Exploring Confusion and Frustration as Non-linear Dynamical Systems

Elizabeth B. Cloude
Faculty of Education and Culture
Tampere University
Tampere, Finland
elizabeth.cloude@tuni.fi

Anabil Munshi
Edmentum
New York City, NY, USA
munshi.anabil@gmail.com

J. M. Alexandra Andres
Penn Center for Learning Analytics
University of Pennsylvania
Philadelphia, PA, USA
alexandralandres@gmail.com

Jaclyn Ocumpaugh
Penn Center for Learning Analytics
University of Pennsylvania
Philadelphia, PA, USA
jlocumpaugh@gmail.com

Ryan S. Baker
Penn Center for Learning Analytics
University of Pennsylvania
Philadelphia, PA, USA
ryanshaunbaker@gmail.com

Gautam Biswas
Department of Computer Science and Engineering
Vanderbilt University
Nashville, TN, USA
gautam.biswas@vanderbilt.edu

ABSTRACT

Numerous studies aim to enhance learning in digital environments through emotionally-sensitive interventions. The D'Mello and Graesser (2012) model of affect dynamics hypothesizes that when a learner encounters confusion, the degree to which it is prolonged (and transitions into frustration) or resolved, significantly affects their learning outcomes in digital environments. However, studies yield inconclusive results regarding relations between confusion, frustration, and learning. More research is needed to explore how confusion and frustration manifest during learning and its relation to outcomes. We go beyond past work looking at the rate, duration, and transitions of confusion and frustration by treating these affective states as non-linear dynamical systems consisting of expressive and behavioral components. We examined the frequency and recurrence of facial expressions associated with basic emotions (as automatically labeled by AffDex, a standard tool for analyzing emotions with video data) during confused and frustrated states (as automatically labeled with BROMP-based detectors applied to students' interaction data). We compare these co-occurring patterns to learning outcomes (pre-tests, post-tests, and learning gains) within a digital learning environment, Betty's Brain. Results showed that the frequency and recurrence rate of basic emotions expressed during confusion and frustration are complex and remain incompletely understood. Specifically, we show that confusion and frustration have different relationships with learning outcomes, depending on which basic emotion expressions they co-occur with. Implications of this study open avenues for better understanding these emotions as complex and non-linear dynamical systems, in the long-term enabling personalized feedback and emotional support within digital learning environments that enhance learning outcomes.

LAK '24, March 18–22, 2024, Kyoto, Japan

ACM ISBN 979-8-4007-1618-8/24/03.
https://doi.org/10.1145/3636555.3636875

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments**; **Computer-assisted instruction**.

KEYWORDS

Confusion, Frustration, Non-linear Dynamical Systems, Learning Outcomes, Digital Learning Environments

ACM Reference Format:
Elizabeth B. Cloude, Anabil Munshi, J. M. Alexandra Andres, Jaclyn Ocumpaugh, Ryan S. Baker, and Gautam Biswas. 2024. Exploring Confusion and Frustration as Non-linear Dynamical Systems. In *The 14th Learning Analytics and Knowledge Conference (LAK '24), March 18–22, 2024, Kyoto, Japan.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3636555.3636875

1 INTRODUCTION

Emotions are not mere byproducts of learning; they are integral components of the learning process affected by a learner's knowledge and goals [40]. Numerous studies aim to create emotionally sensitive interventions to improve engagement and learning outcomes in digital learning environments [23]. The D'Mello and Graesser [18] model of affect dynamics is a popular contemporary framework for studying emotions with digital learning environments. This cognitive model is centered around the notion of cognitive disequilibrium [51], a state of uncertainty when a learner is confronted with an obstacle in assimilating knowledge into their schema. The degree to which cognitive disequilibrium is restored determines whether an affective state is advantageous or detrimental to learning processes.

In this model, confusion emerges when cognitive disequilibrium is initially experienced. Restoring cognitive equilibrium requires problem solving and reasoning to resolve the impasse in understanding and is hypothesized to benefit learning [19]. However, prolonged confusion and cognitive disequilibrium can transition to frustration, which is theorized to be harmful to information processing and learning [18]. Many studies have utilized diverse methods to investigate the role of emotions on learning with digital learning environments, including leveraging physiological sensors [26], concurrent self-report instruments [12], retrospective interviews [19], emote-aloud protocols [16], behaviors [14], and observations [31].

However, the empirical landscape surrounding the interplay between confusion, frustration, and their influence on the process of learning remains inconclusive [35, 38]. A systematic review of 39 studies [30] revealed that the transition from confusion to frustration, a key assumption of D'Mello and Graesser [18], was a rare occurrence during learning with digital learning environments. Moreover, the findings emphasized that there is no clear consensus regarding the role of confusion and frustration and their subsequent impact on information processing and learning [30].

For instance, [12] found that learners who frequently reported experiencing confusion and/or frustration during a learning task tended to have poorer learning outcomes. This finding suggests a negative relationship between the frequency of confusion and/or frustration and learning outcomes. Yet, when delving further into the temporal dimension of this relationship, another study [11] found that prolonged confused and frustrated facial expressions were positively (and moderately) associated with the time learners engaged in information processing. This result suggests that, at times, confusion and frustration may enhance cognitive processing. Conversely, [4] found no discernible relationship between the transition from prolonged confusion to frustration and its impact on outcomes (pre, post, learning gains), while [38] found positive relationships for both transitions from confusion to frustration and frustration to confusion with learning outcomes.

In sum, most studies fail to yield supporting evidence for the hypothesized relationships between confusion, frustration, and learning [30]. The inconclusive results may stem from a variety of factors, including theoretical, methodological, and analytical challenges. First, many approaches hypothesize that confusion and frustration can be simplified into distinct emotion categories [30]. Yet, emotions in general may not be distinct states with clear boundaries [7]. An emotion is a multi-faceted system that comprises affective processes emerging from multiple, interacting subsystems involving psychophysiological, subjective, behavioral, facially expressive, and neurological components [55].

Second, there are many approaches to defining confusion and frustration [32] and a variety of methods to measure it (e.g., automated detection of combined facial action units to classify emotions [11], classroom observations [31], or interaction-based log detection [53]. Finally, a common technique for modeling relations between confusion, frustration, and learning is through using mechanistic, reductionist methods that decompose a whole into the sum of its parts, often losing information on the dynamic, multi-level affective processes [4, 11, 12]. Yet, many real-world multi-faceted systems exhibit nonlinear dependencies. These systems have benefited from different types of statistical analyses than are typically used in the learning analytics (LA) community to study patterns of epistemic emotions.

Thus, we argue that studying confusion and frustration as a multi-faceted, non-linear dynamical system (NLDS) (NLDS) could be more comprehensively understood by using multiple sources of data that capture underlying affective processes that co-occur during periods of confusion and frustration during learning with digital learning environments. This novel approach is aligned with the notion that emotions comprise multiple subsystems [55] and seeks to answer two research questions:

- **RQ1:** What is the frequency and recurrence rate of basic emotions (as detected automatically using facial recognition tracking [43]) during confusion and frustration (as detected using established, interaction-based detectors built using BROMP labels of confusion and frustration [29]), and do they align with the theorized relationships between basic and epistemic emotions that can be found in the literature?
- **RQ2:** Are learning outcomes associated with the frequency and recurrence rate of basic emotions during confusion and frustration? We investigate these questions within the context of a digital learning environment called Betty's Brain [34] with the goal of developing a more holistic understanding of emotions regarding their impact on learning with digital learning environments.

