The Implications of a Subtle Difference in the Calculation of Affect Dynamics

Shamya KARUMBAIAH, Juliana Ma. Alexandra L. ANDRES, Anthony F. BOTELHO, Ryan S. BAKER, Jaelyn OCUMPAUGH

aUniversity of Pennsylvania, USA
bWorcester Polytechnic Institute, USA

Abstract: D'Mello and Graesser’s (2012) highly-cited model of affect dynamics proposes a sequence of theoretically-grounded transitions between affective states during learning. However, empirical studies in a range of contexts have not produced the predicted results. Several factors may explain this lack of replication, including the demographics of the populations studied, the degree of authenticity of the learning setting (e.g. classrooms versus laboratory studies), the grain-size of observation, affect data collection procedure, duration of research sessions, and the methodological choices used to analyze transitional patterns. In particular, whereas D'Mello and Graesser (2012) exclude self-transitions (when a student remains in the same affective state across two observations) in calculations of transition probabilities, most other published works in this area, including some of their earlier publications, do not. This paper investigates the impact of this seemingly minor methodological choice by applying both analyses to previously collected data from a study of the Physics Playground system. In particular, this paper investigates whether this difference is sufficient to produce (or suppress) the transitions theorized in D'Mello and Graesser's theoretical model.

Keywords: Affect Dynamics, Student Emotion, D’Mello’s L, Engagement, Virtual Learning

1. Introduction

Affect within intelligent tutors and other types of adaptive and artificially intelligent educational systems has been shown to correlate with a range of other important constructs including self-efficacy (McQuiggan & Lester, 2009), analytical reasoning (D'Mello, Person, & Lehman, 2009), motivation (Rodrigo et al., 2008), and learning (Bosch & D'Mello, 2017; D'Mello et al., 2012; D'Mello & Graesser, 2010). Consequently, affect-sensitive interventions have been designed to improve student learning gains (D'Mello et al., 2010; DeFalco et al., 2018) and overall experience (Karumbaiah et al., 2017). Developing effective interventions that occur in real-time depends on understanding how affect develops and manifests over time, an area of research termed affect dynamics (i.e. Kuppens, 2015), with a large body of research examining how students transition from one affective state to the next during learning activities (i.e., Andres & Rodrigo, 2014; Baker et al., 2007; Bosch & D'Mello, 2013; Bosch, & D'Mello, 2017; D'Mello & Graesser, 2012; D'Mello et al., 2009; D'Mello et al., 2007; D'Mello & Graesser, 2010; Guia et al., 2011, 2013; McQuiggan et al., 2008, 2010; Ocumpaugh et al., 2017; Rodrigo, et al., 2008, 2011, 2012). The most commonly-cited model of affect dynamics in this context, D’Mello & Graesser (2012), postulates that a specific set of affect transitions will be particularly prominent, but few empirical studies have matched that model’s predictions, an issue which this paper investigates.

Research has shown that affect plays three primary roles in learning and education: signaling, evaluation, and modulation. These roles refer to the ability of affective states to draw attention to learning challenges (Schwarz, 2012), appraise learning (Izard, 2010), and guide cognitive focus (Barth & Funke, 2010; D’Mello & Graesser, 2015; Fredrickson & Branigan, 2005; Schwarz, 2012). These roles play a key function within the D’Mello and Graesser (2012) model of affective dynamics during learning, which hypothesizes transitions between the educationally-important affective states of engaged concentration, confusion, frustration, and boredom (e.g., Fig.
1). That model builds upon another theoretical model that includes delight and surprise, but this alternate model (in the same paper) has not received the same degree of attention in the literature and will not be the focus of this paper. The primary model cited from the paper predicts that students who detect an impasse during the flow state will transition to a state of disequilibrium, experienced as the affective state of confusion. If the students resolve this impasse, they are predicted to transition back to flow. If, however, the confusion is not resolved, students are hypothesized to become “stuck” (experienced as frustration). If the frustration persists, the model suggests the learner will disengage, transitioning to boredom. Two other links in this highly cited model (confusion→frustration and boredom→frustration) are also hypothesized as likely, but the justification for these transitions is not discussed as thoroughly.

