Psychometrika (2025), **0**, 1–32
doi:10.1017/psy.2025.10031

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Bayesian Transition Diagnostic Classification Models with Polya-Gamma Augmentation

Joseph Resch[1], Samuel Baugh[2], Hao Duan[1], James Tang[1], Matthew J. Madison[3], Michael Cotterell[4] and Minjeong Jeon[5]

[1]Department of Statistics & Data Science, University of California, Los Angeles, CA, USA; [2]Department of Statistics, The Pennsylvania State University, University Park, PA, USA; [3]Department of Education, University of Georgia, Athens, GA, USA; [4]Department of Computer Science, University of Georgia, Athens, GA, USA; [5]Department of Education, University of California, Los Angeles, CA, USA

Corresponding author: Minjeong Jeon; Email: mjjeon@ucla.edu

(Received 13 February 2025; revised 15 June 2025; accepted 17 June 2025)

Abstract

Diagnostic classification models assume the existence of latent attribute profiles, the possession of which increases the probability of responding correctly to questions requiring the corresponding attributes. Through the use of longitudinally administered exams, the degree to which students are acquiring core attributes over time can be assessed. While past approaches to longitudinal diagnostic classification modeling perform inference on the overall probability of acquiring particular attributes, there is particular interest in the relationship between student progression and student covariates such as intervention effects. To address this need, we propose an integrated Bayesian model for student progression in a longitudinal diagnostic classification modeling framework. Using Pòlya-gamma augmentation with two logistic link functions, we achieve computationally efficient posterior estimation with a conditionally Gibbs sampling procedure. We show that this approach achieves accurate parameter recovery when evaluated using simulated data. We also demonstrate the method on a real-world educational testing data set.

Keywords: diagnostic classification models; Gibbs sampling; intervention effects; Pòlya-gamma augmentation; transition analysis

1. Introduction

Recent research has developed longitudinal diagnostic classification models (DCMs; Rupp et al., 2010) as psychometric options for diagnostic assessments administered over multiple occasions. These models examine how respondents change in their attribute proficiency statuses over time and have been used to evaluate intervention effects (e.g., Wang et al., 2018). This study focuses on the transition diagnostic classification model (TDCM; Madison & Bradshaw, 2018a) framework, and its direct extension to examine multiple-group differential rates of change in attribute proficiency in a mathematics education research study (Madison & Bradshaw, 2018b). In their analysis, Madison and Bradshaw did not account for covariate effects, which could indicate the differential effectiveness of the instructional intervention. As noted in their discussion, classroom effects were not analyzed because the methodology to do so in a general DCM framework was not available.

To overcome this limitation of the multiple-group TDCM in evaluating interventions, we propose a reformulation of the standard TDCM using hierarchical logistic regressions. The method presented can be thought of in terms of two interrelated components. The first is a binomial logistic measurement model (e.g., log-linear cognitive diagnosis model (LCDM); Henson et al., 2009), forming the relationship between attribute proficiency status to the item response. The second is a multinomial logistic structural model, relating a set of chosen covariates to change in attribute proficiency over time. Thus, the latent transition analysis (LTA) model of the standard TDCM is converted to a full regression formulation. The generalization that is introduced by the proposed reformulation allows for the relaxation of strict conditions that are fundamental to existing DCMs while ensuring proper interpretation and result convergence. At the same time, intervention effects become easy to interpret using log-odds. Most importantly, a highly adjustable covariate structure can be applied to attribute transition trajectories of interest. These benefits in model reformulation serve as the key contribution provided by our proposed model.

Note that the resulting extended TDCM model represents an advanced longitudinal DCM that offers more flexibility than the currently available DCMs. For instance, while DCMs have been adapted for longitudinal assessments, often incorporating a regression-like structure for covariates, most advancements have focused on hidden Markov structures (Liu et al., 2023; Wang et al., 2018; Yamaguchi & Martinez, 2024; Yigit & Douglas, 2021; Zhang & Chang, 2020). Although hidden Markov DCMs are computationally efficient, they can lead to biased conclusions when the underlying Markov assumption is violated (e.g., Pan et al., 2020). More detailed reviews of existing models are provided in Section 2.

The proposed extended TDCM model is accompanied by inference complexity. That is, model inference proceeds over three sets of parameters; the item parameters that link attributes to response probabilities (denoted as $\boldsymbol{\beta}$), the person parameters (i.e., attribute classifications) at each time (denoted as $\boldsymbol{\alpha}^{(t)}$, where t indicates discrete time points), and the multinomial regression coefficients for the "progression" of attributes over time (represented as $\boldsymbol{\gamma}$). To deal with the computational complexity, we propose a Bayesian framework using Pòlya-gamma augmentation (Polson et al., 2013) to be used for both the response and transition logistic models. This procedure allows efficient Gibbs sampling for the TDCM and serves as our second significant contribution. Pòlya-gamma augmentation has been utilized for estimating DCMs. For example, Balamuta et al. (2022) applied the augmentation scheme for the restricted latent class model (RLCM) at a single time point. This was subsequently extended by Jimenez et al. (2023) to multiple time points for the RLCM, although their model does not incorporate a latent transition structure, unlike TDCMs. Further, the multinomial version of Pòlya-gamma augmentation for logistic regressions used in our application has not been applied to longitudinal DCMs, including TDCMs, or in a hierarchical structure for both the item-attribute and transition elements. It is worth stressing that deriving Pòlya-gamma augmentation is neither trivial nor straightforward, given the hierarchical complexity of the proposed TDCM, which includes the item-attribute and transition partitions.

The rest of the article is structured as follows: Section 2 offers a review of the relevant models presented in the literature. Sections 3 and 4 detail the proposed model and its framework for Bayesian inference. Following that, empirical and simulation studies are provided in Sections 5 and 6 with a goal in mind to show the flexibility of the proposed TDCM extension in comparison to the standard TDCM. We conclude our article in Section 7 with a summary and discussion on limitations and avenues for future research.

2. Review of related models

In this section, we review models that form the basis for the proposed development. Specifically, we provide both conceptual and technical background on the DCM and the TDCM, which is the motivation of our proposed work. Additionally, we review hidden Markov DCMs in comparison with the proposed TDCM framework.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

2.1. DCMs

DCMs (Rupp et al., 2010), are psychometric models designed to classify respondents into ordered proficiency categories according to specified categorical latent traits, or attributes. In the dichotomous case, DCMs provide classifications according to each measured attribute, probabilistically placing respondents into one of two groups, typically termed mastery and non-mastery or proficient and non-proficient. Statistically, DCMs are confirmatory and constrained latent class models because 1) they require the specification of a Q-matrix (Tatsuoka, 1983), which delineates the item-attribute alignment; 2) the latent class space is specified a priori as the set of attribute proficiency patterns; and 3) they place constraints on the parameters in a traditional latent class model. Several DCMs have been developed that differ in the way they relate attribute mastery to item response probabilities.

2.2. TDCM

To accommodate longitudinal data in a DCM framework, a latent transition model (Collins & Wugalter, 1992) framework can be utilized. The latent transition model is a longitudinal extension of the latent class model, designed to simultaneously classify respondents into latent classes and model their transitions to and from different latent classes over time. Analogous to a DCM being a constrained latent class model, the TDCM; Madison & Bradshaw, 2018a is a constrained latent transition model that classifies respondents into attribute profiles and models their transitions to and from different attribute mastery statuses over time. The TDCM provides measures of student growth on a discrete scale in the form of attribute mastery transitions. In this way, the TDCM supports categorical and criterion-referenced interpretations of growth. On an individual level, growth in the TDCM framework is defined as transitions between different mastery statuses (e.g., non-mastery to mastery). On a group level, growth in the TDCM framework is defined as growth in attribute mastery proportion. For example, if a group went from 20% mastery of Standard 1 at time point 1 to 80% mastery at time point 2, that would be an indication of student learning.

Consider respondent i responding to J items over T testing occasions. In the general form of the TDCM, the probability of the item response vector $\boldsymbol{y}_i \in \{0,1\}^{J\times T}$ for respondent i is given by

$$P\left(Y_{ijt}=y_{ijt}\right)=\sum_{\boldsymbol{\alpha}_i^T}^{\mathcal{A}}\cdots\sum_{\boldsymbol{\alpha}_i^{(t)}}^{\mathcal{A}}\kappa_{\boldsymbol{\alpha}_i^{(1)}}\tau_{\boldsymbol{\alpha}_i^{(2)}|\boldsymbol{\alpha}_i^{(1)}}\cdots\tau_{\boldsymbol{\alpha}_i^{(T)}|\boldsymbol{\alpha}_i^{(T-1)}}\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\pi_{j\boldsymbol{\alpha}_i^{(t)}}\right)^{y_{ijt}}\left(1-\pi_{j\boldsymbol{\alpha}_i^{(t)}}\right)^{1-y_{ijt}}. \tag{1}$$

In the above equation, $\boldsymbol{\alpha}_i^{(t)} \in \{0,1\}^K$ represents the attribute mastery status of respondent i at time t and K is the number of attributes. For attribute $k \in \{1,\ldots,K\}$, $\mathcal{A}$ denotes the collection of attribute vectors $\boldsymbol{\alpha}_i^{(t)} \in \{0,1\}^K$. $\pi_{j\boldsymbol{\alpha}_i^{(t)}}$ represents the probability of respondent i with attribute mastery status $\boldsymbol{\alpha}_i^{(t)}$ answering item j at time t correctly. The TDCM has the same components as a standard latent class model, except the structural model has an extra component: transition probabilities denoted by $\tau_{\boldsymbol{\alpha}_i^{(t)}|\boldsymbol{\alpha}_i^{(t-1)}}$. Each $\tau_{\boldsymbol{\alpha}_i^{(t)}|\boldsymbol{\alpha}_i^{(t-1)}}$ represents the probability of transitioning between different attribute mastery statuses between testing occasions for respondent i. The other component of the structural model, $\kappa_{\boldsymbol{\alpha}_i^{(1)}}$, represents the probability of having attribute status $\boldsymbol{\alpha}_i$ for respondent i at time point 1.

To evaluate intervention effects in a DCM framework, the TDCM was extended to account for multiple groups (Madison & Bradshaw, 2018b). The multiple-group TDCM is designed to assess differential growth in attribute mastery between a treatment group and a control group in a pre-test/post-test designed experiment. The general form of the multiple-group TDCM is given by

$$\begin{aligned}P\left(Y_{ijt}=y_{ijt}|G=g\right)=&\sum_{\boldsymbol{\alpha}_i^T}^{\mathcal{A}}\cdots\sum_{\boldsymbol{\alpha}_i^{(t)}}^{\mathcal{A}}\kappa^{(g)}_{\boldsymbol{\alpha}_i^{(1)}}\tau^{(g)}_{\boldsymbol{\alpha}_i^{(2)}|\boldsymbol{\alpha}_i^{(1)}}\cdots\tau^{(g)}_{\boldsymbol{\alpha}_i^{(T)}|\boldsymbol{\alpha}_i^{(T-1)}}\\&\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\pi^{(g)}_{j\boldsymbol{\alpha}_i^{(t)}}\right)^{y_{ijt}}\left(1-\pi^{(g)}_{j\boldsymbol{\alpha}_i^{(t)}}\right)^{1-y_{ijt}},\end{aligned} \tag{2}$$

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

where $\kappa^{(g)}_{\boldsymbol{\alpha}_i^{(1)}}$ represents the probability of having attribute status $\boldsymbol{\alpha}_i$ for respondent i from group g at time point i, $\tau^{(g)}_{\boldsymbol{\alpha}_i^{(2)}|\boldsymbol{\alpha}_i^{(1)}}$ represents the probability of transitioning between different attribute mastery statuses between testing occasions for respondent i from group g, and $\pi^{(g)}_{j\boldsymbol{\alpha}_i^{(t)}}$ represents the probability of respondent i from group g with attribute mastery status $\boldsymbol{\alpha}_i^{(t)}$ answering item j at time t correctly. The multiple-group TDCM is similar to the single-group TDCM in Equation 1, but the structural parameters are conditional on group membership G, which represents the potential for respondents in different groups (e.g., treatment versus control) to have different mastery transition patterns and probabilities. If the treatment group shows significantly more growth in attribute mastery than the control group, this would be evidence of a successful intervention.

While the multiple-group TDCM presents a promising methodology for evaluating intervention effects with interpretive benefits over other longitudinal psychometric modeling options, it has limitations with respect to estimation. In preliminary explorations using commercially available software (Mplus; Muthén & Muthén, 2017), estimation time and data requirements were not feasible for applications with more than two groups, more than four dichotomous attributes, more than two time points, or more complex covariate structures such as nested effects or interaction effects. This significantly limits the full utilization of the TDCM in practice as experimental designs for evaluating intervention effects often include more than two measurement occasions, more than two groups, and complex experimental designs. To provide a more flexible framework for evaluating intervention effects in a DCM framework, a modified, more flexible TDCM with a practical estimation approach is necessary.

2.3. Hidden Markov DCMs

For longitudinal assessment data, the hidden Markov model (HMM) posed significant advancements by accounting for dependence in attribute mastery at each time point. For instance, Wang et al. (2018) used a HMM of higher order combined with a constrained deterministic-inputs, noisy-and-gate (DINA) model to evaluate the efficacy of different learning interventions. This logistic model included covariates to account for individual differences in skill transitions. Similarly, Zhang & Chang (2020) introduced a DINA-integrated multilevel logistic HMM with random effects to account for variability in instructional methods. Transition for this model is represented by a random effects model that accounts for a student's overall learning ability to acquire all attributes. Liu et al. (2023) proposed identifiability conditions to ensure that the DINA parameters for hidden Markov DCMs can be reliably estimated. Computationally, Yamaguchi & Martinez (2024) improved the efficiency of the hidden Markov DCM using DINA by providing a variational Bayesian inference framework. In another form of DCM, Yigit & Douglas (2021) presented a first-order HMM integrated with the generalized DINA (G-DINA) using expectation maximization to track learning trajectories, where G-DINA has provided much-desired flexibility in handling complexity such as compensatory relationships compared to the standard DINA.

Existing hidden Markov DCMs have experienced significant developments in recent years. However, strict conditions and constraints have limited their applications. Namely, the first-order Markov property for attributes only depends on the immediately preceding state and may fail to realize nuances of attribute change such as capturing long-term dependencies, accounting for non-linear learning trajectories, or other external factors (Pan et al., 2020). Although this strict assumption may not pose any issues in the scenario with two time points, the ignorance of time points other than the direct previous may not be desirable in cases with more than two time points.

In this study, we focus on the longitudinal extension of the LCDM (Henson et al., 2009), the most general and flexible model that offers a unified framework through which many previously developed DCMs, such as DINA, can be specified by placing constraints on the LCDM parameters. A hidden Markov DCM built on LCDM can be comparable to our proposed model in longitudinal assessment settings. However, a hidden Markov LCDM is restricted by the Markovian assumption for time points greater than two, as is well-known for HMMs in general (Glennie et al., 2023). Hence, our model is more

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

flexible in interpreting the entire length of a learning trajectory (e.g., how an attribute state at the first time point relates to its third time point state) as opposed to a limited Markovian perspective of a time point and its direct preceding. In other words, a hidden Markov LCDM may lead to biased conclusions regarding respondents' progress and transitions when the Markovian assumption is not met, whereas our model can be applied without these limitations (e.g., Pan et al., 2020).

3. Proposed model: extended Bayesian TDCM

We expand upon the multiple-group TDCM framework of Madison & Bradshaw (2018b) by adjusting the model formulation and offering a feasible estimation framework for model inference. Here, we utilize an updated regression-based formulation to incorporate additional covariates outside group-level treatment intervention effects.

The extended TDCM is composed of two partitions of the complete model that interact with one another: the item-attribute model and the latent transition model. For the item-attribute model, we specify LCDM to mirror the standard TDCM exactly. The proposed framework's primary contribution to the TDCM can be found in the chosen parameterization for the latent transition model. Departing from the transition analysis for the standard TDCM shown in Equation 2, we model profile transitions in the form of discrete attribute transition types between two adjacent time points using a multinomial logistic regression for each attribute. The key appeal of the extended TDCM lies in the flexible nature of regressions to include individual-level covariates for different transition types, as demonstrated in the following empirical and simulation studies. In the following, notation is first set up in Section 3.1, then the extended TDCM formulation is described formally in Section 3.2.

