Affect Sequences and Learning in Betty’s Brain

Juliana Ma. Alexandra L. Andres
University of Pennsylvania
3700 Walnut Street
Philadelphia, PA 19104
aandres@gse.upenn.edu

Jaclyn Ocumpaugh
University of Pennsylvania
3700 Walnut Street
Philadelphia, PA 19104
ojaclyn@upenn.edu

Ryan S. Baker
University of Pennsylvania
3700 Walnut Street
Philadelphia, PA 19104
rybaker@upenn.edu

Stefan Slater
University of Pennsylvania
3700 Walnut Street
Philadelphia, PA 19104
slater.research@gmail.com

Luc Paquette
University of Illinois at Urbana-Champaign
1310 S. Sixth St.
Champaign, IL 61820
lpaq@illinois.edu

Yang Jiang
Columbia University
116th St & Broadway
New York, NY 10027
yj2211@tc.columbia.edu

Shamya Karumbaiah
University of Pennsylvania
3700 Walnut Street
Philadelphia, PA 19104
shamya@gse.upenn.edu

Nigel Bosch
University of Illinois at Urbana-Champaign
1310 S. Sixth St.
Champaign, IL 61820
pnb@illinois.edu

Anabil Munshi
Vanderbilt University
2201 West End Ave,
Nashville, TN 37235
anabil.munshi@vanderbilt.edu

Allison Moore
Vanderbilt University
2201 West End Ave,
Nashville, TN 37235
allison.l.moore@vanderbilt.edu

Gautam Biswas
Vanderbilt University
2201 West End Ave,
Nashville, TN 37235
gautam.biswas@vanderbilt.edu

ABSTRACT

Education research has explored the role of students’ affective states in learning, but some evidence suggests that existing models may not fully capture the meaning or frequency of how students transition between different states. In this study we examine the patterns of educationally-relevant affective states within the context of Betty’s Brain, an open-ended, computer-based learning system used to teach complex scientific processes. We examine three types of affective transitions based on similarity with the theorized D’Mello and Graesser model, transition between two affective states, and the sustained instances of certain states. We correlate the frequency of these patterns with learning outcomes and our findings suggest that boredom is a powerful indicator of students’ knowledge, but not necessarily indicative of learning. We discuss our findings within the context of both research and theory on affect dynamics and the implications for pedagogical and system design.

CCS CONCEPTS
• Applied computing—Computer-managed instruction
• Applied computing—Computer-assisted instruction

KEYWORDS
Affect dynamics, affect, learning analytics

1 INTRODUCTION

The affective processes associated with learning are increasingly of interest to researchers who are trying to understand how students regulate their learning processes [7, 11, 34]. Many studies have examined how students transition between affective states while engaged in learning activities [1, 11, 23, 24, 27, 29, 32, 33], including several that compare these patterns to performance outcomes and learning gains [5, 11, 12, 14, 22]. Most studies have
focused on affective states that are thought to be educationally
relevant [cf. 7, 30] including boredom (BOR), confusion (CON),
engaged concentration (ENG; sometimes labeled flow), delight
(DEL), frustration (FRU), and surprise.

One of the more influential papers in this field, D’Mello and
Graesser [11], hypothesized two main pathways that students
follow when transitioning between affective states: one that
encourages learning (the ENG-CON-DEL-ENG cycle) and one that
inhibits it (the ENG-CON-FRU-BOR cycle). However, based on
analyses of their data (calculating the likelihood of a transition
from one affective state to another), they then proposed a more
succinct model.

Since this pioneering work, a substantial body of research has
investigated these issues, many following the same
methodological approach that analyzes the frequency of two-step
transitions, (e.g., from confusion to frustration; denoted as CON-
FRU) rather than patterns involving multiple transitions (e.g., the
ENG-CON-FRU-BOR cycle).

The current study uses sensor-free, interaction-based
detectors (e.g., [2]) previously developed for the Betty’s Brain [18]
to explore the relationship between affective patterns and
learning outcomes (measured using pre- and post-tests of content
knowledge). Specifically, we compare the four-step patterns
originally proposed in D’Mello and Graesser [11] to shorter
patterns involving three steps, but only two affective states (e.g.,
BOR-BOR-CON). We then compare these results to patterns
involving off-task behavior as well as to sustained instances of
two affective states (boredom and delight) that appear repeatedly
in statistically significant multi-state patterns. These results
suggest that affective patterns in Betty’s Brain may more strongly
reflect prior knowledge than learning, which has implications for
the design of affective-driven interventions.

