles [9, 25], suggesting a potential interplay between self-efficacy and students' emotional responses. To explore these differences, we categorized students based on their responses to the self-efficacy scale [8]. Students whose scores fell within one standard deviation of the mean (avg=3.78, SD=1.05) were excluded from the analysis. We compared students with the lowest self-efficacy (SE<2.73, N=18) and highest self-efficacy (SE>4.83, N=29). For each group, we created individual ONA models and then subtracted the connection weights to obtain a difference model that contrasts the two groups.

## 4 Results

This section analyzes student epistemic emotion across time, examining both attrition and motivation.

### 4.1 Descriptive Results

Preliminary analysis examined 10-minute intervals across the two-day implementation. Table 2 shows the two most prevalent emotions during the first 10 minutes—engaged concentration and confusion. As these results show, engaged concentration rates were high (89-100%) during the first ten minutes and were initially supplemented primarily by confusion. In fact, only the group at 100% engaged concentration did not experience confusion in the first ten minutes. Otherwise, the rates of confusion ranged from 4-9% during that time interval. However, as we look at later time intervals, the students' emotions clearly become more diverse. Between 30 and 40 minutes into the game, engaged concentration drops to a maximum of 86% of student data (for those who persisted more than 60 minutes), and confusion is no longer the only alternative. For example, if we look at the data for minutes 40 and 50 minutes into the game, the group of students who are about to drop out are experiencing 22% confusion (the highest rate of confusion for any time interval/student attrition group), but only 47% engaged concentration.

Table 3 shows how these emotions diversify over time. Interestingly, these data show that students who persist longer are more likely to show emotions that might be classified as negatively valanced, like boredom and frustration (though see Gee's [18] work on pleasurable frustration). Rates for frustration are relatively high (11-28%) later in the game, compared to earlier in the game, with a strong divergence between students who dropped after 60 minutes and those who persisted longer. The first group shows increasing levels of boredom, reaching almost a third of their affective experiences (29%) in their final 10 minutes of gameplay, while the second group shows no boredom. Instead, those who continued to persist after 60 minutes were more likely to see increases in delight.

### 4.2 ONA Results

*4.2.1 ONA by Time.* The ONA model, segmented by units of time (20-minute intervals), shows a strong separation between the transitions that are most common earlier versus later in the two-day implementation. In ONA, the distance between the position of the nodes and unit means can be interpreted as a measure of similarity between them [35]. Differences between each 20-minute intervals can be seen in the means squares units in the center of Figure 3, which models affect transitions of all four intervals. The mean for the first twenty minutes (a red square) almost overlaps with concentration, but the means shift rightward as concentration rates drop. Although the mean for the second twenty-minute interval (blue square) is immediately to the right of the first mean, the means shift more dramatically over time, with the third interval's mean

**Table 2: Rates of Engaged Concentration & Confusion Across Time, Sorted by Student Attrition**

| | | Attrition after 10 min | Attrition after 20 min | Attrition after 30 min | Attrition after 40 min | Attrition after 50 min | Attrition after 60 min | Persisted | Total |
|---|---|---|---|---|---|---|---|---|---|
| Eng | 10min | 100% | 96% | 94% | 91% | 89% | 95% | 95% | 93% |
| | 20min | | 89% | 91% | 85% | 93% | 85% | 98% | 91% |
| | 30min | | | 84% | 88% | 80% | 81% | 95% | 86% |
| | 40min | | | | 85% | 68% | 54% | 86% | 72% |
| | 50min | | | | | 47% | 38% | 74% | 58% |
| | 60min | | | | | | 25% | 64% | 57% |
| | 70min | | | | | | | 43% | 43% |
| Conf | 10min | 0% | 4% | 6% | 9% | 9% | 5% | 5% | 6% |
| | 20min | | 11% | 5% | 11% | 2% | 6% | 2% | 5% |
| | 30min | | | 8% | 7% | 10% | 3% | 4% | 6% |
| | 40min | | | | 6% | 14% | 8% | 3% | 8% |
| | 50min | | | | | 22% | 9% | 5% | 8% |
| | 60min | | | | | | 11% | 6% | 7% |
| | 70min | | | | | | | 10% | 10% |