3.1. Notation

Here, we establish the setting and notation to be used for model formulation (Table 1). We consider $N \in \mathbb{Z}^+$ different respondents. Denote each respondent as $i \in [N]$, where $[N]$ is defined as $\{1,2,\ldots,N\}$ for all $N \in \mathbb{Z}^+$. We consider $T \in \mathbb{Z}^+$ time points. Denote each time point as $t \in [T]$. At time point t, there are $J_t \in \mathbb{Z}^+$ questions given to respondents. Let $J = \sum_{t=1}^{T} J_t$. Denote each question as $j \in [J]$ and $t_j \in [T]$ as the time when the question j is given. The response matrix is denoted as $Y \in \{0,1\}^{N \times J}$. Denote each element as $Y_{ij} \in \{0,1\}$, which indicates whether the i^{th} respondent answers the question j correctly. We consider Y as a random matrix and y_{ij} as the observed data of Y_{ij}. There are $K \in \mathbb{Z}^+$ possible attributes for each question. Denote each attribute as $k \in [K]$. Q-matrix $Q \in \{0,1\}^{J \times K}$ establishes the relationship between questions and attributes, such that an element $q_{jk} \in \{0,1\}$ indicates whether attribute k is required on the question j. In the current article, Q is assumed to be known and held constant. We use $\boldsymbol{\alpha}_i^{(t)} \in \{0,1\}^K$ to denote the attribute profile of respondent i at time t, and $\mathcal{A}$ to denote the collection of all attribute vectors $\boldsymbol{\alpha}_i^{(t)}$. There is a natural bijection between attribute vector $\boldsymbol{\alpha}_i^{(t)}$ and integer classes $c \in [2^K]$. Define vector $\mathbf{v} = [2^{K-1},\ldots,2^0]$, then we can see that $\mathbf{v}'\boldsymbol{\alpha}_i^{(t)} + 1 \in [2^K]$ for any $\boldsymbol{\alpha}_i^{(t)}$. With a slight abuse of notation, the term $\mathbf{v}^{-1}(c)$ is used to denote the corresponding attribute vector associated with integer class c. We also use $\mathbf{v}_k^{-1}(c)$ to denote the k^{th} element of $\mathbf{v}^{-1}(c)$. For example, suppose $K = 3$, then $\mathbf{v}^{-1}(1) = [0,0,0]$ and $\mathbf{v}^{-1}(8) = [1,1,1]$. Also, we have that $\mathbf{v}_1^{-1}(1) = 0$ and $\mathbf{v}_1^{-1}(8) = 1$. For respondent i and attribute k, let $\boldsymbol{\rho}_{ik} \equiv [\boldsymbol{\alpha}_i^{(1)\prime}\boldsymbol{e}_k,\ldots,\boldsymbol{\alpha}_i^{(T)\prime}\boldsymbol{e}_k] \in \{0,1\}^T$ denote the respondent i's latent transition vector for the attribute k corresponding to the set of profiles $\boldsymbol{\alpha}_i^{(t)}$, where $\boldsymbol{e}_k$ is defined as the k^{th} row vector of the identity matrix I_K. Notice that there is a natural bijection between attribute vector $\boldsymbol{\rho}_{ik}$ and integer classes $r \in [2^T]$. Define vector $\boldsymbol{\nu} = [2^{T-1},\ldots,2^0]$, then we can see that $\boldsymbol{\nu}'\boldsymbol{\rho}_{ik} + 1 \in [2^T]$ for any $\boldsymbol{\rho}_{ik}$. With a slight abuse of notation, we use $\boldsymbol{\nu}^{-1}(r)$ to denote the corresponding transition vector associated with integer class r. For example, suppose $T = 2$, then $\boldsymbol{\nu}^{-1}(1) = [0,0]$ and $\boldsymbol{\nu}^{-1}(4) = [1,1]$. For each question j, we define a matrix $\Delta^{(j)} \in \{0,1\}^{2^K \times 2^K}$. The c^{th} row vector of $\Delta^{(j)}$ represents the design vector for attribute profile $\mathbf{v}^{-1}(c)$. Denote the c^{th} row vector of $\Delta^{(j)}$ as $\boldsymbol{\delta}_c^{(j)} = [\delta_{cl}^{(j)}]_{l \in [2^K]}$, where $\delta_{cl}^{(j)} = 0$ if question j is not testing any of all

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

Table 1. List of notation used throughout the article

Symbol	Description
K	Number of latent attributes.
N	Number of respondents.
T	Number of assessments (time points).
J	Number of items over time.
Y	Item response matrix of size $N \times J$.
Q	Q-matrix of size $J \times K$.
$\boldsymbol{\alpha}_i^{(t)}$	Attribute profile of respondent i at time t. $\mathcal{A}$ is the collection of all $\boldsymbol{\alpha}_i^{(t)}$.
$\boldsymbol{\rho}_{ik}$	Transition pattern of attribtue k for respondent i.
$\Delta^{(j)}$	Design matrix of item j
$\boldsymbol{\beta}_j$	Item parameters.
$\boldsymbol{\gamma}_{rk}$	Transition parameters of transition type r for attribute k.

attributes in $\mathbf{v}^{-1}(l)$ or $\mathbf{v}^{-1}(l)$ contains any attribute that $\mathbf{v}^{-1}(c)$ does not have. Otherwise, $\delta_{cl}^{(j)} = 1$. Note that $\Delta^{(j)}$ may contain columns that are 0 for all entries. In the following discussions, we assume there is no column of zeros in $\Delta^{(j)}$. That is, $\Delta^{(j)} \in \{0,1\}^{2^K \times 2^{\sum_{k=1}^K q_{jk}}}$, and Δ denotes the collection of all $\Delta^{(j)}$. For question j, we have the corresponding item parameters $\boldsymbol{\beta}_j = [\beta_{jp}]_{p \in [P_j]} \in \mathbb{R}^{P_j}$, where $P_j = 2^{\sum_{k=1}^K q_{jk}}$. For ease of interpretation, we assume strict monotonicity such that a latent profile class with more acquired attributes must have a response-correctness probability no less than a profile class with fewer attributes. That is, we assume $\beta_{jp} > 0$ for all $p \geq 2$. We use multinomial logistic regressions to model latent attribute transition types. Γ is defined as the collection of $\boldsymbol{\gamma}_{rk} \in \mathbb{R}^{M_{rk}}$ indicating the regression parameters for the r^{th} transition category of the k^{th} attribute, and $X_{rk} \in \mathbb{R}^{N \times M_{rk}}$ is the design matrix for the attribute k. We denote the standard logistic function as $g(x) = \frac{1}{1+e^{-x}}$.

3.2. Model formulation

We define the likelihood function for the response data Y in the extended TDCM framework as

$$P(Y|\mathcal{B},\Gamma) = \prod_{i=1}^{N} \sum_{c_1=1}^{2^K} \cdots \sum_{c_T=1}^{2^K} P\left(\boldsymbol{\alpha}_i^{(1)} = \mathbf{v}^{-1}(c_1), \ldots, \boldsymbol{\alpha}_i^{(T)} = \mathbf{v}^{-1}(c_T)|\Gamma\right) \prod_{j=1}^{J} P\left(Y_{ij} = y_{ij}|\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c_{t_j}), \boldsymbol{\beta}_j\right). \tag{3}$$

Using Δ and $\mathcal{B}$, we can define the right-side probability in Equation 3 for the correctness of a response conditioned upon the latent profile class and LCDM parameters. The logit link function models the probability to be

$$P\left(Y_{ij} = y_{ij}|\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c), \boldsymbol{\beta}_j\right) = g\left((2y_{ij}-1)\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right) = \frac{\exp\left(y_{ij}\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}{1+\exp\left(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}. \tag{4}$$

With the item-attribute portion of the model defined, the former term of the full joint probability in Equation 3 serves as the conditional probability for the transition model. Four discrete, unordered scenarios ($0 \to 0$, $0 \to 1$, $1 \to 0$, and $1 \to 1$, where 0 indicates non-mastery and 1 indicates mastery and $\to$ indicates transition between two time points) may occur for each binary attribute k from one time

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

point to the next, thus motivating the use of multinomial logistic models. We define the conditional probability to be,

$$P\left(\boldsymbol{\alpha}_i^{(1)} = \mathbf{v}^{-1}(c_1), \ldots, \boldsymbol{\alpha}_i^{(T)} = \mathbf{v}^{-1}(c_T)|\Gamma\right) = \prod_{k=1}^{K} P\left(\boldsymbol{\rho}_{ik} = \boldsymbol{\nu}^{-1}(r)|\Gamma\right), \tag{5}$$

where $\boldsymbol{\nu}^{-1}(r) = [\mathbf{v}^{-1}(c_1)'\boldsymbol{e}_k, \ldots, \mathbf{v}^{-1}(c_T)\boldsymbol{e}_k]$. It follows that the right-hand side of Equation 5 is modeled by multinomial logistic regressions such that

$$P\left(\boldsymbol{\rho}_{ik} = \boldsymbol{\nu}^{-1}(r)|\Gamma\right) = \frac{\exp(\psi_{irk})}{\sum_{r^*=1}^{2^T}\exp(\psi_{ir^*k})}, \tag{6}$$

where $\psi_{irk} = \mathbf{x}'_{irk}\boldsymbol{\gamma}_{rk} \in \mathbb{R}$ and r is the transition type associated with a particular attribute. The covariates in consideration pass through the design matrix $X_{rk} \in \mathbb{R}^{N\times M_{rk}}$ to model the probability of r^{th} transition of attribute k. For identifiability purposes, the chosen default "baseline" level $0 \to 0$ attribute transition is constrained to zero, i.e., $\boldsymbol{\gamma}_{1k} = \mathbf{0}$ for all attributes. With this formulation, the model relaxes the Markovian assumption that hidden Markov DCMs are constrained to. The relaxation of this assumption does not present a problem for the TDCM, and would be able to account for scenarios, where an attribute status is not only dependent on its status at the previous time point. It is worth noting that the specification of Γ can be adjusted so that the full multinomial regression is reduced by omitting certain response levels, which can lower computational burden and improve interpretability. For instance, a study may be more interested in growth, i.e., $0 \to 1$ transition, and regression, i.e., $1 \to 0$ transition, than the two unchanging transition types. For time points greater than 2, Section 3.3 provides a more detailed discussion.

3.2.1. Remark 1

The formulation of this section reintroduces $\mathcal{B}$ for the item-attribute LCDM, and proposes the Γ regression parameter to model latent transition using multinomial logistic regressions. It is clear from the notation that the extended TDCM is capable of relaxing time- and group- measurement invariance. LCDM parameter $\mathcal{B}$ has the option to vary longitudinally, in settings, where the same test is not given at each time point. Transition probabilities modeled by Γ are dependent on treatment covariates and, therefore, vary across groups. While group variability is a necessity of the proposed transition model, time invariance may be applied to $\mathcal{B}$ to ease interpretations. In the following section, we detail a method of simulation-based inference using an augmentation approach to logistic data.

3.2.2. Remark 2

In comparison with relevant methods, the proposed TDCM extension has similar applications to the generalized hidden Markov DCM. The multilevel logistic HMM (Zhang & Chang, 2020) in particular, bears resemblance to the proposed hierarchical logistic regression-based TDCM reformulation. The distinction here is clear in covariate structure, where the HMM limits covariates to focus on learning behavior and attribute acquisition ability rather than having a flexible structure capable of group- and individual-level covariates. Moreover, the generalization that LCDM provides over DINA, often used alongside HMM combined with the Markovian assumption, are the important practical differences between the TDCM and hidden Markov DCM classes.

3.3. Transition types and modeling flexibility

Each additional time point in a longitudinal DCM increases the complexity of the model setup and makes interpretation more challenging. In this work, we showcase the modeling flexibility of the extended TDCM with an increasing number of assessment points. Table 2 presents the 2^T possible transition types per attribute can experience across the $T = 2$, $T = 3$, and $T = 4$ time periods.

The extended TDCM allows for flexible control of the design matrix, enabling the researcher to apply covariates selectively to specific transition types. For instance, with $T = 2$, transition type 2 ($0 \to 1$)

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

Table 2. The 2^T types of transitions (denoted r) for each attribute for (a) $T = 2$, (b) $T = 3$, and (c) $T = 4$

(a) $T = 2$

Transition	1	2	3	4
Time 1	0	0	1	1
Time 2	0	1	0	1

(b) $T = 3$

Transition	1	2	3	4	5	6	7	8
Time 1	0	0	0	0	1	1	1	1
Time 2	0	0	1	1	0	0	1	1
Time 3	0	1	0	1	0	1	0	1

(c) $T = 4$

Transition	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Time 1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Time 2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Time 3	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Time 4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Note: The different transitions (4, 8, and 16 types for $T = 2$, $T = 3$, and $T = 4$, respectively) are the possible ways that a respondent can have for a certain attribute over time.

indicates that a student has mastered the attribute during the transition. In this case, the researcher may choose to apply student-level covariates (e.g., gender) only to this transition, while excluding covariates for the other transition types, as the primary focus is on assessing gender differences in the mastery rates of each attribute.

Transition types increase with $T = 3$ and $T = 4$, offering additional flexibility and opportunities to explore interesting growth patterns. For example, with $T = 4$, transition type 8 ($0 \to 1 \to 1 \to 1$) indicates that a student mastered and retained the attribute, which may be more interesting compared to transition $0 \to 0 \to 0 \to 0$ (type 1), where the student never attains the attribute, or $1 \to 1 \to 1 \to 1$ (type 16), where the student always has the attribute. The transition trajectories $0 \to 0 \to 0 \to 1$ (type 2) and $0 \to 0 \to 1 \to 1$ (type 4) may indicate a delayed intervention effect, where the student gains the attribute some time after the intervention. In contrast, a trajectory such as $1 \to 1 \to 1 \to 0$ (type 15) could suggest a weak long-term effect. Another possible interpretation is to hold all time point states constant except for the first time point. Comparing the trajectories, $0 \to 0 \to 1 \to 1$ (type 4) and $1 \to 0 \to 1 \to 1$ (type 12) could be helpful in determining whether having the attribute at the initial time point matters. More complex transitions, such as $0 \to 1 \to 0 \to 1$ (type 6), are also possible, where a student mastered the attribute between the first and second time points, lost it between the second and third time points, and regained it between the third and fourth time points.

Note that these diverse transition patterns may not be observed with hidden Markov DCMs due to their Markovian assumptions. In contrast, the flexibility of the proposed extended TDCM allows for capturing various nuances in growth patterns and individual differences.

4. Bayesian estimation with Pòlya-gamma sampling

The computational complexity of the transition model's multinomial regression formulation increases significantly for each additional attribute, and more drastically so for each additional time point. Thus, a computationally feasible estimation framework is crucial for inference. We propose a framework

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

for Pòlya-gamma data augmentation (Polson et al., 2013) for both the item-attribute LCDM and the transition regression models, two components that compose the extended TDCM framework. The augmented data allows for Gibbs sampling to provide tractable posterior distributions for the parameters of interest, $\mathcal{B}$ and Γ. Pòlya-gamma sampling has been applied by Jiang & Templin (2019) for the two-parameter logistic item-attribute model and by Zhang et al. (2020) for the confirmatory DINA model. Recently, Balamuta et al. (2022) presented the implementation of Pòlya-gamma data augmentation for a class of LCDM, where the Q matrix is inferred. To our knowledge, this work represents the first implementation of multinomial Pòlya-gamma sampling for longitudinal DCM models.

4.1. Pòlya-gamma sampling procedure

The Pòlya-gamma distribution is used to sample the response model auxiliary variables $\{y^*_{jc} : j \in [J], c \in [2^K]\}$ corresponding to responses for each question and attribute profile. Since there are only 2^K possible attribute profiles, we only need $2^K \times J$ auxiliary variables instead of $N \times J$ auxiliary variables. We also use Pòlya-gamma distribution to sample transition model auxiliary variables ρ^*_{irk} corresponding to latent transition ρ_{irk} for each respondent, transition, and attribute, where we define $\rho_{irk} = \mathbb{1}(\boldsymbol{\rho}_{ik} = \boldsymbol{\nu}^{-1}(r))$.

To start, a Pòlya–gamma random variable $X \sim PG(b,c)$ with $b > 0$ and $c \in \mathbb{R}$ is distributed such that random variable

$$X \overset{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-1/2)^2 + c^2/(4\pi^2)}, \tag{7}$$

where $g_k \sim Ga(b,1)$. This distribution is key to the main result of Polson et al. (2013) that shows the Bernoulli logit link response can be augmented with the integral identity,

$$\frac{\left(e^{\theta}\right)^a}{\left(1+e^{\theta}\right)^b} = 2^{-b} e^{\kappa\theta} \int_0^{\infty} e^{-w\theta^2/2} p(w) \mathrm{d}w, \tag{8}$$

where $a,b > 0$, $\kappa = a - b/2$, and w distributed as a Pòlya–gamma random variable, i.e., $w \sim$ Pòlya-gamma$(b,0)$. Notice that when we have a linear function of predictors such that $\theta = \mathbf{x}^T\boldsymbol{\beta}$ in the case of the logistic regression, the left side of Equation 8 becomes the kernel of $\boldsymbol{\beta}$ likelihood. To be exact, the contribution of a single observation i to $\boldsymbol{\beta}$ likelihood

$$\mathcal{L}_i(\beta) = \frac{\left(\exp\left(\boldsymbol{\beta}^{\top}\mathbf{x}_i\right)\right)^{y_i}}{1+\exp\left(\boldsymbol{\beta}^{\top}\mathbf{x}_i\right)}, \tag{9}$$

uses main Pòlya–gamma distribution property in Equation 8 to result in tractable Gaussian posterior using conjugate Gaussian prior. We adopt a Gibbs sampling algorithm in which $\boldsymbol{\beta}$ and the augmented data are sequentially sampled at each iteration, as described in Section 4.1.2. The integrand of Equation 8 may also be adapted straightforwardly to the multinomial logit link likelihood for the transition model, as seen in the following Section 4.1.4. The added efficiency of Gibbs sampling for logistic regression Bayesian inference makes Pòlya-lgamma data augmentation highly appealing. Rigorous proofs and details of the algorithm can be found in Polson et al. (2013).

