This is a pre-proof version of the following article: Zhang, Y., Paquette, L., & Bosch, N. (In press). Conditional and marginal strengths of affect transitions during computer-based learning. *International Journal of Artificial Intelligence in Education*.

Conditional and marginal strengths of affect transitions during computer-based learning

Yingbin Zhang [a*], Luc Paquette [b], Nigel Bosch [c]

[a] Institute of Artificial Intelligence in Education, South China Normal University, No.55 West of Zhongshan Avenue, Guangzhou, Guangdong 510631, China. ORCID: 0000-0002-2664-3093

[b] Department of Curriculum & Instruction, University of Illinois at Urbana-Champaign, 1310 S Sixth Street, Champaign, IL 61820, United States. ORCID: 0000-0002-2738-3190

[c] School of Information Sciences and Department of Educational Psychology, University of Illinois at Urbana-Champaign, 501 E. Daniel St., Champaign, IL 61820, United States. ORCID: 0000-0003-2736-2899

* Correspondence concerning this article should be addressed to Yingbin Zhang, Institute of Artificial Intelligence in Education, South China Normal University, No.55 West of Zhongshan Avenue, Guangzhou, Guangdong 510631, China. Email: zyingbin@m.scnu.edu.cn

Acknowledgment

We thank Prof. Carolyn Anderson for her early input while developing this work as well as Prof. Ma. Mercedes Rodrigo and Prof. Ryan S. Baker for sharing the affect datasets used in this study.

Funding

This work is partially supported by National Natural Science Foundation of China (62407014).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Yingbin Zhang: Conceptualization, methodology, formal analysis, writing - original draft preparation, review, and editing.

Luc Paquette: Conceptualization, methodology, writing - review and editing, supervision, funding acquisition.

Nigel Bosch: Conceptualization, methodology, writing - review and editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT in order to polish the paper. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Abstract: Understanding the transitions among affective states during computer-based learning may guide the design of affect-responsive learning environments. Current studies have focused on the marginal strength of an affect transition, which is the average transition tendency over possible affective states preceding the transition. However, marginal strength ignores the potential influence of the preceding state on the transition. In contrast, a conditional strength, which is the transition tendency given a particular state preceding the transition, accounts for this influence and may contribute to a more comprehensive understanding of students' learning processes. This paper presents a methodological framework that utilizes the logistic mixed model to compute the conditional strengths of affect transitions and examines whether conditional and marginal strengths are comparable. In three real-world datasets, we found that the conditional and marginal strengths of a transition were not identical in most cases. Prediction analysis indicated that accounting for the state preceding the transitions resulted in better affect prediction performance. In addition, empirical data analyses showed that the framework had higher power in detecting the impact of students' factors on affect and affect transitions. The framework also enables researchers to specify the reference transition when computing a transition strength and handle self-transitions, a critical issue in affect transitions. Empirical data analyses showed that the strength of a transition varied substantially when the reference transition changed, highlighting the careful selection of reference transitions in transition analyses.

Keywords: affect transition, emotion in computer-based learning, modeling human emotion, logistic mixed model

1 Introduction

The role of affect in learning has been widely recognized (Buhr et al., 2019; Chevrier et al., 2019; Theobald et al., 2021). Affect dynamics refers to how affective states change over time (Kuppens, 2015), and related work in computer-based learning focuses on the transition tendency between learners' affective states (Botelho et al., 2018; D'Mello & Graesser, 2012; Karumbaiah et al., 2021). For instance, when students are confused during learning, will they transition to frustration or engagement? Understanding such tendencies during learning guides the design of affect-responsive learning environments that provide timely scaffolding to learners who likely persist in vicious cycles of negative emotions (D'Mello et al., 2007).

Previous studies have focused on the marginal strength of affect transitions (Botelho et al., 2018; Caglar-Ozhan et al., 2022; D'Mello & Graesser, 2012; Karumbaiah et al., 2021; Ocumpaugh et al., 2017; Rodrigo et al., 2012), which is th average transition tendency over all possible affective states before the transition. For example, given four educationally important affective states—engagement, confusion, frustration, and boredom—we may be interested in the transition *confusion* → *frustration*. The average transition probability over the four possible states before confusion is a marginal strength, which does not consider the possible influence of the states before confusion on the transition. In contrast, a conditional strength, which is the transition probability given particular states before confusion, accounts for this influence.

The marginal strengths assume that, in a learning process, the type of the next affective state only relies on the current state, i.e., the affective process during learning is a first-order Markov chain (Bakeman & Gottman, 1997). Such an assumption may hold, but it has not been validated by prior studies. If the assumption does not hold and marginal strengths differs from conditional strengths (Agresti, 2013), ignoring conditional strengths may result in an inaccurate understanding of affect dynamics that may contradict theories. Accounting for the conditional strengths may also guide the development of more accurate and adaptive affect-responsive learning environments. However, the conditional strengths in affect transitions are underexplored.

This paper presents a framework that uses mixed-logistic models to compute both the marginal and conditional strengths of affect transitions and examine whether they are equal. In addition, the proposed approach can account for the impact of external factors (e.g., prior knowledge) on affect and affect transitions, which has been emphasized by prior studies (Karumbaiah et al., 2021; Morais & Jaques, 2024). It also enables researchers to specify the reference transition, the strength of which is the benchmark when computing the strengths of transitions of interest. This flexibility allows the removal of self-transitions, which refers to identical successive affective states and is a critical issue in prior affect transition studies (Karumbaiah et al., 2019; Matayoshi & Karumbaiah, 2020). We applied the method to three affect datasets collected during learning in computer-based environments, which involved different domains and learners at different ages. The results showed that conditional strengths of an affect transition could be heterogeneous and different from the marginal strength of the transition, which supports our call for analyzing conditional strengths of affect transitions for a more comprehensive understanding of the affective process during learning.

2 Literature review

2.1 Computing transition strength

This section introduces metric-based and model-based methods for quantifying the strength of affect transitions (and transitions of events in other modalities). It ends with a discussion on conditional and marginal strengths.

2.1.1 Metric-based methods

Generally, any association indicator for two binary variables or a 2×2 contingency table can quantify the transition strength (Bosch & Paquette, 2021). Assume that we want to compute the strength for transition $A \rightarrow B$ in an affect sequence S containing N+1 states. Table 1 helps us understand the similarities and differences among these metrics. The rows of Table 1 represent whether the first affective state is A in a pair of consecutive states in S, and the columns represent whether the second state is B. Let a, b, c, and d denote the occurrences of $A \rightarrow B$, $A \rightarrow \overline{B}$ (indicating not B), $\overline{A} \rightarrow B$, and $\overline{A} \rightarrow \overline{B}$. The sum of a, b, c, and d is N because there are N+1 pairs of consecutive states in S (assuming that S is an unbroken sequence). Taking *confusion* → *engagement* in sequence *{frustration, boredom, confusion, engagement, confusion, confusion, engagement, engagement}* as an example, $a, b, c,$ and d are 2, 1, 1, and 3, respectively.

The most often used metrics in affect dynamics during learning are L (D'Mello et al., 2007) and L* (Matayoshi & Karumbaiah, 2020). The difference between them is that L* is more accurate than L when consecutive identical affective states are code as a single state. This coding strategy is used when researchers are not interested in self-transitions (e.g., *confusion* → *confusion*). The coding removes self-transitions and prevents the suppression of self-transitions on the other transitions (Matayoshi & Karumbaiah, 2020). Appendix A1 describes the computational details for these and the other metrics.

Table 1

The Occurrences of Pairs of Consecutive Events

First event \ Second event	B	$\overline{B}$
A	a	b
$\overline{A}$	c	d

2.1.1.1 The impact of short sequence and event imbalance. The values of transition metrics derived from observed data are estimates of true values (Dagne et al., 2002). The observed values have measurement errors, which rely on sequence length and event balance (Bosch & Paquette, 2021; Dagne et al., 2002). Thus, when sequences are short and contain imbalanced event types, transition metrics are less accurate as indicators of event associations (Bosch & Paquette, 2021; Matayoshi & Karumbaiah, 2021). However, current studies tend to ignore measurement errors when examining the statistical significance of affect transitions. For example, Botelho et al. (2018) computed L per affect transition per learner. The t-test was conducted to examine the significance of each transition without considering the measurement errors of L. One way of controlling the measurement errors is creating a weight variable based on the measurement errors and using a weighted test. Alternatively, we may utilize regression models to address the inaccuracy issues (Dagne et al., 2002; Matayoshi & Karumbaiah, 2021), discussed in the next section.

2.1.2 *Model-based methods*

Researchers have applied generalized linear mixed models (GLMM) to overcome the inaccuracy of transition metrics (Dagne et al., 2003, 2007; Howe et al., 2005; Ozechowski et al., 2007). The GLMM approach accounts for the measurement error when estimating a transition strength. Both logistic and log-linear mixed models have been used (Dagne et al., 2007; Howe et al., 2005; Ozechowski et al., 2007). The non-mixed version of these models can also quantify the transition strength. However, the non-mixed version requires pooling all sequences into a single sequence and cannot account for the nested property of the data (e.g., affect observations are nested within sequences or learners). Pooling data may cause misleading estimates (Wickens, 1993). Thus, the non-mixed version is not discussed.

Matayoshi and Karumbaiah (2021) applied a generalized estimating equation (GEE) to model the marginal strengths of affective transition. GEE also accounts for the measurement error and nested property between affect observations and sequences. The rationale for modeling transition strength is the same between the GEE approach and the logistic mixed model, but the GEE approach does not estimate the variation of transition strengths across individuals (i.e., the random effect). Although this variation itself may not be appealing, affect dynamics researchers have shown interest in how this individual difference is related to environmental and learners' characteristics (Andres et al., 2019; Rodrigo et al., 2012). Therefore, this paper does not discuss the GEE approach in detail.

2.1.2.1 Logistic mixed model. Let $A \rightarrow B$ denotes a transition of interest and assume there are M sequences. Let $S_m = \{e_1, e_2, \ldots, e_{n_m}\}$ denotes sequence m, where n_m is the number of affective states in sequence *m* and varies across learners. Let X_{mj} denotes whether A is the first state in the j^{th} pair of consecutive states in S_m. If the first event is A, X_{mj} = 1; otherwise X_{mj} = 0. For instance, for *confusion* →*engagement* in $S_m =$

$\{frustration, boredom, confusion,\ \ engagement\}$, X_{m1} and X_{m2} = 0 but X_{m3} = 1. Let $p(B|X_{mj})$ denotes the probability that the second event is B conditional on X_{mj}. This probability can be modeled by the logistic mixed model (Ozechowski et al., 2007), as shown in Equation (1).

Level 1: pair of events

$$ln\frac{p(B|X_{mj})}{1-p(B|X_{mj})} = \beta_{0m} + \beta_{1m}X_{mj}$$

Level 2: sequence

$$\beta_{0m} = \gamma_{00} + U_{0m}$$

$$\beta_{1m} = \gamma_{01}\ . \tag{1}$$

The term γ_{00} is the fixed effect of intercept and represents the expected log odds of the second state being B, given that $U_{0m} = 0$ and X_{mj}=0 (i.e., the first state is A). The term U_{0m} is the random effect of the intercept and represents the random variation of this log odds across learners. The term γ_{01} is the fixed effect of the slope and represents the difference in the log odds between the conditions that the second state is B when the first state is A versus $\bar{\text{A}}$. If γ_{01} > 0, $A \rightarrow B$ more likely occurs than $\bar{A} \rightarrow B$. That is, B is more likely to occur after A than $\bar{\text{A}}$. In contrast, if γ_{01} < 0, B is less likely to occur after A than $\bar{\text{A}}$. If γ_{01}= 0, B is independent of whether the first state is A.

The logistic mixed model can only estimate transitions with the same target state at a time. Thus, it is suitable when we are interested in specific transitions and have a priori assumption.

2.1.2.2 Log-linear mixed model. We need to convert each sequence into a $K \times K$ contingency table to apply the log-linear mixed model for estimating the transition strengths, where K is the number of affect categories. For instance, sequence *{frustration, boredom, confusion, engagement, confusion, confusion, engagement, engagement}* can be converted into Table 2. The row indicates the category of the first state in a pair of consecutive states, and the column indicates the category of the second state. Cell values are the occurrences of corresponding transitions, i.e., the occurrences of transitions from the row state to the column state. The strengths of all transitions between the row and column states can be estimated within a single log-linear mixed model (see Appendix A2 for details). Thus, for an exploratory analysis with the goal of finding any significant transitions, the log-linear mixed models are a better option.

Table 2

Contingency Table for the Example Sequence {frustration, boredom, confusion, engagement, confusion, confusion, engagement, engagement}

Frequencies (μ_{mrc})	frustration	boredom	confusion	engagement
frustration	0	1	0	0
boredom	0	0	1	0
confusion	0	0	1	2
engagement	0	0	1	1

Note. Cell values are the occurrences of transitions from the row state to the column state.