**Table 3: Rates of Frustration, Boredom, and Delight Across Time, Sorted by Student Attrition**

| | | Attrition after 10 min | Attrition after 20 min | Attrition after 30 min | Attrition after 40 min | Attrition after 50 min | Attrition after 60 min | Persisted | Total |
|---|---|---|---|---|---|---|---|---|---|
| Frus | 10min | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| | 20min | | 0% | 1% | 1% | 1% | 0% | 0% | 0% |
| | 30min | | | 1% | 5% | 6% | 4% | 1% | 4% |
| | 40min | | | | 8% | 9% | 18% | 6% | 11% |
| | 50min | | | | | 11% | 28% | 13% | 19% |
| | 60min | | | | | | 26% | 16% | 18% |
| | 70min | | | | | | | 26% | 26% |
| Bored | 10min | 0% | 0% | 0% | 0% | 2% | 0% | 3% | 0% |
| | 20min | | 0% | 3% | 3% | 4% | 8% | 0% | 3% |
| | 30min | | | 6% | 0% | 3% | 13% | 0% | 4% |
| | 40min | | | | 1% | 6% | 20% | 0% | 8% |
| | 50min | | | | | 9% | 24% | 0% | 12% |
| | 60min | | | | | | 29% | 0% | 10% |
| | 70min | | | | | | | 0% | 2% |
| Delight | 10min | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| | 20min | | 0% | 1% | 0% | 0% | 0% | 0% | 0% |
| | 30min | | | 0% | 0% | 0% | 0% | 0% | 0% |
| | 40min | | | | 0% | 4% | 1% | 3% | 2% |
| | 50min | | | | | 11% | 1% | 4% | 3% |
| | 60min | | | | | | 9% | 8% | 8% |
| | 70min | | | | | | | 19% | 19% |

halfway to delight (purple square) and the fourth interval's mean (green square) right of that.

Other details of this progression can be seen more clearly when data is divided by time interval (Figure 4, Table 4). In the initial interval (Fig. 4a), students primarily experience sustained concentration, with small numbers of transitions to and from confusion, frustration, and boredom, as well as a few between confusion and frustration. In the second interval (Fig. 4b), concentration continues to be the primary experience, but its reduction makes way for several of the transitions to intensify (e.g., ENG→FRUS, FRUS→ENG, ENG→BOR) and for new links to emerge (i.e., BOR→ENG, ENG→DEL, DEL→ENG). In the third interval (Figure 4c), these transitions continue to intensify as the transitions involving delight become more diverse and the self-transitions of boredom and confusion are also enhanced. After

**Table 4: Line Weights of ONA models for each 20-min interval, with the highest shown in bold. Grayscale shows differences between the 1st and 3rd intervals that remain statistically significant after a Benjamini-Hochberg correction to a Mann-Whitney U test.**

| Transition | Interval 1 (N=120) | Interval 2 (N=111) | Interval 3 (N=62) | Interval 4 (N=23) | Diff. 3rd -1st |
|---|---|---|---|---|---|
| ENG→ENG | **0.971** | 0.915 | 0.684 | 0.556 | -0.287 |
| CONF→CONF | 0.061 | 0.067 | **0.091** | 0.076 | 0.030 |
| FRUS→FRUS | 0.005 | 0.047 | 0.204 | **0.213** | 0.199 |
| BOR→BOR | 0.028 | 0.062 | **0.136** | 0.051 | 0.108 |
| DEL→DEL | - | 0.008 | 0.046 | **0.222** | 0.046 |
| ENG→ CONF | 0.018 | 0.014 | **0.021** | 0.004 | 0.003 |
| CONF→ENG | 0.017 | 0.014 | 0.037 | **0.053** | 0.020 |
| ENG→ FRUS | 0.001 | 0.012 | 0.016 | **0.015** | 0.015 |
| FRUS → ENG | - | 0.008 | 0.019 | **0.020** | 0.019 |
| ENG →DEL | - | 0.003 | 0.014 | **0.039** | 0.014 |
| DEL→ ENG | - | 0.002 | **0.016** | 0.008 | 0.016 |
| CONF → FRUS | 0.001 | 0.007 | **0.009** | 0.004 | 0.008 |
| FRUS → CONF | - | 0.001 | **0.009** | 0.004 | 0.009 |
| ENG→BOR | 0.001 | 0.004 | **0.006** | - | 0.005 |
| BOR→ENG | - | 0.003 | **0.003** | - | 0.003 |
| BOR → FRUS | - | 0.001 | 0.002 | **0.018** | 0.002 |
| FRUS→BOR | - | 0.001 | **0.002** | **-** | 0.002 |
| BOR→CONF | - | 0.002 | **0.004** | 0.003 | 0.004 |
| CONF→BOR | - | 0.001 | 0.003 | **0.003** | 0.003 |
| DEL→CONF | - | 0.001 | 0.001 | **0.012** | 0.001 |
| DEL→FRUS | - | - | 0.001 | **0.002** | 0.001 |
| FRUS→DEL | - | - | 0.001 | **0.002** | 0.001 |
| DEL→BOR | - | - | **0.001** | **-** | 0.001 |
| BRO→DEL | - | - | **0.001** | - | 0.001 |