![Figure 1. D’Mello & Graesser’s (2012) model of affect dynamics](image)

D’Mello and Graesser’s model has been widely referenced (with nearly 250 citations) by various research studies on affect dynamics, including many which have used the likelihood statistic advanced in D’Mello et al., (2007; 2012) to evaluate how probable a transition is, given the base rate of the affective states involved. However, empirical studies across a range of learning environments have not consistently found results that align with the model’s proposed affective transitions. A number of factors may be contributing to divergence between the theoretical model and these empirical results. These include population differences as well as variation involving the learning context and the methodology used to examine it.

However, another key difference between D’Mello and Graesser (2012) and other research is how the data are represented when a student remains in the same affective state across several observation points. In D’Mello and Graesser (2012), only transitions between differing states were considered, whereas in many other studies (including earlier work by the same authors), a student remaining in the same affective state was considered to exhibit a self-transition that was included in calculations. The current study explores how this subtle difference may impact results of affect dynamics analyses by re-analyzing data from a previously published study (Andres et al.’s (2015b) study of Physics Playground) using both methods of calculation. This paper seeks to address whether the difference in reported results and the apparent lack of agreement with D’Mello and Graesser’s model are simply due to the treatment of self-transitions in the analysis of affect dynamics.

2. Affect Dynamics in Previous Research

Prior to investigating this methodological difference, we offer a broader review of the past literature on affect dynamics and how it varies both in terms of this methodological choice and other factors. We focus on fourteen past studies that represented changes in affect using D’Mello’s L (2007), a metric used to determine the likelihood of an affective transition. The current study will focus primarily on the affective states included in the D’Mello and Graesser model (i.e. boredom, flow, frustration, and confusion), but as Table 1 summarizes, a range of other emotions have been included in these previously published papers (i.e., anger, anxious, confusion, curiosity, delight, disgust, eureka, excitement, fear, happiness, neutral, sadness, and surprise).

These studies have yielded a range of results. From the 14 studies considered, transitions that are both significantly more likely to occur than chance and align with the model of affect dynamics have been found predominantly in studies by D’Mello and his colleagues. Flow→confusion was reported in multiple D’Mello studies (2007; 2010; 2012; 2013; 2017) as well as in McQuiggan (2008; 2010) and Ocumpaugh (2017). Frustration→boredom was reported in
D’Mello studies (2012; 2017) and was marginally significant in one Rodrigo study (2008). 
Confusion→flow was reported in two D’Mello studies (2010; 2012; 2017) and in one study by Ocumpaugh (2017). Boredom→frustration was reported by in studies by D’Mello (2007; 2012) and in one study by Rodrigo and colleagues (2012). Transitions of frustration→confusion (in D’Mello et al, 2009; 2013, 2017) and confusion→frustration (in D’Mello et al., 2010; 2012; 2013; 2017) were reported in exclusively in studies by D’Mello and his colleagues. However, as Table 2 summarizes, there are a variety of methodological and population differences that may have influenced these findings.

Table 1

Affective States studied in Previous Research on Affect Dynamics. Categories studied in D’Mello & Graesser’s Model are Highlighted in Gray. (BORed, FLOw, DELight, FRUstration, SURprise, NEUtral, CONfused, ANXious, ANGer, DISgust, SADness, EUReka, CURious, FEAr, EXCited)

<table>
<thead>
<tr>
<th>Studies</th>
<th>BOR</th>
<th>FLO</th>
<th>DEL</th>
<th>FRU</th>
<th>SUR</th>
<th>NEU</th>
<th>ANX</th>
<th>ANG</th>
<th>DIS</th>
<th>SAD</th>
<th>EUR</th>
<th>CUR</th>
<th>FEA</th>
<th>EXC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andres &amp; Rodrigo, 2014</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baker, Rodrigo, &amp; Xolocotzin, 2007</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosch &amp; D’Mello, 2013</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosch, &amp; D’Mello, 2017</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D’Mello &amp; Graesser, 2012</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D’Mello et al., 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D’Mello, Taylor, &amp; Graesser, 2007</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D’Mello &amp; Graesser, 2010</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guia et al., 2011</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guia et al., 2013</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McQuiggan et al., 2008; 2010</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ocumpaugh et al., 2017</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodrigo et al., 2008</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodrigo et al., 2011; 2012</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Summary of the Observed Methodological Differences across 14 Studies on Affect Dynamics