2.1.3 Conditional and marginal strengths of a transition

This section illustrates the conditional and marginal strengths of a transition, how they may differ, and under what conditions they are equal. Assume that we are interested in *confusion* → *engagement*. All states before the transition may influence the strength, but how many states before the transition are considered does not influence the definitions of marginal and conditional strengths. Moreover, from a computational perspective, accounting for too many states before the transition requires a much larger dataset, which is difficult to collect in the real world. Thus, to avoid unnecessary complexity in distinguishing the marginal and conditional strengths, this paper focuses on the first state before the transition. We discuss how to investigate the impact of earlier states in the first paragraph of section 5.2.

The transition strength derived from the first two rows of Table 3 is a conditional strength because this value is under the condition that the first affective state is engagement. This is the same for the transition strength derived from the second two rows. The derived strength is marginal if we do not consider the first affective state, i.e., computing the transition based on the last two rows.

In contingency table analyses, a table composed of the first or second two rows is a partial table, while a table consisting of the last two rows is a two-way marginal table (Agresti, 2013). The first affective state is a confounding variable that needs to be controlled because it may influence the association between the second and third affective states. For instance, when the first state is engagement, the transition metric L for *confusion* → *engagement* is 1, and when the first state is confusion, L is 0.342. However, when aggregating over the first states, L for *confusion* → *engagement* becomes

-0.689. This phenomenon is known as the Simpson's paradox (Agresti, 2013). Cornfield et al. (1954; as cited in Greenhouse, 2009) demonstrated that the Simpson's paradox occurs when the confounding variable (the first affective state) is strongly related to the variables of interest (the second and third affective states). It is unknown whether there are strong associations in one dataset, and it may be better to examine this question rather than ignore it.

Table 3

Artificial Data for the Transition Confusion → Engagement

First state	Second state	Third state		Total
		Engagement	Not engagement	
Engagement	Confusion	10	0	10
	Not confusion	199	10	209
Confusion	Confusion	29	20	49
	Not confusion	1	30	31
Total	Confusion	39	20	59
	Not confusion	200	40	240

Moreover, the conditional and marginal strengths of a transition may still differ, even though all conditional strengths of this transition are identical (i.e., homogeneous associations). The conditional and marginal associations between the two variables of interest are the same only under the collapsibility condition (Agresti, 2013), where two assumptions must hold:

1) *Homogeneous assumption*: *The conditional associations between the two variables of interest are the same*. This requirement is named homogeneous

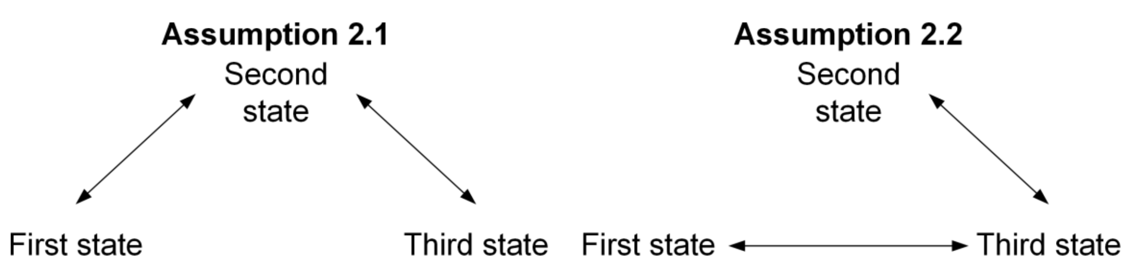
associations. However, affect transition research has not examined this assumption.

2) *Independence assumption*: *The confounding variable is conditionally independent of one of the variables of interest*. For example, in Table 3, this assumption entails that: 2.1) given the third affective state, the second affective state is independent of the first affective state; or 2.2) given the second affective state, the third affective state is independent of the first affective state. Figure 1 illustrates the two assumptions. Assumption 2.1 does not make sense because the third affective state occurs after the first and second states. A future affective state cannot impact past affective states. Assumption 2.2 seems reasonable, but it assumes that no association at lag two, i.e., no association between the first and third affective states.

Thus, the assumptions of the collapsibility condition may not hold in the transition analysis of educational data. We should not ignore the conditional strengths when investigating affect transitions.

Figure 1

The First Affective State is Conditionally Independent of the Second or Third State

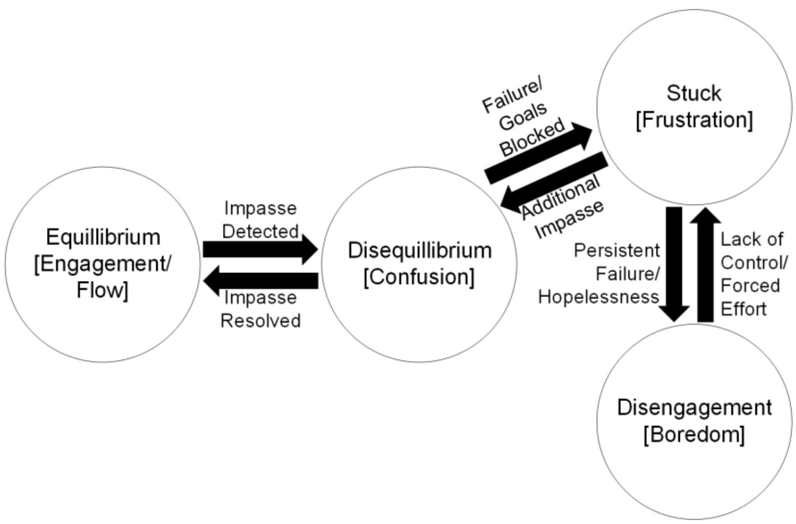


2.2 Affect transitions in computer-based learning

Affect dynamics research in computer-based learning focuses on the transition between affective states (Botelho et al., 2018; D'Mello & Graesser, 2012; Karumbaiah et al., 2021). The most cited theory in this direction is D'Mello and Graesser's model (2012), a state transition network that covers four educationally important emotions (Figure 2): engagement/flow, confusion, frustration, and boredom.

Figure 2

D'Mello and Graesser's (2012) Theoretical Model of Affect Dynamics



This model emphasizes the role of cognitive disequilibrium in the affect trajectory. Cognitive disequilibrium arises when there is an impasse, which is a conflict between new information and a learner's current knowledge structure (Mandler, 1990). Confusion is the affective sign of cognitive disequilibrium. A learner processes and assimilates encountered information when they are engaging in a task. If their knowledge structure cannot assimilate the new information, i.e., an impasse occurs, the learner transitions to a confused state. If the learner solves the impasse, they transition back to flow. By contrast, if the learner cannot resolve the impasse, the prolonged confusion transitions to frustration. Additional impasses induce the transition from frustration to confusion. Persistent frustration causes disengagement (boredom). Boredom may transition back to frustration if the learner must endure the task.

Despite high citations, empirical studies have only partially supported this model (Karumbaiah et al., 2021). For example, in Botelho et al. (2018), there was no significant transition from engagement to confusion and from frustration to confusion. The researchers explained that the reason might be there were not enough occurrences of confusion. However, other transitions also did not follow the theoretical model. The transitions from confusion to frustration and from frustration to boredom occurred at a rate lower than chance, meaning that these transitions were unlikely to happen. There were also significant transitions not predicted by the model, such as engagement to boredom and frustration to engagement. Karumbaiah et al. (2021) reanalyzed twelve datasets used by prior studies with transition metric L^* and aggregated the results from individual datasets to overcome the issue of small sample sizes. Nevertheless, the only significant transition in line with D'Mello and Graesser's model was *engagement → confusion*. The researchers aggregated the results across datasets from the same country and compared the results between countries (the United States versus the Philippines). In the data from the United States, there were significant *engagement → confusion*, *engagement → frustration*, *confusion → engagement*, and *boredom → engagement*. In contrast, there was no significant transition in the data from the Philippines. The researchers concluded that culture might be critical in affect dynamics.

In addition to the four emotions in the theoretical model, researchers have investigated the transitions among other emotions during digital learning. For instance, Ocumpaugh et al. (2017) found that, in vMedic, a computer-based environment teaching combat medicine protocols, the transition from surprise to anxiety was less likely than chance, while the transition from anxiety to surprise was at a chance rate. Caglar-Ozhan et al. (2022) examined teachers' affect transitions in a simulated virtual classroom platform. The analysis indicated that eight transitions occurred at a rate more than chance, including four transitions between different affective states (e.g., *fear* → *disgust*) and four self-transitions (e.g., *happiness* → *happiness*).

In summary, the conditional and marginal strengths of affect transitions may not be identical, and investigating the conditional strength contributes to a more comprehensive understanding of affective process during learning. However, prior studies have mainly investigated the marginal strength of affect transition during learning. There is also no methodological framework on how to compute the conditional strength of a transition and examine whether it is equal to its marginal strength. In addition, model-based approaches can better handle short sequence and event imbalance, but how to tackle self-transitions with the approaches is unclear. Moreover, studies have emphasized the impact of external factors on affect transitions (Karumbaiah et al., 2021; Morais & Jaques, 2024), but it is not clear whether we could address this topic with the model-based approaches.

3 A framework for computing transition strengths

This section proposes a framework to compute the conditional strengths of affect transitions, examine whether it is equal to the marginal strength, and analyze the impact of external factors on affect transitions. Affect datasets in education are often collected from multiple students, and researchers are mainly interested in the transition strength across students. If we use transition metrics to quantify transition strengths, we need to compute the metric for each student's sequence and aggregate these metric values over students to determine the association strength and significance. The observed value of a transition metric has measurement errors, so it is not recommended to compute these metrics in each sequence and average them over sequences (Dagne et al., 2007; Matayoshi & Karumbaiah, 2021). Thus, our framework uses GLMM to analyze the strengths of transitions.

It is useful to discuss the data preprocessing for affect transition analyses briefly. Prior studies have included self-transitions when the focus is the persistence of individual affective states but removed self-transitions if focusing on transitions between different affective states, which may be suppressed by self-transitions (Karumbaiah et al., 2018, 2019, 2021). However, with GLMM, the suppression effect of self-transitions can be controlled without excluding self-transitions. Section 3.1 elaborates on this point. Moreover, including self-transitions increases data points and makes GLMM produce more accurate estimates. Thus, self-transitions were not removed.

Below, we introduce the following topics sequentially: computing the marginal transition strength, computing the conditional transition strength, examining the assumption of identical marginal and conditional strengths, and investigating the impact of external factors on transitions.

3.1 Computing marginal strengths of affect transitions

The framework starts by examining the marginal strengths of affect transitions for two reasons. First, GLMM has not been applied to the field of affect transitions in prior work.

The marginal strengths of affect transitions in GLMM enables researchers to check to what extent the results based on GLMM match prior findings, where L and L^* have been mainly used to quantify the transition strengths. Second, a GLMM for modeling the marginal strength of a transition is much simpler than that for conditional strengths. Starting with a simpler model allows for easier diagnosis in cases of any estimation issue.

Both mixed logistic and log-linear models have been used for transition analyses, although not in the field of affect dynamics (Dagne et al., 2007; Ozechowski et al., 2007). Estimating the conditional strengths of affect transitions entails converting a student's affect sequence to a three-way contingency table. With the log-linear approach, the data size of a study may not be sufficient for analyzing the conditional strength. For example, estimating the conditional strengths of transitions among engagement, confusion, frustration, and boredom requires converting an affect sequence into a 4×4×4 table containing 64 cells. However, the average length of affect sequences is typically in the tens (e.g., Karumbaiah et al., 2021). Such a length means that the counts in a large proportion of cells would be smaller than five and even be zero. In addition, affective states typically follow a largely imbalanced distribution, and states like boredom and frustration may rarely occur in a sequence with tens of affective observations. Consequently, marginal zeros, which refer to the total counts in a row or column being zero, are likely to appear in the contingency table. Fitting a log-linear model to such sparse contingency tables with marginal zeros would encounter such issues as convergence failures and inaccurate estimates (Fienberg, 1979; Wickens, 1989). By contrast, a mixed logistic model suffers less from these issues. With the mixed logistic model, we convert an affect sequence to a 4×4×2 table (see Table 4), which contains half the number of cells compared to a mixed log-linear model. In addition, marginal zeros impact weakly the parameter estimation for a mixed logistic model. Thus, the framework adopts the logistic approach.

In Table 4, the first, second, and third affective states refer to the state in the corresponding position of a three-affect subsequence. Thus, the cell value represents the occurrences of a subsequence of three affective states. For instance, the subsequence consisting of three engagement observations occurred seven times, which is the cell value in the first row and first column. Our framework focuses on whether the transition strength between the second and third states is conditional on the first state. Thus, for clarity, the remainder of this paper named the first state as the conditional state, the second state as the given state, and the third state as target state.

