Yang, J. C. et al. (Eds.) (2018). Proceedings of the 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education

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

Shamya KARUMBAIAH[a], Juliana Ma. Alexandra L. ANDRES[a], Anthony F. BOTELHO[b], Ryan S. BAKER[a], Jaclyn OCUMPAUGH[a]
[a]*University of Pennsylvania, USA*
[b] *Worcester Polytechnic Institute, USA*
*shamya@gse.upenn.edu

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.

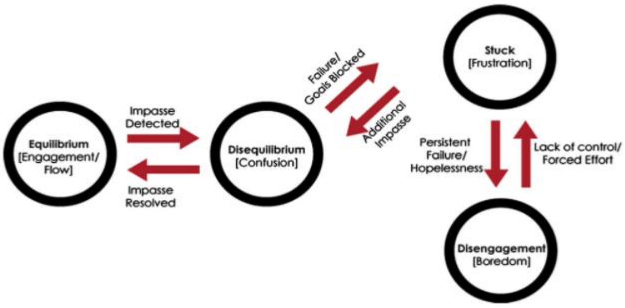

Figure. 1. D'Mello & Graesser's (2012) model of affect dynamics

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)

Studies	BOR	FLO	DEL	FRU	SUR	NEU	CON	ANX	ANG	DIS	SAD	EUR	HAP	CUR	FEA	EXC
Andres & Rodrigo, 2014	x	x	x	x	x		x									
Baker, Rodrigo, & Xolocotzin, 2007	x	x	x	x	x	x	x									
Bosch & D'Mello, 2013	x	x		x			x									
Bosch, & D'Mello, 2017	x	x		x	x	x	x	x	x	x	x		x	x	x	
D'Mello & Graesser, 2012	x	x	x	x	x	x	x									
D'Mello et al., 2009	x			x	x	x	x	x	x	x	x	x	x	x		
D'Mello, Taylor, & Graesser, 2007	x	x	x	x	x		x									
D'Mello & Graesser, 2010	x	x	x	x	x		x									
Guia et al., 2011	x	x	x	X	x	x	x									
Guia et al., 2013	x	x	x	x	x	x	x									
McQuiggan et al., 2008; 2010	x	x	x	x			x	x	x		x				x	x
Ocumpaugh et al., 2017	x	x		x	x		x	x								
Rodrigo et al., 2008	x	x	x	x	x	x	x									
Rodrigo et al., 2011; 2012	x	x	x	x	x	x	x									

Table 2

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

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

* PH: Philippines, QFO: Qualitative field observation, RJP: Retrospective judgment protocol, SRI: self-report based on interactions, Inc: self transitions included, Exc: self transitions excluded

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 -∞.

$$L(prev \rightarrow next) = \frac{P(next \mid prev) - P(next)}{1 - P(next)} \qquad (1)$$

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 → 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.

$$P(\boldsymbol{next} \mid \boldsymbol{prev}) = \frac{\boldsymbol{Count}\,(\boldsymbol{prev} \rightarrow \boldsymbol{next})}{\boldsymbol{Count}\,(\boldsymbol{prev})} \qquad (2)$$

The cases below illustrate situations where transition calculations may not be straight forward:

1. *L* is 0 for any transition going into a state that did not occur in a student's affect sequence. In that case, *P*(*next*) = 0 and *P(next* | *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 α 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%±14% flow, 6%±6% confusion, 5%±6% frustration, 3%±6% boredom and 13%±11% NA.

It is worth noting that in some of the previously published papers on affect dynamics listed above, *P(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(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 are significantly above chance were not hypothesized, including two with relatively small *L*s (*flow→frustration, flow→boredom*), but also two with relatively large *L*s (*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

L Values with and without Self-transitions. Self-transitions (st) and Transitions Hypothesized (h) in D'Mello & Graesser's Model are noted. L Values that Flipped Sign (i.e, a less likely transition became more likely by removing self-transitions or vice-versa) are Highlighted in Gray, while those that are Significantly Above Chance are given in Bold. All Significance After B&H is marked with Asterisk