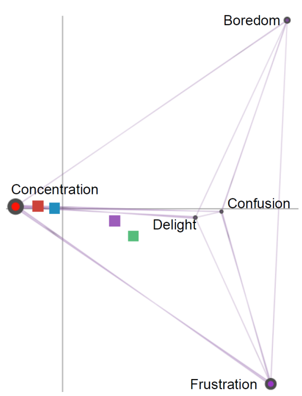
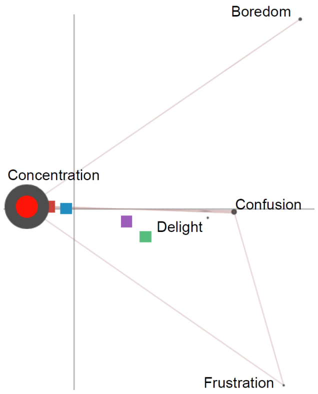



**Figure 3: ONA Across All Units of Time. Squares show the differences in unit means of each 20-minute interval: Red 0-20 min, N=120; Blue 20-40 min; Purple 40-60 min, N=62; and Green +60 min).**

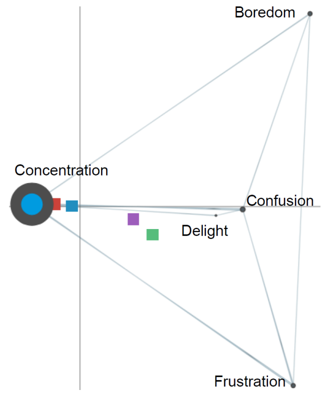
an hour of game play (Fig. 4d), concentration levels reduce further as transitions involving frustration and delight become more common. Notably, two transitions disappear entirely (BOR→ENG, BOR→DEL). Meanwhile, there is a non-parallel change in transitions between boredom and frustration; BOR→FRUS increases, but FRUS→BOR stops altogether.

*4.2.2 ENA of Time by Self Efficacy.* Further exploration shows that although students who differ in self-efficacy start with relatively equal levels of confusion and concentration, their experiences diverge across time. In the first 20-minute interval (Fig. 5a), only students with low self-efficacy transition from ENG→BOR and their rates for ENG→CONF and CONF→ENG are nearly twice those with high self-efficacy. In contrast, only students with high self-efficacy transition from ENG→FRUS. In the third interval, while there are still more than half a dozen students with low self-efficacy working on the game, the differences are even more extreme. As Figure 5b shows, sustained concentration is being experienced primarily by the high self-efficacy students, while the low self-efficacy students are more likely to experience sustained boredom and confusion. Likewise, the high self-efficacy students are more likely to experience delight or frustration, and the transitions between these two emotions and concentration are substantially more common for these learners. Line weights for the models are given in Table 5.

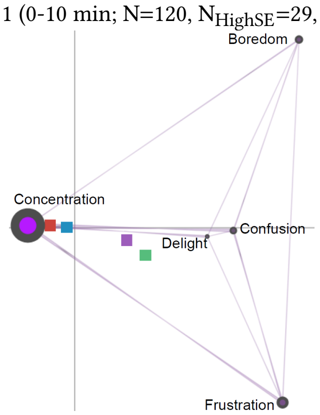
Interestingly, however, the transitions between concentration and confusion are not parallel; ENG→CONF is more common among high self-efficacy students, but this is heavily obscured by the much larger number of transitions in the opposite direction (CONF→ENG) among low self-efficacy learners, who started experiencing confusion earlier (first 40 minutes) and also keep experiencing confusion (self-transitions) at a higher rate than their high-self-efficacy peers. Moreover, the high self-efficacy learners



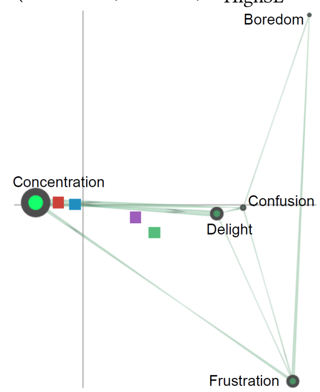

4a. Interval 1 (0-10 min; N=120, $N_{HighSE}$=29, $N_{LowSE}$=18)




4b. Interval 2 (20-40min; N =111, $N_{HighSE}$=26, $N_{LowSE}$=16)

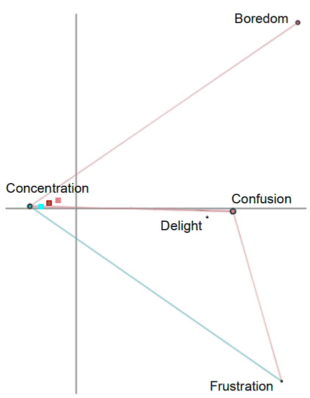



4c. Interval 3 (40-60min; N= 62, $N_{HighSE}$=21, $N_{LowSE}$=9)