We investigate these questions within the context of a digital learning environment called Betty's Brain [34] with the goal of developing a more holistic understanding of emotions regarding their impact on learning with digital learning environments.

2 PREVIOUS RESEARCH ON CONFUSION AND FRUSTRATION DURING LEARNING

Confusion and frustration are commonly described as "epistemic" emotions because they deal with knowledge-related properties [40]. In this way, both are linked by cognitive and affective processes [50], but there can be notable differences. Confusion signifies that a learner is grappling to understand new and challenging concepts. Thus, it arises based on a learner's appraisal of feeling uncertain, which may include varying degrees of novelty, complexity, conflict, and unfamiliarity, with low comprehensibility of the materials [8, 22, 56]. However, variation within this category has been noted in the literature, including "productive" and "unproductive" types of confusion, which may be defined by the learner's threshold of tolerance for this emotion and their ability to regulate it [36]. According to Lodge et al. [39], confusion exists on a spectrum and its impact on learning changes depending on factors related to task difficulty, prior knowledge, emotion regulation, and the feedback/support available.

On the other hand, frustration is theorized to emerge when a learner's motivation or goal pursuit is blocked, sometimes following a state of "hopeless confusion" [18]. A learner might reach a state of hopeless confusion when they repeatedly cannot resolve the impasse and important learning goals remain blocked with no available plan or strategy to move forward. Thus, frustration differs from confusion because it involves a feeling of being "stuck" or a conflict of motives [10]. However, Amsel [3] argues that not all frustration is the same. Some frustration can boost a learner's motivation and direct their attention, while other forms can suppress learning. The dominance of one effect over another hinges on the individual learner's anticipatory goal response, degree of agency in moving forward, and how they approach conflict (e.g., avoid conflicts due to fear of failure) [10].

The anticipatory goal response is influenced by early theories in animal learning literature, more commonly known as reward schedules. In other words, the anticipation a learner has for achieving a goal stems from their prior experiences of successes (rewards) or failures (non-rewards) specific to the learning context [3]. Primary frustration, also known as frustration drive, is a temporary state that arises when a learner encounters a situation where they expected a reward (e.g., making progress toward a learning goal), but the reward is not received promptly, is delayed, or reduced [2]. This type of frustration is hypothesized to have motivational and energizing effects on learning. When learners experience at least some intermittent or inconsistent rewards during the learning process, they tend to persist despite obstacles and frustration. In contrast, if primary frustration remains unresolved, it may lead to conditioned frustration.

Conditioned frustration is a learned response that occurs when a learner anticipates frustration based on prior experiences in specific learning situations. This anticipated or conditioned frustration can have an adverse impact on motivation and engagement that may inhibit learning and memory [27]. Conditioned frustration is believed to increase as a function of repeated failures, which is fundamentally linked to the "frustration-aggression hypothesis" [17]. The frustration-aggression hypothesis underscores the aggressive tendencies of conditioned frustration. However, few studies in educational research have investigated whether variation in the type of frustration impacts learning differently with digital learning environments.

Historically, research in affective computing within education has often overlooked variability in how confusion or frustration experienced by learners manifests with digital learning environments, beyond considering their frequency, sequence, and duration [4, 11, 12, 30]. To advance the quantification of various forms of confusion and frustration, it may be essential to gather data that capture the affective processes underlying multiple subsystems that make up an episode of confusion or frustration during learning with digital learning environments. Specifically, data channels that capture components corresponding to the psychophysiological, neurological, expressive (e.g., facial expressions), behavioral, and subjective feelings subsystems of confusion and frustration [55] may offer a more comprehensive measure of how confusion and frustration manifest during learning with digital learning environments.

2.1 Facial Expressions and Affective Behaviors

Recent years within the LA community have witnessed a surge of educational studies utilizing facial expression analysis via facial recognition tracking systems. The most common method involves classifying what have been called "basic emotions" [20] in a moment-by-moment fashion during learning activities with digital learning environments: anger, disgust, enjoyment, surprise, fear, and sadness. These basic emotions are detected based on specific facial expressive configurations [20].

To facilitate facial expression analysis, researchers rely on algorithms built from the Facial Action Coding System (FACS), which is an anatomy-based coding system enabling human coders to assess basic emotions using 46 observable facial action units (FAUs) that correspond to facial expressions of emotions [20]. This method has been used to classify epistemic emotions [11]; however, there is much debate about whether facial expression configurations cleanly map onto epistemic emotions [21], which are influenced by cultural and individual differences [54].

Notably, [47] found a substantial overlap between the facial expressions tied to basic emotions and the behaviors associated with epistemic emotions, including confusion and frustration. This overlap was substantial enough to make it possible to adapt a basic emotion FAU detector through machine learning to effectively identify epistemic emotions. For example, a combination of elevated likelihoods of anger and disgust being present (according to FAU detectors), coupled with a high likelihood of sadness or low contempt, successfully predicted confusion (according to the interaction-based detectors built using BROMP), while frustration was predicted by either a high likelihood of disgust presence paired with low fear, or low disgust, the potential presence of sadness, and low contempt.

These findings challenge traditional definitions that draw clear distinctions between the two epistemic emotions in question—namely that confused behaviors are typically associated with facial expressions of disgust and anger, while frustration behaviors are more likely to be associated with disgust and fear [3, 17]. Thus, the objective of this paper is to further investigate the findings from [47] by examining confusion and frustration as multi-faceted, dynamical systems. Specifically, we apply a non-linear dynamical systems technique called recurrence quantification analysis [59] to explore the degree to which facial expressions of basic emotions co-occur during confusion and frustration (measured via interaction-based detectors built using BROMP [29]). We examined the frequency and recurrence of facial expressions of basic emotions that co-occur during confusion and frustration and their association to learning outcomes with Betty's Brain.

2.2 Recurrence Quantification Analysis: Auto-Recurrence Quantification (aRQA) & Multidimensional Recurrence Quantification (MdRQA)

Recurrence quantification analysis (RQA) is a method for studying nonlinear dynamical systems (NLDS). Even when applied to categorical time series data, RQA is versatile, as it simplifies the recurrence matrix into distance comparisons to binary decisions (1=recurrence; 0=non-recurrence), and it can be used to analyze time series with at least 20 data points [59]. Within the LA community, it has been used to study text comprehension [37] and eye-gaze dynamics during collaborative learning [45].

Generically, RQA identifies instances when a system revisits a state it has encountered before, taking into account historical patterns by assessing the similarity between time-series data points using a recurrence plot, which builds a recurrence matrix using a phase-space trajectory analysis [41]. Within a phase space, every parameter of the system is represented as an axis of a multidimensional space, and recurrence is defined by a threshold—often referred to as the radius—that measures the "closeness" between elements within a time series. The radius defines the window in which recurrence is computed along the phase space trajectory. Figure 1 shows four different recurrence plots, which plot individual time series points against one another, allowing for comparisons along the x and y axes to identify the temporal structure of the time series.