<table>
<thead>
<tr>
<th>Studies</th>
<th>Region</th>
<th>Age</th>
<th>N</th>
<th>School/Grade Population</th>
<th>Learning System</th>
<th>Class v. Lab</th>
<th>Obs. Type/ Grain Size</th>
<th>Obs. Session</th>
<th>Self- trans</th>
<th>Aligned Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andres &amp; Rodrigo, 2014</td>
<td>Quezon City, PH</td>
<td>13-16</td>
<td>60</td>
<td>Public school</td>
<td>Physics Playground</td>
<td>C</td>
<td>QFO</td>
<td>2hrs</td>
<td>Inc</td>
<td>0</td>
</tr>
<tr>
<td>Baker et al., 2007</td>
<td>Manila, PH</td>
<td>14-19</td>
<td>36</td>
<td>High school</td>
<td>Inc. Machine</td>
<td>C</td>
<td>QFO ev. 60s</td>
<td>10min</td>
<td>Inc</td>
<td>0</td>
</tr>
<tr>
<td>Bosch &amp; D’Mello, 2013</td>
<td>US</td>
<td>--</td>
<td>29</td>
<td>Undergrads</td>
<td>Unnamed</td>
<td>L</td>
<td>RJP on 100 fixed points</td>
<td>25min</td>
<td>Exc</td>
<td>3</td>
</tr>
<tr>
<td>Bosch, &amp; D’Mello, 2017</td>
<td>Midwestern US</td>
<td>17-21</td>
<td>99</td>
<td>Undergrads</td>
<td>Unnamed</td>
<td>L</td>
<td>RJP on 100 fixed points</td>
<td>25min</td>
<td>Exc</td>
<td>5</td>
</tr>
<tr>
<td>D’Mello &amp; Graesser, 2012</td>
<td>Southern US</td>
<td>--</td>
<td>28</td>
<td>Undergrads</td>
<td>Auto-Tutor</td>
<td>L</td>
<td>RJP every 20s; fixed points</td>
<td>32min; 35min</td>
<td>Exc</td>
<td>4;5</td>
</tr>
<tr>
<td>D’Mello et al., 2007</td>
<td>Southern US</td>
<td>--</td>
<td>30</td>
<td>Undergrads</td>
<td>Auto-Tutor</td>
<td>L</td>
<td>RJP ev. 20s</td>
<td>32min; Inc</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>D’Mello et al., 2009</td>
<td>Southern US</td>
<td>--</td>
<td>41</td>
<td>Undergrads</td>
<td>Unnamed</td>
<td>L</td>
<td>RJP on fixed points</td>
<td>35min</td>
<td>Exc</td>
<td>1</td>
</tr>
<tr>
<td>D’Mello &amp; Graesser, 2010</td>
<td>Southern US</td>
<td>--</td>
<td>28</td>
<td>Undergrads</td>
<td>Auto-Tutor</td>
<td>L</td>
<td>RJP ev. 20s; fixed points</td>
<td>32min; Exc</td>
<td>3;3</td>
<td></td>
</tr>
<tr>
<td>Guia et al., 2011; 2013</td>
<td>Quezon City, PH</td>
<td>18-20</td>
<td>60</td>
<td>Undergrads</td>
<td>SQL Tutor</td>
<td>C</td>
<td>QFO ev. 200s</td>
<td>1hr</td>
<td>Inc</td>
<td>0</td>
</tr>
<tr>
<td>McQuiggan et al., 2008; 2010</td>
<td>US</td>
<td>21-60</td>
<td>35</td>
<td>Grad students</td>
<td>Crystal Island</td>
<td>L</td>
<td>SRI</td>
<td>35min</td>
<td>Inc</td>
<td>1</td>
</tr>
<tr>
<td>Ocumpaugh et al., 2017</td>
<td>New York, US</td>
<td>18-22</td>
<td>108</td>
<td>West Point</td>
<td>vMedic</td>
<td>C</td>
<td>QFO ev. 122s</td>
<td>--</td>
<td>Inc</td>
<td>2</td>
</tr>
<tr>
<td>Rodrigo et al., 2008</td>
<td>Quezon City &amp; Cavite Prov., PH</td>
<td>9-13</td>
<td>180</td>
<td>Private school</td>
<td>Ecolab</td>
<td>C</td>
<td>QFO</td>
<td>40min</td>
<td>Inc</td>
<td>1</td>
</tr>
<tr>
<td>Rodrigo et al., 2011; 2012</td>
<td>Quezon City, PH</td>
<td>12-14</td>
<td>126</td>
<td>High school</td>
<td>Scatterplot Tutor</td>
<td>C</td>
<td>QFO ev. 200s</td>
<td>80min</td>
<td>Inc</td>
<td>1</td>
</tr>
</tbody>
</table>
2.1 Demographic Differences in Previous Work Examined