In the extended TDCM case, computations of posterior distributions for the parameters in $\mathcal{B}$ and Γ loosely follow the Pòlya-gamma framework of Balamuta et al. (2022). Consistent with that framework, the priors of the model are defined to be

$$\beta_{jp} \sim N(0, \sigma^2_{\text{prior};\beta_{jp}}), \tag{10}$$

$$\boldsymbol{\gamma}_{rk} \sim N(0, \Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}}), \tag{11}$$

where we choose $\mathcal{B}$ priors $\sigma^2_{\text{prior};\beta_{jp}}$ for each $1 \le p \le P_j$ and Γ priors $\Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}} \in \mathbb{R}^{M_{rk} \times M_{rk}}$. In accordance with the monotonicity condition for $\mathcal{B}$, elements of $\boldsymbol{\beta}_j$ are updated sequentially. Monotonicity here is based

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

on an "intuitive plausible assumption" (Rupp et al., 2010), and allows the modeler to decide whether to impose the constraint based on its appropriateness in context or even additional efforts required for model calibration. To be consistent with the motivating setting, the posterior distributions we use are truncated such that $\beta_{jp} > L_{jp}$ at each iteration, where L_{jp} is set such the full conditional distribution for intercept and interaction terms are unrestricted with only main effects restricted (Balamuta et al., 2022). These exact posteriors distributions are defined in Section 4.1.2.

The MCMC procedure for the extended TDCM is hierarchical, reflecting the hierarchy of the response model, which depends on the transition models in Section 3.2. The main Gibbs sampling procedure for the response LCDM samples $\mathcal{B}$ and $\mathcal{A}$, while the transition parameter Γ is sampled in the second-level procedures within the main MCMC for each of the K transition regressions using the same Pòlya-gamma scheme. To decrease the drastic computational cost of an additional attribute, the K second-level procedures are set to have drastically fewer iterations (m) compared to the LCDM Gibbs sampling iterations (M) such that $m \ll M$. As long as M is large enough to account for burn-in, our efforts show that posterior samples achieve convergence for small m. The minimum iteration $m = 1$ is used in all following empirical and simulation settings to reach convergence. However, m may be increased to a greater value (e.g., $m = 10$) with increased complexity to transition parameter Γ.

The following procedure samples three sets of parameters defined in the previous section: $\mathcal{A}$,$\mathcal{B}$, and Γ. The first step to sample $\mathcal{A}$ is dependent on both the LCDM and transition regressions, while the augmentation scheme to sample $\mathcal{B}$ and Γ depends on the sampled $\mathcal{A}$. Since each step relies on conditional distributions given the other parameters, each variable within $\mathcal{A}$,$\mathcal{B}$, and Γ should be randomly initialized in advance. The procedure is fully implemented using R and can be accessed through GitHub at [anonymized]. A single full iteration of the Gibbs sampler for $t \in [T]$ is described in the three steps below. Additional details are provided in Appendix A on the derivation of the posterior distribution for the proposed MCMC procedure. A summary of wall-clock runtimes for all empirical and simulation models is provided in Table 10 in the Supplementary Material, illustrating the practical efficiency of the proposed sampling procedure.

4.1.1. Step 1: sample $\mathcal{A}$

Each of the M iterations begins with sampling latent attribute profiles $\mathcal{A}$. For each i and t, $\boldsymbol{\alpha}_i^{(t)}$ is sampled from the conditional distribution

$$P(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)|Y,\mathcal{A}_i^{(-t)},\mathcal{B},\Gamma) = \frac{P\left(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)|\mathcal{A}_i^{(-t)},\Gamma\right)\prod_{j:t_j=t} g\left((2y_{ij}-1)\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}{\sum_{c_*} P\left(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c_*)|\mathcal{A}_i^{(-t)},\Gamma\right)\prod_{j:t_j=t} g\left((2y_{ij}-1)\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}, \tag{12}$$

where $j : t_j = t$ being question from time t. We define the set of profiles of respondent i excluding time t as

$$\mathcal{A}_i^{(-t)} = \{\boldsymbol{\alpha}_i^{(1)},\ldots,\boldsymbol{\alpha}_i^{(t-1)},\boldsymbol{\alpha}_i^{(t+1)},\ldots,\boldsymbol{\alpha}_i^{(T)}\}, \tag{13}$$

Further, the conditional probabilities can be calculated with the implicit attribute transition probabilities from one time point to its next using transition regression parameter Γ. Equation (12) is computed such that

$$P(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)|\mathcal{A}_{i(-t)},\Gamma) = \frac{P(\boldsymbol{\alpha}_i^{(1)} = \mathbf{v}^{-1}(c_1),\ldots,\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c),\ldots,\boldsymbol{\alpha}_i^{(T)} = \mathbf{v}^{-1}(c_T)|\Gamma)}{\sum_{c_t=1}^{2^K} P(\boldsymbol{\alpha}_i^{(1)} = \mathbf{v}^{-1}(c_1),\ldots,\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c_t),\ldots,\boldsymbol{\alpha}_i^{(T)} = \mathbf{v}^{-1}(c_T)|\Gamma)}, \tag{14}$$

which involves the calculation of the joint probabilities

$$P\left(\boldsymbol{\alpha}_i^{(1)} = \mathbf{v}^{-1}(c_1),\ldots,\boldsymbol{\alpha}_i^{(T)} = \mathbf{v}^{-1}(c_T)|\Gamma\right) = \prod_{k=1}^{K} P\left(\boldsymbol{\rho}_{ik} = [\mathbf{v}_k^{-1}(c_1),\ldots,\mathbf{v}_k^{-1}(c_T)]|\boldsymbol{\gamma}_{rk}\right), \tag{15}$$

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

where $\mathbf{v}_k^{-1}(\cdot)$ is defined in Section 3.1. The probabilities in the product are calculated according to the transition model defined in Equation 6. The resulting $\mathcal{A}$ carries attribute profiles for respondents at each of the T time points.

4.1.2. Step 2: sample $\mathcal{B}$

Following $\mathcal{A}$, we sample item-attribute LCDM parameters $\mathcal{B}$ by first sampling Pòlya–gamma auxiliary response variables $y_{jc}^* \sim PG(n_{jc}, \boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)$, where $n_{jc} = \sum_{i=1}^N \mathbb{1}\left(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)\right)$ for each question j and profile c. Then, β_{jp} can be sampled from the truncated normal posterior distribution $\beta_{jp}|\cdots \sim N(\mu_{\text{post};\beta_{jp}}, \sigma^2_{\text{post};\beta_{jp}})\mathbb{1}(\beta_{jp} \geq L_{jp})$, where $L_{j0} = -\infty$ and is set to zero for all other p for monotonicity constraint. The posterior has the derived parameters

$$\sigma^2_{\text{post};\beta_{jp}} = (\sigma^{-2}_{\text{prior};\beta_{jp}} + \boldsymbol{\Delta}_p^{(j)\prime}\text{diag}\left([y_{jc}^*]_{c=1}^{2^K}\right)\boldsymbol{\Delta}_p^{(j)})^{-1}, \tag{16}$$

$$\mu_{\text{post};\beta_{jp}} = \sigma^2_{\text{post};\beta_{jp}}\boldsymbol{\Delta}_p^{(j)\prime}\text{diag}\left([y_{jc}^*]_{c=1}^{2^K}\right)\tilde{\boldsymbol{z}}_j, \tag{17}$$

where $\boldsymbol{\Delta}_p^{(j)}$ is the p^{th} column vector of $\Delta^{(j)}$, namely, $\boldsymbol{\Delta}_p^{(j)} = [\delta_{1p}^{(j)}, \delta_{2p}^{(j)}, \ldots, \delta_{2^Kp}^{(j)}]$. $\tilde{\boldsymbol{z}}_j = \boldsymbol{z}_j - \Delta_{-p}^{(j)}\boldsymbol{\beta}_{j,-p}$ with $\boldsymbol{z}_j = [\frac{\kappa_{j1}}{y_{j1}^*}, \frac{\kappa_{j2}}{y_{j2}^*}, \ldots, \frac{\kappa_{j2^K}}{y_{j2^K}^*}]$ and $\kappa_{jc} = -\frac{1}{2}n_{jc} + \sum_{i=1}^N y_{ij}\mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c))$. Here, we denote $\Delta_{-p}^{(j)}$ is $\Delta^{(j)}$ with the p^{th} column removed and $\boldsymbol{\beta}_{j,-p}$ is $\boldsymbol{\beta}_j$ with the p^{th} entry removed.

Note the sampling of $\mathcal{B}$ here is the general case, where we specify a distinct set of LCDM parameters for each time point t, thus allowing tests at different time points to have different Q. In the more common setting, where Q remains the same over time, the following passage provides the posterior distributions such that $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J_t}$ encompasses model parameters over all time points. The unconstrained version of the proposed model is comparable to that of the standard TDCM, and the time-constrained adjustment to sampling $\mathcal{B}$ is straightforward to implement as well.

4.1.3. Time-constrained $\mathcal{B}$

The general case for sampling $\mathcal{B}$ in Section 4.1.2 allows for the flexibility of varying Q matrices over time, i.e., a unique set of questions at each time point. However, the motivating setting of Madison & Bradshaw (2018b) focuses on assuming a constant Q matrix for all time points, thus fixing the β_{jp} constant over time. We show here that the time-varying $\mathcal{B}$ may be straightforwardly altered to account for that. Let us consider the alternate case, where the J questions are repeated over time. In this case, we have $J \times T$ questions in total. With a slight abuse of notation, we use y_{ijt} to denote the correctness of respondent i's response to question j at time t. Then, $\beta_{jp} \sim N(\mu_{\text{post};\beta_{jp}}, \sigma^2_{\text{post};\beta_{jp}})$ for

$$\sigma^2_{\text{post};\beta_{jp}} = \left(\sigma^{-2}_{\text{prior};\beta_{jp}} + \boldsymbol{\Delta}_p^{(j)\prime}\text{diag}\left([y_{jc}^*]_{c=1}^{2^K}\right)\boldsymbol{\Delta}_p^{(j)}\right)^{-1}, \tag{18}$$

where $y_{jc}^* \sim PG(n_{jc}, \boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)$ with $n_{jc} = \sum_{i=1}^N\sum_{t=1}^T \mathbb{1}\left(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)\right)$. $\boldsymbol{\Delta}_p^{(j)}$ is defined in the same way as in the previous setting.

$$\mu_{\text{post};\beta_{jp}} = \sigma^2_{\text{post};\beta_{jp}}\boldsymbol{\Delta}_p^{(j)\prime}\text{diag}\left([y_{jc}^*]_{c=1}^{2^K}\right)\tilde{\boldsymbol{z}}_j, \tag{19}$$

where $\tilde{\boldsymbol{z}}_j = \boldsymbol{z}_j - \Delta_{-p}^{(j)}\boldsymbol{\beta}_{j,-p}$ with $\boldsymbol{z}_j = [\frac{\kappa_{j1}}{y_{j1}^*}, \frac{\kappa_{j2}}{y_{j2}^*}, \ldots, \frac{\kappa_{j2^K}}{y_{j2^K}^*}]$ and

$$\kappa_{jc} = -\frac{1}{2}n_{jc} + \sum_{i=1}^N\sum_{t=1}^T y_{ijt}\mathbb{1}(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)).$$

The posterior distribution here is directly adapted to Step 2 of the above procedure, with the remaining steps held constant. Implementations of this article default to constraining the extended TDCM by fixing $\mathcal{B}$, while time-varying $\mathcal{B}$ is an option that can be easily called upon as well.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

4.1.4. Step 3: sample Γ

In the final step, Γ parameters for transition multinomial logistic regressions are sampled. This is done by first sampling the Pòlya-gamma auxiliary variables $\rho^*_{irk} \sim PG(1, \eta_{irk})$ where $\eta_{irk} = \psi_{irk} - \log(\sum_{r_* \neq r} \exp \psi_{ir_*k})$, and recalling that $\psi_{irk} = \mathbf{x}'_{irk}\boldsymbol{\gamma}_{rk}$. Also recall that we constrain $\boldsymbol{\gamma}_{1k} = \mathbf{0}$ for all k to serve as the "baseline" transition. It follows that $\boldsymbol{\gamma}_{rk}$ can be sampled from the posterior distribution $\boldsymbol{\gamma}_{rk}|\cdots \sim N(\mu_{\text{post};\boldsymbol{\gamma}_{rk}}, \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}})$ with parameters

$$\Sigma_{\text{post};\boldsymbol{\gamma}_{rk}} = \left(\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}} + X'_{rk}\text{diag}\left([\rho^*_{irk}]_{i=1}^N\right)X_{rk}\right)^{-1}, \tag{20}$$

$$\mu_{\text{post};\boldsymbol{\gamma}_{rk}} = \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}}X'_{rk}\left(\boldsymbol{\zeta}_{rk} - \text{diag}\left([\rho^*_{irk}]_{i=1}^N\right)\mathbf{c}_{rk}\right), \tag{21}$$

where

$$\mathbf{c}_{rk} = [\log(\sum_{r_* \neq r} \exp \psi_{1r_*k}), \ldots, \log(\sum_{r_* \neq r} \exp \psi_{Nr_*k})], \tag{22}$$

and

$$\boldsymbol{\zeta}_{rk} = [\rho_{1rk} - 1/2, \ldots, \rho_{Nrk} - 1/2]. \tag{23}$$

We note here that the step to sample Γ is repeated for each of the K regression models, and a single taken sample is embedded within each M Gibbs iteration. The three MCMC inference steps sampling each latent attribute $\mathcal{A}$, item-attribute $\mathcal{B}$, and transition regression Γ are iterated M times to produce full posterior samples.

While the number of Gibbs sampling iterations required is case-specific, the efficiency that Pòlya-gamma provides flexibility in increasing iterations without a costly sacrifice to computation. In each of the following empirical and simulation studies, $M = 3{,}000$ with 500 burn-in iterations is used and often proved to be more than enough for the model to reach convergence. Convergence is assured in each of the following scenarios through careful checking of trace plots.

5. Empirical validation

For empirical validation, we first apply the extended TDCM to the two cases examined in Madison & Bradshaw (2018a) and Madison & Bradshaw (2018b), with the single-group and multiple-group settings, respectively. In both cases, the results presented here are validated by comparing them with the previously reported findings, assuming the prior results are considered valid. We showcase the additional insights provided by the proposed reformulation of the TDCM, which is primarily reflected in the transition regression parameter Γ. Further, we use a separate data set with the same assessment to demonstrate the extended TDCM's main contribution: its flexibility in including covariates in a multiple-group setting. Across all three scenarios, we assess model fit using posterior predictive checks, which represent an additional advantage of the proposed approach over the existing TDCM framework.

The main goal of the empirical section is to ensure that the proposed model provides the exact results that the standard TDCM was able to, while opening a window to the more complex simulation scenarios that emphasize the extended TDCM's added contributions. As we find in the following, the empirical data used here have no or a simple covariate structure under only two time points. The extended TDCM produces satisfactory results and demonstrates its suitability for the more complex settings that follow.

5.1. Single-group setting

5.1.1. Data and analysis

We use the mathematics assessment data from the studies of Bottge et al. (2014) and Bottge et al. (2015). The studies aim to evaluate the effectiveness of an instructional method (enhanced anchored instruction [EAI]) on mathematics exams over the course of $T = 2$ time points one year apart. The total

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

Table 3. *Q* matrix for empirical data (Bottge et al., 2014, 2015) indicating attribute requirements for each $J = 21$ test item questions

Item No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
RPR	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
MD	0	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
NF	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
GG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1

Note: The four attributes indicated are: ratios and proportional relationships (RPR), measurement and data (MD), number systems (NF), and geometry and graphing (GG).

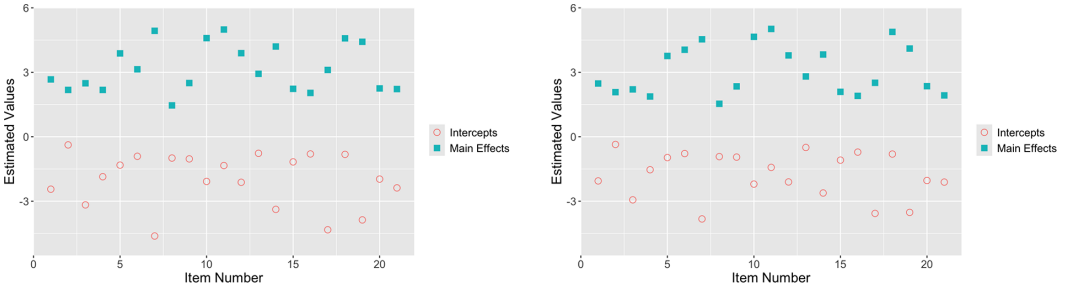

Figure 1. *Comparison of $\mathcal{B}$ point estimates between the standard TDCM(left) and the extended TDCM (right) for the single-group setting.*

number of students is $N = 849$, of which $N = 423$ received EAI and act as the treatment group while the control group of 456 students did not receive the instructional treatment. The mastery of four ($K = 4$) attributes of interest is implicitly measured: RPR, MD, NF, and GG. Each with $J = 21$ questions, the two tests measuring attributes are defined by the same *Q* matrix matching items to attributes, shown in Table 3. We see here that none of the items measure more than one attribute, meaning the item-attribute regression parameter $\mathcal{B}$ contains only main effects and no interaction terms.