Transition			D'Mello's L with self-transitions					D'Mello's L without self-transitions				
from	to		Mean L	Std	T	p		Mean L	Std	t	p	
FLO	FLO	(st)	**0.13**	0.16	10.81	0.00	*	-	-	-	-	
	CON	(h)	-0.01	0.02	-4.18	0.00	*	**0.15**	0.17	11.16	0.00	*
	FRU		-0.01	0.04	-3.97	0.00	*	**0.09**	0.13	9.57	0.00	*
	BOR		-0.02	0.04	-4.63	0.00	*	**0.03**	0.07	6.1	0.00	*
CON	FLO	(h)	-0.71	2.85	-3.04	0.00	*	**0.52**	0.48	13.12	0.00	*
	CON	(st)	**0.09**	0.17	6.6	0.00	*	-	-	-	-	
	FRU	(h)	**0.04**	0.19	2.5	0.01	*	-0.01	0.22	-0.38	0.7	
	BOR		-0.02	0.07	-2.49	0.01	*	-0.02	0.08	-2.77	0.01	*
FRU	FLO		-0.42	1.5	-3.42	0.00	*	**0.47**	0.52	11.06	0.00	*
	CON	(h)	0	0.11	-0.02	0.98		-0.07	0.16	-4.94	0.00	*
	FRU	(st)	**0.07**	0.16	5.81	0.00	*	-	-	-	-	
	BOR	(h)	**0.03**	0.16	2.38	0.02	*	0.02	0.18	1.58	0.12	
BOR	FLO		-0.77	1.96	-3.37	0.00	*	**0.56**	0.49	9.94	0.00	*
	CON		-0.05	0.09	-4.54	0.00	*	-0.1	0.15	-5.61	0.00	*
	FRU	(h)	0.01	0.14	0.86	0.39		-0.01	0.17	-0.74	0.46	
	BOR	(st)	**0.23**	0.25	8.02	0.00	*	-	-	-	-	

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

Andres, J. M. A. L., Andres, J. M. L., Rodrigo, M. M. T., Baker, R. S., & Beck, J. B. (2015a). An investigation of eureka and the affective states surrounding eureka moments. *In Proceedings of the 23rd International Conference on Computers in Education*. China: Asia-Pacific Society for Computers in Education.

Andres, J.M.L., & Rodrigo, M.M.T. (2014). The Incidence and persistence of affective states while playing Newton's playground. *7th IEEE International Conference on Humanoid, Nanotechnology, Information Tech., Communication and Control, Environment, and Management*.

Andres, J.M.L., Rodrigo, M.M T., Sugay, J., Banawan, M., Paredes, Y., Cruz, J., Palaoag, T. (2015b). More Fun in the Philippines? Factors Affecting Transfer of Western Field Methods to One Developing World Context. In *AIED Workshops*.

Baker, R.S., Rodrigo, M.M.T., Xolocotzin, U. (2007). The dynamics of affective transitions in simulation problem-solving environments. *International Conf. on Affective Computing and Intelligent Interaction*. Springer Berlin Heidelberg, 666-67.

Baker, R.S., D'Mello, S., Rodrigo, M.M.T., Graesser, A. (2010) Better to be frustrated than bored: The incidence, persistence, & impact of learners' cognitive-affective states during interactions with 3 different computer-based learning environments. *Int'l J. Hum.-Comp. Stu., 68* (4),223-41.

Banawan, M. P., Rodrigo, M. M. T., & Andres, J. M. L. (2017). Predicting Student Carefulness within an Educational Game for Physics using Support Vector Machines. *Proc. of the 25th International Conference on Computers in Education,*62-67. New Zealand: Asia-Pacific Society for Computers in Education.

Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition and Emotion, 24*(7), 1259-1268.

Bosch, N., & D'Mello, S. (2013). Sequential patterns of affective states of novice programmers. In *The 1st Workshop on AI-supported Education for Computer Science (AIEDCS 2013)*, 1-10.

Bosch, N., & D'Mello, S. (2017). The Affective Experience of Novice Computer Programmers. *International Journal of Artificial Intelligence in Education*, 1-26.