2 PREVIOUS RESEARCH

D’Mello and Graesser [11] investigated theoretically-
motivated affective cycles by studying the two-step transitions
within them (i.e., ENG-CON, CON-DEL, DEL-ENG, CON-FRU, and
FRU-BOR). After synthesizing this data, their revised theoretical
model (shown in Figure 1) no longer included delight, but instead
focused on transitions between engaged concentration, confusion,
frustration, and boredom.

![Figure 1: D'Mello & Graesser's [11] model of affect dynamics](image)

Other studies examine patterns of longer lengths. In Cognitive
Tutor Algebra, Liu et al. [22] explored three-step affect transitions
involving the presence or absence of confusion and frustration
(e.g., CON-Not-CON or FRU-FRU-Not). In this study, patterns
where all three steps were labeled with one of these affective
states (e.g., CON-CON-CON or FRU-FRU-FRU) were negatively
correlated with learning, but so was the absence of either emotion
(i.e., Not-Not-Not).

More recent work in affect dynamics [19] has looked at how
the operationalization of these measurements may influence
which transitions emerge as statistically significant, focusing on
the implementation of D’Mello’s L—the metric introduced
by D’Mello and colleagues [9, 11], to calculate the likelihood that a
transition from one affective state to the next will occur, given
the base rates observed for each. Previous analyses showed that the
exclusion of self-transitions (e.g., CON-CON) when calculating
D’Mello’s L can inflate the number of statistically significant
transitions. However, even when this adjustment was made, not
all hypothesized patterns were seen, either in Karumbiah et al.
[19] or in D’Mello and Graesser’s original paper [9].

Moreover, the use of D’Mello’s L, as currently defined, limits
research to investigations of two-step patterns. Using an
expanded list of affective patterns and longer sequences allows for
the discovery of meaningful affective patterns while
accounting for the varying lengths of emergence of different
states.

3 Betty’s Brain

Betty’s Brain is an open-ended, computer-based learning system
that uses a learning-by-teaching paradigm to teach complex
scientific processes [21]. Betty’s Brain asks students to teach a
virtual agent (Betty) about scientific phenomena (e.g., climate
change, ecosystems, thermoregulation) by constructing concept
maps that demonstrate the causal relationships involved (see
Figure 2.)

The learning process required by Betty’s Brain necessitates
high levels of self-regulation. As students construct their map,
they must navigate through multiple hypermedia information
sources where they can read about a variety of subthemes. They
choose how often to test Betty’s knowledge, and they may elect
to interact with a virtual mentor agent (an experienced teacher
named Mr. Davis) if they are having trouble teaching Betty.
Because of these design factors, strong self-regulated learning
behaviors are crucial for succeeding within Betty’s Brain.
4 Methods

4.1 Study Design

This study uses data from 93 sixth graders who used Betty’s Brain during the 2016-2017 school year during their science classes in an urban public school in Tennessee. Data were collected over the course of seven school days. On the first day of the study, students completed a 30-45-minute paper-based pre-test that measured knowledge of scientific concepts and causal relationships. On day 2, students participated in a 30-minute training session that familiarized them with the learning goals and user interface of the software. Following the pre-test and training, students used the Betty’s Brain software on days 2 through 6, for approximately 45-50 minutes each session, using concept maps to teach Betty about the causal relationships involved in the process of climate change. On day 7, students completed a post-test that was identical to the pre-test, in order to assess changes in knowledge based on working with Betty’s Brain for the week.

4.2 BROMP Observations & Detectors

Students’ affect was assessed using a model-based approach that employed the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) [26]. These were developed using established methods and that have been previously published.

4.2.1 BROMP Observations. The Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) is a method for conducting Quantitative Field Observations (QFOs) of student behavior and affect [26]. BROMP is primarily used to develop sensor-free detectors [3], but has been used for other purposes as well (e.g., [2]).

Within BROMP, students are observed using a momentary time sampling method [25], where students are repeatedly coded individually, in a predetermined order. This sampling method is designed to achieve a representative sample of behavior and affect among the students while reducing the tendency to focus on more extreme events. Both affect and behavior are recorded separately, but simultaneously, in each observation. Typical coding categories include boredom, confusion, engaged concentration, delight, and frustration (for affective states) and on task, on task-conversation, and off-task (for behaviors).