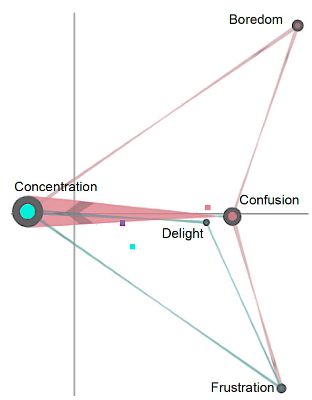



4d. Interval 4 (60+min; N= 23 Students, $N_{HighSE}$=3, $N_{LowSE}$=1)

**Figure 4: ENA Differences Across Time**




5a. Minutes 0-20 ($N_{HighSE}$=29, $N_{LowSE}$=18)




5b. Minutes 40-60 ($N_{HighSE}$=21, $N_{LowSE}$=9)

**Figure 5: Difference Model for High and Low Self Efficacy. Pink = Low Self Efficacy, Blue = High Self Efficacy**

are the only ones who shift back and forth between concentration and frustration (ENG→FRUS and FRUS→ENG), a result that contrasts with the low-self-efficacy learners' tendency to instead transition between frustration and confusion (FRUS→CONF and CONF→FRUS), and between confusion and boredom (CONF→BOR and BOR→CONF). Additionally, we see that the means shift across the two time periods, with the mean for high self-efficacy learners (blue square) shifting from a near overlap with concentration in the first twenty minutes to a position that is halfway to delight in minutes 40-60. The shift in the low self-efficacy learner's mean (pink triangle) is even more extreme, moving from a position near concentration to one adjacent to confusion. These changes indicate divergence between the two groups.

**Table 5: Line Weights of ONA models for the 1st and 3rd 20 min interval comparing high and low self-efficacy groups. Only connections with a weight greater than 0.01 for at least one self-efficacy group are shown in the table. Bold shows |difference| >.05**

| Transition | 0-20 Min (N=47) | | | 40-60 Min (N=30) | | |
|---|---|---|---|---|---|---|
| | High SE (N=29) | Low SE (N=18) | Diff. | High SE (N=21) | Low SE (N=9) | Diff. |
| ENG→ENG | 0.998 | 0.939 | **0.059** | 0.633 | 0.289 | **0.344** |
| CONF→CONF | 0.026 | 0.103 | **-0.077** | 0.098 | 0.381 | **-0.283** |
| FRUS→FRUS | 0.002 | 0.002 | 0.000 | 0.289 | 0.173 | **0.116** |
| BOR→BOR | - | 0.050 | **-0.050** | 0.058 | 0.219 | **-0.161** |
| DEL→DEL | - | - | - | 0.069 | - | **0.069** |
| ENG→CONF | 0.011 | 0.020 | -0.009 | 0.018 | - | 0.018 |
| CONF→ENG | 0.011 | 0.018 | -0.007 | 0.006 | 0.215 | **-0.209** |
| ENG→FRUS | 0.001 | - | 0.001 | 0.010 | - | 0.010 |
| FRUS→ENG | - | - | - | 0.011 | - | 0.011 |
| CONF→FRUS | - | 0.002 | -0.002 | 0.011 | 0.035 | -0.024 |
| FRUS→CONF | - | 0.002 | -0.002 | 0.012 | 0.035 | -0.023 |
| ENG→BOR | - | 0.001 | -0.001 | - | 0.018 | -0.018 |
| BOR→ENG | - | - | - | - | 0.018 | -0.018 |
| BOR→FRUS | - | 0.001 | -0.001 | - | - | - |
| ENG→DEL | - | - | - | 0.012 | - | 0.012 |
| DEL→ENG | - | - | - | 0.026 | - | 0.026 |
| CONF→BOR | - | - | - | - | 0.018 | -0.018 |
| BOR→CONF | - | - | - | - | 0.015 | -0.015 |
| DEL→FRU | - | - | - | 0.005 | - | 0.005 |
| FRU→DEL | - | - | - | 0.003 | - | 0.003 |