In this study, we employed two RQA methods: auto-recurrence quantification (aRQA), which assesses whether repetitive patterns exist within a *single* time series by comparing it to itself, and multi-dimensional recurrence quantification (MdRQA), which extends recurrence analysis to involve *multiple* time series [58]. In other words, aRQA is a technique used to measure the presence of self-similar (recurring) patterns within a single (unidimensional) time series (e.g., how repetitive a learner interacts with game elements such as reading research articles during game-based learning [15]), while MdQRA—as its name suggests—characterizes self-similarity for more than two time series (e.g., collective patterns of regularity across team members' signals of speech rate, body movements, and team interactions during collaborative problem solving [1]). hile aRQA is useful for determining the recurrence of a single basic emotion time series, it requires data aggregation, a limitation that could destroy information that may be key to understanding the affective dynamics underlying confusion or frustration. In contrast, MdRQA provides a more comprehensive analysis that does not require aggregation of data across the time series.

To define the **rate of occurrence** (i.e., how often facial expressions of basic emotions co-occurred during confused and frustrated actions), ***frequency*** **(FREQ)**, was used to measure the proportion times an event occurred; but this metric is simply a count of the events that happened within the whole time series and does not account for recurring patterns within slices of a phase-space trajectory. Thus, three common RQA metrics [41] were used to measure the rate of recurrence patterns over time. First, ***recurrence rate*** **(RR)**, uses a sliding time window (or radius) to divide the time series, and then calculates the proportion of times a specific event reoccurs within those windows (shown as black dots in Figure 1) out of all the events. In simpler terms, a high RR means there are more repeating points, while a low RR means the pattern is more irregular. Like RR, ***determinism*** **(DET)** uses the same sliding time window, but unlike RR, DET calculates the proportion of events that form diagonal patterns in the recurrence plot (shown as black lines in Figure 1; this happens when the system is more deterministic and less stochastic). As such, higher DET values (longer diagonals) indicate stronger recurring patterns across those windows, while lower DET means there is more irregularity. Finally, the third metric, ***diagonal entropy*** **(DENTR)**, defines the chaos in the system by looking at how far these points are from the diagonal line, such that irregular distances from the diagonal pattern equal high entropy [41]. In other words, higher entropy means the system is more irregular.

For the MdRQAs, the entropy metric was extended to better define the multidimensional structure of all facial expressions of basic emotions (six time series) during confused and frustrated moments. Specifically, ***vertical entropy*** **(VENTR)**, in addition to DENTR, was used to compute all conceivable diagonal and vertical lines within a recurrence matrix using a frequency distribution [1]. Similar to DENTR, which measures the irregularity of a system using diagonal lines, VENTR assesses the degree of irregularity of the system returning to a specific state using vertical lines [41].

[1]Shannon entropy quantifies the level of irregularity within a system. When the distribution of a set of observations is evenly spread out, Shannon entropy reaches its maximum value. In contrast, when all observations have the same value, Shannon entropy is minimized and equals zero.

3 METHODS

3.1 Data Collection and Study Design

This study uses the same data as [47], which included 74 sixth-graders in a classroom at a large, urban public school in the south-eastern USA. Participants were provided with webcam-equipped laptops, and the study lasted 7 days. On day one, participants completed a 30-45 minute paper-based, pre-test content assessment. On day two, participants engaged in a 30-minute training session to understand the learning goals and how to utilize the features built into the Betty's Brain interface. Over the next four days, participants were instructed to teach Betty about the causal relationships involved in the process of climate change using concept maps in Betty's Brain software (roughly 50-minute sessions). On the last day, participants completed a post-test assessment similar to the content assessment administered at pre-test.

Prior to recruitment and data collection, an ethics review board approved this study. Assent and informed consent were obtained prior to data collection, and to maintain the privacy of participants, only de-identified data were shared following post-hoc processing procedures. The digital labels generated from the video files using the facial recognition algorithm built by AffDex [43] were the only data included in our data analysis (i.e., we did not have access to raw video showing faces). Due to IRB requirements, demographic information was not collected on participants; but sixth graders in this part of the USA are typically 11 to 12 years old. Inclusion in current analysis required participants to have full data channels for our variables of interest, with more than 20 data points (requirements for conducting RQA) [59] for both basic emotions via facial expressions and behavioral instances of confused and frustrated states during learning with Betty's Brain. Exclusion due to data loss from movement or technical problems resulted in 5 students being removed from the confusion analysis, and 33 from the frustration analysis.

3.2 The Betty's Brain Learning Environment

Betty's Brain [34] is an open-ended learning environment [33], where middle-schoolers learn about complex science topics by building causal (cause-and-effect) models of scientific processes (e.g., climate change). Betty's Brain uses a learning-by-teaching paradigm, where learners teach a virtual pedagogical agent named Betty [9]. The system provides learners with resources and tools to construct and evaluate their causal models. A science book provides resource pages embedded within the system, while the causal map interface has a drag-and-drop menu to help learners build the causal maps they developed to teach Betty and provides a visual representation of their current causal map, with tools to add, delete, and modify if needed (Figure 2). In addition, a quiz tool allows learners to probe Betty's domain knowledge. The mentor agent in Betty's Brain, Mr. Davis, administers and grades the quizzes, providing strategic feedback when needed via adaptive conversational scaffolds [46].

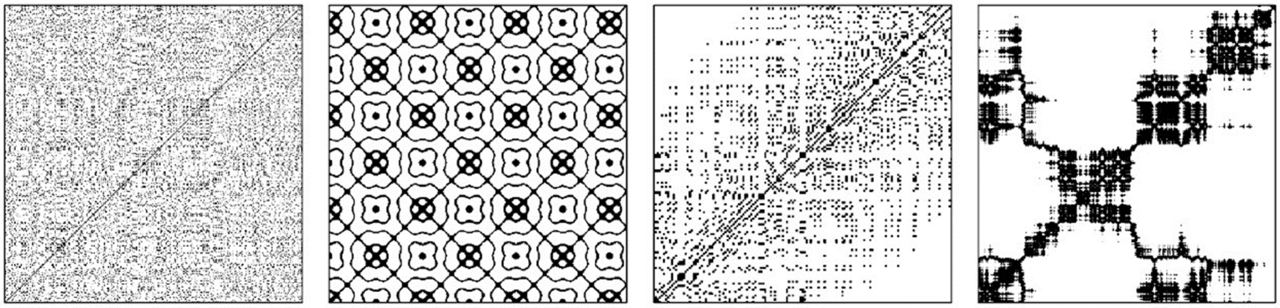

Figure 1: Different recurrence plots are shown to emphasize diverse temporal structures. From left to right, we see 1) uniform homogeneity (also known as white noise, characterized by the presence of uniformly distributed noise), 2) harmonic periodicity–comprising the overlay of harmonic oscillations, 3) drifting patterns–emerging from a logistic map influenced by a linearly increasing component, and 4) disrupted sequences–notable in the context of Brownian motion. Each of these are examples of auto-RQA and are symmetric. Multi-dimensional RQA are between two or more time series and generally not symmetric (image obtained from [41, 52]).