4.2.2 BROMP-based Detectors. BROMP-based detectors have now been developed for over a dozen intelligent tutoring systems. In this process, codes generated by BROMP-certified observers are matched to student log files, and then machine-learning techniques are used to generate classifiers for each affective state or behavioral category based on the students’ interactions with the log files (e.g., [2]).

The detectors used in this study were previously developed for Betttys Brain by Jiang et al. [18]. These detectors, generated using either logistic or step regression, were evaluated using A’. Their performance under 10-fold student-level cross-validation using AUC ROC (summarized in Table 1).

<table>
<thead>
<tr>
<th>Affect/Behavior</th>
<th>AUC ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boredom</td>
<td>0.682</td>
</tr>
<tr>
<td>Confusion</td>
<td>0.568</td>
</tr>
<tr>
<td>Delight</td>
<td>0.570</td>
</tr>
<tr>
<td>Engaged Concentration</td>
<td>0.624</td>
</tr>
<tr>
<td>Frustration</td>
<td>0.634</td>
</tr>
<tr>
<td>Off-Task Behavior</td>
<td>0.725</td>
</tr>
</tbody>
</table>

For the purposes of this study, which focuses on affective transitions, only one behavior detector (off-task) is applied to the data. It is included so that we can compare negative effects of boredom to this detector for additional validity in our findings.

4.2.2 Field Observations vs. Detectors. Detector performance is generally not as high as interrater reliability between two trained human coders. However, it does offer a significant advantage in terms of grain-size and scale. That is, with BROMP-based detectors, model labels are generated every 20 seconds for each detector. In contrast, BROMP field observers use a momentary time sampling procedure rather than observing the same student rapidly in succession. Thus, the time between field observations of a single student can sometimes be as high as several minutes, depending on how many students are being observed and what is happening in the classroom.
4.3 Prevalence of Affective Patterns

The frequency of specific affective and behavioral patterns was calculated for each student. This process required several steps.

4.3.1 Applying Detectors. First, the BROMP-based detectors were applied to the log data for every student in the study. Each of the automated detectors was applied at the level of 20-second clips in the log data, producing a confidence value for each clip. Clips with a confidence value of at least 50% were labelled with the corresponding affective state. In the case where multiple detectors provided positive labels to a single log-file clip, all labels were applied, and multiple patterns were considered.

Of the 53,087 clips in this study, 17,585 were labeled as bored (33%), 8,683 were labeled as confused (16%), 34,934 were labeled as engaged concentration (66%), 25,069 were labeled as delighted (47%), 6,200 were labeled as frustrated (12%), and 2,335 received no affective labels (4%). Predictions for off-task behavior were made using the same method, but separately from affective state labels, resulting in 22,537 clips labeled as off-task (42%). After all six detectors were applied, each 20-second clip contained a prediction from each detector (e.g., bored or not; confused or not; off-task or not; etc.); as such, the sum of the labels across all six detectors is greater than the total number of clips.

4.3.2 Pattern Types. In our analyses we focus on three types of affect pattern. Each involved a sequence of either three or four 20-second log-file clips. First, we looked at sequences that mirror the two cycles outlined by D’Mello & Graesser [11]: the ENG-COND-ENG cycle and the ENG-COND-FRU-BOR cycle. For the purposes of this study, we have limited the analysis to 80 second (four-clip) versions of these cycles.

Next, we considered transitions between two states. For these analyses, we looked for a student having at least two consecutive clips with the same affective state predictions, before then transitioning to a second state (e.g., ENG-ENG-BOR or CON-COND-FRU). These durations allow us to explore the potential effect that a longer duration (two or more steps) of any given antecedent might have on the subsequent steps in a sequence. Thus, we are able to explore the possibility that affective states of a longer duration (more than one successive step) might be influencing the results seen for sequences involving multiple transitions without testing all possible durations.

Finally, we consider sustained instances of two affective states that seemed to be driving the other patterns of statistical significance in this study. These are operationalized as 4-clip sequences (BOR-BOR-BOR-BOR and DEL-DEL-DEL-DEL), which we compare to sustained off-task behavior (OFF-OFF-OFF-OFF).