The same data set and *Q* matrix were used for the single-group TDCM in Madison & Bradshaw (2018a). In the single-group setting, where no group membership is considered, the effect of the educational EAI treatment is not of interest and thus the transition regressions are simplified as explained in the following Γ posteriors. The Gibbs sampling procedure for the extended TDCM defined in Section 4 is followed here. We use $M = 3{,}000$ as the number of iterations for the MCMC procedure, with the first 500 burn-in samples removed from posterior estimates. The number of iterations appears adequate, given the speedy convergence of Gibbs samplings as seen in trace plots in the Supplementary Material. For Bayesian inference, the standard deviations of the normal distribution priors on $\mathcal{B}$ and Γ are $\sigma_{\text{prior};\beta_{jp}} = 1$ and $\Sigma_{\text{prior};\gamma_{rk}} = 0.5$, respectively, such that the priors are approximately non-informative with reasonable constraint.

5.1.2. Replication

We replicate key results shown by the original study of Madison & Bradshaw (2018a) and present additional results demonstrating extended TDCM's added value. The comparison begins with point estimates for $\mathcal{B}$ in Figure 1. The standard TDCM results on the left were obtained using the Mplus code from Madison & Bradshaw (2018a) (these results are identical to those presented in Figure 2 of Madison & Bradshaw (2018a)). The extended TDCM point estimates are the averages of the 2,500 burned-in posterior samples outputted by the Pòlya-Gamma algorithm. The results here are approximately identical for both the intercepts and main effects for each of the 21 questions.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

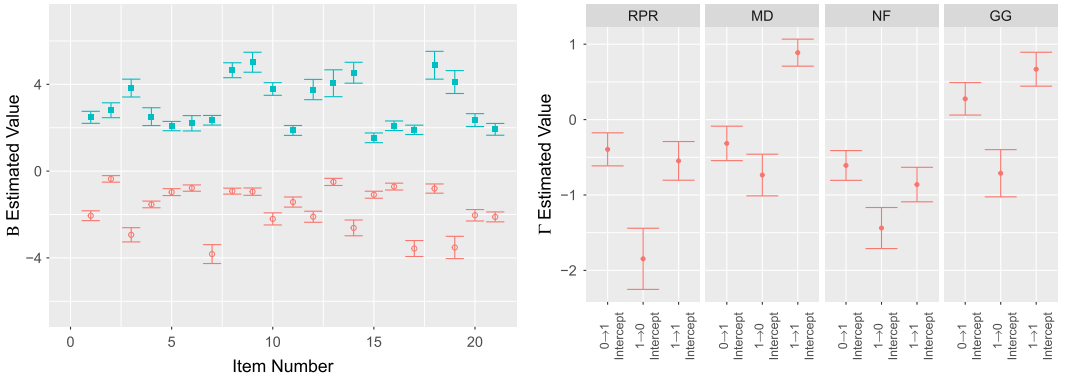

Figure 2. *Posterior sample distributions for $\mathcal{B}$ (left) and Γ (right) bounded by two standard deviations.*
Note: For $\mathcal{B}$, each question item contains an intercept in red and a main effect in blue. The four segments for Γ plot correspond to the four attributes: RPR, MD, NF, and GG. The indices within each attribute segment denote transition $0 \rightarrow 1$ intercept, $1 \rightarrow 0$ intercept, and $1 \rightarrow 1$ intercept, respectively.

5.1.3. *Further flexibility*

While this is encouraging, we further show the appeal of extended TDCM by noting that the proposed model offers an additional set of regression parameters Γ for transition, on top of LCDM parameters $\mathcal{B}$. For both sets of parameters, full posterior samples are available and show convergence evidenced in the Supplementary Material. Point estimates and their 95% credible intervals are shown for both sets of parameters in Figure 2 for the burned-in samples. We see that the variability of parameters remains roughly constant among all parameters, with $\mathcal{B}$ variability increasing slightly as point estimates depart from zero.

The full posterior distributions of Γ, shown on the right side of Figure 2, serve as a main appeal of our TDCM extension. For the single-group case, we note that the simplified transition regressions can be considered null models in that the multinomial regression for each attribute contains only the intercept. The figure shows four segments denoted by the labels on the top showing Γ for each of the $K = 4$ attributes. Within each segment, the indices on the bottom denote the intercepts for the transition $0 \rightarrow 1$, $1 \rightarrow 0$, and $1 \rightarrow 1$, respectively, with the $0 \rightarrow 0$ transition being the baseline level and 0 indicates non-mastery, 1 indicates mastery, and $\rightarrow$ indicates transition between two time points. The converged estimates here show approximate constant variance. It is also seen that the attribute regress transition $1 \rightarrow 0$, the loss of an attribute between the two time points, is lower in log-odds than other transition types. This makes sense in context given the underlying behavior that some respondents received treatment that helped them answer the questions between the two time points. These estimates are greatly useful in the context of regression, and they further lead to direct conversion into transition probabilities conditioned on whether an attribute is acquired at the initial time point.

The conditional transition probability is a direct function of Γ posteriors by nature of the logistic link. In the original study (Madison & Bradshaw, 2018a) where latent transition were not modeled by regressions, the transition probabilities were attained by comparing latent profiles $\mathcal{A}$ between the two time points. These are provided on the left side of Table 4, with the probability of transitioning conditioned on the presence of attributes at the initial, pre-test time point. In comparison, the transitioning probabilities posterior distributions derived from Γ posteriors are presented on the right side of Table 4, with point estimates and attached standard deviations in parentheses. These posteriors here are not naturally conditional; therefore, conditioning by force leads to the same standard deviation for estimates that share common attribute status at the initial time point (e.g., standard deviation given no attribute 1 are both 0.025 in Table 4). Precisely, in comparison, the extended TDCM provides the advantage of full posteriors for transition probabilities, whereas the standard TDCM only produces point estimates.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

Table 4. Comparison of implicit conditional transition probabilities for each $K = 4$ attributes of the single-group empirical study for the standard TDCM (left) and extended TDCM (right) posterior means and standard deviations in parenthesis

RPR		Post-test	
		0	1
Pre-test	0	.52	.48
	1	.21	.79
MD		**Post-test**	
		0	1
Pre-test	0	.57	.43
	1	.16	.84
NF		**Post-test**	
		0	1
Pre-test	0	.59	.41
	1	.31	.69
GG		**Post-test**	
		0	1
Pre-test	0	.42	.58
	1	.19	.81

RPR		Post-test	
		0	1
Pre-test	0	.598(.025)	.402(.025)
	1	.195(.037)	.805(.037)
MD		**Post-test**	
		0	1
Pre-test	0	.572(.028)	.428(.028)
	1	.163(.019)	.837(.019)
NF		**Post-test**	
		0	1
Pre-test	0	.651(.022)	.349(.022)
	1	.361(.038)	.639(.038)
GG		**Post-test**	
		0	1
Pre-test	0	.428(.025)	.572(.025)
	1	.201(.024)	.799(.024)

Note: The matrices are labeled with 1 being attribute mastery and 0 otherwise.

5.1.4. Goodness-of-fit

The Bayesian estimation approach adopted here enables the evaluation of model fit through posterior predictive checks, offering an additional advantage over the standard TDCM framework. The use of posterior predictive checks in this empirical study is loosely grounded in the framework proposed by Gelman et al. (1996). In psychometric applications, such checks have general precedent in educational measurement settings (Sinharay et al., 2006) and have been more specifically applied within the context of CDMs (Park et al., 2015). Although there are no treatments or covariates for the single-group case, null models themselves can be used to perform this predictive check. Here, the posterior predictive LCDM data are simulated and compared with the actual item-attribute data. We take the final 1,000 post-burn-in posterior samples and simulate posterior predictive data for each iteration. The one-to-one match provides percentage matching between the real data and 1,000 sets of posterior predictive data, which can be summarized in that the posterior predictive data matches on average 70.84% of actual data in the first time point, and on average 70.62% of actual data in second time point. The matching probabilities can further be broken down by question in Table 5 and by person in Figure 3. Alternative to percentage matching and done similarly in Sinharay et al (2006, Figure 6), Figure 4 uses the 1,000 posterior predictive data sets to observe the percentage of respondents who answer correctly for each question. These posterior predictive distributions align reasonably well with the observed data, though they reflect variation due to simulation noise. However, the percentage matching at around 70% also indicates room for improvement. Overall, the moderate match rate of approximately 70% suggests that the model captures key response patterns, though improvements may be possible by incorporating individual-level covariates, which are not available in the present dataset.

To further evaluate the predictive performance of the extended TDCM beyond posterior predictive checks, we report the area under the curve (AUC) and Brier Score at each time point. These predictive checks are commonly used by CDMs, such as Liang et al. (2023) using both in combination. In our

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

Table 5. Probabilities of extended TDCM fit to match the empirical response data are averaged over all 849 respondents for each of the 21 test items, done for both time points of the study in the single-group setting

Item No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Time 1	.71	.64	.71	.65	.69	.67	.84	.57	.63	.84	.85	.77	.62	.81	.61	.59	.88	.77	.74	.64	.66
Time 2	.66	.69	.66	.62	.73	.73	.79	.58	.67	.86	.88	.76	.69	.76	.62	.63	.75	.85	.67	.62	.59

Note: Posteriors for latent attribute α are used to identify profiles at each time point for the respondents, after which $\mathcal{B}$ posteriors infer logistic fits to be compared with the empirical data of interest.

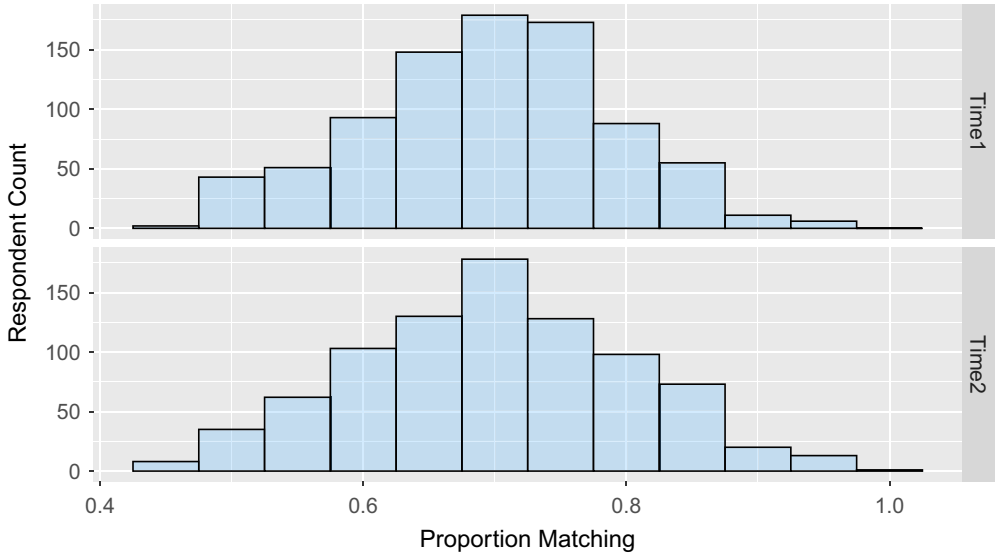

Figure 3. *Distribution of probability of extended TDCM fit to match the empirical response data average for* 849 *respondents averaged over* 21 *total questions, done for both time points of the study in the single-group setting.*

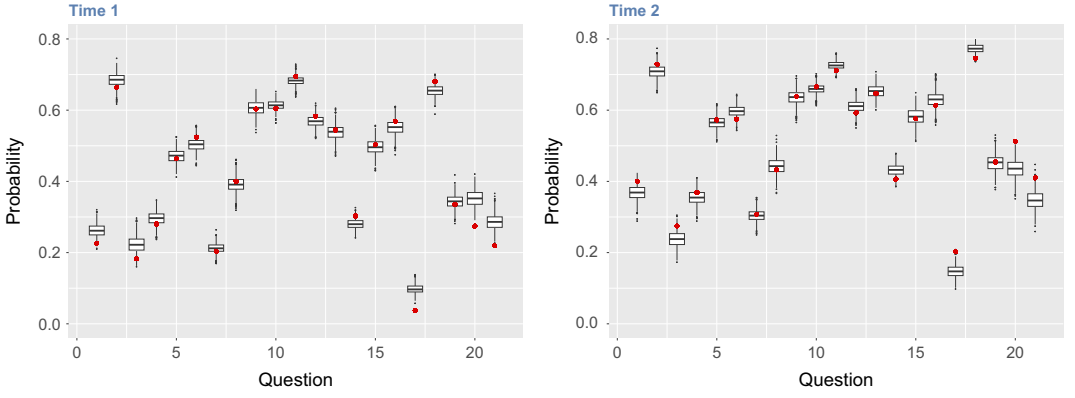

Figure 4. *Percentage of* 849 *respondents that answered each of the* 21 *questions correctly, with the value for the original data shown by the red points and simulated posterior predictive data shown by distributions.*
Note: Results for both time points are shown for the single-group setting.

analysis, the AUC across 21 test items was 0.868–0.87 (Time 1) and 0.877–0.886 (Time 2). The average Brier scores were 0.141 for Time 1 and 0.139 for Time 2. Taken together with the posterior predictive checks reported above, these results provide converging evidence for the model's predictive adequacy across both time points.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

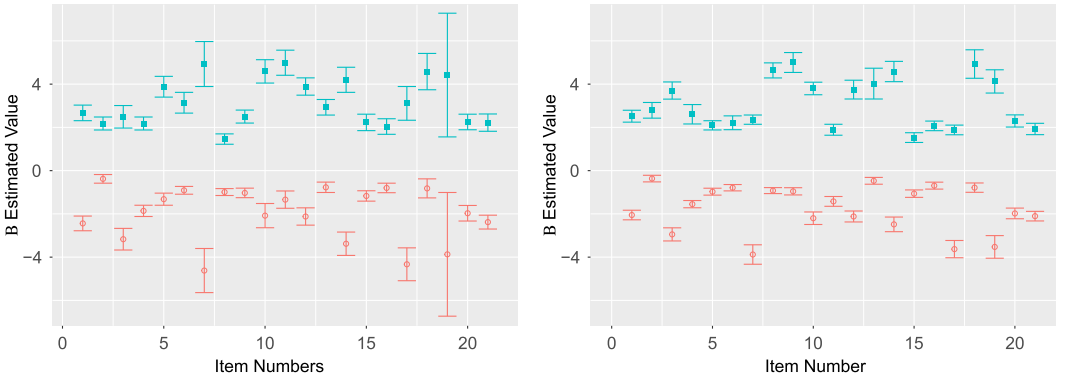

Figure 5. *Posterior $\mathcal{B}$ distributions plotted using the TDCM results (obtained using the Mplus code from Madison & Bradshaw (2018b)) bounded by two reported standard errors (left) is compared to the extended TDCM bounded by two posterior standard deviations (right). Note*: Red indicates the question intercepts, while blue indicates the main effects.

5.2. Multiple-group setting

5.2.1. Data and analysis

We now consider a multiple-group setting, where the key results are compared against those from the standard multiple-group TDCM (Madison & Bradshaw, 2018b). The same mathematical education data set is considered, and now with the educational intervention considered between the two time points. That is, we consider the group-level treatment covariate such that 423 respondents receive the EAI intervention (treatment group), while the remaining 456 respondents receive no intervention (control group). By treating the mathematical education intervention as a covariate, it can be applied to each of the $K = 4$ transition regressions. It follows that we have the option to apply the covariate for select transition types, intuitively the changing transition $0 \to 1$ and $1 \to 0$, while ignoring the covariate for unchanging transition levels $0 \to 0$ (baseline), and $1 \to 1$. This is done in our case for computational simplicity, as seen further in detail from the structure of Γ. Priors remain the same as those used by the single-group case, with $\sigma_{\text{prior};\beta_{jp}} = 1$ and $\Sigma_{\text{prior};\gamma_{rk}} = 0.5$. The same number of $M = 3{,}000$ Gibbs samples were computed, with the first 500 removed as burn-in samples. Trace plots for random indices of $\mathcal{B}$ and Γ are provided in the Supplementary Material showing proper convergence.

5.2.2. Replication

The resulting estimates for $\mathcal{B}$ are shown in Figure 5 for both the standard multiple-group TDCM (left) and its extension (right). The results on the left for the standard multiple-group TDCM were obtained using the Mplus code from Madison & Bradshaw (2018b) (which are identical to the graphical representation of Madison & Bradshaw (2018b, Table 2)). Note that the results of the extended TDCM shown here are nearly identical to the extended TDCM $\mathcal{B}$ posteriors for the single-group data set in Figure 2 in the previous section.