Table 4

An Example of the Contingency Table for Analyzing Transitions with Engagement as the Target Affective State

$\boldsymbol{\mu_{mcgy}}$		**Third/target state (y)**	
First/conditional state (c)	Second/given state (g)	Engagement ($y = 1$)	Others ($y = 0$)
Engagement	Engagement	7	14
	Confusion	4	3
	Frustration	4	1
	Boredom	3	1
Confusion	Engagement	2	2
	Confusion	2	1
	Frustration	6	0
	Boredom	1	0

Frustration	Engagement	2	4
	Confusion	5	1
	Frustration	10	5
	Boredom	1	1
Boredom	Engagement	1	1
	Confusion	1	0
	Frustration	4	0
	Boredom	3	0

Table 4 is constructed for analyzing transitions where the target affective state is engagement. Similar tables can be used for analyzing transitions to other target affective states. The remainder of this section uses transitions to engagement as an example to illustrate the modeling procedure. We use Equation (2) to estimate the marginal strengths of transitions between given and target affective states. $\mu_{m\bullet g1}$ is the sum of the corresponding cell values in student *m*'s table, i.e., the sum of cell values where the given affective state is g and the target affective state is engagement, regardless of the conditional affective state. By contrast, $\mu_{m\bullet g0}$ is the sum of cells where the given affective state is g but the target affective state is not engagement. The intercept β_{00m} equals ln $(\mu_{m\bullet 01}/\mu_{m\bullet 00})$, where $\mu_{m\bullet 01}$ is the sum of cells where the given affective state is the reference state and the target affective state is engagement. We explain the reference state below. The terms γ_{00} and U_{0m} are the fixed and random effects of intercept, respectively.

Level 1: cell

$$ln\frac{\mu_{m\bullet g1}}{\mu_{m\bullet g0}} = \beta_{00m} + \sum_{g=1}^{G} \beta_{m\bullet g} X_{m\bullet g}$$

Level 2: student

$$\beta_{00m} = \gamma_{00} + U_{00m}$$
$$\beta_{m\bullet g} = \gamma_{\bullet g} \,. \tag{2}$$

The term $X_{\mathrm{m}\bullet\mathrm{g}}$ is the dummy variable for the given affective state. The dummy coding determines the reference state and should be based on research purposes and context. For instance, when the target affective state is engagement, the dummy coding in Table 5 can be used to examine D'Mello and Graesser's model (2012) and control the influence of self-transitions. First, based on this model, only confusion is likely to transition into engagement. Thus, $X_{\mathrm{m}\bullet 2}$ is used to indicate whether the given affective state is confusion, and $\gamma_{\bullet 2}$ models the strength of *confusion* → *engagement* (we will explain this point in the next paragraph). Second, self-transitions may suppress the transitions between different affective states (Karumbaiah et al., 2021). If using all affective states other than confusion as the reference, the strength of *confusion* → *engagement* may be underestimated. Thus, we use $X_{\mathrm{m}\bullet 1}$ to indicate whether the given affective state is engagement and control the self-transition of engagement. The affective states not indicated by $X_{\mathrm{m}\bullet 1}$ and $X_{\mathrm{m}\bullet 2}$, including frustration and boredom, are the reference given states. *G* in Equation (2) is the number of non-reference given states. For the dummy coding in Table 5, *G* equals two because non-reference given states include confusion and engagement. For an exploratory study without theoretical guidance, all states except for the first state in the transition can be the reference state. We will return to this point in section 5.1.1.

Table 5

Dummy Coding for the Given Affective State When the Target Affective State Is Engagement

Dummy variable $\boldsymbol{X_{m \bullet g}}$	Given affective state			
	Engagement	Confusion	Frustration	Boredom
$X_{m \bullet 1}$	1	0	0	0
$X_{m \bullet 2}$	0	1	0	0

By putting $\beta_{00m} = \ln(\mu_{m \bullet 01}/\mu_{m \bullet 00})$ into level 1 of Equation (2) and shifting items, we get $\gamma_{\bullet 2} = \ln\,(\mu_{m \bullet 21} * \mu_{m \bullet 00}/(\mu_{m \bullet 20} * \mu_{m \bullet 01}))$, which is a log-odds ratio. Thus, $\gamma_{\bullet 2}$ represented the natural log of the ratio of the odds for *confusion → engagement* to the odds for the transition from the reference states (frustration and boredom) to engagement. That is, the transition from frustration and boredom to engagement is the reference.

3.2 Computing conditional strengths of affect transitions

The conditional affective state and its interactions with the given affective state are added to Equation (2) to analyze conditional transitions. The model becomes Equation (3).

Level 1: cell

$$ln\frac{\mu_{mcg1}}{\mu_{mcg0}} = \beta_{00m} + \sum_{c=1}^{C} \beta_{mc \bullet} X_{mc \bullet} + \sum_{g=1}^{G} \beta_{m \bullet g} X_{m \bullet g} + \sum_{g=1}^{G} \sum_{c=1}^{C} \beta_{mcg} X_{mc \bullet} X_{m \bullet g}$$
$$= \beta_{00m} + \sum_{c=1}^{C} \beta_{mc \bullet} X_{mc \bullet} + \sum_{g=1}^{G} \left(\beta_{m \bullet g} + \sum_{c=1}^{C} \beta_{mcg} X_{mc \bullet}\right) X_{m \bullet g}$$

Level 2: student

$$\beta_{00m} = \gamma_{00} + U_{00m}$$
$$\beta_{mc \bullet} = \gamma_{c \bullet}$$

$$\beta_{m\bullet g} = \gamma_{\bullet g}$$
$$\beta_{mcg} = \gamma_{cg}\ . \tag{3}$$

Taking Table 4 as an example, μ_{mcg1} and μ_{mcg0} are the counts when the target affective state is engagement and not, respectively, in the case that the conditional affective state is c, and the given affective state is g. The term $X_{mc\bullet}$ is the dummy variable for the conditional affective state, and Table 6 displays the dummy coding. Because the association between conditional and target affective states is not the focus of our framework, we use boredom as the reference conditional state.

Table 6

Dummy Coding for the Conditional Affective State

Dummy variable	Conditional affective state			
$\boldsymbol{X_{mc\bullet}}$	Engagement	Confusion	Frustration	Boredom
$X_{m1\bullet}$	1	0	0	0
$X_{m2\bullet}$	0	1	0	0
$X_{m3\bullet}$	0	0	1	0

The term β_{mcg} represents the interaction effect between the conditional and given affective states. The term $\gamma_{\bullet g}$ is the transition strength from the given affective state g to engagement when the conditional affective state is boredom, while γ_{cg} represents the strength difference of this transition when the conditional affective state is c versus boredom. For example, $\gamma_{\bullet 2}$ represents the transition strength from confusion to engagement when the conditional affective state is boredom because $X_{\mathrm{m}\bullet 2}$ is the dummy variable indicating whether the given affective state is confusion (see Table 5). By contrast, γ_{12} represents the difference in the transition strength when the conditional affective state is engagement versus boredom because $X_{m1\bullet}$ represents whether the conditional affective state is engagement. That is, when the conditional affective state is engagement, the transition strength from confusion to engagement is $\gamma_{\bullet 2} + \gamma_{12}$.

3.3 Examining the assumption of identical conditional and marginal affect transitions

Recall that the conditional and marginal strengths of a transition are equal only if the following assumptions hold:

1) *The homogeneous transition assumption entails that the transition strength from the given affective state to the target affective state does not change by the conditional affective states*. In Model (3), this assumption can be represented as $\gamma_{cg} = 0$. To examine whether the assumption holds in this study, we can analyze the data with a simpler version of Model (3), where all γ_{cg} are set to zero, i.e., removing variable $X_{mc\bullet}$. Since the two models are nested, a likelihood ratio test (LRT) can inform whether the simpler model fits the data as well as Model (3). In addition, GLMM researchers recommend information criteria as a complementary approach to the LRT for model comparisons, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Bolker et al., 2009). Unlike the LRT, which focuses on estimating *p*-values, the difference in information criterion between two models quantifies the extent that the assumption does not hold. The smaller the difference in information criterion, the more evidence the simpler model fits the data as well as Model (3). If the LRT generates a *p*-value lower than the researcher-specified significance level (e.g.,

0.05), and the difference in information criterion is small, we may conclude the first assumption holds. An intuitive explanation for this assumption is that the conditional state does not moderate the effect of given state on the target state

2) *The target affective state is independent of the conditional affective state when the given affective state is controlled*. In Model (3), the assumptions can be represented as $\gamma_{cg} = 0$ and $\gamma_{c\bullet} = 0$, respectively. If the homogeneous association assumption holds, and we want to test the second assumption, we can compare a version of Model (3) without variable $X_{mc\bullet}$ and a version of Model (3) without variables $X_{mc\bullet}$ and $X_{mc\bullet}$. Again, the LRT and information criteria can be used to examine whether the second assumption holds. An intuitive explanation for this assumption is no need to use the conditional state as a covariate in the transition analysis because it does not influence the target state.

If both assumptions hold, we may safely ignore conditional transitions. Otherwise, conditional transitions should be reported together with the marginal transition.

3.4 Investigating the impact of external factors

To examine whether the transition strength is related to an external factor Z (e.g., whether learners with different prior domain knowledge show differences in the strength of *confusion → engagement*), we can add the factor into equations (2) and (3) by letting $\beta_{00m} = \gamma_{00} + \gamma_{01} Z_m + U_{00m}$ and $\beta_{m\bullet g} = \gamma_{\bullet g} + \gamma_{11} Z_m$. γ_{01} represents the main effect of Z on the target state, i.e., the extent that Z influences the probability of target state. γ_{11}

represents the moderation effect of Z on the transition strength, i.e., the extent that the transition strength relies on Z.

4 Empirical examples

This section applies our framework to three datasets used by prior affect dynamics research to illustrate the application of the framework. Specifically, we examined four research questions: (1) Whether conditional and marginal strengths are equal in real datasets? (2) Does a model with both conditional and given states have better prediction performance on target states than a model with only given states? (3) Is prior knowledge related to transition strengths? (4) How does the reference transitions influence the strength estimate of transitions of interest?

4.1 Datasets

The datasets were from three computer-based learning environments, Physics Playground, vMedic, and Betty's Brain[1]. We used these datasets based on the availability and the considerations of sequence lengths and observation intervals. Long sequences ensure more accurate strength estimates. With small observation intervals, the probability that a student's affect transitioned twice between two successive observations was low. For example, given that two consecutive observations were confusion and frustration, it was unlikely that the student also experienced boredom between confusion and

[1] To access to these datasets, readers may contact Ma. Mercedes Rodrigo for the Physics Playground dataset, Ryan S. Baker for the vMedic dataset, and Gautam Biswas for the Betty's Brain dataset.

frustration because the interval between confusion and frustration was too small to allow two transitions (i.e., *confusion* $\rightarrow$ *boredom* and *boredom* $\rightarrow$ *frustration*).

All datasets were collected using the Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP; Ocumpaugh, Baker, & Rodrigo, 2015). BROMP was developed for quantitative field observations of affect and on-task behaviors. The tool was implemented in the Android App HART (Ocumpaugh, Baker, & Rodrigo et al., 2015), which added a timestamp to each observation. While students were learning, certified observers recorded both students' affect and behaviors (not the focus of the current study).

The Physics Playground dataset was collected from 120 eighth-grade students in two cities and 60 tenth-grade students in a third city in the Philippines (Andres et al., 2015). Physics Playground is a game designed to teach middle school students about qualitative physics. The students played a computer game for 30 to 90 minutes in classrooms. Each student was observed on average once every 32 seconds (SD = 15.48) and 155 times (SD = 56), resulting in 27,918 observations in total.

The vMedic dataset was from 119 West Point cadets in USA (Ocumpaugh et al.; 2017). vMedic is a simulation system that provided training in combat medicine and battlefield doctrine around medical first response. The cadets used vMedic for up to 25 minutes. Each was observed on average once every 99 seconds (SD = 66.89) and 20 times (SD = 14.93), resulting in 2,643 observations in total.

The Betty's Brain dataset was from 93 sixth-grade students in an urban public school in Tennessee, USA (Munshi et al., 2018). Betty's Brain is an open-ended environment that teaches scientific phenomena . The students used the system for 160 to

200 minutes over four days. Each was observed on average once every 258 seconds (SD = 141.06) and 54 times (SD = 43.73), resulting in 5,177 observations in total.

4.2 Data preprocessing

The marginal strengths of affect transitions in Physics Playground and vMedic datasets have been investigated by Karumbaiah et al. (2021) and Matayoshi and Karumbaiah (2021). These studies analyzed all possible transitions among engagement, confusion, frustration, and boredom. Unlike these studies, the current analysis focused on the six transitions in D'Mello and Graesser's model (2012). Correspondingly, we removed transitions involving affective states not in D'Mello and Graesser's model (2012), except for the vMedic dataset. We did not remove the other states in the vMedic dataset because these states accounted for 1,061 observations, and removing these observations would decrease the number of transitions from 2336 to 1275, resulting in less accurate estimates of transition strengths. Thus, we included the other states and coded these states as *others* in the analysis of vMedic data.