4.3.3 Prevalence Calculations. Once the data from Betty’s Brain was fully labeled, we calculated the prevalence of each affective pattern within each student’s log files. Because we are interested in multi-step transitions, we use a prevalence calculation (as opposed to D’Mello’s L, which is more common in the affect dynamics literature, but which is designed to study only 2-clip patterns).

In this study, prevalence is the total number of times a pattern occurred within a given student’s data (Oact) divided by the total number of times it could have occurred in that data (Opos). Because we consider the possibility of overlapping sequences, we calculate Opos by considering the total number of clips in a pattern (Cpat) and the total number of clips in the student’s data (Cstu).

\[
\text{prevalence} = \frac{\text{Oact}}{\text{Opos}} = \frac{\text{Oact}}{\text{Cstu} - (\text{Cpat} - 1)}
\]

\text{Equation 1.}

For example, consider the calculation of Sustained BOR (operationalized as BOR-BOR-BOR-BOR) for a student whose total data consisted of the following 7-clip sequence: FRU-BOR-BOR-BOR-BOR-CON. For this student, Oact = 4, since Sustained BOR is observed two times, from the 2nd to the 5th clip and from the 3rd to the 6th clip. Meanwhile, Cstu = 7 (the total number of clips in the student’s data) and Cpat = 4 (since it is a 4-clip pattern), so that Opos = 3. Therefore, prevalence of Sustained BOR in this student’s data is 2/4 or 50%.

Were we interested in a 3-clip pattern like BOR-BOR-CON for the same students’ data, the calculation would be as follows:

\[
\frac{\text{Oact}}{\text{Cstu} - (\text{Cpat} - 1)} = \frac{1}{7 - (3 - 1)} = \frac{1}{5} = 20\%
\]

\text{Equation 2.}

Because it is possible for one clip to be labeled with more than one affect, it is possible to have multiple different affective patterns within the same number of clips. For instance, BOR-BOR-BOR-FRU would be counted as both Sustained BOR and the affective transition BOR-FRU.

4.4 Statistical Analysis

Spearman’s Rho was used to correlate the prevalence of each affective state pattern to pre-test, post-test, and normalized gain scores calculated as (post-pre)/(1-pre) [15]. While a gain score measures the difference between pre-test and post-test, the normalized gain score measures this difference relative to the maximum possible improvement given a student’s pre-test score. Spearman’s Rho is a non-parametric correlation coefficient, and it is commonly used in analyses where assumptions of normality are not met across an entire feature space.

Because correlating each affective pattern against each of our three outcome measures resulted in 81 separate statistical tests, Benjamini and Hochberg’s [4] post-hoc FDR correction was used to adjust the significance criterion across multiple tests. P-values in the results section are only marked as significant if they remained significant after this correction was applied.

5 RESULTS AND DISCUSSION

5.1 Pattern Frequencies

Affective patterns are described in Table 3, which summarizes the observation period and frequency for patterns involving boredom.
(BOR), confusion (CON), delight (DEL), engaged concentration (ENG), and frustration (FRU).