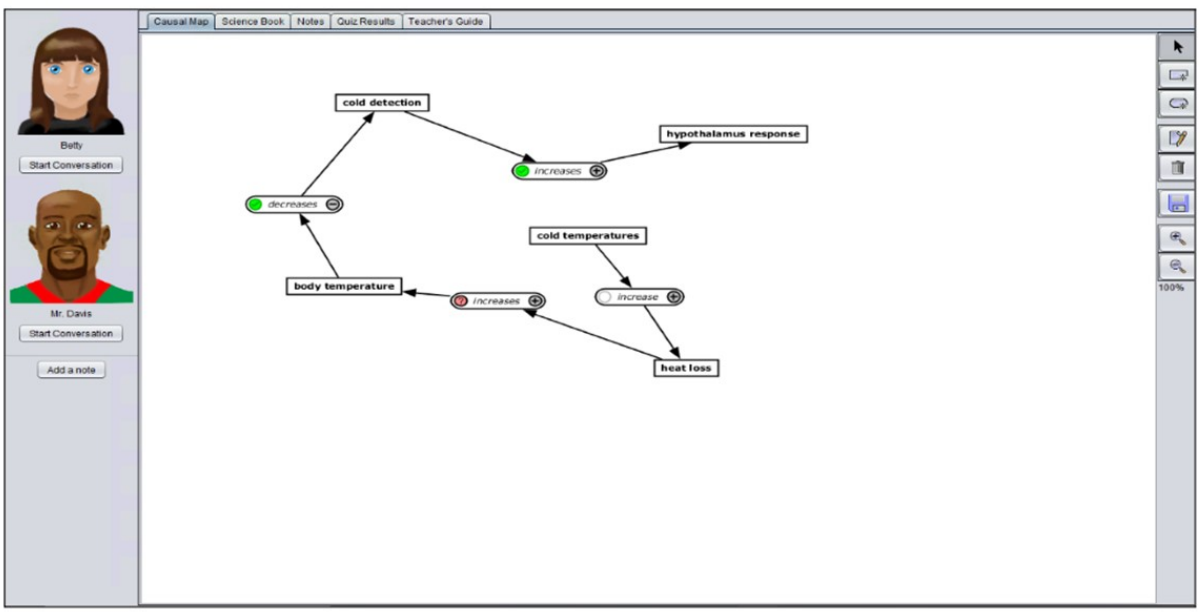

Figure 2: The "causal map" interface in the Betty's Brain thermoregulation unit (image from [46]).

3.3 Data Processing, Coding and Scoring

3.3.1 Automated Facial Expression Detectors. Videos were collected at a sampling rate of 30Hz and processed post hoc using the AffDex module in iMotions software [28]. AffDex detects facial landmarks and applies a set of rules built by Affectiva Inc. [44] that is based on FACs [20]. To classify momentary basic emotions from facial expressions, AffDex generates time series data for each participant that provides the log-likelihood values of a human coder rating a basic emotion as present. A validation study [57] demonstrated that the AffDex algorithm achieved an acceptable accuracy for prototypical facial expressions in laboratory settings; however, there is no evidence regarding the validity for facial expressions that occur in natural settings such as classrooms. The digital labels generated by the algorithm classify the log likelihood that a human coder labeled a facial expression, based on an exact configuration of FAUs mapped onto facial structures, as present or absent, for six basic emotions: joy, sadness, anger, fear, disgust, and surprise. The values produced ranged from negative, 0, and positive; it is important to note that these values represent the probability of absence or presence of the basic emotion rather than its intensity. We defined the presence of six basic emotions: anger, disgust, joy, surprise, fear, sadness, based on values at or above 1. All negative values were turned into zeros to indicate that the basic emotion was not labeled as present by the algorithm.

3.3.2 Interaction-Based Detectors. Confused and frustrated states were detected using interaction-based detectors [6], which were developed using BROMP labels. Observers using BROMP collect quantitative field observations of participants' epistemic behaviors and affect using a momentary time sampling method [49]. In this study, coding categories included boredom, confusion, delight, engaged concentration, frustration, and other task-related behaviors. These labels are then used to train detectors of these affective states that can run in real time. For this paper, we used interaction-based detectors of confusion and frustration that were previously validated for Betty's Brain [29, 48].

3.3.3 Data Synchronization. The OBS (video) output was aligned to the log files by offsetting the time based on the difference between the OBS time server (EST) and interaction log time server (UTC). To align the video data with the log-based affect, we processed the output of the tracking system at every 20 seconds since the log-based affect detectors were built to generate the probability of an emotion every 20 seconds. Thus, when participants indicated they were either confused or frustrated via log-based detectors at 20-second intervals, we included the likelihood values of the presence of six basic emotions. Next, we computed the co-occurrence of each basic emotion separately for 1) confusion or 2) frustration over the course of learning with Betty's Brain (Table 1). In addition, we calculated the recurrence rate of each individual basic emotion during confused or frustrated states, computing the degree of reappearance of the basic emotion after an initial occurrence during learning.

3.3.4 Outcomes Measures. Identical pre- and post-tests were used to define the learner's domain knowledge, based on the percentage of correct responses in 7 multiple choice items and 6 open-response items. Learning gain scores were calculated for each learner using a normalized change equation [42].

3.4 Data Analysis

For RQ1, we used two RQA methods (aRQA and MdRQA) to investigate how basic emotions emerged in confusion and frustration. aRQA was calculated with the 'crqa' package [13], using six separate categorical aRQAs (one for each of the six basic emotions) for each of the two epistemic emotions (confusion and frustration), for a total of 12 aRQAs per participant. We set parameters for both the embedding dimension (the number of dimensions to which a unidimensional signal was promoted) and the delay (the sampling distance along that unidimensional signal at which successive embedded dimensions were estimated). To ensure that temporal structures of small granularity were detected, we set parameters to low thresholds (delay=1, radius=0.1, and embedding=2), using a Euclidean distance. This distance of the radius was set to a fixed and low threshold of 0.1 to ensure that each window captured all states and any possible recurrences to allow for symmetrical comparison [41]. In addition, since the aRQA variables require aggregation, we averaged the recurrence variables for each basic emotion during 1) confusion and 2) frustration.

MdRQA analyses were conducted by utilizing a function built by [58]. Two separate MdRQAs were calculated, one for confusion and another for frustration, using all six time series of facial expressions of basic emotions during learning. The parameters were again set to a low threshold to detect any granular temporal structures across the six time series (delay=1, radius=0.1, embedding=2), using a Euclidean distance. MdRQA does not require aggregation and thus we did not average the recurrence variables for this analysis.

Finally, for RQ2, a series of Pearson correlations were calculated to examine associations between the frequency and recurrence variables of aRQA and MdRQA metrics with learning gain and assessment knowledge scores (pre and post). A Benjamini and Hochberg posthoc correction was applied to control for false discoveries due to multiple testing.

4 RESULTS

4.1 RQ1. What is the frequency and recurrence rate of basic emotions during confusion and frustration?

Table 1 shows the aRQA analyses of confusion and frustration. Notably, both academic emotions show frequent overlap with surprise (37.07% for confusion, 33.16% for frustration) and with disgust (30% for confusion and 31.1% for frustration. Other aRQA metrics also showed high overlap. For confusion, the basic emotion of disgust showed high recurrence rates (RR=74), sequence patterns (DET=96), and entropy values (ENT=2.42), which demonstrated that confusions' co-occurrence with disgust was relatively stable and regular. For frustration, these numbers were also high (RR=82; DET=98; ENT=3.26). Similar results were found for confusion and surprise (RR=70; DET=95.16, ENT=2.53), but frustration's relationship with surprise was slightly more irregular and unpredictable (RR=79; DET=97, ENT=3.27). Other basic emotions co-occurred and recurred less regularly with confusion and frustration (FREQ>6%; RR>90, DET>99, ENT≥3.11 for confusion; FREQ>7%; RR>81, DET>97, ENT≥3 for frustration), though one might note that frequency rates that hover around 5% (roughly 1 in 20 instances) are not exactly rare.

4.2 RQ2. Are learning outcomes associated with the frequency and recurrence rate of basic emotions during confusion and frustration?