We note, however, that the same is not true between the standard single-group and multiple-group TDCM, with their $\mathcal{B}$ estimates being noticeably different from one another. This is seen by comparing Figure 5 (left) and Figure 2 (left). Intuitively, these results favor the extended TDCM over the standard multiple-group TDCM since the addition of a covariate on the same data set should not affect the attribute-response relationship reflected by $\mathcal{B}$. More so, we see in the standard TDCM results that items 7 and item 19 show particularly large variability in both $\mathcal{B}$ intercept and main effect. The extended TDCM does not reflect this in contrast.

5.2.3. Further flexibility

The complementing set of attribute transition Γ posteriors for the extended TDCM are shown in Figure 6. In the same manner as single-group Γ in Figure 2 (right), the multiple-group Γ is split into quadrants for each attribute. The indices explicitly show the adaptation of the treatment covariate, where

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

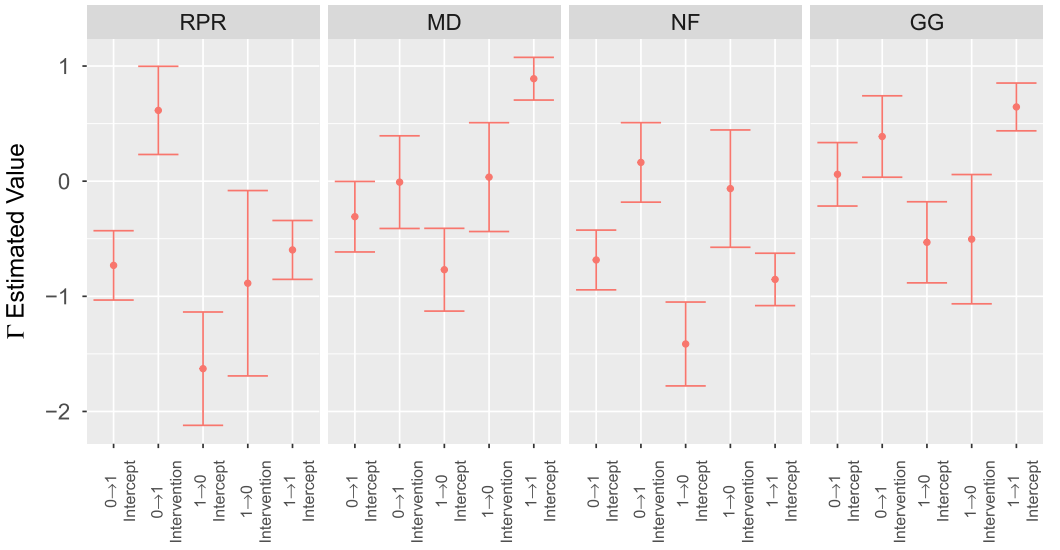

Figure 6. *The four segments for Γ plot correspond to the four attributes: RPR, MD, NF, and GG.*
Note: The indices within each attribute segment denote transition 0 → 1 intercept, 0 → 1 intervention, 1 → 0 intercept, 1 → 0 intervention, and 1 → 1 intercept, respectively.

transitions 0 → 1 and 1 → 0 receive the covariate. As is the goal for the EAI treatment, the treatment effects clearly indicate that the treatment increases the log-odd for the attribute gain transition 0 → 1 and decreases the log-odd for attribute loss transition 1 → 0 for attributes RPR, NF, and GG. The treatment has near-zero effects for attribute NF, and it follows that its intercepts are approximately those of the single-group in Figure 2 (right). These Γ posteriors can be further compared using attribute transition probabilities.

To demonstrate the attribute transition aspect of the treatment effect, the authors of the standard TDCM show the difference in transition probability between the treatment and control groups. Attribute probabilities for non-mastery to mastery (transition 0 → 1) and mastery to non-mastery (transition 1 → 0) are displayed for each attribute in Madison & Bradshaw (2018b, Table 5). The Γ posteriors are used again for the extended TDCM to produce transition probabilities, just as done for single-group previously. The two sets of probabilities for the control and treatment groups are differentiated using two sets of design matrices for Γ for whether treatment is received. Table 6 shows the full comparison in these transition probabilities between the proposed extension and its standard form. The table shows that the results are similar, which is encouraging for the extended TDCM since the standard TDCM derives transition probabilities using sampled attribute posteriors rather than transition regression posteriors.

5.2.4. Goodness-of-fit

The predictive check done for the previous single-group example is applied for the multiple-group model case, by questions in Table 7 and by respondents in Figure 7. Here, we see that on average 70.80% of predictive data matches at the first time point, and 70.60% at the second time point. Looking at the percentage answered correctly by respondents in Figure 8, the posterior predictive data results match approximately the original data results just as seen in Figure 4 for the single-group case. Furthermore, the AUC across 21 test items was 0.867–0.878 (Time 1) and 0.877–0.887% (Time 2), while the average Brier scores were 0.141 (Time 1) and 0.139 (Time 2). We see that AUC and Brier score match closely with those of the single-group case as well. Predictive evidence here indicates a moderately strong fit of the extended TDCM with multiple groups to the data.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

Table 6. Transition probabilities for 0 → 1 and 1 → 0 transitions between the control and treatment groups, compared between the standard multiple-group TDCM (Standard) and the extended TDCM (Extended)

		Control		Treatment	
Transition	Attribute	Standard	Extended	Standard	Extended
	RPR	.37	.329	.59	.472
Non-mastery to	MD	.39	.425	.48	.421
Mastery (0 → 1)	NSF	.38	.334	.45	.373
	GG	.48	.512	.69	.607
	RPR	.27	.272	.14	.134
Mastery to	MD	.17	.161	.15	.166
Non-mastery (1 → 0)	NSF	.34	.364	.28	.346
	GG	.24	.238	.15	.151

Table 7. Probabilities of extended TDCM fit to match the empirical response data are averaged over all 849 respondents for each of the 21 test items, done for both time points of the study in the multiple-group setting

Item No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Time 1	.72	.64	.71	.65	.69	.67	.84	.57	.64	.84	.85	.77	.62	.80	.60	.59	.88	.77	.74	.64	.66
Time 2	.67	.69	.66	.63	.73	.73	.79	.58	.67	.85	.88	.76	.69	.76	.62	.62	.75	.85	.68	.61	.59

Note: Posteriors for latent attribute α are used to identify profiles at each time point for the respondents, after which $\mathcal{B}$ posteriors infer logistic fits to be compared with the empirical data of interest.

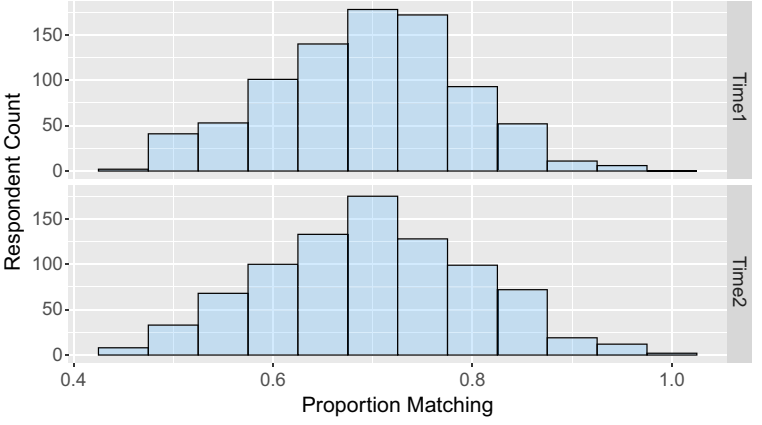

Figure 7. *Distribution of probability of extended TDCM fit to match the empirical response data average for* 849 *respondents averaged over* 21 *total questions, done for both time points of the study in the multiple-group setting.*

5.3. Multiple-group with covariates setting

5.3.1. Data and analysis

Next, we consider a separate dataset from the same studies (Bottge et al., 2014, 2015), which includes two student-level covariates in a multiple-group setting. Incorporating additional student-level covariates, introduces extra complexity into the transition parameter Γ structure for the extended TDCM. Given the same mathematics test, i.e., the same Q-matrix for attributes RPR, MD, NF, and GG, as those used in the single-group and multiple-group settings, this unique set of response data was taken and accompanied by the student-level covariates gender (male and female) and binary membership status in the English

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

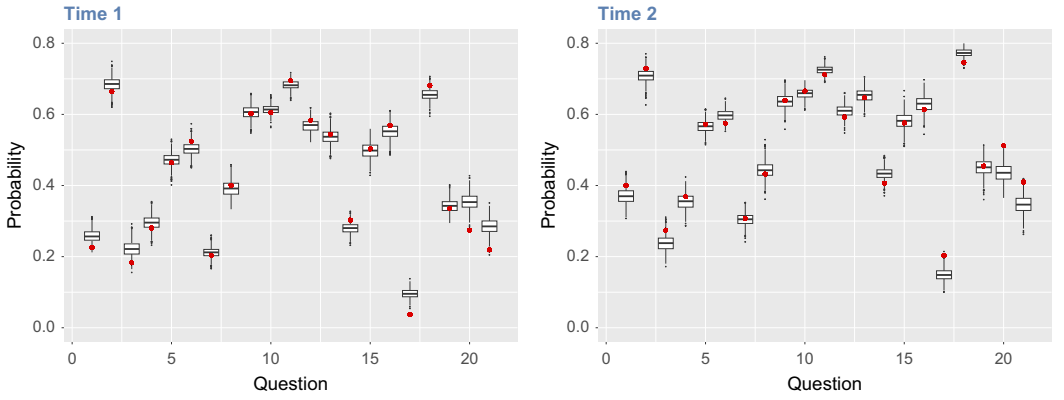

Figure 8. *Percentage of* 849 *respondents that answered each of the* 21 *questions correctly, with the value for the original data shown by the red points and simulated posterior predictive data shown by distributions.*
Note: Results for both time points are shown for the multiple-group setting.

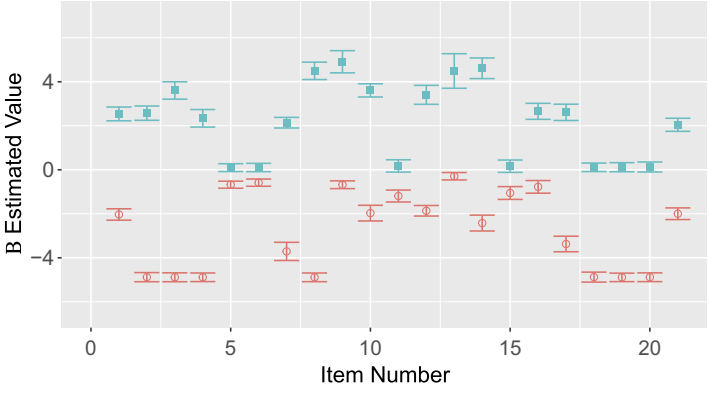

Figure 9. *Posterior $\mathcal{B}$ distributions of extended TDCM bounded by two posterior standard deviations (right).*
Note: Red indicates the question intercepts, while blue indicates the main effects.

as a Second Language (ESL) program. Observations are taken for 755 students in this data set, of which 368 students receive the EAI treatment and 387 do not.

The EAI intervention treatment of interest is treated as the same level as the student-level covariates in the extended TDCM, thus the focus of this setting lies in the additional dimensions of the Γ structure. Intervention treatment and the two covariates are enforced onto the $0 \to 1$ and $1 \to 0$ transitions, as is consistent with the previous empirical examples. In that same respect, 3,000 Gibbs sampler iterations were taken with the first 500 reported as burn-in. Priors used are Γ are $\sigma_{\text{prior};\beta_{jp}} = 1$ and $\Sigma_{\text{prior};\gamma_{rk}} = 0.5$.

5.3.2. Results

Figure 9 shows the posterior distribution of $\mathcal{B}$ parameters. Of more interest is Figure 10, showing the regression log-odds effects of intervention as well as the covariate effects of gender and ESL. The EAI intervention is seen to boost the log-odds of a $0 \to 1$ transition for all attributes except MD, for which its effect does not significantly depart from zero. It is also reasonable here that the intervention does not contribute to losing an attribute, i.e., the $1 \to 0$ transition. For student-level covariates, gender does not seem to be an influential factor for attribute transition. However, being in an ESL program does seem to inhibit students from gaining an attribute (i.e., $0 \to 1$ transition). This is reasonably so as the English ability would factor into how well a student learns in an English environment. It is also worth noting that

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

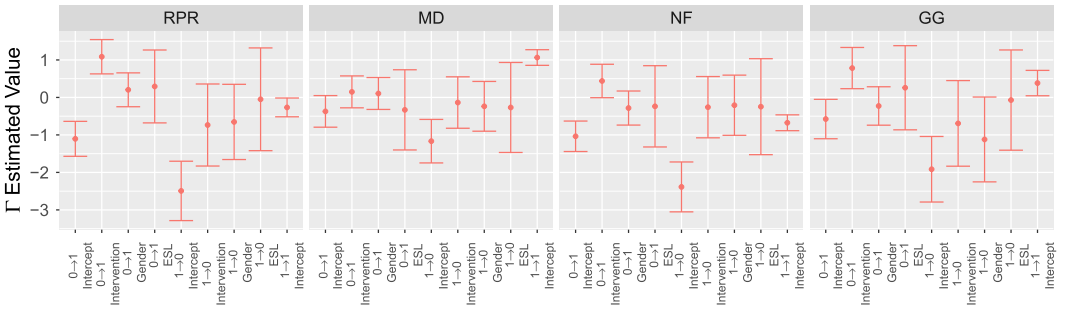

Figure 10. *The four segments for Γ plot correspond to the four attributes: RPR, MD, NF, and GG.*
Note: The indices within each attribute segment denote transition 0 → 1 intercept, 0 → 1 intervention, 0 → 1 gender, 0 → 1 ESL, 1 → 0 intercept, 1 → 0 intervention, 1 → 0 gender, 1 → 0 ESL, and 1 → 1 intercept, respectively.

Table 8. Probabilities of extended TDCM fit to match the empirical response data are averaged over all 755 respondents for each of the 21 test items, done for both time points of the study in the multiple-group covariates setting

Item No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Time 1	.72	.99	.99	.99	.66	.67	.85	.99	.64	.83	.85	.76	.62	.79	.67	.67	.87	.99	.99	.99	.67
Time 2	.64	.99	.99	.99	.70	.73	.80	.99	.67	.85	.88	.75	.69	.73	.68	.70	.72	.99	.99	.99	.59

Note: Posteriors for latent attribute α are used to identify profiles at each time point for the respondents, after which $\mathcal{B}$ posteriors infer logistic fits to be compared with the empirical data of interest.

if a student already has an attribute, being in an ESL program does not cause them to lose the attribute, as seen by the 1 → 0 ESL effects for each attribute in Figure 10.

5.3.3. Goodness-of-fit

The posterior predictive checks are based on response data generated using the original design matrix and posterior draws from the extended TDCM, consistent with the approach used in the two earlier empirical examples. The result here states that the posterior predictive data matches on average 81.81% of the actual data at the first time point, and 81.33% at the second time point. Table 8 provides detailed matching percentages by question, which is reflected by the histograms composed of matching percentages for individual students in Figure 11. The true probabilities of questions answered correctly in Figure 12 show the improvement from time point 1 to time point 2 being captured by the burned-in posterior predictive data sets. In addition, we observe AUC of 0.947–0.953 (Time 1) and 0.947–0.953 (Time 2) with Brier scores of 0.08 and 0.087, respectively, showing reliable posterior fits.

To parallel the comparison of posterior predictive checks for the single-group and multiple-group models, we additionally fit a simplified model without covariates to the data analyzed in this section. That is, we ignore the intervention effect along with student-level covariates for gender and ESL participation from the transition Γ structure. The percentage matching by question produced by this model's posterior predictive data using the final 1,000 post-burn-in samples is shown in Table 9. Compared to full model results in Table 8, we see almost no difference in the posterior fitting (including AUC and Brier score) here. This lack of difference is also seen for the single-group and multiple-group cases when comparing Tables 5 and 7. This may be caused by moderate Γ values associated with the Bernoulli intervention effect and student-level covariates of our empirical data.