The 14 studies summarized in Table 2 differ noticeably in terms of the demographic characteristics of their samples, including age and the region where the research was conducted. Differences in culture influence variation in beliefs and personal dispositions towards emotional expression and moderation (Tsai & Levenson, 1997; Uchida et al., 2009) and the frequency and emergence of certain affective states (Kitayama et al., 2000) while age influences emotional expressivity (Dunn & Brown, 1994; Gross et al., 1997) and inhibition (Cole, 1986). It is possible that differences in results may be due to these factors; if so, this would suggest that D’Mello and Graesser’s model may not be general across cultural contexts.

2.2 Learning Settings

The studies were conducted across multiple instructional settings, including regular classroom environments and laboratory settings. Educational software has covered a variety of educational content, including mathematics (Rodrigo et al., 2011, 2012), biology (McQuiggan et al., 2008; 2010; Rodrigo et al., 2008), emergency medical content (Ocumpaugh et al., 2017), physics (Andres & Rodrigo, 2014; Baker et al., 2007; D’Mello, Taylor, & Graesser, 2007), computer literacy and programming (Bosch & D’Mello, 2013, 2017; D’Mello & Graesser, 2010; Guia et al., 2011, 2013), and analytical problem solving (D’Mello et al., 2009). The learning systems that have been used across these studies have also differed in terms of design. Scatterplot Tutor, SQL-Tutor, AutoTutor and the other researcher-built learning environments used in studies conducted by D’Mello follow more linear designs wherein learners must complete problems before they are able to proceed. On the other hand, environments such as Physics Playground, Crystal Island, Incredible Machine, vMedic, and Ecolab, are open-ended systems that offer learners the opportunity to explore the range of possible solutions.

2.3 Data Collection Procedure, including Observation Grain-Size and Session Duration

Six of the 14 studies use the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP; Ocumpaugh et al., 2015), a momentary time sampling method that uses a holistic coding practice to code for both affect and behavior. In this protocol, students are observed for up to 20 seconds in a round-robin manner throughout the given observation period to ensure uniform frequencies of student observation. The protocol is enforced by an Android application known as the Human Affect Recording Tool (HART, Ocumpaugh, et al., 2015).

By contrast, D’Mello and his colleagues have used self-reporting methods, collecting affect data through retrospective judgment protocols which synchronize webcam video of students’ faces to screen capture of the learning environment (Bosch & D’Mello, 2013, 2017; D’Mello et al., 2007; 2009; D’Mello & Graesser, 2010, 2012). McQuiggan et al., (2008; 2010) also collected self-reported data, but used in-game dialogs to collect spontaneous reports rather than a retrospective technique.

Observation sessions in this research varied in length, ranging from 10 minutes (Baker et al., 2007) to 2 hours (Andres & Rodrigo, 2014), potentially influencing the affect that emerges during observation. Prolonged exposure to similar tasks may produce fatigue or boredom (Gonzalez et al., 2011), decreasing learner performance (Healy et al., 2004). It may also increase students’ susceptibility to what D’Mello et al., (2007) describe as vicious cycles of boredom, where learners are unable to transition to other affective states.