In the Physics Playground and vMedic datasets, we removed transitions involving observations with an interval longer than 3 minutes to mitigate the probability that more than one transition happened within the interval. In the Betty's Brain dataset, we used 4 minutes as the threshold because using 3 minutes would result in so few observations that the model for transitions to confusion did not converge and yielded extremely large standard errors for all estimates. Then, we counted each learner's valid transitions and only kept those with no less than 10 transitions to ensure relatively reliable estimates of transition strengths. Table 7 displays the characteristics of each dataset after the above preprocessing.

Table 7

The dataset characteristics after preprocessing

Dataset	Sequences	Transitions per sequence	Proportion in the dataset			
			Engagement	Confusion	Frustration	Boredom
Physics Playground	180	112.03	80%	8%	7%	5%
vMedic	75	24.51	51%	9%	2%	7%
Betty's Brain	76	20.97	83%	6%	5%	5%

4.3 Analysis

To examine the assumption of identical marginal and conditional strengths, which contains two sub-assumptions: homogeneous and independence assumptions, we fitted three models for each target state: (1) with only given state, (2) with both given and conditional states, and (3) with interactions between given and conditional states. We compared model 2 and 3 to examine the homogeneous assumption based on LRT and information criteria. If model 3 fit the data better than model 2, the assumption is rejected. Otherwise, we continued to compare model 2 and 1. If model 2 fit the data better than model 1, the independence assumption is rejected.

The Betty's Brain dataset contained domain knowledge test scores, which was collected before students used Betty's Brain. We used the scores as an indicator of prior knowledge and added it to the logistic-mixed models to illustrate the analysis of the impact of external factors on transitions. As a comparison, we also computed the linear correlation between the scores and the transition strengths measured by L^*, a commonly used transition metric. Such method has been frequently used to investigate the impact of environmental and learner characteristics on transitions (Andres et al., 2019; Morais & Jaques, 2024).

To compute the model performance in predicting affective states, we used 10-fold cross-validations at the student level, where the model was fitted to 90% of students and used for prediction on the remaining 10% of students. Only the fixed effects were used for prediction, and thus, the model became a logistic model. We used four performance metrics: the accuracy, AUC, the F1 score, and Cohen's kappa. We compared the models with only given states and with both given and conditional states to illustrate the importance of accounting for conditional states.

All the models were implemented via the *glmer()* function of the *lme4* library (version 1.1-27.1; Bates et al., 2015) in *R*[1]. The Nelder-Mead method was used for parameter optimization. Approximate *t*-tests with the residual degrees of freedom were used for inferences on the fixed effects. The cross-validation and the performance metrics were implemented via the *caret* library in *R.* An R script for the analysis and a runnable example dataset are available in GitHub[2].

[1] Some models with interactions between conditional and given affective states failed to converge under the default convergence criterion (0.002). Further inspection showed that the maximum likelihood estimation was close to the Bayesian estimation (obtained via the *brms* library, version 2.17.0; Bürkner, 2017), with differences around 0.01. Thus, the maximum likelihood estimation of these models was reported to be consistent with the other models.
[2] https://github.com/yingbinz/mixed_model_transition/tree/main

4.4 Comparing conditional and marginal affect transitions

4.4.1 The assumption of identical conditional and marginal affect transitions

In four of the 12 cases (four sets of models with different target states × three datasets; see Table 8), the homogeneous transition assumption was rejected based on the combination of LRT and information criteria (see Tables B2.2, B2.4, B3.1, and B3.2). The results mean that, in these cases, the conditional state moderated the transition strengths between given and target states. For instance, in the Physics Playground dataset, the coefficient of given state engagement on target state confusion (i.e., *engagement → confusion*) was 0.36 when the conditional state was engagement and 0.83 when not (see Tables B2.2). This implies that, for this sample and learning environment, when students were in successive engaged states, they were less likely to transition into a confused state. By contrast, when they just engaged, they were more likely to transition into a confused state.

Among the eight cases where the homogeneous assumption was not rejected, the independence assumption was examined and rejected in five cases based on the combination of LRT and information criteria. This meant that, although the conditional state did not moderate the transition strengths between given and target states, it influenced target states and was a covariate that should be controlled when estimating the transition strengths. Otherwise, its effect on the target states may be incorrectly attributed to the given states, resulting in inaccurate estimates of the transition strengths. For example, in the Betty's Brain dataset, the coefficients of conditional states engagement and confusion on target state engagement were statistically significant (γ = 1.17 & 1.07, SE = 0.28 & 0.38, $p < 0.01$). When the model did not contain these conditional states, the coefficient of given state confusion on target state engagement (i.e., the marginal strength of *confusion* → *engagement*) was statistically significant (γ = 0.60, SE = 0.30, p = 0.04). Adding the conditional decreased the coefficient of given state confusion and made it insignificant (i.e., the conditional strength; γ = 0.49, SE = 0.29, p = 0.11). Similar differences in the transition strengths between controlling conditional states and not controlling them existed in the other cases where the independence assumption was rejected, although the conditional strengths were still significant in some of these cases.

Table 8

The homogeneous transition and independence assumptions in the datasets

Target state	Dataset	Homogeneous transition assumption	Independence assumption
Engagement	Physics playground	√	×
	vMedic	×	/
	Betty's Brain	√	×
Confusion	Physics playground	×	/
	vMedic	×	/
	Betty's Brain	√	√
Frustration	Physics playground	√	×
	vMedic	√	√
	Betty's Brain	√	×
Boredom	Physics playground	×	/
	vMedic	√	×
	Betty's Brain	√	√

Note. √: not rejected. ×: rejected.

4.4.2 *Prediction performance with and without conditional affective states*

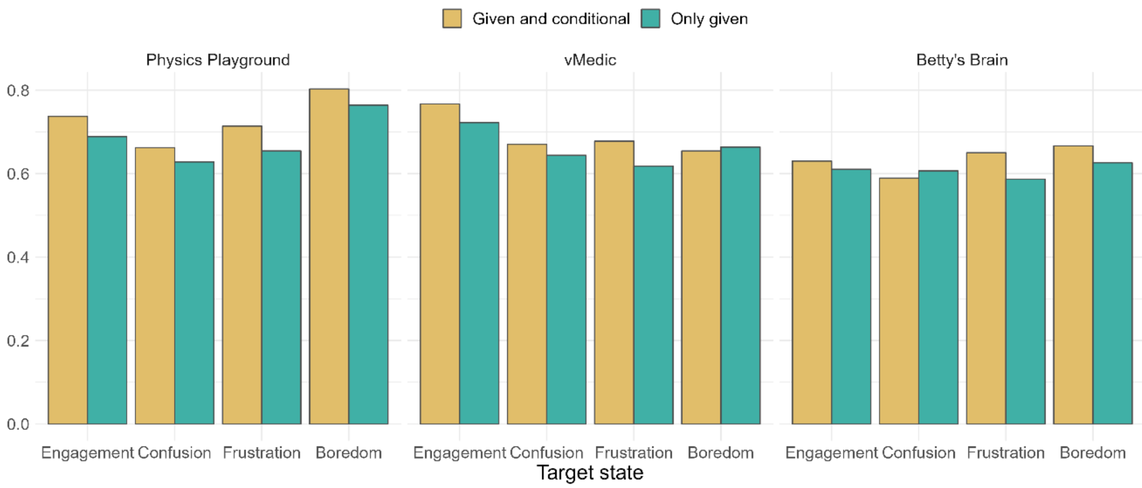
The prediction performance between the models with and without conditional affective states were almost identical in accuracy and F1 (see Appendix B5). Adding conditional states improved AUC with a magnitude ranging from 3% ~ 11% in 10 cases and slightly decreased AUC by 1% and 3% in two cases (see Figure 3). Note that AUC evaluates prediction performance across all thresholds of the predicted probabilities, while accuracy and F1 are based on a fixed threshold (0.5). Thus, the increase in AUC and no increase in accuracy and F1 suggest that adding conditional states resulted more reliable predicted probabilities, but such improvement was not sufficient to change the overall number of correct predictions at the fixed threshold.

When the target state was engagement, the model with given state predicted all samples as engagement in the Physics playground and Betty's Brain datasets, and thus, kappa was zero. Adding conditional states to the model changed the invariant prediction and improved kappa by 0.09 and 0.15. When the target state was confusion, frustration, or boredom, kappa was zero, regardless of adding conditional states.

In summary, the assumption of identical conditional and marginal strengths of affect transitions did not hold in nine of the 12 cases. Accounting for the conditional strengths of transitions improved the affect prediction in some cases.

Figure 3

The prediction performance in AUC



4.5 The impact of external factors on affect transitions

We selected the models fitting the Betty's Brain data best in Table 8 as the base model and added prior knowledge to these models to investigate the impact of prior knowledge on affect transitions. Tables 9 and 10 display the results when the target state was engagement, confusion, frustration, and boredom, respectively. Since the best-fitting models in Table 8 varied by the target state, the models in Tables 9 and 10 for different target states contained different variables. Prior knowledge had a positive main effect on engagement and a negative effect on frustration and boredom. This is in line with expectation because students with more prior knowledge were less likely to encounter impasses during learning that could not be solved, and thus, they experienced less frustration and boredom.

As for the impact of prior knowledge on transitions, it moderated the effect of given state confusion on engagement (*confusion* → *engagement*; see Table 9) and the effect of given state confusion on frustration (*frustration* → *confusion*; Table 10). The moderation on *frustration* → *confusion* was negative. This moderation effect could be explained by D'Mello and Graesser's model (2012), which claims impasses during frustration induce the transition from frustration to confusion. Students with more prior knowledge were less likely to encounter impasses, and thus, they were less likely to transition to confusion during frustration than those with lower prior knowledge. The moderation on *confusion* → *engagement* was negative. This means that, when students were confused, those with higher knowledge were less likely to transition to engagement. This is unexpected as students with higher knowledge are expected to resolve confusion and become engaged.

Table 9

The impact of prior knowledge on engagement and confusion

Target-engagement	Estimate (SE)	Target-confusion	Estimate (SE)
Fixed effect			
Intercept	−0.24 (0.31)	Intercept	−3.44 (0.74) ***
c-Engaged	1.12 (0.27) ***	g-Engaged	0.28 (0.75)
c-Confusion	1.06 (0.38) **	g-Confusion	1.40 (0.82)
c-Frustration	0.39 (0.37)	g-Frustration	0.67 (0.94)
g-Engaged	1.22 (0.20) ***	Knowledge	0.19 (0.10)
g-Confusion	0.39 (0.30)	g-Engaged*Knowledge	−0.16 (0.11)
Knowledge	0.11 (0.03) **	g-Frustration*Knowledge	−0.58 (0.22) **
g-Confusion*Knowledge	−0.25 (0.09) **		
Random effect			
Intercept	0.42	Intercept	0.74

Note. The reference given state was boredom and frustration when the target state was engagement, and was boredom when the target state was confusion. g-: given state. c-:

conditional state. The units of coefficients were the natural log of odds between cell counts, and values in the parentheses were standard errors. Prior knowledge was mean-centered.

For the linear correlation between prior knowledge and transition strengths measured by L^*, prior knowledge was only related to the L^* of *frustration* → *confusion* (r = -0.27, p = 0.02) but not to any other transitions, including *confusion* → *engagement* (r = 0.02, p = 0.85). The negative correlation between prior knowledge and the L^* of *frustration* → *confusion* was in line with the negative moderation effect yielded by the proposed mixed-logistic model.

Table 10

The impact of prior knowledge on frustration and boredom

Target-frustration	Estimate (SE)	Target-boredom	Estimate (SE)
Fixed effect		**Fixed effect**	
Intercept	−3.38 (0.17) ***	Intercept	−4.50 (0.44) ***
c-Frustration	1.48 (0.34) ***	g-Frustration	0.98 (0.63)
g-Confusion	0.56 (0.43)	g-Boredom	0.97 (0.33) **
g-Frustration	1.04 (0.38) **	Knowledge	−0.28 (0.12) *
g-Boredom	1.22 (0.52) *	g-Frustration* Knowledge	−0.20 (0.26)
knowledge	−0.14 (0.05) **		
g-Confusion*Knowledge	0.16 (0.14)		
g-Boredom*Knowledge	0.06 (0.22)		
Random effect			
Intercept	0.37	Intercept	1.53

Note. The reference given states were boredom.