Chaachoua, H., Nicaud, J., Bronner, A., Bouhineau, D. (2004). Aplusix, a learning environment for algebra, actual use and benefits. *10th International Congress on Mathematical Education*, 8.

Cole, P. (1986). Children's spontaneous control of facial expression. *Child development*, 1309-1321.

D'Mello, S., Craig, S., Gholson, B., Franklin, S., Picard, R., Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. *Affective Interactions: The Computer in the Affective Loop Workshop at International Conf. on Intelligent User Interfaces*, NY:AMC P. 7-13.

D'Mello, S., Person, N., Lehman, B. (2009). Antecedent-Consequent Relationships and Cyclical Patterns between Affective States and Problem Solving Outcomes. *In AIED*, 57-64.

D'Mello, S. Graesser, A. (2012). Dynamics of Affective States during Complex Learning. *Learning and Instruction, 22*, 145-157.

D'Mello, S., Taylor, R., & Graesser, A. (2007). Monitoring Affective Trajectories during Complex Learning. D. McNamara & J. Trafton (Eds.), *Proc. 29th Annual Cognitive Science Soc.*, 203-8.

D'Mello, S., Graesser, A. (2010). Modeling cognitive-affective dynamics with Hidden Markov Models. *Proceedings of the 32nd Annual Cognitive Science Society*, 2721-2726.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Lydia, P., & Graesser, A. (2010, June). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In *International conference on intelligent tutoring systems*. Springer, Berlin, Heidelberg, pp. 245-254.

D'Mello, S., Graesser, A. (2015). Feeling, thinking, & computing with affect-aware learning technologies. Calvo, D'Mello, Gratch, Kappas (eds.) *Handbook of Affective Computing*. Oxford UP.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L. Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. *Intelligent Tutoring Systems* (pp. 245-254). Springer Berlin/Heidelberg.

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education*, *28*(2), 152-193.

Dunn, J., & Brown, J. (1994). Affect expression in the family, children's understanding of emotions, and their interactions with others. *Merrill-Palmer Quarterly* (1982-), 120-137.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion, 19*(3), 313-332.

Gonzalez, C., Best, B., Healy, A., Kole, J. A., Bourne Jr, L. (2011). A cognitive modeling account of simultaneous learning and fatigue effects. *Cognitive Systems Research, 12*(1), 19-32.

Graesser, A., Lu, S., Jackson, G., Mitchell, H., Ventura, M., Olney, A., Louwerse, M. (2004) AutoTutor: A tutor with dialogue in natural language. *Beh. Res. Methods, Instr., & Comp., 36*(2), 180-92.

Guia, T.F.G., Rodrigo, M.M.T., Dagami, M., Sugay, J., Macam, F., Mitrovic, A. (2013) An exploratory study of factors indicative of affective states of students using SQL-Tutor. *Research & Practice in Technology Enhanced Learning, 8*(3), 411-430.

Guia, T.F.G., Sugay, J., Rodrigo, M.M.T., Macam, F., Dagami, M., Mitrovic, A. (2011). Transitions of Affective States in an Intelligent Tutoring System. *Proc. Philippine Computing Soc.* 31-5.

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., & Hsu, A. Y. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12*(4), 590.

Healy, A., Kole, J., Buck-Gengler, C., Bourne Jr, L., (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied*, 10, 188–199.

Izard, C. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review, 2*(4), 363-370.

Karumbaiah, S., Lizarralde, R., Allessio, D., Woolf, B. P., Arroyo, I., & Wixon, N. (2017). Addressing Student Behavior and Affect with Empathy and Growth Mindset. In *Educational Data Mining*.

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition & Emotion, 14*(1), 93-124.

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Rev., 7*(4), 297-300.

Lehman, B., D'Mello, S., & Person, N. (2008). All alone with your emotions: An analysis of student emotions during effortful problem solving activities. In *Workshop on Emotional and Cognitive Issues in ITS at the 9th International Conference on Intelligent Tutoring Systems*.

Luckin, R., & Du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education, 10*(2), 198-220.