**Table 3: Affect pattern frequencies**

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Obs. Period (sec)</th>
<th>N</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG-CON-FRU-BOR</td>
<td>20-20-20-20</td>
<td>173</td>
<td>0.30%</td>
</tr>
<tr>
<td>ENG-CON-DEL-ENG</td>
<td>20-20-20-20</td>
<td>1,678</td>
<td>3.20%</td>
</tr>
<tr>
<td>ENG-BOR</td>
<td>40-20</td>
<td>7,964</td>
<td>15.00%</td>
</tr>
<tr>
<td>ENG-CON</td>
<td>40-20</td>
<td>3,774</td>
<td>7.11%</td>
</tr>
<tr>
<td>ENG-DEL</td>
<td>40-20</td>
<td>15,834</td>
<td>29.83%</td>
</tr>
<tr>
<td>ENG-FRU</td>
<td>40-20</td>
<td>1,972</td>
<td>3.71%</td>
</tr>
<tr>
<td>BOR-CON</td>
<td>40-20</td>
<td>1,871</td>
<td>3.52%</td>
</tr>
<tr>
<td>BOR-DEL</td>
<td>40-20</td>
<td>5,866</td>
<td>11.05%</td>
</tr>
<tr>
<td>BOR-ENG</td>
<td>40-20</td>
<td>7,789</td>
<td>14.67%</td>
</tr>
<tr>
<td>BOR-FRU</td>
<td>40-20</td>
<td>1,006</td>
<td>1.90%</td>
</tr>
<tr>
<td>CON-BOR</td>
<td>40-20</td>
<td>2,096</td>
<td>3.95%</td>
</tr>
<tr>
<td>CON-DEL</td>
<td>40-20</td>
<td>3,423</td>
<td>6.45%</td>
</tr>
<tr>
<td>CON-ENG</td>
<td>40-20</td>
<td>3,902</td>
<td>7.35%</td>
</tr>
<tr>
<td>CON-FRU</td>
<td>40-20</td>
<td>2,224</td>
<td>4.19%</td>
</tr>
<tr>
<td>DEL-BOR</td>
<td>40-20</td>
<td>5,755</td>
<td>10.84%</td>
</tr>
<tr>
<td>DEL-CON</td>
<td>40-20</td>
<td>3,094</td>
<td>5.83%</td>
</tr>
<tr>
<td>DEL-ENG</td>
<td>40-20</td>
<td>14,329</td>
<td>26.99%</td>
</tr>
<tr>
<td>DEL-FRU</td>
<td>40-20</td>
<td>1,318</td>
<td>2.48%</td>
</tr>
<tr>
<td>FRU-BOR</td>
<td>40-20</td>
<td>1,189</td>
<td>2.24%</td>
</tr>
<tr>
<td>FRU-CON</td>
<td>40-20</td>
<td>2,421</td>
<td>4.56%</td>
</tr>
<tr>
<td>FRU-DEL</td>
<td>40-20</td>
<td>1,736</td>
<td>3.27%</td>
</tr>
<tr>
<td>FRU-ENG</td>
<td>40-20</td>
<td>2,614</td>
<td>4.92%</td>
</tr>
<tr>
<td>Sustained BOR</td>
<td>80</td>
<td>11,759</td>
<td>22.15%</td>
</tr>
<tr>
<td>Sustained DEL</td>
<td>80</td>
<td>16,219</td>
<td>30.55%</td>
</tr>
<tr>
<td>Sustained OFF</td>
<td>80</td>
<td>10,155</td>
<td>19.13%</td>
</tr>
<tr>
<td>OFF-ENG</td>
<td>40-20</td>
<td>7,750</td>
<td>14.60%</td>
</tr>
</tbody>
</table>

The most frequently observed affective pattern in our data is sustained delight (33%), followed by ENG-DEL (29.83%), DEL-ENG (26.99%), Sustained BOR (22.15%), and Sustained OFF (19.3%). The two cycles hypothesized by D’Mello & Graesser [10] occur relatively infrequently (0.3% and 3.2%). In fact, all but three of the patterns occur more frequently than these two cycles: BOR-FRU (1.9%), DEL-FRU (2.48%), and FRU-BOR (2.24%). The relative infrequency of the hypothesized cycles, however, may be due to methodological differences between these studies, specifically our decision to look at 20-second intervals rather than to collapse all repeated instances of a given emotion into a single unit of observation.

It is worth noting, however, that these analyses are being conducted on indirect measures of affective states. While our detectors have undergone a careful cross-validation process, these results should be interpreted carefully. In particular, the detector for delight appears to over-predict the prevalence of this affective state [18]. However, as the results presented below suggest, the student actions associated with delight appear to be highly relevant to understanding knowledge and learning behaviors.

**5.2 Correlations with Pre-test Scores**

As Table 4 shows, 9 of the 26 affective patterns in this study were significantly negatively correlated with pre-test scores. These include the inhibitory cycle hypothesized by D’Mello & Graesser (ENG-CON-FRU-BOR, $p = -0.239$), though it was rare, accounting for only 0.3% of the 4-clip sequences in the data.