4.2.1 Confusion, Basic Emotions, and Learning Measures. Table 2 shows a series of Pearson correlations between the aRQA metrics of each basic emotion expressed during a confusion label and student learning gains. This analysis revealed a marginally significant, and moderate, positive association between recurring patterns of sadness (DET) co-occurring during confusion and learning gains (r=.27, p=.04, adjusted α=.01; Table 2), despite the relatively low the frequency of sadness co-occurring with confusion (6%); thus,

Table 1: The frequency and aggregated aRQA metrics for each basic emotion during confusion and frustration.

	Confusion (n=69)				Frustration (n=41)			
	FREQ	RR	DET	DENTR	FREQ	RR	DET	DENTR
Anger	3.94%	94.28	99.71	3.43	3.01%	94.37	99.73	3.29
Sad	6.07%	92.01	99.46	3.32	5.57%	92.17	99.18	3.38
Disgust	30%	73.8	96.27	2.42	31.1%	81.58	97.79	3.26
Joy	4.06%	81.16	97.22	2.99	2.96%	92.99	99.85	3.11
Surprise	37.07%	70.3	95.16	2.53	33.16%	78.95	96.78	3.27
Fear	6.23%	92.41	99.53	3.37	5.77%	89.46	99.18	3.32

Note. All RQA variables represent averages.

this result indicated that the recurring patterns (DET) of sadness during confusion were positively associated with learning gains.

No significant relationships were found for post-tests (Table 3), but Table 4 shows there was also a moderate—though marginally significant—correlation between the frequency of facial expressions of sadness during confusion and pre-test scores (r=.22, p=.07, adjusted α=.01).

Finally, we examine the results of our MdRQA (Table 5), which examines these patterns with less data loss. This analysis did not show any significant associations between learning gains, post-, and pre-test scores with the remaining frequency and recurrence variables of basic emotions expressed during confusion (ps>.05; Table 3 & Table 5). Thus, for confusion, only low levels of sadness (6%) are positively associated with the learner's prior knowledge and only with an aRQA analysis.

4.2.2 Frustration, Basic Emotions, and Learning Measures. The same aRQA analyses conducted for confusion were also applied to explore the relationships between frustration's recurrence with basic emotions and learning outcomes. As Table 6 shows, Pearson correlations found a marginally significant, and moderate, negative association between the average DET of disgust during frustration and learning gains (r=-.33, p=.04, adjusted α=.01). In other words, the longer recurring patterns of disgust during frustration were associated with lower learning gains. Similarly, there was a marginally significant, and moderate, positive association between the average frequency of disgust (31%) during frustration and learning gains (r=.3, p=.06, adjusted α=.01). This indicated that when disgust co-occurred with frustration, but did not recur over time, it was positively associated with learning gains.

For post-test scores (Table 7), there were only marginally significant results. The average ENT of sadness during frustration was moderately and negatively correlated with post-test scores (r=-.44, p=.04, adjusted α=.01). In other words, the more irregular recurring patterns of sadness was expressed during frustration was associated with worse knowledge scores on the post-test assessment.

Correlations with pre-test scores showed both significant and marginally significant correlations (Table 8). The average DET of disgusted facial expressions during frustration was significantly, moderately, and positively correlated with pre-test scores (r=.38, p=.02, adjusted α=.03), while the average RR of disgust during frustration showed a marginally significant, and positive correlation with pre-test scores (r=.28, p=.08, adjusted α=.04). In contrast, the frequency of disgusted facial expressions during frustration was significantly, moderately, and negatively correlated with pre-test scores (r=-.46, p=.002, adjusted α=.01). This indicated that the more often disgust was facially expressed during frustration (FREQ), the less prior knowledge the learner had about the learning domain. In contrast, the more recurring patterns of disgusted facial expressions during frustrated actions (both RR and DET), the more prior knowledge the learner had about the domain.

Finally, Table 9 presents the results of our MdRQA analysis of frustration. As with confusion, this analysis did not yield any significant associations between the remaining recurrence metrics of basic emotions during frustration and outcomes (ps>.05).

5 DISCUSSION

In this study, we investigated confusion and frustration as non-linear dynamical systems by evaluating the underlying facially expressive affective processes that co-occur during confusion and frustration [55]. Specifically, we evaluated the frequency and recurrence rate of facial expressions of basic emotions during confusion and frustration and its relationship with outcomes (learning gain, post, pre) during learning with Betty's Brain.

In **RQ1**, we found that confusion, on average, co-occurred with facial expressions of disgust and surprise most frequently (roughly 30% and 40% of the time respectively, while it showed less co-occurrence with the facial expressions of anger, sadness, joy, and fear (> 5% of the time on average). Other metrics from our aRQA analysis corresponded with these findings, including the recurrence rate and length of the recurrence. These findings partially supported our hypothesis, where the results were aligned with [47] given that the high frequency of disgusted facial expressions were associated with confusion. However, our findings showed that confusion corresponded with a high frequency of surprised facial expressions, yet low frequency of angry and sad facial expressions [47]. Notably, the higher frequency of surprise has sometimes been found to occur during confusion [56], a finding aligned with how learners might react to feelings of uncertainty paired with high novelty when an existing schema was disrupted [51]. It is interesting to also note that when surprised facial expressions recurred regularly during confusion, it may indicate a possible lack of effective problem solving or emotion regulation if the learner was repeatedly surprised during learning.

Similarly, we found that frustration, on average, co-occurred with facial expressions of disgust and surprise (roughly 31% and

Table 2: Learning gains vs aRQA metrics of basic emotions during confusion; significant correlations given in black, bold font.

	LEARNING GAINS																	
	Anger			Sad			Disgust			Joy			Surprise			Fear		
	r	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*
RR	.17	.23	.04	.16	.24	.04	-.07	.58	.04	-.01	.94	.05	-.16	.24	.04	-.01	.99	.05
DET	.21	.12	.01	**.27**	**.04**	**.01**	-.13	.29	.03	-.06	.63	.01	-.1	.47	.05	-.03	.77	.04
DENTR	.09	.53	.05	-.03	.83	.05	.01	.93	.05	.03	.81	.03	-.16	.24	.01	-.05	.54	.03
FREQ	-.18	.14	.03	-.16	.19	.03	.13	.28	.01	.03	.79	.04	.11	.38	.03	.19	.12	.01

Note. All aRQA variables represent averages; α=adjusted alpha.

Table 3: Post-test scores vs aRQA metrics of basic emotions during confusion; significant correlations given in black, bold font.

	POST-TEST SCORES																	
	Anger			Sad			Disgust			Joy			Surprise			Fear		
	r	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*
RR	.17	.23	.04	.16	.24	.04	-.07	.58	.04	-.01	.94	.05	-.16	.24	.04	-.01	.99	.05
DET	.18	.20	.03	.16	.22	.01	-.14	.26	.01	-.06	.63	.01	.06	.64	.04	-.02	.81	.04
DENTR	.27	.15	.01	.02	.88	.05	.04	.75	.0	-.01	.94	.04	-.15	.25	.01	-.07	.44	.03
FREQ	-.10	.42	.05	-.02	.84	.04	.08	.54	.04	.02	.99	.05	.06	.62	.03	.13	.30	.01

Note. All aRQA variables represent averages; α=adjusted alpha.

Table 4: Pre-test scores vs aRQA metrics of basic emotions during confusion; significant correlations given in black, bold font.