This model fits the data better than the simpler single-group and multiple-group models discussed earlier, as seen in the goodness-of-fit analysis. However, a direct comparison between this and the other two cases is not ideal, as different datasets were used. Additionally, the goodness-of-fit assessment was conducted for each case to demonstrate the capabilities of the proposed extended Bayesian TDCM framework. The focus of the evaluation is not on the actual fit of the models, as 1) these are existing

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

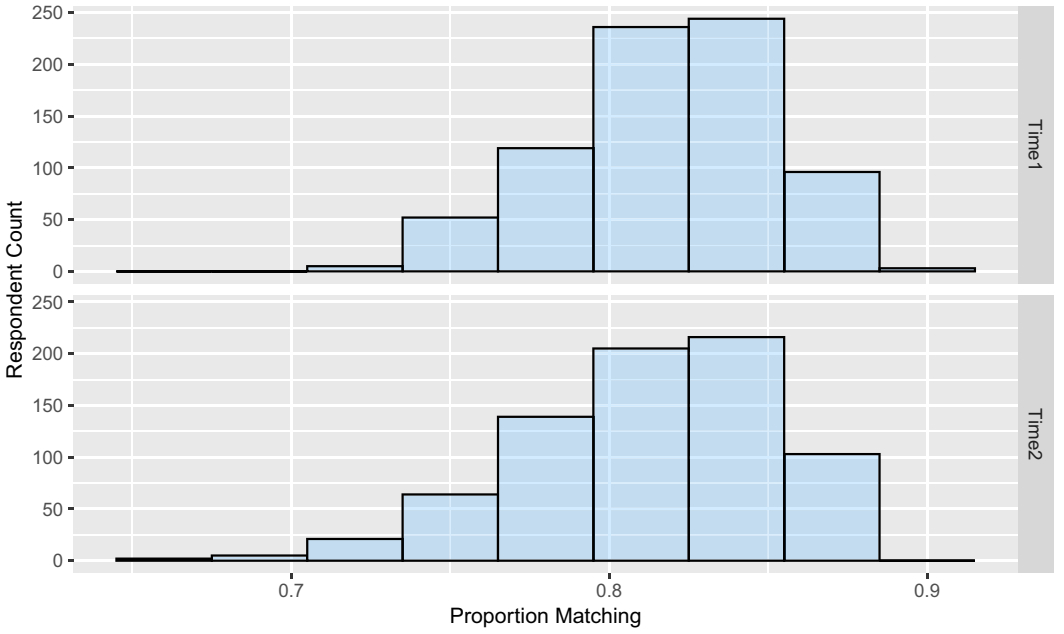

Figure 11. *Distribution of probability of extended TDCM fit to match the empirical response data average for* 849 *respondents averaged over* 21 *total questions, done for both time points of the study in the multiple-group covariates setting.*

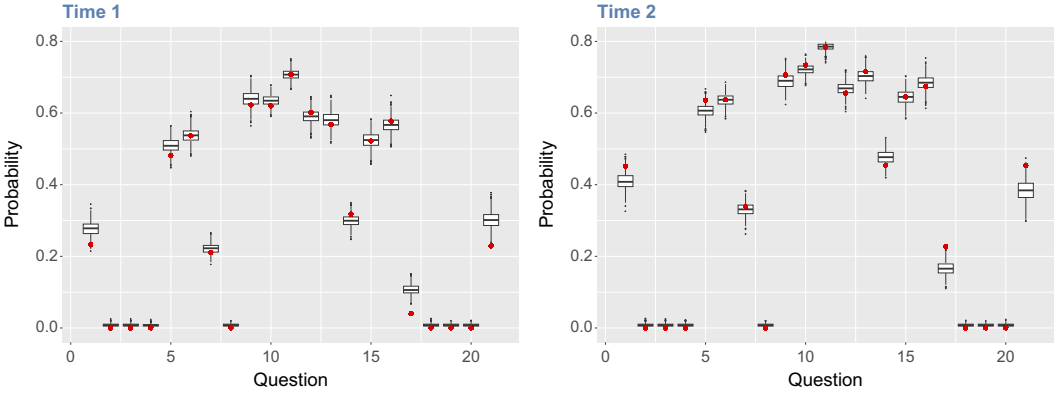

Figure 12. *Percentage of* 755 *respondents that answered each of the* 21 *questions correctly, with the value for the original data shown by the red points and posterior predictive data shown by distributions.*
Note: Results for both time points are shown for the multiple-group with covariates setting.

Table 9. Probabilities of extended TDCM fit to match the empirical response data are averaged over all 755 respondents for each of the 21 test items, done for both time points of the study in the single-group, no-covariate setting

Item No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Time 1	.72	.99	.99	.99	.66	.67	.85	.99	.64	.83	.85	.76	.62	.79	.67	.67	.87	.99	.99	.99	.67
Time 2	.64	.99	.99	.99	.70	.73	.80	.99	.67	.85	.88	.75	.69	.73	.68	.70	.72	.99	.99	.99	.59

Note: Posteriors for latent attribute α are used to identify profiles at each time point for the respondents, after which $\mathcal{B}$ posteriors infer logistic fits to be compared with the empirical data of interest.

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

models, and 2) the goal is to promote the extended, flexible modeling framework, rather than the models themselves.

6. Simulation study

In the simulation studies for the extended Bayesian TDCM, four simulation settings are used to evaluate the performance of the proposed extended TDCM. These scenarios demonstrate the potential for TDCM to model data with additional complexity compared to the empirical data analyzed in Section 5. In each setting, we provide the posteriors for $\mathcal{B}$ and Γ as is done in the previous section. With true parameter values that simulated data used available to us, we draw focus on parameter recovery rates, or credible interval coverage probability, shown for each setting. In the favorable scenario, the extended TDCM would be able to produce posterior samples with credible intervals having high coverage of true simulation parameters.

The simulation settings are organized as follows: We begin with a non-complex two time point scenario that includes no item-attribute interaction terms with only an intervention treatment, mirroring the previous empirical set-up (Setting 1). The model then becomes more flexible and increases the number of parameters by introducing item-attribute interaction effects (Setting 2). This is followed by the inclusion of additional covariates as is the appeal of logistic regressions for attribute transition (Setting 3). We further push the extended TDCM to account for data with three time points, which drastically increases the number of transition regression parameters (Setting 4).

In each simulation setting, data is generated using 800 respondents ($N = 800$), 21 questions ($J = 21$), and 3 attributes ($K = 3$). A total of 100 data sets, each of which includes item-correct response data Y, and respondent profiles α, are simulated from fixed $\mathcal{B}$ item-attribute parameters and Γ transition regression parameters. We set prior for each parameter with intention to be approximately non-informative, such that $\mathcal{B}$ and Γ have the respective standard deviation priors $\sigma_{\text{prior};\beta_{jp}} = 2.5$ and $\Sigma_{\text{prior};\gamma_{rk}} = 1$ for all four settings. Each set of data Y and attribute profiles α exhausts 3,000 Gibbs sampling iterations with 500 burn-in samples. The remaining 2,500 posterior samples are taken in mean, and the distribution of 100 simulation means are reported. These distributions aim to estimate the true values of $\mathcal{B}$ and Γ, thus the coverage rate of each simulation's accounting of the true parameter values is reported as well.

The results following indicate high credible interval coverage by posteriors produced in all simulation cases, even for three time points. This indicates the extended TDCM and its inference framework to be computationally proficient. In addition to model fit, we report total computation times for all simulation settings in Table 10 in the Supplementary Material. In one comparison, the extended TDCM for single-group (9 min 52 sec) requires less than half of the standard TDCM (20 min 54 sec). Further results show that the proposed approach achieves efficient convergence even under complex model structures.

6.1. Setting 1: with treatment covariate only

In the simplest simulation, a single intervention covariate is used, where half of the respondents receive the intervention. The Q matrix is specified such that $\mathcal{B}$ contains active parameters of 36 main effects without any interaction effect, as shown in Section 3.1 of the Supplementary Material. Intervention here is distributed evenly amongst the total of 800 respondents, with both the control and treatment groups having a size of 400. For computational ease and at no cost to the results, the intervention treatment is applied only to the $0 \to 1$ transition.

The resulting posterior point estimates and variation bound of two standard deviations can be seen in Figure 13 for $\mathcal{B}$ (left) and Γ (right). For $\mathcal{B}$ indices, the first 21 indicate intercepts and are followed by the 21 main effects. Transition Γ's are divided into three sections, corresponding to $K = 3$ multinomial logistic regressions for each attribute. For each attribute, the transition types are indexed by $0 \to 1$ intercept, $0 \to 1$ main effect, $1 \to 0$ intercept, and $1 \to 1$ intercept, respectively. Here, the "true values" of both sets of parameters used to simulate the data modeled are also given by the solid blue and gold points for $\mathcal{B}$ and teal points for Γ to validate our results directly. We see that the coverage of the estimated

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

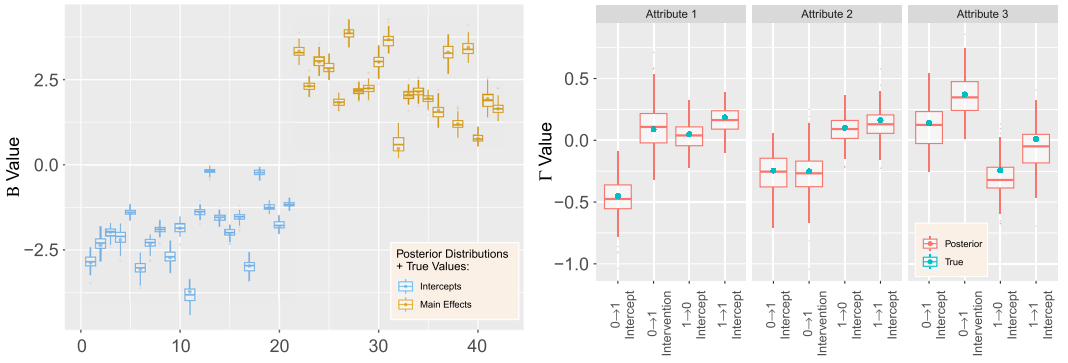

Figure 13. *Posterior $\mathcal{B}$ (left) and Γ (right) distributions bound by two standard deviations for the $K = 3$ simulation setting.*
Note: The four indices for each attribute denote transition 0 → 1 intercept, 0 → 1 treatment (single group), 1 → 0 intercept, and 1 → 1 intercept, respectively.

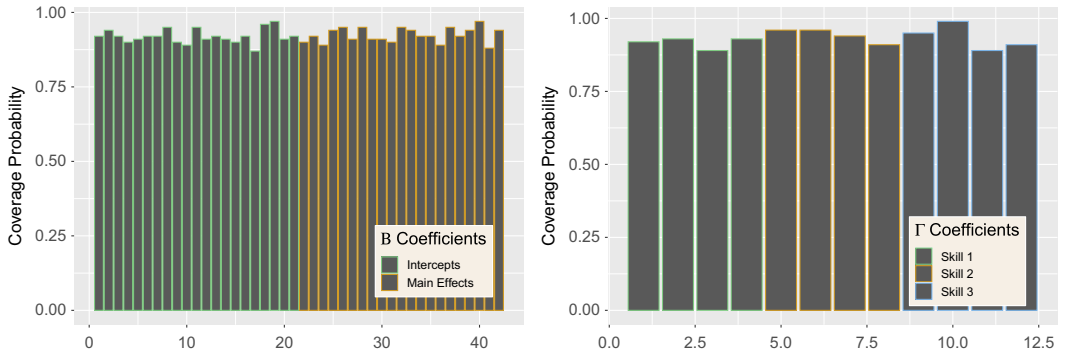

Figure 14. *True value coverage rates for* 95% *credible intervals of $\mathcal{B}$ (left) and Γ (right) distributions bound for the $K = 3$ simulation setting.*
Note: The four indices for each attribute denote transition 0 → 1 intercept, 0 → 1 treatment (single group), 1 → 0 intercept, and 1 → 1 intercept, respectively.

posteriors in Figure 14 indicate high rate of recovery for both sets of parameters, with an average of 92.17%(SD = 2.39) for the left figure $\mathcal{B}$ and 93.17%(SD = 3.01) for the right figure Γ. That is, almost all of the 100 simulations produced posteriors that covered true parameter values.

This simulation setting is closely related to the empirical set-up of Section 5, and serves as the foundational case of these simulation studies. Complexity is then further added to explore how the current efficacy persists.

6.2. Setting 2: with item-attribute interaction terms

The following simulation setting serves to additionally include LCDM item-attribute interaction terms to the structure of $\mathcal{B}$. This is done by updating the Q matrix to have questions impacted by multiple attributes, as shown in Section 3.1 of the Supplementary Material. Recall in Section 4.1.2 the $\mathcal{B}$ posteriors are truncated such that $\beta_{jp} > L_{jp}$ for main item-attribute effects, but not intercept effects. Interaction terms included in this study are treated the same as intercepts and are not truncated in the posterior since having full normal posteriors does not violate the monotonicity condition.

Figure 15 shows posterior distributions of $\mathcal{B}$ interactions in green to the right of 21 intercept terms in blue and 36 main effect terms in gold. The structure and results of Γ remain the same as seen in the previous Section 6.1. Notice in the figure that $\mathcal{B}$ interaction terms exhibit differences compared to those of the intercepts and main effects. These interaction posteriors remain unbiased, but have large variance as the model expresses uncertainty about these estimates. Coverage of the estimated posteriors

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

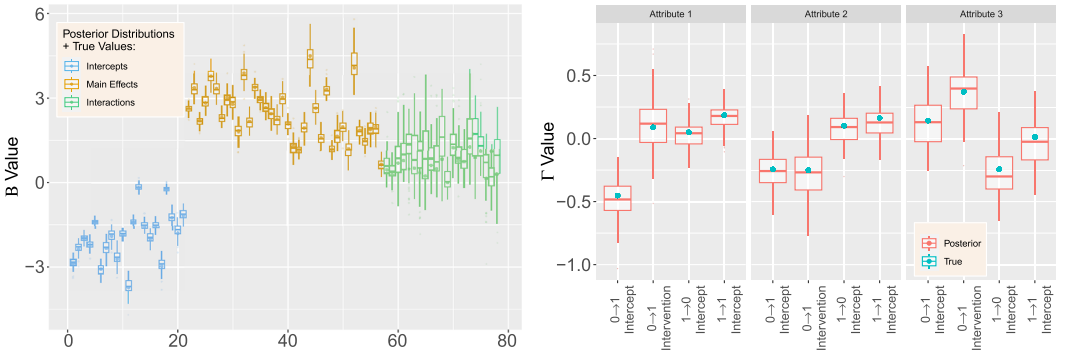

Figure 15. *Posterior $\mathcal{B}$ (left) and Γ (right) distributions bound by two standard deviations for the simulation setting with four covariates.*

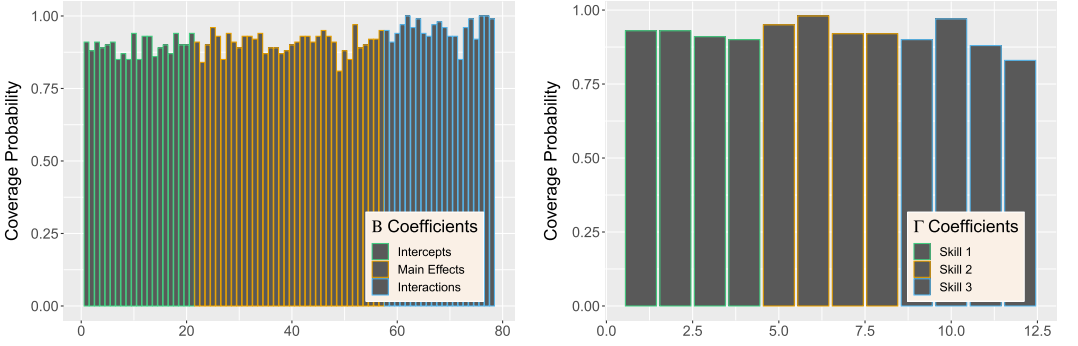

Figure 16. *True value coverage rates for* 95% *credible intervals of $\mathcal{B}$ (left) and Γ (right) distributions bound for the $K = 3$ with additional covariates simulation setting.*
Note: The first five estimates correspond to the $0 \rightarrow 1$ transition (intercept plus four covariate effects), followed by the next five for $1 \rightarrow 0$, and the intercept for $1 \rightarrow 1$.

in Figure 16 again report high recovery for both sets of parameters, with an average of 91.7% (SD = 4.15) for the left figure $\mathcal{B}$ and 91.8% (SD = 4.01) for the right figure Γ. Note that the large posterior variances on $\mathcal{B}$ interactions have led to slightly greater coverage of 95.57% (SD = 3.79) compared to other terms.

Item-attribute interactions are an aspect in most empirical settings. Although the effects themselves may not be impactful, this section shows how well the extended TDCM does with an increased number of parameters. In the following, the flexibility of the newly introduced transition regressions is updated.

6.3. Setting 3: with additional covariates

The inclusion of respondent covariates is a direct extension to the standard LCDM and is easily done given the updated regression formulation for LTA. This serves to be a crucial aspect for the DCMs, as individual-level covariates are nontrivial in determining how a respondent's attribute status changes, outside the intervention treatment. The simulation setting adjustment of this section addresses this limitation of the standard TDCM and demonstrates the flexible covariate structure of Γ by including additional covariates.

Keeping our intervention treatment the same, we draw two variables from the uniform distribution (i.e., $\mathcal{U}[0,1]$), and the normal distribution (i.e., $\mathcal{N}(30,10)$) as toy covariates to be appended to the respondent design matrix in addition to the treatment covariate. These respondent-level covariates of the model are again only applied to the $0 \rightarrow 1$ transitions for computational simplicity and ease of display.