4.6 The influence of reference states

Table 11 presents the marginal strengths of the transitions with different reference transitions. When all the other transitions were references, only five of the 18 cases (six transitions × three datasets) were statistically significantly positive, and four transitions were statistically significantly negative. Removing self-transitions resulted in positively stronger transition strengths. The four negative transitions became null or positive, and six null transitions became significantly positive. This was in line with prior studies (Karumbaiah et al., 2019), which have suggested that self-transitions may suppress the strengths of transitions of interest.

The change between the model without self-transitions and the model with only one reference transition was inconsistent between the vMedic dataset and the others. In the Physics playground and Betty's Brain datasets, the changes in the strength coefficients were positive, with a magnitude ranging from 0.01 to 1.09. While in the vMedic dataset, the changes in the strength coefficients were negative, with a magnitude ranging from -0.01 to -1.03 and three positive transitions became null. The inconsistency was perhaps because the analyses for the vMedic dataset included the unknown affective state, which were observations that were not or could not be classified into the four states (engagement, confusion, frustration, and boredom). The transition from the given unknown state to the target state were statistically significantly negative (see Table B3.1), and including them in the reference weakened the overall strength of the reference transitions and inflated the strengths of transitions of interest. For example, the strength of *unknown* → *engagement* was -1.59 (SE = 0.20, $p < 0.001$). When including *unknown* → *engagement* in the reference, the strength of *confusion* → *engagement* was 0.97 (SE =

0.19, $p < 0.001$). In contrast, when excluding *unknown* → *engagement* in the reference, the strength of *confusion* → *engagement* became 0.09 (SE = 0.24, $p = 0.70$).

Table 11

Transition strength estimates varied by reference transitions

γ (SE)	Dataset	Reference transitions		
		All others	No self-transition	One reference [a]
ENG_CON	Physics playground	-0.69 (0.06) ***	0.10 (0.09)	0.69 (0.17) ***
	vMedic	0.29 (0.18)	0.96 (0.23) ***	0.03 (0.34)
	Betty's Brain	−0.87 (0.25) ***	−0.59 (0.34)	0.50 (0.74)
CON_ENG	Physics playground	-0.74 (0.06) ***	0.55 (0.07) ***	0.93 (0.09) ***
	vMedic	−0.10 (0.18)	0.97 (0.19) ***	−0.09 (0.24)
	Betty's Brain	−0.53 (0.25) *	0.60 (0.30) *	0.61 (0.35)
CON_FRU	Physics playground	0.18 (0.10)	0.50 (0.10) ***	0.55 (0.10) ***
	vMedic	1.63 (0.46) ***	1.65 (0.46) ***	1.30 (0.47) **
	Betty's Brain	0.25 (0.43)	0.41 (0.43)	0.54 (0.43)
FRU_CON	Physics playground	0.00 (0.10)	0.27 (0.11) *	0.91 (0.19) ***
	vMedic	1.04 (0.55)	1.18 (0.53) *	0.84 (0.61)
	Betty's Brain	0.92 (0.38) *	1.08 (0.38) **	1.56 (0.81)
FRU_BOR	Physics playground	0.33 (0.12) **	1.15 (0.12) ***	1.16 (0.12) ***
	vMedic	−0.16 (0.77)	0.07 (0.75)	0.06 (0.75)
	Betty's Brain	1.20 (0.45) **	1.37 (0.44) **	1.36 (0.45) **
BOR_FRU	Physics playground	-0.00 (0.13)	0.35 (0.13) **	0.43 (0.13) ***
	vMedic	0.36 (0.63)	0.37 (0.63)	0.24 (0.66)
	Betty's Brain	1.31 (0.39) ***	1.41 (0.38) ***	1.46 (0.38) ***

Note. ENG: engagement. CON: confusion. FRU: frustration. BOR: boredom. *, **, *** p < .05, .01, .001.a: When the target state is engagement, confusion, frustration, and boredom, the reference transition is *boredom* → *engagement*, *boredom* → *confusion*, *engagement* → *frustration*, and *engagement* → *boredom*, respectively.

5 Discussion

This study proposes a framework that utilizes the logistic mixed model to compute the conditional strengths of affect transitions, examine whether marginal and conditional strengths are equal, and investigate the impact of external factor on transitions. We applied the framework to three affect sequence datasets and found: (1) marginal and conditional strengths were not equal in 9 of 12 cases; (2) accounting for conditional strengths improved the performance of predicting affective states; (3) prior knowledge negatively moderated the strengths of *confusion* → *engagement* and *frustration* → *confusion*; (4) the affective state used as the reference level substantially influenced the transition strength estimates. This section discusses the implications of the findings for affect dynamics and transition analyses of events in other modalities, highlights the limitations of this study, and suggests future directions.

5.1 Implications for affect dynamics

5.1.1 Methodological implications for affect dynamics and transition analyses of events in other modalities

The finding that the marginal and conditional strengths of affect transitions were not equal in most cases suggests that the next affective state (target states) depends on both the current (given states) and previous states (conditional states) during the learning process (i.e., a higher-order Markov chain; Bakeman & Gottman, 1997). In addition, the previous states may moderate the transition strength between the current and next states. The role of the previous state is similar to a covariate that may influence the next state and the relationship between the current and next states. Without controlling the previous state, the marginal strength would be a mix of the effects of previous and current states on the next state, rather than the pure effect of the current state on the next state. Consequently, the marginal strength estimates derived from a dataset may greatly rely on the affect state distribution in this dataset. For example, in the Physics Playground dataset, the conditional strength of *engagement* → *confusion* was 0.36 when the conditional state was engagement (i.e., *engagement* → *engagement* → *confusion*) and 0.83 when not (i.e., *not engagement* → *engagement* → *confusion*). If most observations were engagement in the dataset, the marginal strength would be close to 0.36. In contrast, if most observations were not engagement, the marginal strength would be close to 0.83. The proportion of engagement in the dataset was 80%, and thus, the marginal strength was 0.69, much closer to 0.83 rather than 0.36. Therefore, it is recommended to control the conditional state whenever possible, or at least, check the assumption of identical conditional and marginal transition strengths.

The empirical analysis for the impact of prior knowledge on transition strengths in the Betty's Brain dataset suggests that model-based transition analysis approaches, such as the proposed framework, may have a higher power in detecting the impact of students' and environmental factors on affect transitions than metric-based approaches. This is likely due to that the model-based approaches account for measurement errors in transition strengths. Prior studies have called for investigating the impact of students' and environmental factors on affect and affect transitions (Andres et al., 2019; Karumbaiah et al., 2021; Morais & Jaques, 2024). Future research may use the model-based approaches to do so.

When estimating the transition of interest, the choice of which transition to be the reference significantly impacts the estimate. Prior studies have found that including self-transitions in the references deflated the estimate (Karumbaiah et al., 2018; Karumbaiah et al., 2021; Matayoshi & Karumbaiah, 2021). The findings in this study further suggested that including the other transitions in the reference may inflate or deflate the estimate, depending on the strength of these transitions. For example, the changes in the estimate between the model without self-transitions and the model with only one reference transition were positive in the Physics Playground and Betty's Brain datasets because positive transitions were excluded from the reference, while the changes was negative in the vMedic dataset because negative transitions were excluded from the reference. These results imply that researchers must carefully operationalize the significance or strength of an affect transition based on the context and research purpose.

For selecting the reference transitions, we recommend excluding self-transitions if they are not of interest. This is because they are typically strong and positive, and thus, including them in the reference may deflate the strength estimates of positive transitions and inflate the strength estimates of negative transitions. If there is a theoretical model that hypothesizes which transitions likely happen, using the other transitions as reference would be a reasonable choice. With such reference setting, researchers can examine whether the transitions hypothesized by the theory are more likely to happen than the others. If the analysis is exploratory, researchers may estimate one transition each time with all others as the reference. This reference setting allows identifying the strongest transitions. Regardless of the reference setting, the key point is to clearly indicate what it is and articulate the reasons.

The above implications are applied to the transition analysis of learning events in other modalities, such as action logs or discourses during learning. For example, lag-sequential analysis (LSA) has been widely used in identifying the transitions between coded discourses in collaborative learning. LSA is based on log-linear model but can only model marginal strengths between events. It does not allow users to incorporate the students and environmental factors into the model directly. As such, the proposed framework is a promising alternative to LSA in the transition analysis of discourses.

5.1.2 Practical implications

The difference between conditional and marginal transition strengths has practical implications for designing adaptive scaffolding. For example, resolved confusion benefits learning, but unresolved confusion may harm learning (D'Mello et al., 2014). If a learner is not likely to resolve their confusion and become engaged, providing scaffolding for confusion resolution will benefit the learner. In the vMedic dataset, the marginal strength of *confusion* → *engagement* was not statistically significant, and the conditional strength was the same when the conditional state was not *others*. But when the conditional state was *others*, the conditional strength was statistically significant. Based on the information of the marginal strength, we may infer that learners have difficulty in resolving their confusion in this environment and provide scaffolding when they are confused. However, based on the information of conditional strength, we also know that learners may resolve confusion when the state before confusion was *others*. As such, the system may not need to provide scaffolding when the state before confusion was *others*, given that unnecessary scaffolding may annoy the learner and result in the expertise reversal effect (Kalyuga, 2007).

Moreover, the effects of conditional state on the target state and the transition strength between given and target states, as well as the improvement in affect prediction performance by accounting for conditional strengths, imply that affect prediction and detectors are likely more accurate when incorporating more historical states. Note that this study only considered the conditional and given states for simplicity and due to the sequence length requirement of the mixed logistic model. It does not mean that states earlier than the conditional state are not useful for prediction. What minimal historical states are sufficient for accurate prediction is beyond the scope of this study, but the main point is that, when developing an affect prediction or detection model in computer-based learning environments, it may be useful to incorporate historical states into the model.

5.2 Limitations and future directions

The current study considered only the impact of the previous affective state on the next state and the transition from the current state to the next state, which implied a second-order Markov chain. A higher-order Markov chain model, which considers the impact of earlier affective states, is possible. However, such investigation entails a dataset with long affect sequences that may be difficult to collect. For example, to include the previous two affective states before the current affective state in the analysis, we need to convert a sequence to a 4×4×4×2 table containing 128 cells. Wickens (1989) noted that the total counts of a table should be at least four times the number of cells to ensure relatively accurate estimation. This requirement on the sample size of individual tables may be lowered for GLMM because it leverages information across individual tables to compute the strength estimates of transitions (Dagne et al., 2007). However, we may not expect 100 affect observations per sequence to guarantee accurate strength estimates of the transition from the current state to the next state when considering the impact of the previous two states. Even though two counts in a cell may allow relatively accurate estimation, a sequence still needs to contain 256 affect observations. Collecting so many affect observations is challenging via human observation (Matayoshi & Karumbaiah, 2021). For instance, the typical observation window for affective states is 20 seconds in the literature (Andres et al., 2019; Rodrigo et al., 2012). Collecting an affect sequence of 256 observations take 1.42 hours. The requirement on the number of sequences is easier to meet, e.g., 50 sequences, but collecting 50 sequences would still take 71.11 hours of observation. Moreover, even with such an amount of time on data collection, the data size would only allow for investigating how the previous two affective states influence the transition from the current to the next affective states.

There are two directions to address the limitation of GLMM in analyzing the impact of earlier affective states. One is using deep learning-based event sequence models to capture the long-term dependency between affective states. For example, transformer-based event sequence models can effectively learn the dependency between two states with many other states between them, where the dependency is indicated by the attention weights (Zhang et al., 2020; Zuo et al., 2020). Although such approach may not require many observations per sequence, it may require many sequences to return reliable dependencies between affective states. Also, attention weights are less interpretable than the coefficients in the mixed logistic model (Bai et al., 2021; Bibal et al., 2022).

Another direction is taking advantage of machine learning and multimodal data analytics to collect a large amount of affect data with fine granularity (Calvo & D’Mello, 2010). For instance, both sensor-based data (e.g., posture and eye movement) and sensor-free data (e.g., action log) can be useful in affect detection (Henderson et al., 2020; Sims & Conati, 2020). Deep learning models, such as convolutional neural network, long-short-term memory, and their combination, is promising in affect detection (Wang et al., 2022). Nevertheless, there are two caveats when using affect datasets collected by machine learning tools. First, if an observation only lasts a few seconds, many successive observations would be the same affective state. In this case, self-transitions must be controlled. Otherwise, the other transitions would be null or negative. Second, the accuracy of machine learning tools influences the analysis. Taking *confusion* → *engagement* as an example. If the precision on detecting engagement is low, many observed instances of *confusion* → *engagement* would be *confusion* → $\overline{\textit{engagement}}$. Consequently, the strength estimate *confusion* → *engagement* may be inflated or deflated, depending on the relative strength of *confusion* → *engagement* to *confusion* → $\overline{\textit{engagement}}$.