	PRE-TEST SCORES																	
	Anger			Sad			Disgust			Joy			Surprise			Fear		
	r	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*
RR	-.09	.5	.03	-.18	.19	.03	-.05	.69	.04	-.08	.49	.01	.1	.45	.03	-.05	.56	.03
DET	-.06	.69	.05	-.16	.22	.04	-.04	.74	.05	-.05	.65	.04	.15	.25	.01	-.04	.66	.05
DENTR	-.07	.62	.04	-.1	.46	.05	.12	.31	.03	.02	.84	.05	.01	.94	.05	.04	.62	.04
FREQ	.09	.48	.01	**.22**	**.07**	**.01**	-.13	.28	.01	-.06	.6	.03	-.08	.53	.04	-.11	.38	.01

Note. All aRQA variables represent averages; α=adjusted alpha.

Table 5: MdRQA metrics of basic emotions during confusion and outcomes; significant correlations given in black, bold font.

	LEARNING GAINS			POST-TEST SCORES			PRE-TEST SCORES		
	r	*p*	*α*	*r*	*p*	*α*	*r*	*p*	*α*
RR	-.02	.85	.05	-.01	.96	.05	-.1	.41	.01
DET	-.08	.52	.04	-.01	.91	.04	-.07	.54	.03
DENTR	-.1	.41	.01	-.12	.34	.03	-.02	.85	.05
VENTR	-.08	.51	.03	-.14	.24	.01	-.06	.63	.04

Note. α=adjusted alpha.

33% of the time respectively), compared to anger, joy, sadness, and fear (> 6% of the time). In addition, on average, there was a high recurrence rate of facial expressions of disgust and surprise during frustration, with similar regularity compared to the other basic emotions. In contrast, while the frequency of joy, anger, sadness, and fear was very low, there were relatively high recurring patterns during frustration, with slightly more irregularity during learning. Again, these results partially supported our hypothesis, where we did not find evidence that there was a high frequency of sadness, as found by [47]. In addition, we did not find that anger was associated with frustration, as suggested by the Frustration-aggression hypothesis [3, 10], which may indicate that learners did not experience conditioned frustration. Moreover, the findings partially

Table 6: Learning gains vs aRQA metrics of basic emotions during frustration; significant correlations given in black, bold font.

LEARNING GAINS																		
	Anger			Sad			Disgust			Joy			Surprise			Fear		
	r	p	α	r	p	α	r	p	α	r	p	α	r	p	α	r	p	α
RR	-.20	.41	.04	-.23	.29	.03	-.2	.22	.04	-.09	.58	.03	-.24	.38	.01	-.42	.2	.03
DET	-.21	.38	.03	-.12	.58	.05	**-.33**	**.04**	**.01**	-.05	.78	.05	-.18	.49	.04	-.35	.28	.04
DENTR	.1	.69	.05	-.34	.11	.01	.1	.54	.05	.1	.55	.01	-.1	.71	.05	-.43	.19	.01
FREQ	-.22	.17	.01	-.14	.4	.04	**.3**	**.06**	**.03**	.08	.61	.04	.12	.46	.03	.04	.78	.05

Note. All aRQA variables represent averages; α=adjusted alpha.

Table 7: Post-test scores vs. aRQA metrics of basic emotions during frustration; significant correlations given in black, bold font.

POST-TEST SCORES																		
	Anger			Sad			Disgust			Joy			Surprise			Fear		
	r	p	α	r	p	α	r	p	α	r	p	α	r	p	α	r	p	α
RR	-.08	.75	.05	-.17	.45	.03	-.06	.73	.03	-.03	.86	.04	-.08	.76	.04	-.34	.3	.01
DET	-.11	.67	.03	-.07	.74	.04	-.19	.25	.01	.01	.96	.05	-.11	.67	.03	-.24	.48	.04
DENTR	.09	.73	.04	**-.44**	**.04**	**.01**	.01	.95	.05	.04	.82	.03	.01	.98	.05	-.28	.4	.03
FREQ	-.24	.13	.01	.01	.97	.05	.01	.94	.04	.23	.15	.01	.07	.66	.01	.06	.72	.05

Note. All aRQA variables represent averages; α=adjusted alpha.

Table 8: Pretest scores vs. aRQA metrics of basic emotions during frustration; significant correlations given in black, bold font.

PRE-TEST SCORES																		
	Anger			Sad			Disgust			Joy			Surprise			Fear		
	r	p	α	r	p	α	r	p	α	r	p	α	r	p	α	r	p	α
RR	.02	.99	.05	.06	.78	.03	**.28**	**.08**	**.04**	.14	.42	.03	.26	.34	.04	-.09	.8	.03
DET	.21	.39	.03	.04	.84	.04	**.38**	**.02**	**.03**	.08	.64	.04	.00	.1	.01	.02	.96	.05
DENTR	-.37	.12	.01	-.01	.96	.05	-.01	.97	.05	.00	.1	.01	.35	.19	.03	.19	.57	.01
FREQ	.08	.62	.04	.24	.13	.01	**-.46**	**.002**	**.01**	.06	.70	.05	-.08	.63	.05	.03	.86	.04

Note. All aRQA variables represent averages; α=adjusted alpha.

Table 9: MdRQA metrics of basic emotions during frustration and outcomes; significant correlations given in black, bold font.

	LEARNING GAINS			POST-TEST SCORES			PRE-TEST SCORES		
	r	p	α	r	p	α	r	p	α
RR	-.1	.55	.03	-.09	.58	.03	.11	.48	.03
DET	-.19	.25	.01	-.12	.46	.01	.13	.42	.01
DENTR	-.04	.81	.04	.06	.69	.05	.03	.84	.05
VENTR	-.02	.92	.05	.08	.6	.04	.04	.82	.04

Note. α=adjusted alpha.

matched [47], in that there was a high frequency of disgusted, yet low frequency of fearful facial expressions related to frustration.

A possible explanation for the results could be due to learners' individual differences, such as differences in emotion-regulation skills, prior experiences, intrinsic motivation, problem-solving skills, etc. In addition, some learners may have suppressed their facial expressions during learning, as other research studies find that adolescents tend to suppress negative emotions as a maladaptive emotion-regulation strategy during learning [24, 25].

In **RQ2**, we examined the relationship between facial expressions of basic emotions during confusion and frustration with outcome measures (learning gain, pre, post; RQ2). Notably, we found that while sadness rarely occurred during confusion, when sadness recurred regularly during confusion (DET), it was beneficial to learning gains. A similar correlation was found between the frequency of sadness during confusion and pre-test scores. In other words, learners with low prior knowledge, more frequently facially expressed sadness during confused actions, and, on average, regularly recurring patterns of this co-occurrence across the learning task, was associated with higher learning gains of the material. Thus, a "productive" type of confusion [39] may involve some degree of more frequent and regularity recurring (DET) facial expressions of sadness during confused actions; however, more research is needed.

For frustration, the regularity of recurring patterns (DET) of disgusted facial expressions was associated with lower learning gains, even though the overall frequency of disgusted facial expressions during frustration was associated with higher learning gains. In other words, more bursts of disgust during frustration may benefit learning, but when disgusted facial expressions are regularly expressed during frustration, it was harmful to learning. Possibly, what we are seeing is that learners may need more time to recover from whatever is causing recurring patterns of disgust to emerge with their frustration, since these results suggest that longer breaks in between those co-occurrences may have a beneficial impact on the learning process. In contrast, if disgust and frustration are regularly co-occurring repeatedly across the learning task without effective resolution, it may indicate an ineffective problem solving or emotion regulation strategy that is harmful to information processing and learning. Notably, learners tended to score lower on the post-test assessment, the more irregular their facial expressions of sadness (DENTR) repeatedly co-occurred with frustration.