**Table 4: Correlations between affect patterns and pre-test scores**

<table>
<thead>
<tr>
<th>Pattern</th>
<th>$\rho$</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG-CON-FRU-BOR</td>
<td>-0.239*</td>
<td>0.026</td>
</tr>
<tr>
<td>ENG-CON-DEL-ENG</td>
<td>0.209</td>
<td>0.052</td>
</tr>
<tr>
<td>ENG-BOR</td>
<td>-0.434**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ENG-CON</td>
<td>-0.002</td>
<td>0.985</td>
</tr>
<tr>
<td>ENG-DEL</td>
<td>0.194</td>
<td>0.081</td>
</tr>
<tr>
<td>ENG-FRU</td>
<td>-0.011</td>
<td>0.921</td>
</tr>
<tr>
<td>BOR-CON</td>
<td>-0.239*</td>
<td>0.026</td>
</tr>
<tr>
<td>BOR-DEL</td>
<td>-0.336**</td>
<td>0.001</td>
</tr>
<tr>
<td>BOR-ENG</td>
<td>-0.449**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BOR-FRU</td>
<td>-0.262*</td>
<td>0.014</td>
</tr>
<tr>
<td>CON-BOR</td>
<td>-0.137</td>
<td>0.207</td>
</tr>
<tr>
<td>CON-DEL</td>
<td>0.081</td>
<td>0.458</td>
</tr>
<tr>
<td>CON-ENG</td>
<td>0.063</td>
<td>0.561</td>
</tr>
<tr>
<td>CON-FRU</td>
<td>-0.036</td>
<td>0.739</td>
</tr>
<tr>
<td>DEL-BOR</td>
<td>-0.106</td>
<td>0.345</td>
</tr>
<tr>
<td>DEL-CON</td>
<td>0.063</td>
<td>0.574</td>
</tr>
<tr>
<td>DEL-ENG</td>
<td>0.197</td>
<td>0.076</td>
</tr>
<tr>
<td>DEL-FRU</td>
<td>0.05</td>
<td>0.657</td>
</tr>
<tr>
<td>FRU-BOR</td>
<td>-0.17</td>
<td>0.116</td>
</tr>
<tr>
<td>FRU-CON</td>
<td>-0.045</td>
<td>0.687</td>
</tr>
<tr>
<td>FRU-DEL</td>
<td>-0.076</td>
<td>0.495</td>
</tr>
<tr>
<td>FRU-ENG</td>
<td>0.06</td>
<td>0.579</td>
</tr>
<tr>
<td>Sustained BOR</td>
<td>-0.450**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sustained DEL</td>
<td>0.195</td>
<td>0.07</td>
</tr>
<tr>
<td>Sustained OFF</td>
<td>-0.376**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>OFF-ENG</td>
<td>-0.220*</td>
<td>0.041</td>
</tr>
</tbody>
</table>
Like D’Mello and Graesser’s [11] inhibitory cycle (ENG-CON-FRU-BOR), six of the other significant patterns that are negatively correlated with pre-test scores also include boredom (ENG-BOR, BOR-CON, BOR-DEL, BOR-ENG, BOR-FRU, and Sustained BOR). This suggests that students who begin the activity with low knowledge are more likely to experience boredom. In fact, boredom patterns are even more likely (ρ = -.239 to -.450) among low-knowledge learners than off-task behaviors (Sustained OFF, ρ = -.376).

5.3 Correlations with Post-test Scores

Correlations between affective state transitions and post-test scores are presented in Table 5, which shows patterns similar to those found in the pre-test results. However, this time, the role of delight is more pronounced.