2.4 Differences in the Treatment of Self-transitions Between Studies

All of the studies considered in this section analyze time series data (e.g., the order of the occurrences of each affective state), but they have been inconsistent in their treatment of self-transitions, which occur when a student remains in the same affective state over two consecutive observations. In more recent studies, D’Mello and colleagues have removed self-transitions during the data preparation.
stage (Bosch & D’Mello, 2013; 2017; D’Mello et al., 2009; D’Mello & Graesser, 2010; 2012). For example, a sequence of confusion, flow, flow, boredom has one self-transition (from flow to flow). However, this practice is not followed in all work. Nearly a dozen other studies conducted in this field do not report discarding self-transitions in their data processing (Baker et al., 2007; 2012; Guia et al., 2011; 2013; McQuiggan et al., 2008; 2010; Ocumpaugh et al., 2017; Rodrigo et al., 2008; 2011), including early work by D’Mello and his colleagues (e.g. D’Mello et al., 2007). As we will demonstrate, this seemingly small step may have disproportionate effects on study outcomes.

3. Methods

In this section, we discuss the method for affect dynamics analysis used in this paper, and present the previously published dataset (from Andres et al., 2015b) that is used to assess the impact of how self-transitions are considered (see section 2.4).

3.1 Dataset: Physics Playground

We investigate the implications of how self-transitions are represented in affect dynamics analyses using a previously published data set with which none of the current authors were previously involved in collecting or analyzing (Andres et al., 2015b). In this study, 120 8th graders and 60 10th graders spent 2 hours using Physics Playground, a learning environment that teaches qualitative physics to secondary students (Shute & Ventura, 2013). In this 2-dimensional game, students sketch different objects like pendulum, ramp, lever, and springboard to guide a ball to touch a balloon. Laws of physics apply to all the objects on the screen.

This data was collected in 2015 in schools in Baguio, Cebu and Davao, Philippines (Andres et al., 2015b) using BROMP (Ocumpaugh et al., 2015). This data set was selected because the study had multiple observers, resulting in a high rate of sampling. Each student was observed approximately once per minute, for an average of 135 observations per student (24,330 total). While affective states have been studied in Filipino classrooms using Physics Playground (Andres et al., 2014), this data has not been previously used for such purposes. Previously published analyses on this data have involved the development of a wheel-spinning model (Palaoag et al., 2016), challenges encountered in field studies (Andres et al., 2015b), and the development of a model of student carefulness (Banawan et al., 2017), but none specifically considered affect beyond its relationship to eureka moments (Andres et al., 2015a).

3.2 L statistics and Affect Dynamics Analysis

The studies cited above have employed the D’Mello L statistic (Equation 1) to calculate the likelihood that an affective state (prev) will transition to a subsequent (next) state, given the base rate of the next state occurring. L values greater than 0 indicate that a transition is more likely than chance, and L values less than 0 indicate that a transition is less likely than chance; 0 indicates chance. The value of L varies from 1 to -∞.

\[ L(prev \rightarrow next) = \frac{P(next|prev) - P(next)}{1 - P(next)} \]  

(1)

The L value for each affect combination is calculated individually per student. For m affective states, there are \( m^2 \) possible types of transitions if self-transitions are included, and \( m^2 - m \) possible types of transitions if self-transitions are omitted. The probability \( P(next) \) of an affective state is the percentage of times that the state had occurred as a next state. Thus, the first affective state in the sequence of a student must be excluded from this calculation since this state cannot take the role of a next state. Similarly, the calculation of the prev state excludes the last state in the sequence. The term \( P(next|prev) \) is a conditional probability calculated using Equation 2, where \( Count(prev \rightarrow next) \) is the number of times the prev state transitioned to the next state, and \( Count(prev) \) is the number of times the state in prev occurred as the previous state.
The cases below illustrate situations where transition calculations may not be straight forward:

1. \( L = 0 \) for any transition going into a state that did not occur in a student's affect sequence. In that case, \( P(\text{next}) = 0 \) and \( P(\text{next} | \text{prev}) = 0 \), and thus, \( L = 0 \).