There is variability in the transition strengths across datasets. For instance, the assumption of identical marginal and conditional strengths was rejected in models for any target state in the Physics Playground dataset and but not rejected in models for confusion and boredom. These datasets differed in many characteristics, such as cultures, the learning platform, learners’ ages, time intervals between observations, and sequence lengths. It is hard to say what dataset characteristics caused the variability, and it requires datasets with fewer differences to explore this question.

5.3 Conclusion

This paper presents a methodological framework for analyzing affective transitions in computer-based learning. By considering both the marginal and conditional strengths of affect transitions, we unveiled a more nuanced understanding of how emotions unfold during learning processes. The findings challenge the assumption that the affective process is a first-order Markov chain because the strengths of some affect transitions were dependent on preceding affective states, implying a higher-order Markov chain. By enhancing the accuracy of estimating affect transitions, the proposed framework may allow the realization of more precise affect interventions and support the design of affect-responsive learning environments. In addition, the empirical data analyses suggests that the framework is promising in detecting the impact of students' and environmental factors on transitions. Our research also highlights the significance of selecting reference transitions in transition analyses.

References

Agresti, A. (2013). *An introduction to categorical data analysis* (3rd ed.). John Wiley & Sons.

Andres, J. M. A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A., & Biswas, G. (2019). Affect Sequences and Learning in Betty's Brain. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 383-390). Tempe, AZ, USA. http://doi.org/10.1145/3303772.3303807

Andres, J. M. L., Rodrigo, M. M. T., Sugay, J. O., Banawan, M. P., Paredes, Y. V. M., Cruz, J. S. D., & Palaoag, T. D. (2015). More fun in the Philippines? Factors affecting transfer of western field methods to one developing world context. In *the sixth international workshop on culturally-aware tutoring systems* (pp. 31-40). CEUR Workshop Proceedings.

Bai, B., Liang, J., Zhang, G., Li, H., Bai, K., & Wang, F. (2021). Why attentions may not be interpretable? In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 25-34). ACM. http://doi.org/10.1145/3447548.3467307

Bakeman, R., Mcarthur, D., & Quera, V. (1996). Detecting group differences in sequential association using sampled permutations: Log odds, kappa, and phi compared. *Behavior Research Methods, Instruments, & Computers*, *28*(3), 446-457. http://doi.org/10.3758/BF03200524

Bakeman, R., & Dabbs Jr, J. M. (1976). Social interaction observed: Some approaches to the analysis of behavior streams. *Personality and Social Psychology Bulletin, 2*(4), 335-345. https://doi.org/10.1177/014616727600200403

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge university press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1), 1-48. http://doi.org/10.18637/jss.v067.i01

Bibal, A., Cardon, R. E. M., Alfter, D., Wilkens, R., Wang, X., Fran C C Ois, T., & Watrin, P. (2022). Is attention explanation? An introduction to the debate. In

Proceedings of the 60th annual meeting of the association for computational linguistics (pp. 3889-3900). ACL. http://doi.org/10.18653/v1/2022.acl-long.269

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127-135. http://doi.org/https://doi.org/10.1016/j.tree.2008.10.008

Bosch, N., & Paquette, L. (2021). What' s next? Sequence length and impossible loops in state transition measurement. *Journal of Educational Data Mining, 13*(1), 1-23. http://doi.org/10.5281/zenodo.5048423

Botelho, A. F., Baker, R. S., Ocumpaugh, J., & Heffernan, N. T. (2018). Studying affect dynamics and chronometry using sensor-free detectors. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the seventh international conference on educational data mining* (EDM'2018) (pp. 157-166). International Educational Data Mining Society.

Buhr, E. E., Daniels, L. M., & Goegan, L. D. (2019). Cognitive appraisals mediate relationships between two basic psychological needs and emotions in a massive open online course. *Computers in Human Behavior, 96*, 85-94. http://doi.org/https://doi.org/10.1016/j.chb.2019.02.009

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1-28. http://doi.org/10.18637/jss.v080.i01

Caglar-Ozhan, S., Altun, A., & Ekmekcioglu, E. (2022). Emotional patterns in a simulated virtual classroom supported with an affective recommendation system.

British Journal of Educational Technology, 53(6), 1724-1749. http://doi.org/10.1111/bjet.13209

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18-37. http://doi.org/10.1109/T-AFFC.2010.1

Chevrier, M., Muis, K. R., Trevors, G. J., Pekrun, R., & Sinatra, G. M. (2019). Exploring the antecedents and consequences of epistemic emotions. *Learning and Instruction, 63*, 101209. http://doi.org/10.1016/j.learninstruc.2019.05.006

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multilevel analysis. *Small Group Research, 36*(5), 600-631. http://doi.org/10.1177/1046496405279309

Dagne, G. A., Brown, C. H., & Howe, G. W. (2003). Bayesian hierarchical modeling of heterogeneity in multiple contingency tables: An application to behavioral observation data. *Journal of Educational and Behavioral Statistics, 28*(4), 339-352. http://doi.org/10.3102/107699860280043

Dagne, G. A., Brown, C. H., & Howe, G. W. (2007). Hierarchical modeling of sequential behavioral data: Examining complex association patterns in mediation models. *Psychological Methods*, *12*(3), 298-316. http://doi.org/10.1037/1082-989X.12.3.298

Dagne, G. A., Howe, G. W., Brown, C. H., & Muthén, B. O. (2002). Hierarchical modeling of sequential behavioral data: An empirical Bayesian approach. *Psychological Methods, 7*(2), 262-280. http://doi.org/10.1037/1082-989X.7.2.262

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153-170. http://doi.org/10.1016/j.learninstruc.2012.05.003

D'Mello, S., Taylor, R. S., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *29*, 203-208. https://doi.org/10.1007/978-1-4419-1428-6_849

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157. http://doi.org/10.1016/j.learninstruc.2011.10.001

Fienberg, S. E. (1972). The analysis of incomplete multi-way contingency tables. *Biometrics*, *28*(1), 177-202. http://doi.org/10.2307/2528967

Greenhouse, J. B. (2009). Commentary: Cornfield, epidemiology and causality. *International Journal of Epidemiology*, *38*(5), 1199-1201. http://doi.org/10.1093/ije/dyp299

Henderson, N., Rowe, J., Paquette, L., Baker, R. S., & Lester, J. (2020). Improving affect detection in game-based learning with multimodal data fusion. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Lecture notes in computer science: Vol 12163. Artificial intelligence in education. AIED 2020*. (pp. 228-239). Springer.

Howe, G. W., Dagne, G., & Brown, C. H. (2005). Multilevel methods for modeling observed sequences of family interaction. *Journal of Family Psychology*, *19*(1), 72-85. http://doi.org/10.1037/0893-3200.19.1.72

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539. http://doi.org/10.1007/s10648-007-9054-3

Karumbaiah, S., Andres, J., Botelho, A. F., Baker, R. S., & Ocumpaugh, J. S. (2018). The implications of a subtle difference in the calculation of affect dynamics. In C. Yang, M. Chang, L. Wong, & M. M. T. Rodrigo (Eds.), *Proceedings of the 26th international conference for computers in education* (pp. 29-38). Asia-Pacific Society for Computers in Education.

Karumbaiah, S., Baker, R. S., Ocumpaugh, J., & Andres, A. (2021). A re-analysis and synthesis of data on affect dynamics in learning. *IEEE Transactions on Affective Computing*. http://doi.org/10.1109/TAFFC.2021.3086118

Karumbaiah, S., Baker, R. S., & Ocumpaugh, J. (2019). The case of self-transitions in affective dynamics. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Lecture notes in computer science: Vol 11625. Artificial intelligence in education. AIED 2019* (pp. 172-181). Springer. http://doi.org/10.1007/978-3-030-23204-7_15

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297-300. http://doi.org/10.1177/1754073915590947

Mandler, G. (1990). Interruption (discrepancy) theory: Review and extensions. In S. Fisher & C. L. Cooper (Eds.), *On the move: The psychology of change and transition* (pp. 13-32). Wiley. http://doi.org/10.1901/jaba.2007.40-501

Matayoshi, J., & Karumbaiah, S. (2020). Adjusting the L statistic when self-transitions are excluded in affect dynamics. *Journal of Educational Data Mining*, *12*(4), 1-23.

Matayoshi, J., & Karumbaiah, S. (2021). Using marginal models to adjust for statistical bias in the analysis of state transitions. In M. Scheffel, N. Dowell, S. Joksimovic, & G. Siemens (Eds.), *Proceedings of the 11th international learning analytics and knowledge conference* (pp. 449-455). SoLAR. http://doi.org/10.1145/3448139.3448182

Morais, F. D., & Jaques, P. A. (2024). The dynamics of Brazilian students' emotions in digital learning systems. *International Journal of Artificial Intelligence in Education, 34*, 519-544. http://doi.org/10.1007/s40593-023-00339-0

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In *Proceedings of the 26th conference on user modeling, adaptation and personalization*. (pp. 131-138). ACM. https://doi.org/10.1145/3209219.3209241

Ocumpaugh, J., Andres, J. M., Baker, R., DeFalco, J., Paquette, L., Rowe, J., Mott, B., Lester, J., Georgoulas, V., Brawner, K., & Sottilare, R. (2017). Affect dynamics in military trainees using vMedic: From engaged concentration to boredom to confusion. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Lecture notes in computer science: Vol 10331. Artificial intelligence in education. AIED 2017* (pp. 238-249). Springer.

Ocumpaugh, J., Baker, R. S., Rodrigo, M. M., Salvi, A., van Velsen, M., Aghababyan, A., & Martin, T. (2015). HART: The human affect recording tool. In *Proceedings of the 33rd annual international conference on the design of communication* (pp. 1-6). ACM. http://doi.org/10.1145/2775441.2775480

Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. B. (2015). *Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual*. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences. Retrieved from http://apsce.net/icce/icce2018/wp-content/uploads/2018/12/C1-04.pdf

Ozechowski, T. J., Turner, C. W., & Hops, H. (2007). Mixed-effects logistic regression for estimating transitional probabilities in sequentially coded observational data. *Psychological Methods, 12*(3), 317-335. http://doi.org/10.1037/1082-989X.12.3.317

Rodrigo, M. M. T., Baker, R. S. J. D., Agapito, J., Nabos, J., Repalam, M. C., Reyes, S. S., & San Pedro, M. O. C. Z. (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), 224-236. http://doi.org/10.1109/T-AFFC.2011.41

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's playground. *The Journal of Educational Research, 106*(6), 423-430. http://doi.org/10.1080/00220671.2013.832970

Sims, S. D., & Conati, C. (2020). A neural architecture for detecting user confusion in eye-tracking data. In *Proceedings of the 2020 international conference on multimodal interaction*. (pp. 15-23). ACM. http://doi.org/10.1145/3382507.3418828

Theobald, M., Breitwieser, J., Murayama, K., & Brod, G. (2021). Achievement emotions mediate the link between goal failure and goal revision: Evidence from digital learning environments. *Computers in Human Behavior, 119*, 106726. http://doi.org/https://doi.org/10.1016/j.chb.2021.106726

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., & Zhang, W. (2022). A systematic review on affective computing: emotion models, databases, and recent advances. *Information Fusion*, 83-84, 19-52. http://doi.org/https://doi.org/10.1016/j.inffus.2022.03.009

Wickens, T. D. (1989). Multiway contingency table analysis for the social sciences. Lawrence Erlbaum Associates.

Wickens, T. D. (1993). Analysis of contingency tables with between-subjects variability. *Psychological Bulletin*, *113*(1), 191-204. http://doi.org/10.1037/0033-2909.113.1.191

Zhang, Q., Lipani, A., Kirnap, O., & Yilmaz, E. (2020). Self-attentive Hawkes process. *Proceedings of the 37th International Conference on Machine Learning, PMLR, 119*, 11183-11193.

Zuo, S., Jiang, H., Li, Z., Zhao, T., & Zha, H. (2020). Transformer Hawkes process. *Proceedings of the 37th International Conference on Machine Learning, PMLR, 119*, 11692-11702.

Appendix A : Transition metrics and mixed log-linear models

A1. Transition metrics

Assume that we want to compute the strength for transition $A \rightarrow B$ in an affect sequence *S* containing *N*+1 states. Let *a*, *b*, *c*, and *d* denote the occurrences of $A \rightarrow B$, $A \rightarrow \overline{B}$ (indicating not *B*), $\overline{A} \rightarrow B$, and $\overline{A} \rightarrow \overline{B}$ (Table A1). The simplest and most intuitive transition metric may be the transition probability (*TP*) used in the Markov chain model (Bakeman & Dabbs, 1976), which is $TP(A \rightarrow B) = a/(a+b)$. The metric $TP(A \rightarrow B)$ considers the occurrence rate of *A* but does not consider that of *B*, and thus, $TP(A \rightarrow B)$ is heavily influenced by the occurrence rate of *B*. Its interpretation is difficult because the same value has quite different meanings when the occurrence rate of B differs greatly (Bakeman & Gottman, 1997).