Table 5: Correlations between affect patterns and post-test scores

<table>
<thead>
<tr>
<th>Pattern</th>
<th>ρ</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG-CON-FRU-BOR</td>
<td>-.308**</td>
<td>0.004</td>
</tr>
<tr>
<td>ENG-CON-DEL-ENG</td>
<td>.214*</td>
<td>0.046</td>
</tr>
<tr>
<td>ENG-BOR</td>
<td>-.443**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ENG-CON</td>
<td>-.003</td>
<td>.976</td>
</tr>
<tr>
<td>ENG-DEL</td>
<td>0.19</td>
<td>.087</td>
</tr>
<tr>
<td>ENG-FRU</td>
<td>0.038</td>
<td>.736</td>
</tr>
<tr>
<td>BOR-CON</td>
<td>-.350**</td>
<td>0.001</td>
</tr>
<tr>
<td>BOR-DEL</td>
<td>-.389**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BOR-ENG</td>
<td>-.468**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BOR-FRU</td>
<td>-.294**</td>
<td>0.006</td>
</tr>
<tr>
<td>CON-BOR</td>
<td>-.224*</td>
<td>0.037</td>
</tr>
<tr>
<td>CON-DEL</td>
<td>0.082</td>
<td>.448</td>
</tr>
<tr>
<td>CON-ENG</td>
<td>0.073</td>
<td>.499</td>
</tr>
<tr>
<td>CON-FRU</td>
<td>-.039</td>
<td>.719</td>
</tr>
<tr>
<td>DEL-BOR</td>
<td>-.096</td>
<td>.393</td>
</tr>
<tr>
<td>DEL-CON</td>
<td>0</td>
<td>.999</td>
</tr>
<tr>
<td>DEL-ENG</td>
<td>0.18</td>
<td>.105</td>
</tr>
<tr>
<td>DEL-FRU</td>
<td>0.092</td>
<td>.411</td>
</tr>
<tr>
<td>FRU-BOR</td>
<td>-.191</td>
<td>.076</td>
</tr>
<tr>
<td>FRU-CON</td>
<td>-.031</td>
<td>.783</td>
</tr>
<tr>
<td>FRU-DEL</td>
<td>-.051</td>
<td>0.65</td>
</tr>
<tr>
<td>FRU-ENG</td>
<td>-.042</td>
<td>.698</td>
</tr>
<tr>
<td>Sustained BOR</td>
<td>-.505**</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sustained DEL</td>
<td>-.300**</td>
<td>0.005</td>
</tr>
<tr>
<td>Sustained OFF</td>
<td>-.512**</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Two of these are positive correlations for the patterns ENG-CON-DEL-ENG (ρ = .214) and Sustained DEL (ρ = .300). These positive correlations suggest that increases in the frequency of these patterns would relate to an increase in post-test scores as well. By contrast, ten patterns show negative correlations. Five of these patterns begin with boredom, two with off-task behavior (Sustained OFF, ρ = -.512; OFF-ENG, ρ = -.307), two with engaged concentration (ENG-CON-FRU-BOR, ρ = -.308; ENG-BOR, ρ = -.443), and one with confusion (CON-BOR, ρ = -.224). No significant correlations were found for 3-step patterns beginning with either delight or frustration.

Overall, 8 of the significant patterns involve boredom. These include several patterns that were significantly negatively correlated with pre-test scores, including D’Mello and Graesser’s negative cycle (ENG-CON-FRU-BOR, ρ = -.308) and the six other boredom-related patterns (ENG-BOR, BOR-CON, BOR-DEL, BOR-ENG, BOR-FRU, and Sustained BOR). In these cases, the strength of the correlations have all increased. The same is true for CON-BOR, which was not significantly correlated with pre-test scores, but which is significantly correlated with post-test scores (ρ = -.224).

5.4 Correlations with Learning Gains

Table 6 summarizes the correlations between affect patterns and learning gains. In contrast with pre-and post-test scores, only one significant correlation was found, suggesting that in this case affective states may be more closely tied to student knowledge than to student learning.

Curiously, the only pattern associated with higher learning gains is Sustained DEL (DEL-DEL-DEL-DEL, ρ = .335), which was a pattern that was significantly correlated with higher post-test scores but not with prior knowledge. This result corresponds to the finding that the ENG-CON-DEL-ENG cycle was also correlated with post-test, but not pre-test scores; this cycle was not correlated significantly to learning gain, although there was a trend in that direction that may be worth further investigation.

Table 6: Correlations between affect patterns and learning gains

<table>
<thead>
<tr>
<th>Pattern</th>
<th>ρ</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG-CON-FRU-BOR</td>
<td>-.086</td>
<td>0.397</td>
</tr>
<tr>
<td>ENG-CON-DEL-ENG</td>
<td>0.141</td>
<td>0.162</td>
</tr>
<tr>
<td>ENG-BOR</td>
<td>-.002</td>
<td>0.981</td>
</tr>
<tr>
<td>ENG-CON</td>
<td>-.02</td>
<td>0.847</td>
</tr>
<tr>
<td>ENG-DEL</td>
<td>0.01</td>
<td>0.927</td>
</tr>
<tr>
<td>ENG-FRU</td>
<td>0.043</td>
<td>0.681</td>
</tr>
<tr>
<td>BOR-CON</td>
<td>-.13</td>
<td>0.197</td>
</tr>
<tr>
<td>BOR-DEL</td>
<td>-.036</td>
<td>0.719</td>
</tr>
<tr>
<td>BOR-ENG</td>
<td>-.028</td>
<td>0.782</td>
</tr>
<tr>
<td>BOR-FRU</td>
<td>-.098</td>
<td>0.332</td>
</tr>
</tbody>
</table>
5.5 Summary of Results