McQuiggan, S. W., & Lester, J. (2009). Modeling affect expression and recognition in an interactive learning environment. *International Journal of Learning Technology, 4*(3-4), 216-233.

McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2010). Affective transitions in narrative-centered learning environments. *Educational Technology & Society, 13*(1), 40-53.

McQuiggan, S., Robison, J., & Lester, J. (2008). Affective transitions in narrative-centered learning environments. In *Intelligent Tutoring Systems,* 490-499). Springer Berlin/Heidelberg.

Nicaud, J. F., Bouhineau, D., Mezerette, S., & Andre, N. (2007). *Aplusix II* [Computer software].

Ocumpaugh, J., Andres, J.M., Baker, R., DeFalco, J., Paquette, L., Rowe, J., et al. (2017). Affect Dynamics in Military Trainees using vMedic: From Engaged Concentration to Boredom to Confusion. In *International Conf. on Artificial Intelligence in Ed.*, 238-249. Springer, Cham.

Ocumpaugh, J., Baker, R.S., Rodrigo, M.M.T. (2015) Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical & Training Manual. *Technical Report.* NY, NY: Teachers College, Columbia U. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.

Ocumpaugh, J., Baker, R., Rodrigo, M., Salvi, A., VanVelsen, M., Aghababyan, A., Martin, T. (2015) HART: The human affect recording tool. *Proc. 33rd Ann. Intern.l Conf. Design of Comm.*, 24.

Palaoag, T. D., Rodrigo, M. M. T., Andres, J. M. L., Andres, J. M. A. L., & Beck, J. E. (2016, June). Wheel-spinning in a game-based learning environment for physics. *In International Conference on Intelligent Tutoring Systems* (pp. 234-239). Springer, Cham.

Rodrigo, M.M.T., Anglo, E., Sugay, J., Baker, R. (2008). Use of unsupervised clustering to characterize learner behaviors and affective states while using an intelligent tutoring system. In *International Conference on Computers in Education*, 57-64.

Rodrigo, M.M.T., Baker, R., Agapito, J., Nabos, J., Repalam, M., Reyes Jr, S., San Pedro, M.O.C. (2011). The effects of an embodied conversational agent on student affective dynamics while using an intelligent tutoring system. *IEEE Transactions on Affective Computing*, 2(4), 18-37.

Rodrigo, M.M.T., Baker, R., Agapito, J., Nabos, J., Repalam, M., Reyes, S., San Pedro, M.O.C. (2012). The effects of an interactive software agent on student affective dynamics while using; an intelligent tutoring system. *IEEE Transactions on Affective Computing, 3*(2), 224-36.

Rodrigo, M.M.T., Baker, R., D'Mello, S., Gonzalez, M., Lagud, M., Lim, S., Macapanpan, A., Pascua, S., Santillano, J., Sugay, J., Tep, S., Viehland, N. (2008). Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. *International Conf. on Intelligent Tutoring Systems* (40-49). Springer Berlin Heidelberg.

Rodrigo, M.M.T., Rebolledo-Mendez, G., Baker, R., du Boulay, B., Sugay, J.., Lim, S., Luckin, R. (2008) The effects of motivational modeling on affect in an intelligent tutoring system. *Proc. of International Conference on Computers in Education*, 57, 64.

Schwarz, N. (2012). Feelings-as-Information Theory. In P. Van Lange, A. Kruglanski & T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (pp. 289-308). Thousand Oaks, CA: Sage.

Shute, V., Ventura, M. (2013). *Stealth assessment: Measuring & supporting learning in video games*. MIT Press.

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., Ventura, M., & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education*, 86, 224-235.

Tsai, J., Levenson, R. (1997) Cultural influences on emotional responding: Chinese Am. & European Am. dating couples during interpersonal conflict. *J. Cross-Cultural Psych., 28*(5), 600-25.

Uchida, Y., Townsend, S., Rose Markus, H., Bergsieker, H. (2009). Emotions as within or between people? Cultural variation in lay theories of emotion expression and inference. *Personality and Social Psychology Bulletin, 35*(11), 1427-1439.