Frustration's relationship with disgust also showed significant correlations with prior knowledge. Two metrics exploring the emergence of disgust during frustration—both RR and DET—were correlated with high prior knowledge, while a third (FREQ) was correlated with low prior knowledge. This indicates that regular, recurring bouts of disgust during frustration was associated more prior knowledge, whereas a higher occurrence of disgust during frustration indicated less prior knowledge. Future research should examine the degree to which learners' expectations may be influencing the regularity and recurrence of affective patterns with digital learning environments. For example, it is possible learners with high prior knowledge exhibit stronger emotional control because they expect to perform better even if disgust sometimes still emerges during frustration. However, given that frustration's relationship with the occurrence of disgust was also associated with higher learning gains, we might not necessarily think that suppressing the emergence of disgust is all that important. Instead, given that regularly recurring patterns of disgust during frustration was associated with less learning, looking to how learners are able to regulate their emotions during learning may be fruitful direction forward.

Regardless, this may be indicative of different "types" of frustration based on their relationship to outcome measures [3].

5.1 Threats to Validity

Demographic data on age, gender, and race were not collected, which may limit our understanding regarding the diversity of the population under study. In addition, to align the facial expressions of basic emotions with the interaction-based detectors for confusion and frustration required processing the continuous time series logs into 20 second intervals, based on the initial use of BROMP to develop the interaction-based detectors. These alignments resulted in a low number of recurrences for the continuous facial emotions data (i.e., 30 Hz to 1 second to every 20 seconds), and while these time series yield sufficient recurrences to calculate reliable values for the recurrence measures (>20 data points), the result reflects a weighted average of the continuous dynamics within the 20 second intervals. Moreover, the confusion and frustration detectors were far from perfect, with an AUC ROC of .56 and .63 respectively [29, 58]. Thus, it is possible that some confusion and frustration may not have been accurately detected. It is also important to note that facial recognition algorithms were originally primarily trained using data sets of adult, white male faces; thus, this characteristic of the training data could potentially introduce some measurement errors in facial expression analysis.

5.2 Future Directions and Implications

Future research should investigate additional data channels that may collect other affective processes that underlie the subsystems of an emotion. For example, physiological signals, e.g., heart rate, may further distinguish between different types of confusion and frustration. This also requires building multidimensional emotional labels like "regularly-recurring-disgusted-frustration" to offer a more nuanced understanding of learners' affective states.

Moreover, utilizing mixed methods, such as data-driven interviews, can uncover valuable qualitative insights into learners' emotional experiences, shedding light on hidden emotional complexities and cultural influences to complement quantitative emotion labels [5]. Investigating affective nuances within different cultural contexts, e.g., Western and Eastern cultures, who may interpret similar emotional states differently, is vital for recognizing emotional variability and designing culturally sensitive digital learning environments.

The implications of this research provide a deeper understanding on the different types of confusion and frustration that learners face with digital learning environments. The ability to detect and distinguish between types of confusion and frustration opens opportunities for personalized feedback. A better understanding of affect can lead to better affective support, eventually enabling educators to provide targeted guidance to learners based on their emotional responses, helping them overcome obstacles to improve their learning outcomes.

ACKNOWLEDGMENTS

This research was supported by NSF ECR Award (1561676). Any conclusions expressed in this material do not necessarily reflect the views of NSF. We would also like to thank other project members for their ideas and contributions. ChatGPT 4 was also used to assist in editing the text and building scripts in RStudio.

REFERENCES

[1] M. J. Amon, H. Vrzakova, and S. K. D'Mello. 2019. Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive Science* 43, 10 (2019), 1–22.