2. The \( L \) value is undefined for any transition out of a state that does not occur for a student, as we do not know what would have followed that state if it had occurred.

3. When a student remains in one affective state throughout an observation period, all transitions to states other than that state are 0, and all transitions to the single affective state seen have undefined \( L \), as the denominator of the equation is 0 in that case.

4. When self-transitions are discarded from the data, an affect sequence consisting of a single state is reduced to a single state. In this case, since there would be no affective state in the next value, \( L \) is undefined for all states.

   In all cases where \( L \) is undefined, those values are discarded from further analysis.

   This procedure is implemented in a software package which can be found at https://github.com/Shamya/L-Statistic-for-Transition-Likelihood

Two tailed \( t \)-tests are conducted on the calculated \( L \) values to measure whether each transition is significantly more or less likely than chance (i.e., is \( L \) across students significantly more or less than 0, the chance level). A Benjamini-Hochberg post-hoc correction procedure is used to control for false positive results (using an initial \( \alpha \) of 0.05) since the set of hypotheses involves multiple comparisons. As in previous studies, the present analysis considers any value that is statistically significantly higher than zero, however small, to indicate that a transition is more likely than chance.

### 3.3 Data Analysis

In conducting this analysis, affective states that were not incorporated in the D’Mello and Graesser model have been merged into NA, a dummy state. Across students in the Physics Playground dataset, the mean and standard deviation of the proportion of occurrence of the affect states are \(-73\%\pm14\%\) flow, \(6\%\pm6\%\) confusion, \(5\%\pm6\%\) frustration, \(3\%\pm6\%\) boredom and \(13\%\pm11\%\) NA.

It is worth noting that in some of the previously published papers on affect dynamics listed above, \( P(\text{next}) \) within the D’Mello’s \( L \) metric was calculated across all students whereas, as elaborated in the detailed data analysis steps in the previous section, we are calculating \( P(\text{next}) \) for each student while computing the corresponding \( L \) value. By calculating these individually, percentages are normalized per student rather than across the entire dataset, avoiding outlier estimates of \( L \) where a specific student’s prevalence of a specific affect state varies considerably from the rest of the sample.

### 4. Result

Results comparing the 2 methods for calculating \( L \) (with vs. without including self-transitions) show that this difference has a substantial effect on both \( L \) values and, to a lesser degree, the number of transition patterns that are found to be statistically significant. As Table 3 shows, both techniques yield the same number of transitions that occur significantly above chance, but the inclusion of self-transitions yields far more transitions that occur less often than chance. These lead to very different result patterns, but neither provides substantial evidence for D’Mello and Graesser’s (2012) model.

#### 4.1 Including Self Transitions

When self-transitions are included, 14 of 16 possible transitions are statistically significantly different than chance, but only 6 of these transitions are more likely than chance. Two of these are hypothesized by D’Mello & Graesser (confusion→frustration and frustration→boredom), but four of these are self-transitions that were not hypothesized (flow→flow, confusion→confusion, frustration→frustration, and boredom→boredom).
Overall, though, these results do not support the hypothesized model when self-transitions are included. The primary cycle of flow→confusion and confusion→flow, thought to be critical for learning, is shown to occur significantly below chance. Likewise, the transition from frustration→confusion has an L value of 0, and the transition from boredom→flow is significantly below chance.

4.2 Excluding Self Transitions

If self-transitions are removed, the proportion of L values that occur above chance increases. Six transitions previously found to be less likely than chance with the previous method flip sign (i.e., becoming more likely than chance) upon the exclusion of self-transitions, but only two of these were hypothesized in the model (flow→confusion, and confusion→flow). The other four transitions that were significantly above chance were not hypothesized, including two with relatively small Ls (flow→frustration, flow→boredom), but also two with relatively large Ls (frustration→flow, and boredom→flow).

The hypothesized model fares only slightly better when self-transitions are excluded. The cycle from flow→confusion and confusion→flow becomes more likely than chance upon the exclusion of self-transitions, but frustration→confusion goes from being non-significant to significantly below chance. The remaining three hypothesized links (confusion→frustration, frustration→boredom, and boredom→frustration) showed L values that were very close to chance and were not found to be statistically significant.