Figure 17 shows the estimated posteriors with the true simulation values for our two sets of parameters. Indices for $\mathcal{B}$ remain the same, whereas Γ are organized such that for each attribute, the

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

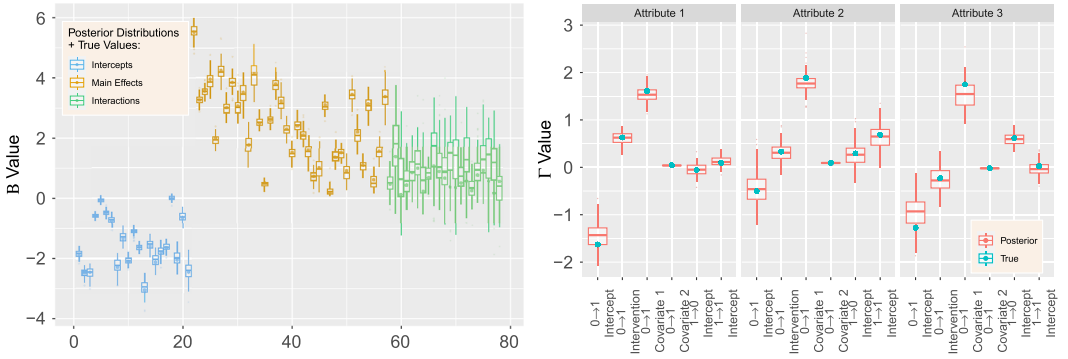

Figure 17. *Posterior $\mathcal{B}$ (left) and Γ (right) distributions bound by two standard deviations for the simulation setting with four covariates.*

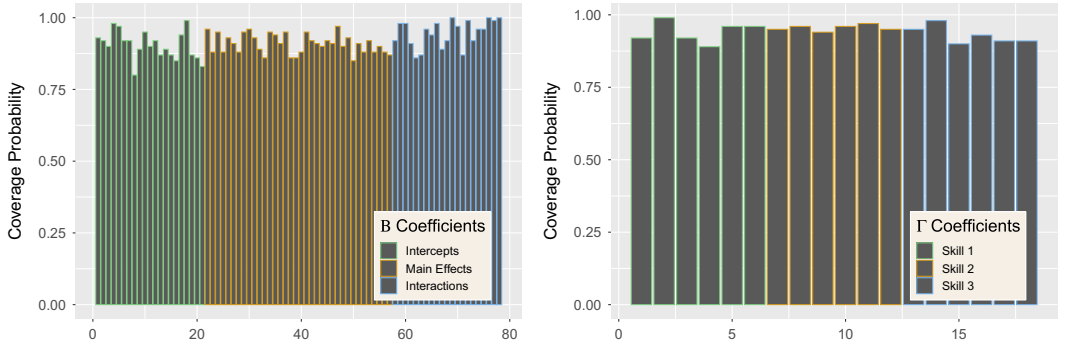

Figure 18. *True value coverage rates for* 95% *credible intervals of $\mathcal{B}$ (left) and Γ (right) distributions bound for the $K = 3$ with additional covariates simulation setting.*
Note: The first five estimates correspond to $0 \rightarrow 1$ transition (intercept plus four covariate effects), followed by the next five for $1 \rightarrow 0$, and the intercept for $1 \rightarrow 1$.

indices correspond to the $0 \rightarrow 1$ transition intercept, intervention effect, uniform distribution covariate effect, normal distribution covariate effect, followed by the intercepts for $1 \rightarrow 0$, and $1 \rightarrow 1$. In this right side Γ figure, we can see that the posterior distributions corresponding to the Gaussian distributed covariate ("$0 \rightarrow 1$ Covariate 2") do not spread as widely as those of other covariates. This is in accordance with our expectation as the normal distribution we are drawing from has much larger mean than the other covariates, thus resulting in smaller standard error. Figure 18 finds that coverage rates for both $\mathcal{B}$ and Γ remain favorable and are not noticeably different from those of the previous simulation settings, with an average of 91.8% (SD = 4.46) for the left figure $\mathcal{B}$ and 94.17% (SD = 4.01) for the right figure Γ.

The results here are encouraging for the extended TDCM, as seen from the unwavering ability of the model to recover true parameter values as complexity increases. The coverage rates do not experience much, if any, change as the structural complexity as well as the number of parameters increase.

6.4. Setting 4: $T = 3$ time points

We further evaluate the performance of the extended TDCM in the scenario of $T = 3$ time points while keeping the remaining settings consistent with those of the previous simulations, with only the intervention treatment covariate. Note that the previously discussed empirical studies (Madison & Bradshaw, 2018a, 2018b) in Section 5 were based on two time points only. Thus, this simulation demonstrates the capacity of the extended TDCM that goes beyond the standard TDCM applications.

Here, the transition multinomial response for each logistic regression increases to 2^T levels, or distinct ways an attribute can change between the three sequential time points. The number of responses

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

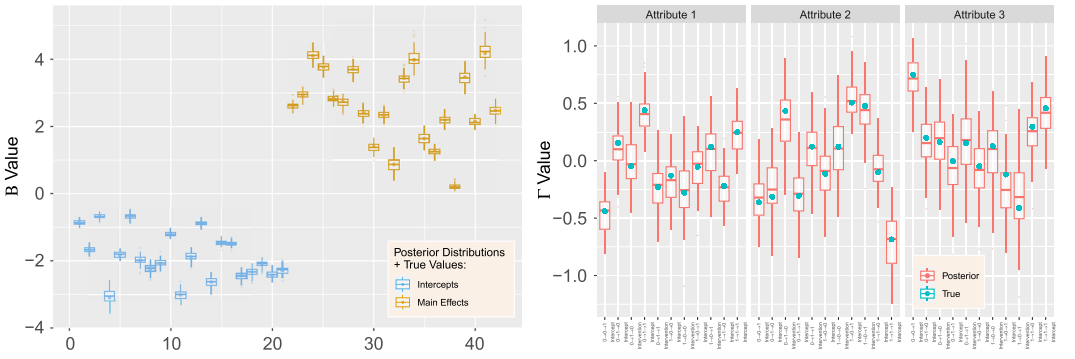

Figure 19. *Posterior $\mathcal{B}$ (left) and Γ (right) distributions bound by two standard deviations for the $T = 3$ simulation setting.*

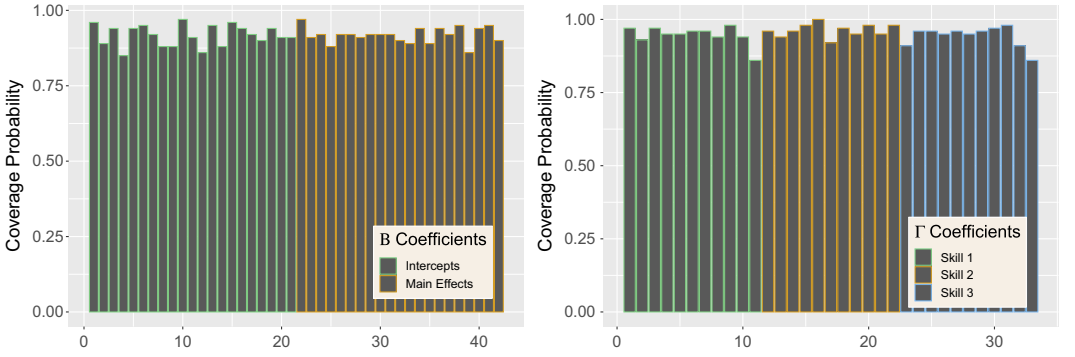

Figure 20. *True value coverage rates for* 95% *credible intervals of $\mathcal{B}$ (left) and Γ (right) distributions bound for the $K = 3$ simulation setting.*
Note: The four indices for each attribute denote transition 0 → 1 intercept, 0 → 1 treatment (single group), 1 → 0 intercept, and 1 → 1 intercept, respectively.

for the multinomial regression doubles from the previous simulation and empirical settings, where $T = 2$. That is, for each of the $K = 3$ attributes, Table 2(b) shows the possible transitions ($r \in [1,8]$) that can happen for $T = 3$.

For three time points, assume the intervention treatment takes place between the first time point and the second time point, but not between the second and third time points. Intuitively, this is to look at the long-term effect of a single intervention occurrence. The treatment covariate is applied to the transitions $0 \to 1 \to 0$ (type 3), $0 \to 1 \to 1$ (type 4), $1 \to 0 \to 0$ (type 5), and $1 \to 0 \to 1$ (type 6) (corresponding to $r = \{3,4,5,6\}$) in Table 2(b) while the first transition type $0 \to 0 \to 0$ serves as the baseline constraint. Note that this setup is consistent with applying covariates to transition $0 \to 1$ and $1 \to 0$ for $T = 2$.

For Γ, each transition with the treatment covariate contains both an intercept and slope, while the remaining transition only has an intercept excluding the baseline transition. This totals $2^T + 4 - 1 = 11$ values of transition type γ_k for each attribute. These 11 parameters for each attribute are explicitly labeled as indices in the Γ posteriors on the right side of Figure 19. The structure of Γ here is indexed following the order of Table 2(b). Figure 20 shows that overall the estimate intervals demonstrate proficient coverage of true values, with a mean of 91.74% (SD = 3.01) for the left figure $\mathcal{B}$ and 95.06% (SD = 3.08) for the right figure Γ.

Useful interpretations can be inferred for specific posteriors in the figure. For example, the inclusion of intervention on students would favor the learning trajectory $0 \to 1 \to 1$ (type 4) for each of the three attributes, based on the positive intervention effect for each. This implies that the intervention would help a student who did not originally have an attribute gain the attribute in the second time point and retain it at the third time point. In addition, none of the attributes showed a positive tendency

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

toward the $1 \rightarrow 0 \rightarrow 0$ (type 5) trajectory. Naturally, the intervention here is not designed for a student who has an attribute to lose the attribute. The goal of aiding the acquisition of attributes, or at least retaining an attribute, is supported by these interpretations. We shed light on the interpretation of learning trajectories that seem less straightforward in the next section. It is also worth noting that in both of the trajectories discussed, the start state and end state are of particular interest and provide much insight. The hidden Markov DCM, however, would not allow for such interpretation due to its Markovian assumption.

7. Discussion

7.1. Summary

This study introduced a more flexible extension to the recent TDCM. The TDCM is a statistical model measuring change in latent attribute mastery over time in educational testing studies. A focus of an intervention treatment effect leads to a multiple-group context, where a treatment group receives the intervention, and a control group does not receive it. Our extension to the TDCM reformulates the standard TDCM by choosing multinomial logistic regressions to model latent transitions directly, while the LCDM item-attribute model remains the same as used for the TDCM. The primary advantage of our extension lies in the full predictive posteriors for the transition parameter Γ. Regression formulation allows for a full consideration of respondent-level covariates, which is not a function of the standard TDCM but an important aspect of cognitive diagnostic models. The intervention effect can then be treated as an additional covariate in implementation. Using this set-up, posteriors for attribute transition probabilities can be functionally obtained from transition Γ and provide a measure of the effectiveness of an intervention treatment, along with log-odds interpretability of logistic regressions.

An appealing method of Bayesian inference is crucial for a complex psychometric model such as the proposed. Whereas the standard TDCM relied on the proprietary `Mplus` (Muthén & Muthén, 2017), the extended TDCM offers a Gibbs sampling framework made efficient by Pòlya-gamma data augmentation. In empirical studies, we reused the mathematical education data set analyzed by the standard TDCM. Approximately same LCDM results were achieved, while the LTA regressions of the extended TDCM resulted in similar attribute transition probabilities as the standard results. In simulation studies, we considered settings to further encompass the capabilities of the extended TDCM, including additional challenges outside the scenarios explored by standard TDCM. This led to the inclusion of LCDM interaction terms, various respondent-level covariates to transition regressions, and an additional time point to demonstrate how transition regression functions in more than two time points. The results showed high coverage rates in both LCDM and transition regression parameters for each case discussed.

7.2. Outlook

In the current study, we assumed measurement invariance for $\mathcal{B}$ across time in all analyses. This was set for the comparisons with previous work based on the same dataset. In addition, previous studies in comparison reported no issue with the assumption. We would like to stress, though, that the proposed algorithm is flexible, and an unconstrained model without the invariance assumption can be estimated. In practice, it is recommended that users test measurement invariance before proceeding with additional analysis.

As illustrated in this article, the flexibility provided in our proposed model presents many advantages; it also serves as a point of limitation, with appropriate model specification necessary. For example, an increase in the number of time points calls for a careful choice of which transition trajectories the researcher believes matter versus which to sacrifice. The decision to model too many transition types may lead to issues with identifiability as Equation 6's multinomial regression likelihood changes in denominator for even a few nonzero covariate slopes. This choice is also crucial to the interpretability of the model, which can gain clarity by selecting transition types that tend to be applicable to real research

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

questions. Restricted for the standard TDCM, the proposed generalization and extension to the TDCM allows modelers to design for increased time points.

In comparison to existing longitudinal DCM models based on hidden Markov structures, the extended TDCM is able to relax the Markovian assumption that is foundational to hidden Markov DCMs. The relaxation of this assumption provides flexibility to accommodate real applications, but may lead to greater complexity for a high number of time points with poor modeling specification as discussed above. It would be of interest to compare the hidden Markov DCM with the proposed model, but an extensive comparison may be warranted in future works.

On another note, the Gibbs sampling framework of the extended TDCM alleviates some computational complexity required by standard TDCM. However, in uncommon settings, where we have a large number of time points, the parameters for transition regression increase drastically and may hinder extended TDCM's computation performance. The flexible structure and inference method of the extended TDCM provide plentiful opportunities for future works in longitudinal DCMs.

We also acknowledge that assuming independence among the transition trajectories of different attributes can be restrictive. Considering the dependence between attributes will be a useful extension. However, it is common to have similar assumptions in previous works of literature. Li et al. (2016) proposed a LTA model that assumes the independence between attributes. Given K attributes and T time points, the number of possible learning trajectories is 2^{KT}, which grows exponentially with respect to both K and T. Due to the very large number of learning trajectories and parameters, it can be very challenging to capture all those learning trajectories without enough data. Therefore, many approaches are proposed to decrease the number of learning trajectories. For example, Chen et al. (2018) suggests to only model nondecreasing learning trajectories. Similar to these ideas, we assume the independence among attributes to reduce the number of parameters to $\mathcal{O}(K2^T)$ from $\mathcal{O}(2^{KT})$.

In conclusion, this study presented a reparametrized latent transition aspect to the TDCM and provided efficient model inference via Gibbs sampling. This extension allows for further flexibility and eases computational requirements for longitudinal (pretest/posttest) design in educational research. We are hopeful that the method we provided will assist educators in assessing growth over time and thereby better support students' learning and improvement.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2025.10031.

Funding statement. This research was supported by funding from the Institute of Education Sciences, U.S. Department of Education, through Grant R305D220020. The content of this publication does not necessarily reflect the views or policies of the Institute of Education Sciences or the U.S. Department of Education.

Competing interests. The authors declare that they have no financial or personal interests that could have appeared to influence the work presented in this article.

References

Balamuta, J. J., & Culpepper, S. A. (2022). Exploratory restricted latent class models with monotonicity requirements under Pólya-Gamma data augmentation. *Psychometrika*, *87*(3), 903–945. https://doi.org/10.1007/s11336-021-09815-9

Bottge, B. A., Ma, X., Gassaway, L., Toland, M. D., Butler, M., & Cho, S. J. (2014). Effects of blended instructional models on math performance. *Exceptional Children*, *80*(4), 423–437. https://doi.org/10.1177/0014402914527240

Bottge, B. A., Toland, M. D., Gassaway, L., Butler, M., Choo, S., Griffen, A. K., & Ma, X. (2015). Impact of enhanced anchored instruction in inclusive math classrooms. *Exceptional Children*, *81*(2), 158–175. https://doi.org/10.1177/0014402914551742

Chen, Y., Culpepper, S. A., Wang, S., & Douglas, J. (2018). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills. *Applied Psychological Measurement*, *42*(1), 5–23. https://doi.org/10.1177/0146621617721250

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, *27*(1), 131–157. https://doi.org/10.1207/s15327906mbr2701_8

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*(2), 403–417. https://doi.org/10.1080/10618600.2016.1172487

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–760.

Glennie, R., Adam, T., Leos-Barajas, V., Michelot, T., Photopoulou, T., & McClintock, B. T. (2023). Hidden Markov models: Pitfalls and opportunities in ecology. *Methods in Ecology and Evolution*, *14*(1), 43–56. https://doi.org/10.1111/2041-210X.13801

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5

Jiang, Z., & Templin, J. (2019). Gibbs samplers for logistic item response models via the Pólya-Gamma distribution: A computationally efficient data-augmentation strategy. *Psychometrika*, *84*(2), 358–374. https://doi.org/10.1007/s11336-018-9641-x

Jimenez, A., Balamuta, J. J., & Culpepper, S. A. (2023). A sequential exploratory diagnostic model using a Pólya-gamma data augmentation strategy. *British Journal of Mathematical and Statistical Psychology*, *76*(3), 513–538. https://doi.org/10.1111/bmsp.12307

Li, F., Cohen, A., Bottge, B., & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, *76*(2), 181–204. https://doi.org/10.1177/0013164415588946

Liang, M. Z., Chen, P., Knobf, M. T., Molassiotis, A., & Ye, Z. J. (2023). Measuring resilience by cognitive diagnosis models and its prediction of 6-month quality of life in Be Resilient to Breast Cancer (BRBC). *Frontiers in Psychiatry*, *14*, 1102258. https://doi.org/10.3389/fpsyt.2023.1102258

Liu, Y., Culpepper, S. A., & Chen, Y. (2023). Identifiability of hidden Markov models for learning trajectories in cognitive diagnosis. *Psychometrika*, *88*(2), 361–386. https://doi.org/10.1007/s11336-023-09904-x

Madison, M. J., & Bradshaw, L. P. (2018a). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*(4), 963–990. https://doi.org/10.1007/s11336-018-9638-5

Madison, M. J., & Bradshaw, L. P. (2018b). Evaluating intervention effects in a diagnostic classification model framework. *Journal of Educational Measurement*, *55*(1), 32–51. https://doi.org/10.1111/jedm.12162

Muthén, B., & Muthén, L. (2017). Mplus. In W. J. van der Linden (Ed.), *Handbook of item response theory* (pp. 507–518). Chapman and Hall/CRC.