Table A1

The Occurrences of Pairs of Consecutive Events

First event \ Second event	B	$\overline{B}$
A	a	b
$\overline{A}$	c	d

Other metrics control both occurrence rates of *A* and *B*, such as the log-odds ratio (Bakeman et al., 1996), *L* (D'Mello et al., 2007), and *L** (Matayoshi & Karumbaiah, 2020). The odds ratio of $A \rightarrow B$ is calculated as. :

$$LOR(A \rightarrow B) = ln\frac{ad}{bc}. \tag{A4}$$

LOR(A→B) represents the natural log of the ratio of the odds of *B* after *A* to that of *B* after $\overline{A}$. If *LOR(A→B)* > 0, *B* is more likely to occur after *A* than $\overline{A}$; if *LOR(A→B)* < 0, *B* is less likely to occur after *A* than $\overline{A}$; if *LOR(A→B)* = 0, there is no relation between *A* and *B*. For instance, in the beforementioned affect sequence, $LOR(confusion \rightarrow engagement)$ is $\ln(2 * 3/(1 * 1)) = \ln(6)$, which means that the odds of engagement after confusion are six times the odds of engagement after the other affective states.

The metric *L* quantifies the transition strength of *A* → *B* differently from the log-odds ratio, as shown in Equation (A5) (D'Mello et al., 2007). It controls the occurrence rate of *B* (i.e., $(a + c)/(a + b + c + d)$) by subtracting it from *TP(A→B)* and scales the difference so that the maximum value is one. If *L(A→B)* = 0, no relationship exists between *A* and *B*. If *L(A→B)* > 0, *B* occurs after *A* at a rate more than chance; if *L(A→B)* < 0, it means that B occurs after A at a rate lower than chance.

$$L(A \rightarrow B) = \frac{\frac{a}{a+b} - \frac{a+c}{a+b+c+d}}{1 - \frac{a+c}{a+b+c+d}}. \tag{A5}$$

The two affective states of a transition may be the same, e.g., *confusion* → *confusion*. Such transitions are named self-transitions (Matayoshi & Karumbaiah, 2020). Researchers may not be interested in self-transitions and thus combine consecutive identical affective states into a single state. For instance, consecutive observations of confusion may be recoded as a single observation of confusion. In this case, *B* can only be a state different from *A*, and *TP(A→B)* does not necessarily equal the occurrence rate of *B* when *B* is independent of *A*. Consequently, *L* becomes biased. Another metric, *L**, mitigates the bias by deriving *a, b, c,* and *d* only from pairs of consecutive states where the second state is not A (Matayoshi & Karumbaiah, 2020). With the new counts, the computation of *L** follows Equation (A5).

A2. Mixed log-linear model

Recall that cell values in Table A2 are the occurrences of transitions from the row state to the column state.

Table A2

Contingency Table for the Example Sequence {frustration, boredom, confusion, engagement, confusion, confusion, engagement, engagement}

Frequencies (μ_{mrc})	**frustration**	**boredom**	**confusion**	**engagement**
frustration	0	1	0	0
boredom	0	0	1	0
confusion	0	0	1	2
engagement	0	0	1	1

Note. Cell values are the occurrences of transitions from the row state to the column state.

The following two-level model can fit the cell values :

Level 1 : cell

$$ln\mu_{mrc} = \beta_{0m} + \beta_{rm}X_{mr\bullet} + \beta_{cm}X_{m\bullet c} + \beta_{rcm}X_{mrc}$$

Level 2 : learner

$$\beta_{0m} = \gamma_{00} + U_{0m}$$
$$\beta_{rm} = \gamma_{r0}$$
$$\beta_{cm} = \gamma_{c0}$$
$$\beta_{rcm} = \gamma_{rc0} \,. \quad \text{(A6)}$$

The term μ_{mrc} is the cell value in the r^{th} row and c^{th} column of the m^{th} contingency table. The terms $X_{mr\bullet}$ and $X_{m\bullet c}$ are dummy variables for the r^{th} row state and the c^{th} column state, respectively, while γ_{r0} and γ_{c0} are their fixed effects. The term X_{mrc} are dummy variables for interactions between rows and columns, and γ_{rc0} represents its fixed effect, i.e., the average association between the r^{th} row and the c^{th} column states. If γ_{rc0} is significantly different from 0, there is an association between the r^{th} row and the c^{th} column states. γ_{00} is the fixed effect of the intercept, while U_{0m} is the random effect of the intercept.

Taking Table A2Table 2 as an example, r = 1, 2, 3, 4, and c = 1, 2, 3, 4. If we let the engagement row and the engagement column be the references, $X_{m4\bullet}$ will always be 0 if the cell is in row engagement, and $X_{m\bullet 4}$ will always be 0 if the cell is in column engagement. The term $X_{m1\bullet}$ = 1 if row frustration otherwise 0, and $X_{m\bullet 1}$ = 1 if column frustration otherwise 0. The term X_{m11} = 1 if row frustration and column frustration. If γ_{110} is significantly different from 0, it means that a student in a frustrated state stays at this state at a rate more than transitioning into engagement. Other coding techniques, such as contrast coding, can also be used to investigate specific questions, such as whether students in a confused state are more likely to resolve confusion and transition into engagement than transition into frustration and boredom. Readers may refer to Dagne et al. (2007) for details.

Appendix B : Marginal and conditional transition strengths

B1. Dummy Coding of the Given Affective State

Table B1.

Dummy Coding of the Given State at Different Target States

Dummy variable $\boldsymbol{X_{m\bullet g}}$	Given affective state			
	Engagement	Confusion	Frustration	Boredom
Target affective state is engagement				
$X_{m\bullet 1}$	1	0	0	0
$X_{m\bullet 2}$	0	1	0	0
Target affective state is confusion				
$X_{m\bullet 1}$	1	0	0	0
$X_{m\bullet 2}$	0	1	0	0
$X_{m\bullet 3}$	0	0	1	0
Target affective state is frustration				
$X_{m\bullet 1}$	0	1	0	0
$X_{m\bullet 2}$	0	0	1	0
$X_{m\bullet 3}$	0	0	0	1
Target affective state is boredom				
$X_{m\bullet 1}$	0	0	1	0
$X_{m\bullet 2}$	0	0	0	1

Note. In the analysis of the vMedic dataset, the reference given state did not include the state *others*, regardless of the target state.

B2. Marginal and Conditional Transitions in the Physics Playground dataset

Table B2.1

The Marginal and Conditional Transitions to Engagement

Model	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	0.47 (0.09) ***	-0.35 (0.10) ***	-0.35 (0.10) ***
g-Engaged [a]	1.64 (0.05) ***	1.35 (0.05) ***	1.35 (0.05) ***
g-Confusion [a]	0.55 (0.07) ***	0.36 (0.08) ***	0.46 (0.34)
c-Engaged		1.21 (0.08) ***	1.23 (0.08) ***
c-Confusion		0.58 (0.10) ***	0.54 (0.11) ***
c-Frustration		0.42 (0.10) ***	0.42 (0.10) ***
c-Engaged : g-Confusion			-0.17 (0.35)
c-Confusion : g-Confusion			0.03 (0.36)
c-Frustration : g-Confusion			-0.04 (0.40)
Random effect			
Intercept	0.87	0.68	0.68
Other information			
# Parameters	4	7	10
AIC	16953.55	16610.80	16614.73
BIC	16985.35	16666.45	16694.22
Deviance	16945.55	16596.80	16594.73

Note. a* the reference affective states were boredom and frustration. g-* given affective state. c-* conditional affective state. The units of coefficients were the natural log of odds between transitions, and values in the parentheses were standard errors. For example, in the model with only the given affective state, the coefficient of *confusion* → *engagement* was 0.55, which means that the odds of *confusion* → *engagement* was 1.73 (= $e^{0.55}$) times that of *boredom/frustration* → *engagement*.

Table B2.2

The Marginal and Conditional Transitions to Confusion

Model	Only given	Given and conditional	All interactions	One interaction
Fixed effect				
Intercept	−3.90 (0.19) ***	−4.04 (0.21) ***	−4.07 (0.24) ***	−4.12 (0.21) ***
g-Engaged	0.69 (0.17) ***	0.57 (0.17) **	0.68 (0.31) *	0.83 (0.19) ***
g-Confusion	1.91 (0.17) ***	1.68 (0.18) ***	1.68 (0.19) ***	1.63 (0.18) ***
g-Frustration	0.91 0.19) ***	0.77 (0.20) ***	0.74 (0.48)	0.73 (0.19) ***
c-Engaged		0.24 (0.16)	0.40 (0.25)	0.51 (0.17) **
c-Confusion		0.89 (0.17) ***	0.78 (0.26) **	0.88 0.17) ***
c-Frustration		0.22 (0.19)	0.21 (0.31)	0.22 (0.19)
c-Engaged : g-Engaged			−0.27 (0.34)	−0.47 (0.13) ***
c-Confusion : g-Engaged			0.22 (0.36)	
c-Frustration : g-Engaged			0.07 (0.41)	
c-Engaged : g-Frustration			0.10 (0.51)	
c-Confusion : g-Frustration			−0.10 (0.56)	
c-Frustration : g-Frustration			−0.06 (0.56)	
Random effect				
Intercept	1.25	1.12	1.09	1.09
Other information				
# Parameters	5	8	14	9
AIC	9592.52	9524.24	9521.89	9512.96
BIC	9632.26	9587.83	9633.18	9584.50
Deviance	9582.52	9508.24	9493.89	9494.96

Note. In the model with all interactions, the coefficients of the interactions between conditional states and given frustration were small and might be unnecessary. Moreover, the interaction between conditional and given engagement was negative, while the interactions between the other conditional states and given engagement were positive. This suggests that the strength of *engagement* → *confusion* may be weaker when the conditional state was engagement than when the conditional state was the others. Thus, we implemented a model with only the interaction between conditional and given engagement.

Table B2.3

The Marginal and Conditional Transitions to Frustration

Model	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	−3.43 (0.10)***	−3.15 (0.16)***	−3.15 (0.19)***
g-Confusion	0.55 (0.10) ***	0.46 (0.10) ***	0.32 (0.56)
g-Frustration	1.52 (0.08) ***	1.30 (0.08) ***	1.28 (0.08) ***
g-Boredom	0.43 (0.13) ***	0.25 (0.14)	0.28 (0.23)
c-Engaged		−0.37 (0.14) **	−0.38 (0.17) *
c-Confusion		0.11 (0.16)	0.15 (0.20)
c-Frustration		0.71 (0.14) ***	0.74 (0.18) ***
c-Engaged : g-Confusion			0.28 (0.57)
c-Confusion : g-Confusion			−0.07 (0.59)
c-Frustration : g-Confusion			0.01 (0.61)
c-Engaged : g-Boredom			−0.01 (0.33)
c-Confusion : g-Boredom			0.54 (0.49)
c-Frustration : g-Boredom			−0.27 (0.35)
Random effect			
Intercept	1.28	1.04	1.03
Other information			
# Parameters	5	8	14
AIC	8737.50	8576.73	8583.71
BIC	8777.24	8640.32	8695.00
Deviance	8727.50	8560.73	8555.71

Table B2.4

The Marginal and Conditional Transitions to Boredom

Model	Only given	Given and conditional	All interactions	One interaction
Fixed effect				
Intercept	−4.58 (0.16) ***	−3.20 (0.18) ***	−3.26 (0.18) ***	−3.21 (0.18) ***
g-Frustration	1.15 (0.12) ***	0.99 (0.13) ***	1.52 (0.26) ***	1.13 (0.14) ***
g-Boredom	2.70 (0.09) ***	2.22 (0.09) ***	2.24 (0.09) ***	2.21 (0.09) ***
c-Engaged		−1.37 (0.10) ***	−1.31 (0.10) ***	−1.38 (0.10) ***
c-Confusion		−1.41 (0.17) ***	−1.38 (0.19) ***	−1.42 (0.17) ***
c-Frustration		−1.02 (0.15) ***	−0.79 (0.17) ***	−0.83 (0.17) ***
c-Engaged : g-Frustration			−0.59 (0.31)	
c-Confusion : g-Frustration			−0.36 (0.45)	
c-Frustration : g-Frustration			−0.98 (0.35) **	−0.60 (0.29) *
Random effect				
Intercept	2.38	1.95	1.94	1.96
Other information				
# Parameters	4	7	10	8
AIC	5978.24	5793.89	5791.75	5791.38
BIC	6010.04	5849.54	5871.24	5854.97
Deviance	5970.24	5779.89	5771.75	5775.38