Even after the use of Benjamini and Hochberg’s [4] post-hoc control, 22 of the 81 tested correlations produced significant results. Eighteen of these results emerged from only nine affective patterns (BOR-CON, BOR-DEL, BOR-ENG, BOR-FRU, FLO-CON-FRU-BOR, ENG-BOR, OFF-ENG, Sustained BOR, Sustained OFF), which were negatively correlated with both pre- and post-test scores. One other pattern (CON-BOR) was also negatively associated with post-test scores. None of these patterns, however, were correlated with learning gains, a result which highlights the role of prior knowledge in the manifestation of affective patterns.

Only three of the significant correlations were positively associated with our outcome measures, and notably both involved the affective state of delight. D’Mello’s ENG-CON-DEL-ENG was positively associated with post-test scores (but not pre-test scores or learning gains). Sustained DEL was significantly correlated with both the post-test scores and learning gains (but not pre-test scores), suggesting that experiences of delight may increase the likelihood of improved learning outcomes.

6 CONCLUSIONS

Overall, the findings in this study suggest that boredom is a powerful indicator of students’ knowledge, but not necessarily indicative of learning. In particular, it is interesting that Sustained BOR correlates (negatively) more strongly to pre-test values than Sustained OFF does. This relationship changes for post-test values (where their rho values are nearly equally negative). This seems to suggest that students may learn despite high levels of boredom, but that low-knowledge learners are more prone to boredom.

Other significant patterns involving boredom (ENG-BOR, BOR-CON, BOR-DEL, BOR-ENG, BOR-FRU, CON-BOR and the negative cycle ENG-CON-FRU-BOR) show comparable trends. In general, their correlation with the pre-test appears to be weaker than their correlation with the post-test, and none are significantly associated with learning gains.

Two patterns involving delight are more weakly associated with knowledge, and one of them (Sustained DEL) is the only pattern correlated with learning gains. However, none of the sequences that transition from delight to another affective state are statistically significant, and even BOR-DEL, which is the only significant transition to delight, is, like other sequences involving boredom, negatively associated with pre- and post-test scores.

Interestingly, even though they are relatively infrequent, the two hypothesized cycles from D’Mello and Graesser [11] appear to have a disproportionate effect on knowledge measures. Both are among the most infrequent patterns seen in the data (0.3% and 3.2%), and both are correlated with post-test scores. Neither is associated with learning gains, which might suggest that their relatively infrequent occurrence may be more strongly associated with prior knowledge than with learning. While both correlate to post-test scores, only the inhibitory cycle (ENG-CON-FRU-BOR) is significantly correlated to pre-test scores.

The limited prevalence (and somewhat limited effect) of the hypothesized cycles, in particular, may be attributed to methodological differences and suggests the need for further research.

The results from this study have been calculated from learning outcomes, measured from pre-test and post-test scores, and predictions from the detectors. When interpreting these results, it is important to remember that the study may be limited by the relatively lower performance of the affect detectors (A’ of the detectors ranges from 0.568 to 0.725) in comparison to those in other learning environments, which we hypothesize is related to the open-ended nature of the Betty’s Brain environment [18]. Although the results do not strongly indicate correlations with the actual emotion, the findings still indicate that there may be a correlation between the tests scores and the features of student behaviors that are associated to the different affective states.

Other methodological issues could also be impacting these results. It is possible that researchers would benefit from varying the number of steps by increasing the number of self-transitions involved in these cycles as well as by increasing the length of observation periods. Increasing the clips and amount of time considered in the analysis could allow for the discovery of more meaningful affect patterns. However, the fact that the shorter patterns showed so few statistically significant results may indicate that longer-stepped patterns will be of limited value.

If the latter is true, then researchers may need to look more carefully for specific learning contexts where these patterns are most effectual. For example, these results also suggest that, at least within the context of this study, affective patterns may be more related to prior knowledge than to learning gains. If this is true, interventions or design changes that help support students with relevant knowledge gaps may disrupt the boredom patterns that were prevalent among these students and increase the prevalence of other patterns thought to be relevant to learning.
REFERENCES


LAK’18, March 2018, Sydney, Australia