[2] A. Amsel. 1958. The Role of Frustrative Nonreward in Noncontinuous Reward Situations. *Psychological Bulletin* 55 (1958), 102–119.
[3] A. Amsel. 1992. *Frustration theory: An analysis of dispositional learning and memory.* Number 11. Cambridge University Press.
[4] J. M. A. L. Andres, J. Ocumpaugh, R. S. Baker, S. Slater, L. Paquette, Y. Jiang, and G. Biswas. 2019. Affect sequences and learning in Betty's Brain. In *LAK'19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge.* 383–390.
[5] Ryan S Baker, Nidhi Nasiar, Jaclyn L Ocumpaugh, Stephen Hutt, Juliana MAL Andres, Stefan Slater, Matthew Schofield, Allison Moore, Luc Paquette, Anabil Munshi, et al. 2021. Affect-targeted interviews for understanding student frustration. In *International Conference on Artificial Intelligence in Education.* Springer, 52–63.
[6] Ryan S.J.d Baker and Jaclyn Ocumpaugh. 2014. Interaction-Based Affect Detection in Educational Software. In *The Oxford handbook of affective computing*, D'Mello S. Gratch Jonathan Matthew Kappas Arvid Calvo, Rafael A. (Ed.). Oxford University Press, 233–245.
[7] L. F. Barrett. 2006. Emotions as natural kinds? *Perspectives on Psychological Science* 1 (2006), 28–58.
[8] D. E. Berlyne. 1960. *Conflict, arousal, and curiosity.* McGraw-Hill Book Company. https://doi.org/10.1037/11164-000
[9] G. Biswas, J. R. Segedy, and K. Bunchongchit. 2016. From design to implementation to practice a learning by teaching system: Betty's brain. *International Journal of Artificial Intelligence in Education* 26 (2016), 350–364.
[10] Judson S Brown and Isabel E Farber. 1951. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychological bulletin* 48, 6 (1951), 465.
[11] E. B. Cloude, D. A. Dever, D. L Hahs-Vaughn, A. J. Emerson, R. Azevedo, and J. Lester. 2022. Affective dynamics and cognition during game-based learning. *IEEE Transactions on Affective Computing* 13, 4 (2022), 1705–1717.
[12] E. B. Cloude, F. Wortha, D. Dever, and R. Azevedo. 2020. How do emotions change during learning with an intelligent tutoring system? Metacognitive monitoring and performance with MetaTutor. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.* 423–429.
[13] M. I. Coco, D. Mønster, G. Leonardi, R. Dale, and S. Wallot. 2021. Unidimensional and multidimensional methods for recurrence quantification analysis with crqa. *The R Journal* 13 (2021), 14.
[14] J. A. DeFalco, J. P. Rowe, L. Paquette, V. Georgoulas-Sherry, K. Brawner, B. W. Mott, and J. C. Lester. 2018. Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education* 28 (2018), 152–193.
[15] D. A. Dever, M. J. Amon, H. Vrzakova, M. D. Wiedbusch, E. B. Cloude, and R. Azevedo. 2022. Capturing sequences of learners' self-regulatory interactions with instructional material during game-based learning using auto-recurrence quantification analysis. *Frontiers in Psychology* 13 (2022), 813677.
[16] I. Di Leo, K. R. Muis, C. A. Singh, and C. Psaradellis. 2019. Curiosity… Confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology* 58 (2019), 121–137.
[17] J. Dollard, L. W. Doob, N. E. Miller, O. H. Mowrer, R. R. Sears, C. S. Ford, C. I. Hovland, and R. T. Sollenberger. 1939. *Frustration and aggression.* Yale University Press.
[18] S. D'Mello and A. Graesser. 2012. Dynamics of affective states during complex learning. *Learning and Instruction* 22 (2012), 145–157.
[19] S. D'Mello, B. Lehman, R. Pekrun, and A. Graesser. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29 (2014), 153–170.
[20] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.
[21] Paul Ekman. 1994. Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. (1994).
[22] Phoebe C Ellsworth. 2003. Confusion, concentration, and other emotions of interest: Commentary on Rozin and Cohen (2003). (2003).
[23] A. C Graesser. 2020. Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction* 70 (2020), 101212.
[24] James J Gross. 2015. Emotion regulation: Current status and future prospects. *Psychological inquiry* 26, 1 (2015), 1–26.
[25] J. T Gross and J. Cassidy. 2019. Expressive suppression of negative emotions in children and adolescents: Theory, data, and a guide for future research. *Developmental Psychology* 55, 9 (2019), 1938.
[26] J. M. Harley, F. Bouchet, M. S. Hussain, R. Azevedo, and R. Calvo. 2015. A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior* 48 (2015), 615–625.
[27] Clark Leonard Hull. 1943. *Principles of behavior: an introduction to behavior theory.* Appleton-Century.
[28] iMotions. 2016. *Facial expression analysis: The definitive guide.* https://imotions.com/facialexpression-guide-ebook/
[29] Y. Jiang, N. Bosch, R. S. Baker, L. Paquette, J. Ocumpaugh, J. M. A. L. Andres, and G. Biswas. 2018. Expert feature-engineering vs. deep neural networks: which is better for sensor-free affect detection?. In *Proceedings of the 19th International Conference on Artificial Intelligence in Education.* 198–211. https://doi.org/10.1007/978-3-319-93843-1_15
[30] Shamya Karumbaiah, Ryan Baker, Yan Tao, and Ziyang Liu. 2022. How does students' affect in virtual learning relate to their outcomes? A systematic review challenging the positive-negative dichotomy. In *LAK22: 12th International Learning Analytics and Knowledge Conference.* 24–33.
[31] S. Karumbaiah, R. S. Baker, J. Ocumpaugh, and A. Andres. 2021. A re-analysis and synthesis of data on affect dynamics in learning. *IEEE Transactions on Affective Computing* (2021).
[32] P. R. Kleinginna Jr and A. M. Kleinginna. 1981. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion* 5, 4 (1981), 345–379.
[33] S. Land. 2000. Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development* 48 (2000), 61–78.
[34] K. Leelawong and G. Biswas. 2008. Designing learning by teaching agents: The betty's brain system. *International Journal of Artificial Intelligence in Education* 18, 3 (2008), 181–208.
[35] 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.
[36] Blair Lehman, Sidney D'Mello, and Arthur Graesser. 2013. Who Benefits from Confusion Induction During Learning? An Individual Differences Cluster Analysis. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, Kalina Yacef, H. Chad Lane, Jack Mostow, and Philip Pavlik (Eds.). Springer, Berlin, 51–60.
[37] Aaron D Likens, Kathryn S McCarthy, Laura K Allen, and Danielle S McNamara. 2018. Recurrence quantification analysis as a method for studying text comprehension dynamics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge.* 111–120.
[38] Zhongxiu Liu, Visit Pataranutaporn, Jaclyn Ocumpaugh, and Ryan Baker. 2013. Sequences of frustration and confusion, and learning. In *Educational data mining 2013.*
[39] J. M. Lodge, G. Kennedy, L. Lockyer, A. Arguel, and M. Pachman. 2018. Understanding difficulties and resulting confusion in learning: An integrative review. In *Frontiers in Education*, Vol. 3. 49.
[40] G. Mandler. 1976. *Mind and Emotion.* Wiley, New York.
[41] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. 2007. Recurrence plots for the analysis of complex systems. *Physics reports* 438, 5-6 (2007), 237–329.
[42] J. D. Marx and K. Cummings. 2007. Normalized change. *American Journal of Physics* 75, 1 (2007), 87–91.
[43] D. McDuff, M. Amr, A. Mahmoud, J. Turcot, M. Mavadati, and R. el Kaliouby. 2016. AFFDEX SDK: A Cross-platform Real-time Multi-face Expression Recognition Toolkit. In *CHI 2016 Extended Abstracts.* San Jose, CA, USA.
[44] D. McDuff, El K. R., K. Kassam, and R. Picard. 2010. Affect valence inference from facial action unit spectrograms. In *Proceedings from 2010 I.E. Computer Society Conference on Computer Vision and Pattern Recognition — Workshops.* 17–24.
[45] R. Moulder, B. Booth, A. Abitino, and S. D'Mello. 2023. Recurrence quantification analysis of eye gaze dynamics during team collaboration. In *Proceedings from the 13th International Learning Analytics and Knowledge Conference.* Solar, 430–440.
[46] A. Munshi. 2023. *An Adaptive Scaffolding Framework for Self-Regulated Learning in an Open-Ended Learning Environment.* Ph. D. Dissertation. PhD Dissertation. Vanderbilt University.
[47] A. Munshi, S. Mishra, N. Zhang, L. Paquette, J. Ocumpaugh, R. S. Baker, and G. Biswas. 2020. Modeling the relationships between basic and achievement emotions in computer-based learning environments. In *21st International Conference of Artificial Intelligence in Education.* Springer International Publishing, 411–422.
[48] A. Munshi, R. Rajendran, J. Ocumpaugh, G. Biswas, R. S. Baker, and L. Paquette. 2018. Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization.* 131–138.
[49] J. Ocumpaugh, R. S. Baker, and M. M. T Rodrigo. 2015. *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual.* Technical Report. Teachers College, Columbia University, Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
[50] R. Pekrun and E. J. Stephens. 2011. Academic Emotions. In *APA Educational Psychology Handbook: Individual Differences and Cultural and Contextual Factors*, K. R. Harris, S. Graham, T. Urdan, S. Graham Royer, and M. Zeidner (Eds.). Vol. 2. American Psychological Association, 3–31.
[51] J. Piaget. 1952. *The Origins of Intelligence.* International University Press, New York.
[52] Recurrence Plots and Cross Recurrence Plots. [n. d.]. *Recurrence Plots At A Glance.* http://www.recurrence-plot.tk/ Accessed on: 01/10/2023.
[53] J. E. Richey, J. M. L. Andres-Bray, M. Mogessie, R. Scruggs, J. M. A. L. Andres, J. R. Star, R. S. Baker, and B. M. McLaren. 2019. More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education* 139 (2019), 173–190.

[54] J. A. Russell. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110, 1 (2003), 145.

[55] K. R Scherer. 1987. Toward a dynamic theory of emotion: The component process model of affective states. *Geneva studies in Emotion and Communication* 1 (1987), 1–98.

[56] P. J. Silvia. 2010. Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts* 4, 2 (2010), 75.

[57] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson. 2018. Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods* 50 (2018), 1446–1460.

[58] S. Wallot, A. Roepstorff, and D. Mønster. 2016. Multidimensional Recurrence Quantification Analysis (MdRQA) for the analysis of multidimensional time-series: A software implementation in MATLAB and its application to group-level data in joint action. *Frontiers in psychology* 7 (2016), 1835.

[59] C. L. Webber Jr and J. P. Zbilut. 2005. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences* 94 (2005), 26–94.