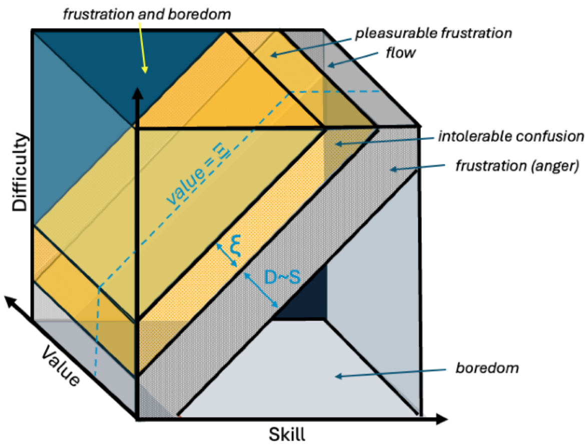
## 5 Discussion and Conclusions

Research on affective learning has a strong theoretical foundation, but thus far it has been difficult to demonstrate that these theories are able to capture the empirical evidence. Based on our findings from this study, we propose a new theoretical model which builds on Pekrun's interpretation of Flow Theory [31], but adds further dimensionality related to self-efficacy and time in game. We also propose two reasons that our methods may have obscured our empirical evidence: (1) because our methods have obfuscated important sequential information that was operating on a different time scale than our analysis and (2) because we were not fully exploring the motivational dimensions of our data.

### 5.1 A new model for adding dimensionalities to affective research

Both previous research and the data from the current study suggest that our current modes of affect dynamics are not fully capturing student experiences. Pekrun's [31] work suggests that value is a strong candidate for adding dimensionality, and we propose that self-efficacy (e.g., [8]) should also be considered. Specifically, adding self-efficacy to the model could explain some of the variability seen in empirical data, since it may clarify why some students immediately experience frustration when an obstacle is blocked while others first experience confusion [9]. In turn, this might also explain why limited amounts of confusion and frustration appear to be better for learning [25], as their occurrence indicates that a student has engaged with material that is above their current skill/knowledge level.

Appraisals, such as those related to self-efficacy, have often been a part of theories of emotion [10; 15]. Ellsworth [15] suggests that appraisal are neural and physiological changes that occur when novelty in a person's stimulus is detected, and that they form the basic ingredients of emotions. These basic appraisals include novelty, valance, goals/needs, agency, norms/values [16]. Notably, self-efficacy appears to be critical to at least one of these basic appraisals: agency. Figure 6 introduces a model that seeks to account for the effect of self-efficacy on the affective space. This model, which we call the *Skill, Difficulty, Value, Efficacy & Time* (SDVET) model, builds upon Pekrun's [31] discussion of flow theory, which defines flow as occurring when skill and difficulty are equally matched and value is high. However, our new model adds an additional plane to account for students' perceived self-efficacy. Our model hypothesizes that confusion—a cognitive experience—is most likely to be experienced when students are working on material that is between the two planes, but that students may experience that confusion differently, depending on whether they value the activity. Students who highly value the learning material are more likely to experience pleasurable frustration (e.g., [18]) when confusion is triggered if they believe they are capable of succeeding (self-efficacy). This may account for emerging evidence that students with high self-efficacy do not emote confusion as robustly as students with low self-efficacy [37]. Meanwhile, those with low value for the learning material are more likely to experience intolerable confusion, which




Assume:
D = difficulty of the learning experience
S = student's skill level
V = value
Ξ = value threshold to be derived empirically
ξ = a coefficient that increases with self-efficacy and decreases with the time spent in the learning experience

Condition (1): If D∼S,
and V< Ξ, then *frustration*
and V> Ξ, then *flow*

Condition (2): If D < S, then *boredom*

Condition (3): If D>S and S<D+ξ
and V< Ξ, then *intolerable confusion*
and V∼Ξ, then *neutral confusion*
and V> Ξ , then *pleasurable frustration*

Condition (4): If S >D+ξ ,
then unpleasant frustration & boredom

**Figure 6: The Skill, Difficulty, Value, Efficacy, & Time (SDVET) Model**

is presumed to trigger canonical, negative experiences frustration. This model should also help to account for interpretations that confusion and frustration as similar and/or overlapping experiences [2, 26].

This model also hypothesizes effects are more likely to occur over longer stretches of time. Specifically, it hypothesizes strong negative feedback over extended periods of time could diminish self-efficacy (or the tolerances for confusion afforded by self-efficacy), lowering the second plane. That is, a student's tolerance for confusion might wane even among those with relatively high self-efficacy, depending on the feedback they are getting from the activity. Likewise, if the learning activity were not rewarding enough, student value might decrease. For example, consider a student with high self-efficacy who highly values the mathematics task they are being asked to complete. If they were repeatedly told that they were wrong, their self-efficacy and/or their situational interest in the task might decrease, causing them to be more likely to experience negative emotions like (canonical) frustration or anger. However, if the same student was making steady progress with occasional obstacles (feedback that intermittent answers were incorrect), they might be more likely to experience pleasurable frustration—and even delight—as they sought to disentangle any confusion they still had about the subject. We might expect similar patterns for students working in learning systems that require sustained engagement with a larger, more complex problem, as seen in online games like Crystal Island. Although the feedback patterns would have different cadences, we should still expect (1) the rates of negatively valanced emotions should increase over time, and (2) these rates will vary based on value and self-efficacy.