Table 3

<table>
<thead>
<tr>
<th>Transition</th>
<th>D’Mello’s L with self-transitions</th>
<th>D’Mello’s L without self-transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean L</td>
<td>Std</td>
</tr>
<tr>
<td>FLO FLO</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>CON (st)</td>
<td>-0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>FRU (st)</td>
<td>-0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>BOR (st)</td>
<td>-0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>CON FLO</td>
<td>-0.71</td>
<td>2.85</td>
</tr>
<tr>
<td>FRU (st)</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>BOR (st)</td>
<td>-0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>FRU FLO</td>
<td>-0.42</td>
<td>1.5</td>
</tr>
<tr>
<td>CON (st)</td>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td>FRU (st)</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>BOR (st)</td>
<td>0.03</td>
<td>0.16</td>
</tr>
</tbody>
</table>

5. Discussion

D’Mello and Graesser’s (2012) model has been one of the most notable theoretical frameworks in affect dynamics research. It postulates how affect develops over time during learning...
and theorizes how the transitions in affect that are hypothesized may contribute to processes of learning and disengagement (D’Mello & Graesser, 2012). However, relatively few studies have found these transitions. By examining previously published research and data from the affect dynamics literature, the current study has attempted to shed light on potential underlying explanations for the observed differences. Specifically, it investigates the degree to which the results of analyses are influenced by the inclusion or exclusion of self-transitions in the calculation of likelihood metrics. Including self-transitions may suppress non-self transitions. If some affective states are particularly persistent (Andres & Rodrigo, 2014; Baker, D’Mello, Rodrigo, & Graesser, 2010), the inclusion of self-transitions could lower the transition probabilities for transitions to new affective states, and/or in some cases, cause them to become non-significant. In contrast, D’Mello and Graesser’s (2012) approach, which excludes self-transitions, may inflate the frequency of seeing transitions between affective states. Yet when applied to the data studied in this paper, neither method resulted in strong evidence for the theoretical model.

While D’Mello and Graesser’s method increases the proportion of transitions that occur above chance (and the number of statistically likely non-self transitions), it does not increase the degree of conformance with their hypothetical model. When including self-transitions, two of the six hypothesized transitions are observed, and no non-hypothesized (non-self) transitions are seen. When excluding self-transitions, two of the six hypothesized transitions (not the same ones) are observed, and four non-hypothesized (non-self) transitions are seen. Therefore, we can conclude that the differences between D’Mello’s hypothesized model and previously published results are not simply due to differences in this analytical method, but are likely related to other factors.

Future research should consider a number of potential reasons for these findings. As we have highlighted, previous studies have differed in several fashions, including variation in student demographics, learning environments, lab versus classroom settings, different grain-sizes of observations, different data collection procedures, and different study session durations. These variations are indicative of the difficulties that emerge in studying affect dynamics. However, unlike the methodological differences analyzed here, such variables are more difficult to control for in a retrospective analysis.

Beyond simply providing evidence as to whether D’Mello and Graesser’s model is accurate, the findings in this study indicate that analytical choices should depend on the research goals and questions of the study. Excluding self-transitions reveals a larger number of affective patterns that might otherwise be suppressed by the presence of persistent affective states. Conversely, including self-transitions in analysis helps us to better understand each state’s persistence, but dilutes any transitions between different affective states. The former is likely important in theoretical models, but the latter might be particularly useful for algorithms being used to trigger interventions, for example.

Future research may help to uncover additional elements that have not yet been recognized in order to better understand affective dynamics, but this study has shown that there are multiple areas of divergence in the methodology that has been employed to date in the affect dynamics literature and the full extent of these differences has yet to be investigated. By studying a broader range of datasets and further investigating which factors are associated with studies matching the predictions in D’Mello and Graesser’s model, we can better understand not just its validity, but its scope of applicability.

Acknowledgements

Our thanks to the Penn Center for Learning Analytics for funding this research, and to Didith Rodrigo and Miguel Andres for sharing this data with us.

References