B3. Marginal and Conditional Transitions in the vMedic dataset

Table B3.1

The Marginal and Conditional Transitions to Engagement

	Only given	Given and conditional	All interactions	One interaction
Fixed effect				
Intercept	0.01 (0.16)	0.06 (0.22)	0.07 (0.22)	0.07 (0.22)
g-Engaged	0.79 (0.18) ***	0.80 (0.18) ***	0.78 (0.18) ***	0.80 (0.18) ***
g-Confusion	−0.15 (0.23)	−0.19 (0.23)	−0.20 (0.65)	−0.28 (0.24)
g-Others	−1.59 (0.20) ***	−1.07 (0.21) ***	−1.05 (0.21) ***	−1.04 (0.21) ***
c-Engaged		0.19 (0.20)	0.22 (0.21)	0.19 (0.20)
c-Confusion		−0.32 (0.25)	−0.45 (0.28)	−0.30 (0.26)
c-Frustration		−0.05 (0.44)	0.00 (0.49)	−0.04 (0.44)
c-Others		−0.96 (0.23) ***	−1.01 (0.24) ***	−1.02 (0.23) ***
c-Engaged : g-Confusion			−0.29 (0.68)	
c-Confusion : g-Confusion			0.44 (0.74)	
c-Frustration : g-Confusion			−0.25 (1.21)	
c-Others : g-Confusion			1.46 (0.94)	1.53 (0.72) *
Random effect				
Intercept	0.10	0.03	0.05	0.04
Other information				
# Parameters	5	9	13	10
AIC	2126.77	2076.11	2076.77	2073.64
BIC	2154.35	2125.76	2148.49	2128.80
Deviance	2066.08	2038.79	2023.53	2029.46

Table B3.2

The Marginal and Conditional Transitions to Confusion

	Only given	Given and conditional	All interactions	One interaction
Fixed effect				
Intercept	−2.47 (0.34) ***	−2.43 (0.41) ***	−2.62 (0.54) ***	−2.36 (0.41) ***
g-Engaged	0.03 (0.34)	0.04 (0.34)	0.50 (0.66)	−0.09 (0.35)
g-Confusion	0.86 (0.38) *	0.82 (0.39) *	0.74 (0.39)	0.79 (0.39) *
g-Frustration	0.84 (0.61)	0.66 (0.62)	−12.93 (1151.30)	0.60 (0.62)
g-Others	−1.71 (0.45) ***	−1.46 (0.48) **	−1.17 (0.50) *	−1.11 (0.49) *
c-Engaged		−0.06 (0.33)	0.36 (0.55)	−0.04 (0.33)
c-Confusion		0.54 (0.39)	0.69 (0.60)	0.53 (0.39)
c-Frustration		0.44 (0.66)	0.76 (1.00)	0.43 (0.66)
c-Others		−0.44 (0.43)	−1.03 (0.74)	−1.35 (0.61) *
c-Engaged : g-Engaged			−0.83 (0.69)	
c-Confusion : g-Engaged			−0.26 (0.77)	
c-Frustration : g-Engaged			−0.56 (1.34)	
c-Others : g-Engaged			0.81 (0.90)	1.45 (0.64) *
c-Engaged : g-Frustration			14.39 (1151.30)	
c-Confusion : g-Frustration			12.55 (1151.30)	
c-Frustration : g-Frustration			−0.49 (2295.76)	
c-Others : g-Frustration			1.68 (2533.12)	
Random effect				
Intercept	0.65	0.48	0.50	0.48
Other information				
# Parameters	6	10	18	11
AIC	983.07	982.55	985.24	979.19
BIC	1016.17	1037.71	1084.54	1039.87
Deviance	887.69	892.80	877.98	887.81

Table B3.3

The Marginal and Conditional Transitions to Frustration

	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	−4.46 (0.39) ***	−3.63 (0.60) ***	−3.43 (0.66) ***
g-Confusion	1.30 (0.47) **	1.20 (0.49) *	0.84 (1.32)
g-Frustration	0.12 (1.11)	0.19 (1.13)	−0.02 (1.18)
g-Boredom	0.24 (0.66)	0.03 (0.68)	−0.59 (1.30)
g-Others	−2.29 (1.04) *	−2.42 (1.09) *	−2.12 (1.28)
c-Engaged		−0.99 (0.57)	−1.20 (0.75)
c-Confusion		−0.54 (0.74)	−0.47 (1.03)
c-Frustration		0.16 (0.97)	−0.19 (1.30)
c-Others		−0.47 (0.74)	−1.23 (1.09)
c-Engaged : g-Confusion			0.30 (1.52)
c-Confusion : g-Confusion			−0.08 (1.74)
c-Frustration : g-Confusion			0.69 (2.11)
c-Others : g-Confusion			1.56 (1.97)
c-Engaged : g-Boredom			0.72 (1.75)
c-Confusion : g-Boredom			−20.85 (94682.12)
c-Frustration : g-Boredom			−19.86 (81777.67)
c-Others : g-Boredom			2.22 (1.94)
Random effect			
Intercept	0.98	0.66	0.76
Other information			
# Parameters	6	10	18
AIC	279.29	283.55	296.75
BIC	312.39	338.71	396.05
Deviance	232.43	237.91	232.21

Table B3.4

The Marginal and Conditional Transitions to Boredom

	Only given	Given and conditional	Interaction	One conditional state
Fixed effect				
Intercept	−2.80 (0.15) ***	−2.46 (0.33) ***	−1.03 (11018.32)	−2.72 (0.15) ***
g-Frustration	0.07 (0.75)	0.05 (0.75)	−15.15 (3203.64)	0.04 (0.75)
g-Boredom	1.70 (0.25) ***	1.73 (0.25) ***	1.73 (0.24) ***	1.76 (0.25) ***
g-Others	−0.64 (0.27) *	−0.21 (0.31)	−0.21 (0.32)	−0.21 (0.31)
c-Engaged		−0.24 (0.30)	−0.29 (0.29)	
c-Confusion		−0.35 (0.43)	−0.33 (0.42)	
c-Frustration		−0.44 (0.80)	−0.43 (0.81)	
c-Others		−1.05 (0.41) *	−1.08 (0.40) **	−0.82 (0.32) **
c-Engaged : g-Frustration			15.92 (3203.64)	
c-Confusion : g-Frustration			0.57 (3813.21)	
c-Frustration : g-Frustration			−0.44 (7144.74)	
c-Others : g-Frustration			−0.96 (7059.00)	
Random effect				
Intercept	0.17	0.10	0.09	0.14
Other information				
# Parameters	5	9	13	6
AIC	872.76	872.54	877.83	867.42
BIC	900.34	922.19	949.54	900.52
Deviance	832.73	835.49	834.16	829.85

Note. In the model with given and conditional states, only the coefficient of c-Others was statistically significant. Thus, we implemented a model with only this conditional state.

B4. Marginal and Conditional Transitions in the Betty's Brain dataset

Table B4.1

The Marginal and Conditional Transitions to Engagement

	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	0.62 (0.21) **	−0.36 (0.32)	−0.28 (0.33)
g-Engaged	1.41 (0.20) ***	1.28 (0.20) ***	1.28 (0.21) ***
g-Confusion	0.60 (0.30) *	0.49 (0.30)	−0.76 (1.18)
c-Engaged		1.17 (0.28) ***	1.09 (0.29) ***
c-Confusion		1.07 (0.38) **	1.07 (0.42) *
c-Frustration		0.45 (0.38)	0.26 (0.39)
c-Engaged : g-Confusion			1.29 (1.22)
c-Confusion : g-Confusion			0.92 (1.33)
c-Frustration : g-Confusion			2.22 (1.47)
Random effect			
Intercept	0.38	0.25	0.25
Other information			
# Parameters	4	7	10
AIC	1325.39	1311.15	1314.18
BIC	1346.88	1348.77	1367.92
Deviance	1234.44	1235.07	1232.15

Table B4.2

The Marginal and Conditional Transitions to Confusion

	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	−3.67 (0.73) ***	−3.34 (0.81) ***	−2.65 (0.82) **
g-Engaged	0.50 (0.74)	0.60 (0.76)	−15.79 (1597.64)
g-Confusion	1.69 (0.80) *	1.77 (0.81) *	2.41 (0.96) *
g-Frustration	1.56 (0.81)	1.63 (0.82) *	1.43 (1.16)
c-Engaged		−0.45 (0.49)	−1.88 (0.86) *
c-Confusion		−0.32 (0.60)	−1.64 (1.00)
c-Frustration		−0.53 (0.67)	−2.21 (1.34)
c-Engaged : g-Engaged			17.18 (1597.64)
c-Confusion : g-Engaged			17.28 (1597.64)
c-Frustration : g-Engaged			16.64 (1597.64)
c-Engaged : g-Frustration			0.71 (1.29)
c-Confusion : g-Frustration			−16.76 (5872.44)
c-Frustration : g-Frustration			2.15 (1.74)
Random effect			
Intercept	0.42	0.42	0.39
Other information			
# Parameters	5	8	14
AIC	683.19	688.24	686.36
BIC	710.06	731.23	761.59
Deviance	625.30	624.62	613.19

Table B4.3

The Marginal and Conditional Transitions to Frustration

	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	−3.34 (0.18) ***	−2.47 (0.41) ***	−2.80 (0.56) ***
g-Confusion	0.54 (0.43)	0.41 (0.44)	1.24 (1.29)
g-Frustration	1.22 (0.36) ***	0.99 (0.38) **	1.02 (0.39) **
g-Boredom	1.46 (0.38) ***	1.07 (0.40) **	1.68 (0.76) *
c-Engaged		−0.97 (0.41) *	−0.63 (0.57)
c-Confusion		−0.81 (0.61)	−0.39 (0.76)
c-Frustration		0.64 (0.49)	1.06 (0.65)
c-Engaged : g-Confusion			−0.66 (1.40)
c-Confusion : g-Confusion			−0.92 (1.74)
c-Frustration : g-Confusion			−1.69 (1.71)
c-Engaged : g-Boredom			−0.91 (0.98)
c-Confusion : g-Boredom			−14.97 (161.91)
c-Frustration : g-Boredom			−0.57 (1.13)
Random effect			
Intercept	0.39	0.25	0.24
Other information			
# Parameters	5	8	14
AIC	639.48	624.19	633.62
BIC	666.35	667.18	708.86
Deviance	587.01	579.22	577.77

Table B4.4

The Marginal and Conditional Transitions to Boredom

	Only given	Given and conditional	Interaction
Fixed effect			
Intercept	−4.47 (0.45) ***	−3.80 (0.57) ***	−3.26 (1.79)
g-Frustration	1.37 (0.44) **	1.32 (0.45) **	1.50 (0.86)
g-Boredom	1.00 (0.34) **	1.00 (0.34) **	1.01 (0.35) **
c-Engaged		−0.59 (0.35)	−0.60 (0.37)
c-Confusion		−0.62 (0.66)	−0.88 (0.74)
c-Frustration		−1.32 (0.72)	−1.54 (0.86)
c-Engaged : g-Frustration			−0.55 (1.05)
c-Confusion : g-Frustration			0.94 (1.65)
c-Frustration : g-Frustration			−0.42 (1.65)
Random effect			
Intercept	3.53	3.12	2.83
Other information			
# Parameters	4	7	10
AIC	555.39	556.64	561.56
BIC	576.88	594.26	615.30
Deviance	441.59	441.25	442.16

B5. Prediction performance

Table B5

The prediction performance in the three datasets

Target state	Model	Accuracy	Kappa	F1	AUC
	Physics Playground dataset				
Engagement	Given and conditional states	0.819	0.161	0.898	0.737
	Only given states	0.804	0.000	0.891	0.689
Confusion	Given and conditional states	0.927	0.000	- [a]	0.663
	Only given states	0.927	0.000	-	0.628
Frustration	Given and conditional states	0.931	0.000	-	0.714
	Only given states	0.931	0.000	-	0.654
Boredom	Given and conditional states	0.946	0.000	-	0.803
	Only given states	0.946	0.000	-	0.764
	vMedic dataset				
Engagement	Given and conditional states	0.706	0.413	0.730	0.767
	Only given states	0.708	0.417	0.714	0.722
Confusion	Given and conditional states	0.500	0.000	-	0.670
	Only given states	0.500	0.000	-	0.644
Frustration	Given and conditional states	0.500	0.000	-	0.678
	Only given states	0.500	0.000	-	0.618
Boredom	Given and conditional states	0.500	0.000	-	0.654
	Only given states	0.500	0.000	-	0.664
	Betty's Brain dataset				
Engagement	Given and conditional states	0.838	0.092	0.909	0.630
	Only given states	0.832	0.000	0.907	0.610
Confusion	Given and conditional states	0.941	0.000	-	0.589
	Only given states	0.941	0.000	-	0.606
Frustration	Given and conditional states	0.948	0.000	-	0.650
	Only given states	0.948	0.000	-	0.587
Boredom	Given and conditional states	0.943	0.000	-	0.667
	Only given states	0.943	0.000	-	0.626

Note. a: The F1 scores could not be computed because the predicted outcome was constant.