As our model changes, so should the statistical assumptions that underlie our analyses. In this model, we hypothesize that analyses of affect dynamics will require (a) the segmentation of single "learning experiences" into appropriate units of time for analysis and (b) the division of learners into groups that reflect differences based on their prior knowledge (to account for the skill/difficulty), their value of the task, and their self-efficacy. We begin this exploration below.

## 5.2 Current Findings

This study has used ONA to examine how students transition from one affective state to another, exploring the degree to which the likelihood of each transition may shift over larger time intervals. It has also investigated whether self-efficacy differences might help to account for the variation that we are seeing in sequential data, even within these larger units of time. By conducting an ONA analysis, which does normalize the data according to the base rates of individual affective states, we are able to see patterns in larger units of time. Specifically, we see that all students begin the game experiencing substantial amounts of engaged concentration, punctuated primarily by instances of confusion. However, as time goes on, the types of affective states that students experience become more variable, with more experiences involving boredom, delight, and frustration.

This sort of information may have been obscured by prior analyses, both because this type of segmentation was not used and because normalizing affective transitions according to the base rate of each affective category may have suppressed transitions that show up within ONA. For example, transitions like ENG⟶BOR or FRUS⟶BOR might have been ignored with L-statistic measures because these transitions appeared less frequently than ENG⟶CONF throughout the learning experience. However, this does not mean that transitions like ENG⟶BOR or FRUS⟶BOR do not occur or are unimportant to understanding student experiences during specific time intervals in a learning experience, as we observed for later moments in the gameplay. This is not to say that ONA is the only method we should be using to study affective states. Instead, we argue that it should be considered alongside other kinds of statistical analyses.

In addition, we have shown that students with higher self-efficacy have more positive affective experiences than those with lower self-efficacy during later stages of the game. These results support the two main hypotheses introduced by our theoretical model. First, time impacts on students' affective experiences. Although most of students began with high levels of engaged concentration, as the game progressed and they faced various challenges (e.g., incorrect hypotheses, the need for collecting additional information, etc.), we observed more frequent instances of confusion, frustration, and boredom. Second, and more importantly, this time effect is closely related to self-efficacy. Even though the initial high level of engaged concentration did not differ between the two self-efficacy groups, in later stages of the game, their experiences diverged. Students with high self-efficacy experienced higher rates of frustration but were more likely to transition between frustration and affective states with positive valance (i.e., engaged concentration and delight). Meanwhile, students with low self-efficacy experienced saw increases in sustained frustration over time but were more likely to transition between frustration and active states with negative valance (i.e., confusion and boredom). This suggests that even though both groups of students were experiencing impasses, the frustration experiences of those with higher self-efficacy may be qualitatively different than the frustration experiences of those with lower self-efficacy. We, therefore, speculate that at least some of the cases of frustration in students could be characterized as Gee's pleasurable frustration, even though the detectors being used in this study are unable to make this differentiation [18].

Note that this interpretation is compatible with the idea that while both frustration and confusion signal a necessary condition for learning, our findings suggest they are not best described as a single term (e.g., "confrustion," [2]). In fact, we would argue that while confusion and frustration both increase over time, they may be qualitatively different experiences for students with high and low self-efficacy, an interpretation that is in keeping with Cloude et al.'s [11] findings that each emotion is more predictive of learning when different facial action units are involved.

This study uses BROMP-based detectors, which are a now-established way of investigating student affect [1]. However, these models produce atomized labels that likely to not capture the full range of experiences that an individual student experiences. In particular, BROMP does not differentiate between (standard) frustration and pleasurable frustration, which might improve our interpretation of this data. Similarly, it does not differentiate between tolerable and intolerable confusion, though the latter is most likely expressed as (standard) frustration.

Our model predicts that pleasurable frustration—possibly punctuated by delight—would be found within the space just above their current skill level, but only when value is also high. These results seem to be compatible with such predictions, though certainly additional measures—including in the moment investigations of the types of frustration students are having, the level of difficulty they perceive, and their current appraisals of the games value/utility, could certainly help us to tease these issues apart. Either text-based probes (e.g., those used to study mind wandering [27] or data-driven classroom interviews [4] could be employed in future research to explore these issues. Similarly, our model predicts two distinct causes for boredom, which could also be explored with these methodologies.

## 5.3 Conclusions

This study has extended established theoretical models to help explain empirical data on epistemic emotions. Specifically, we proposed the SVDET model, which adds student self-efficacy and time spent in the game as factors in the discussion of Flow Theory [31]. We have also shown that by applying ENA to 20-minute intervals of time, we can better capture affect dynamics that are relevant to student learning. Future research should work to help us better document differences between pleasurable and negative experiences of frustration, although these distinctions may be beyond the scope of what we can capture from external observations of the students' experience. Possible next steps could include triggering students to self-report their emotional experience when frustration is detected, which could help us to better understand its valance. We are currently exploring different methodologies for doing so.

Future research should also consider the degree to which we should adjust these time windows in order to better understand how things like in-game milestones and other feedback mechanisms also interact with students' affective experiences. In the case of open-ended systems like Crystal Island, this will complicate analyses, as students who reach different in-game milestones at different times may also differ in their relevant learning and motivational characteristics, including those considered in this study. Fine-tuning our division of both the temporal and student-level characteristics of affect dynamics will take effort, but as we have shown that it has the potential to increase the explanatory power of the data. We hope that future research will seek to explore these mechanisms in more detail.

## Acknowledgments

This research was supported by National Science Foundation Grants IIS-2016943 and IIS-2016993.

## References


[1] Juliana M.A.L. Andres, Stephen Hutt, Jaclyn Ocumpaugh, Ryan S. Baker, Nidhi Nasiar, and Chelsea Porter. 2021. How anxiety affects affect: A quantitative ethnographic investigation using affect detectors and data-targeted interviews. In *Proceedings of the Advances in Quantitative Ethnography: Third International Conference (ICQE 2021)*, Proceedings 3, 268–283.

[2] Ryan S. Baker, Elizabeth Cloude, Juliana M.A.L. Andres, and Zhanlan Wei. 2024. The Confrustion Constellation: A New Way of Looking at Confusion and Frustration.

[3] Ryan S. Baker, Sidney K. D'Mello, Ma Mercedes T. Rodrigo, and Arthur C. Graesser. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68, 4 (2010), 223–241.

[4] Ryan S. Baker, Stephen Hutt, Nigel Bosch, Jaclyn Ocumpaugh, Gautam Biswas, Luc Paquette, Juliana M.A.L. Andres, Nidhi Nasiar, and Anabil Munshi. 2023. Detector-driven classroom interviewing: Focusing qualitative researcher time by selecting cases in situ. *Educational Technology Research and Development*, 1–23.

[5] Ryan S. Baker and Jaclyn Ocumpaugh. 2014. Interaction-based affect detection in educational software. In *The Oxford Handbook of Affective Computing*, 233. Oxford University Press.

[6] Ryan S. Baker, Jaclyn Ocumpaugh, and Juliana M.A.L. Andres. 2020. BROMP quantitative field observations: A review. In *Learning Science: Theory, Research, & Practice*, 127–156.

[7] Albert Bandura and Sebastian Wessels. 1997. *Self-efficacy*. Cambridge University Press, Cambridge, UK.

[8] Shari L. Britner and Frank Pajares. 2006. Sources of science self-efficacy beliefs of middle school students. *J. Res. Science Teach*, 43, 5 (2006), 485–99.

[9] Martin M. Chemers, Li-tze Hu, and Ben F. Garcia. 2001. Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, 93, 1 (2001), 55.

[10] Gerald L. Clore and Andrew Ortony. 2008. Appraisal theories: How cognition shapes affect into emotion.

[11] Elizabeth B. Cloude, Anabil Munshi, Juliana M.A.L. Andres„ Jaclyn Ocumpaugh, Ryan S. Baker, and Gautam Biswas. 2024. Exploring Confusion and Frustration as Non-linear Dynamical Systems. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 241–252.

[12] Mihaly Csikszentmihalyi. 1975. *Beyond Boredom and Anxiety*. Jossey-Bass, San Francisco, CA.

[13] Sidney D'Mello and Arthur Graesser. 2012. Dynamics of affective states during complex learning. *Learning and Instruction*, 22, 2 (2012), 145–157.

[14] Anastasia Efklides. 2011. Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46, 1 (2011), 6–25.

[15] Phoebe C. Ellsworth. 2013. Appraisal theory: Old and new questions. *Emotion Review*, 5, 2 (2013), 125–131.

[16] Phoebe C. Ellsworth and Klaus R. Scherer. 2003. Appraisal processes in emotion.

[17] Johnny R.J. Fontaine, Klaus R. Scherer, Etienne B. Roesch, and Phoebe C. Ellsworth. 2007. The world of emotions is not two-dimensional. *Psychological Science*, 18, 12 (2007), 1050–1057.

[18] James Paul Gee. 2003. Learning about learning from a video game: Rise of Nations. Retrieved August 16, 2003.

[19] Shamya Karumbaiah, Juliana M.A.L. Andres, Anthony F. Botelho, Ryan S. Baker, and Jaclyn S. Ocumpaugh. 2018. The implications of a subtle difference in the calculation of affect dynamics. In *Proceedings of the 26th International Conference for Computers in Education*, 29–38.

[20] Shamya Karumbaiah and Ryan S. Baker. 2021. Studying affect dynamics using epistemic networks. In *Proceedings of Advances in Quantitative Ethnography: Second International Conference (ICQE 2020)*, Proceedings 2, 362–374. Springer International Publishing, 2021.

[21] Shamya Karumbaiah, Ryan S. Baker, Jaclyn Ocumpaugh, and Juliana M.A.L. Andres. 2021. A re-analysis and synthesis of data on affect dynamics in learning. *IEEE Transactions on Affective Computing*.

[22] Shamya Karumbaiah, Ryan S. Baker, Yan Tao, and Ziyang Liu. 2022. How does students' affect in virtual learning relate to their outcomes? A systematic review challenging the positive-negative dichotomy. In *Proceedings of LAK22,* 24–33.

[23] Blair Lehman, Sidney D'Mello, and Arthur Graesser. 2012. Confusion and complex learning during interactions with computer learning environments. *The Internet and Higher Education*, 15, 3 (2012), 184–194.

[24] Lisa Linnenbrink-Garcia, Amanda M. Durik, AnneMarie M. Conley, Kenneth E. Barron, John M. Tauer, Stuart A. Karabenick, and Judith M. Harackiewicz. 2010. Measuring situational interest in academic domains. *Educational and Psychological Measurement*, 70, 4 (2010), 647–671.

[25] Blair Lehman, Sidney D'Mello, and Art Graesser. 2012. Confusion and complex learning during interactions with computer learning environments. *The Internet and Higher Education*, 15, 3 (2012), 184–194.

[26] Zhongxiu Liu, Visit Pataranutaporn, Jaclyn Ocumpaugh, and Ryan Baker. 2013. Sequences of frustration and confusion, and learning. In *EDM*.

[27] Caitlin Mills, Andrew R. Porter, Jessica R. Andrews-Hanna, Kalina Christoff, and Alexander Colby. 2021. How task-unrelated and freely moving thought relate to affect: Evidence for dissociable patterns in everyday life. *Emotion*, 21, 5 (2021), 1029.

[28] Nidhi Nasiar, Andres Felipe Zambrano, Jaclyn Ocumpaugh, Alex Goslen, Jonathan Rowe, Jessica Vandenberg, Jordan Esiason, and Stephen Hutt. 2024. The Influence of Different Measurement Approaches on Student Affect Transitions Using Ordered Networks. In *Proceedings of Advances in Quantitative Ethnography (ICQE 2024)*, In press.

[29] Jaclyn Ocumpaugh. 2015. *Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual*. Technical Report No. 60. Teachers College, Columbia University, New York, NY, USA and Ateneo Laboratory for the Learning Sciences, Manila, Philippines.

[30] Frank Pajares and Dale Schunk. 2001. The development of academic self-efficacy. In *Development of Achievement Motivation*, 1–27.

[31] Reinhard Pekrun. 2006. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18 (2006), 315–341.

[32] Reinhard Pekrun, Nathan C. Hall, Thomas Goetz, and Raymond P. Perry. 2014. Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology*, 106, 3 (2014), 696.

[33] Reinhard Pekrun, Elisabeth Vogl, Krista R. Muis, and Gale M. Sinatra. 2017. Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion*, 31, 6 (2017), 1268–1276.

[34] Paul Rozin and Adam B. Cohen. 2003. High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, 3, 1 (2003), 68.

[35] David Williamson Shaffer, Wesley Collier, and Andrew R. Ruis. 2016. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3, 3 (2016), 9–45.

[36] Yuanru Tan, Andrew R. Ruis, Cody Marquart, Zhiqiang Cai, Mariah A. Knowles, and David Williamson Shaffer. 2022. Ordered network analysis. In Proceedings of the International Conference on Quantitative Ethnography, 101–116. Springer Nature Switzerland, Cham, 2022

[37] Andres Felipe Zambrano, Nidhi Nasiar, Jaclyn Ocumpaugh, Alex Goslen, Jiayi Zhang, Jonathan Rowe, Jordan Esiason, Jessica Vandenberg, and Stephen Hutt. 2024. Says Who? How different ground truth measures of emotion impact student affective modeling. In *Proceedings of the 17th International Conference on Educational Data Mining*, 211–223.

[38] Yingbin Zhang, Luc Paquette, Ryan S. Baker, Jaclyn Ocumpaugh, Nigel Bosch, Gautam Biswas, and Anabil Munshi. 2021. Can Strategic Behaviour Facilitate Confusion Resolution? The Interplay between Confusion and Metacognitive Strategies in Betty's Brain. *Journal of Learning Analytics*, 8, 3.