Pan, Q., Qin, L., & Kingston, N. (2020). Growth modeling in a diagnostic classification model (DCM) framework—A multivariate longitudinal diagnostic classification model. *Frontiers in Psychology*, *11*, 502948. https://doi.org/10.3389/fpsyg.2020.01714

Park, J. Y., Johnson, M. S., & Lee, Y. (2015). Posterior predictive model checks for cognitive diagnostic models. *International Journal of Quantitative Research in Education*, *2*(3–4), 244–264. https://doi.org/10.1504/IJQRE.2015.071738

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349. https://doi.org/10.1080/01621459.2013.829001

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Sinharay, S., Johnson, M. S., & Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, *30*(4), 298–321. https://doi.org/10.1177/0146621605285517

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of educational measurement*, 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Wang, S., Yang, Y., Culpepper, S., & Douglas, J. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics*, *43*(1), 57–87. https://doi.org/10.3102/1076998617719727

Yamaguchi, K., & Martinez, A. J. (2024). Variational Bayes inference for hidden Markov diagnostic classification models. *British Journal of Mathematical and Statistical Psychology*, *77*(1), 55–79. https://doi.org/10.1111/bmsp.12308

Yigit, H. D., & Douglas, J. A. (2021). First-order learning models with the GDINA: Estimation with the EM algorithm and applications. *Applied Psychological Measurement*, *45*(3), 143–158. https://doi.org/10.1177/0146621621990746

Zhang, S., & Chang, H. H. (2020). A multilevel logistic hidden Markov model for learning under cognitive diagnosis. *Behavior Research Methods*, *52*(1), 408–421. https://doi.org/10.3758/s13428-019-01238-w

Zhang, Z., Zhang, J., Lu, J., & Tao, J. (2020). Bayesian estimation of the DINA model with Pólya-gamma Gibbs sampling. *Frontiers in Psychology*, *11*, 525241. https://doi.org/10.3389/fpsyg.2020.00384

Appendix

A. Derivation of posterior distributions in MCMC procedure

The notation in the appendix is the same as stated in Section 3.1. Let $P(\rho^*_{irk}|1,0)$ be the probability density function of a Pòlya-Gamma distribution with parameter $(1,0)$ for all i,r,k. $P(y^*_{jc}|n_{jc},0)$ is the probability density function of a Pòlya-Gamma distribution with parameter $(n_{jc},0)$ for all j,c.

The probability density functions or probability mass functions of $Y|\mathcal{A},\mathcal{B},\mathcal{A}|\Gamma,\mathcal{B}$ and Γ are listed as the following:

$$P(Y|\mathcal{A},\mathcal{B}) = \prod_{i=1}^{N}\prod_{j=1}^{J} P(Y_{ij} = y_{ij}|\boldsymbol{\alpha}_i^{(t_j)},\boldsymbol{\beta}_j)$$
$$= \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{c=1}^{2^K}\left(\frac{\exp(y_{ij}\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)}{1+\exp(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)}\right)^{\mathbb{1}\left(\boldsymbol{\alpha}_i^{(t_j)}=\mathbf{v}^{-1}(c)\right)}, \quad \text{(A.1)}$$

$$
\begin{aligned}
P(\mathcal{A}|\Gamma) &= \prod_{i=1}^{N}\prod_{k=1}^{K} P(\boldsymbol{\rho}_{ik}|\boldsymbol{\gamma}_{rk}) \\
&= \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T}\left(\frac{\exp\psi_{irk}}{\sum_{r^*=1}^{2^T}\exp\psi_{ir^*k}}\right)^{\mathbb{1}\left(\boldsymbol{\rho}_{ik}=\mathbf{v}^{-1}(r)\right)} \\
&= \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T}\left(\frac{\exp\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk}}{\sum_{r^*=1}^{2^T}\exp\boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k}}\right)^{\mathbb{1}\left(\boldsymbol{\rho}_{ik}=\mathbf{v}^{-1}(r)\right)},
\end{aligned} \tag{A.2}
$$

$$
P(\mathcal{B}) = \prod_{j=1}^{J}\prod_{p=1}^{2^{\sum \mathbf{q_j}}} \frac{1}{(1-\Phi(\frac{L_{jp}}{\sigma^2_{\text{prior};\beta_{jp}}}))\sqrt{2\pi\sigma^2_{\text{prior};\beta_{jp}}}}\exp\left(-\frac{\beta_{jp}^2}{2\sigma^2_{\text{prior};\beta_{jp}}}\right)\mathbb{1}(\beta_{jp}\geq L_{jp}), \tag{A.3}
$$

$$
P(\Gamma) = \prod_{k=1}^{K}\prod_{r=1}^{2^T}(2\pi)^{-\frac{M_{rk}}{2}}\det(\Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}})^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\gamma}'_{rk}\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}}\boldsymbol{\gamma}_{rk}\right). \tag{A.4}
$$

The joint likelihood function is:

$$
P(Y,\mathcal{A},\mathcal{B},\Gamma) \tag{A.5}
$$

$$
\begin{aligned}
&= P(Y|\mathcal{A},\mathcal{B})P(\mathcal{A}|\Gamma)P(\mathcal{B})P(\Gamma) \\
&= \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{c=1}^{2^K}\left(\frac{\exp\left(y_{ij}\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}{1+\exp\left(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}\right)^{\mathbb{1}\left(\boldsymbol{\alpha}_i^{(t_j)}=\mathbf{v}^{-1}(c)\right)} \times \prod_{i=1}^{N}\prod_{k=1}^{K}\frac{\prod_{r=1}^{2^T}\exp\left(\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk}\mathbb{1}(\rho_{ik}=\mathbf{v}^{-1}(r))\right)}{\sum_{r^*=1}^{2^T}\exp\left(\boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k}\right)} \\
&\times \prod_{j=1}^{J}\prod_{p=1}^{2^{\sum \mathbf{q_j}}}\frac{1}{(1-\Phi(\frac{L_{jp}}{\sigma^2_{\text{prior};\beta_{jp}}}))\sqrt{2\pi\sigma^2_{\text{prior};\beta_{jp}}}}\exp\left(-\frac{\beta_{jp}^2}{2\sigma^2_{\text{prior};\beta_{jp}}}\right)\mathbb{1}(\beta_{jp}\geq L_{jp}) \\
&\times \prod_{k=1}^{K}\prod_{r=1}^{2^T}(2\pi)^{-\frac{M_{rk}}{2}}\det(\Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}})^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\gamma}'_{rk}\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}}\boldsymbol{\gamma}_{rk}\right)
\end{aligned} \tag{A.6}
$$

We can now find the posterior distributions of all random variables in our model. We use $P(X|\cdot)$ to denote the posterior distribution of X conditional on all other random variables except X. The likelihood function of $\boldsymbol{\beta}_j$ is

$$
L_c(\boldsymbol{\beta_j}) = \left[\prod_{i=1}^{N}\left(\frac{\exp\left(y_{ij}\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}{1+\exp\left(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}\right)^{\mathbb{1}\left(\boldsymbol{\alpha}_i^{(t_j)}=\mathbf{v}^{-1}(c)\right)}\right] \tag{A.7}
$$

$$
\propto \exp\left(\kappa_{jc}\boldsymbol{\delta}_{\boldsymbol{c}}^{(\boldsymbol{j})\prime}\boldsymbol{\beta}_j\right)\int_0^{\infty}\exp\left(-y_{jc}^*\left(\boldsymbol{\delta}_{\boldsymbol{c}}^{(\boldsymbol{j})\prime}\boldsymbol{\beta}_j\right)^2/2\right)P(y_{jc}^*|n_{jc},0)dy_{jc}^*. \tag{A.8}
$$

Therefore, we have the posterior distribution of $\boldsymbol{\beta}_j$

$$
P(\boldsymbol{\beta}_j|y_{jc}^*,\cdot) \propto P(\boldsymbol{\beta}_j)\prod_{c=1}^{2^K}L_c(\boldsymbol{\beta}_j|y_{jc}^*) \tag{A.9}
$$

$$
\begin{aligned}
&\propto \prod_{p=1}^{2^{\sum \mathbf{q_j}}}\exp\left(-\frac{\beta_{jp}^2}{2\sigma^2_{\text{prior};\beta_{jp}}}\right)\mathbb{1}(\beta_{jp}\geq L_{jp})] \\
&\times \prod_{c=1}^{2^K}\exp\left(\kappa_{jc}\boldsymbol{\delta}_{\boldsymbol{c}}^{(\boldsymbol{j})\prime}\boldsymbol{\beta}_j\right)\exp\left(-y_{jc}^*\left(\boldsymbol{\delta}_{\boldsymbol{c}}^{(\boldsymbol{j})\prime}\boldsymbol{\beta}_j\right)^2/2\right)P(y_{jc}^*|n_{jc},0)dy_{jc}^*.
\end{aligned} \tag{A.9}
$$

We can see that the posterior distribution of β_{jp} is $\beta_{jp}\sim N(\mu_{\text{post};\beta_{jp}},\sigma^2_{\text{post};\beta_{jp}})\mathbb{1}(\beta_{jp}\geq L_{jp})$, where

$$
\sigma^2_{\text{post};\beta_{jp}} = (\sigma^{-2}_{\text{prior};\beta_{jp}}+\sum_{c=1}^{2^K}\delta_{cp}^{(j)\,2}y_{jc}^*)^{-1} = (\sigma^{-2}_{\text{prior};\beta_{jp}}+\boldsymbol{\Delta}_p^{(j)\prime}\text{diag}\left([y_{jc}^*]_{c=1}^{2^K}\right)\boldsymbol{\Delta}_p^{(j)})^{-1} \tag{A.11}
$$

$$
\begin{aligned}
\mu_{\text{post};\beta_{jp}} &= \sigma^2_{\text{post};\beta_{jp}}\sum_{c=1}^{2^K}[(\sum_{i=1}^{N}y_{ij}\mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)}=\mathbf{v}^{-1}(c)))\delta_{cp}^{(j)}) - \frac{1}{2}n_{jc}\delta_{cp}^{(j)} - \sum_{p^*\neq p}y_{jc}^*\delta_{cp^*}^{(j)}\beta_{jp^*}\delta_{cp}^{(j)}] \\
&= \sigma^2_{\text{post};\beta_{jp}}\boldsymbol{\Delta}_p^{(j)\prime}\text{diag}(y_{jc}^*)\tilde{\boldsymbol{z}}_j
\end{aligned} \tag{A.12}
$$

$$
\boldsymbol{\Delta}_p^{(j)} = [\delta_{1p}^{(j)},\delta_{2p}^{(j)},\ldots,\delta_{2^Kp}^{(j)}] \tag{A.13}
$$

https://doi.org/10.1017/psy.2025.10031 Published online by Cambridge University Press

$$\tilde{\boldsymbol{z}}_j = \boldsymbol{z}_j - \boldsymbol{\delta}_{-p}^{(j)} \boldsymbol{\beta}_{j,-p} \tag{A.14}$$

$$\boldsymbol{z}_j = [\frac{\kappa_{j1}}{y_{j1}^*}, \frac{\kappa_{j2}}{y_{j2}^*}, \ldots, \frac{\kappa_{j2^K}}{y_{j2^K}^*}] \tag{A.15}$$

$$\kappa_{jc} = -\frac{1}{2} n_{jc} + \sum_{i=1}^{N} y_{ij} \mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c)). \tag{A.16}$$

The likelihood function of $\boldsymbol{\gamma}_{rk}$ is

$$L_i(\boldsymbol{\gamma}_{rk} | \{\boldsymbol{\gamma}_{r^*k}\}_{r^* \neq r}, \boldsymbol{\rho}_{rk}) \tag{A.17}$$

$$= \left(\frac{\exp(\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk})}{\sum_{r^*=1}^{2^T} \exp(\boldsymbol{x}'_{ir^*k} \boldsymbol{\gamma}_{r^*k})} \right)^{\rho_{irk}} \left(\frac{\sum_{r^* \neq r}^{2^T} \exp(\boldsymbol{x}'_{ir^*k} \boldsymbol{\gamma}_{r^*k})}{\sum_{r^*=1}^{2^T} \exp(\boldsymbol{x}'_{ir^*k} \boldsymbol{\gamma}_{r^*k})} \right)^{1-\rho_{irk}} \tag{A.18}$$

$$= \left(\frac{\exp(\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk})}{1 + \exp(\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk})} \right)^{\rho_{irk}} \left(\frac{1}{1 + \exp(\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk})} \right)^{1-\rho_{irk}} \tag{A.19}$$

$$\propto \left[\exp\left((\rho_{irk} - \frac{1}{2})(\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk}) \right) \int_0^{\infty} \exp\left(-\rho_{irk}^* (\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk})^2 / 2 \right) P(\rho_{irk}^* | 1, 0) d\rho_{irk}^* \right], \tag{A.20}$$

where $c_{irk} = \log(\sum_{r_* \neq r} \exp \psi_{ir_*k})$.

The posterior distribution of $\boldsymbol{\gamma}_{rk}$ is

$$P(\boldsymbol{\gamma}_{rk} | \boldsymbol{\rho}_{rk}^*, \cdot) \tag{A.21}$$

$$\propto P(\boldsymbol{\gamma}_{rk}) \prod_{i=1}^{N} L_i(\boldsymbol{\gamma}_{rk} | \{\boldsymbol{\gamma}_{r^*k}\}_{r^* \neq r}, \boldsymbol{\rho}_{rk} | \boldsymbol{\rho}_{rk}^*) \tag{A.22}$$

$$\propto \prod_{i=1}^{N} \left[\exp\left((\rho_{irk} - \frac{1}{2})(\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk}) \right) \exp\left(-\rho_{irk}^* (\boldsymbol{x}'_{irk} \boldsymbol{\gamma}_{rk} - c_{irk})^2 / 2 \right) P(\rho_{irk}^* | 1, 0) \right]$$

$$\times \exp(-\frac{1}{2} \boldsymbol{\gamma}_{rk} \Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}}^{-1} \boldsymbol{\gamma}_{rk}). \tag{A.23}$$

We can see that the posterior distribution of $\boldsymbol{\gamma}_{rk}$ is $\boldsymbol{\gamma}_{rk} \sim N(\mu_{\text{post};\boldsymbol{\gamma}_{rk}}, \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}})$, where

$$\Sigma_{\text{post};\boldsymbol{\gamma}_{rk}} = (\Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}}^{-1} + \sum_{i=1}^{N} \boldsymbol{x}_{irk} \boldsymbol{x}'_{irk} \rho_{irk}^*)^{-1} = (\Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}}^{-1} + X'_{rk} \text{diag}\left([\rho_{irk}^*]_{i=1}^N \right) X_{rk})^{-1} \tag{A.24}$$

$$\mu_{\text{post};\boldsymbol{\gamma}_{rk}} = \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}} [\sum_{i=1}^{N} \boldsymbol{x}_{irk} (\rho_{irk} - \frac{1}{2} - \rho_{irk}^* \log \sum_{r \neq r^*} \exp \boldsymbol{x}'_{ir^*k} \boldsymbol{\gamma}_{r^*k})]$$

$$= \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}} X'_{rk} (\boldsymbol{\zeta}_{rk} - \text{diag}\left([\rho_{irk}^*]_{i=1}^N \right) \boldsymbol{c}_{rk}) \tag{A.25}$$

$$\boldsymbol{\zeta}_{rk} = [\rho_{1rk} - 1/2, \ldots, \rho_{Nrk} - 1/2] \tag{A.26}$$

$$\boldsymbol{c}_{rk} = [c_{irk}, \ldots, c_{Nrk}], \tag{A.27}$$

The posterior distribution of $\boldsymbol{\alpha}_i^{(t)}$ follows a categorical distribution as the following:

$$P(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c) | \cdot) \propto \prod_{j:t_j=t} \exp((y_{ij} - \frac{1}{2}) \boldsymbol{\delta}_c^{(j)\prime} \boldsymbol{\beta}_j) \prod_{k=1}^{K} \exp(X'_{ir^*k} \boldsymbol{\gamma}_{\boldsymbol{r^*k}})$$

$$\propto \exp(\sum_{j:t_j=t} (y_{ij} - \frac{1}{2}) \boldsymbol{\delta}_c^{(j)\prime} \boldsymbol{\beta}_j + \sum_{k=1}^{K} X'_{ir^*k} \boldsymbol{\gamma}_{\boldsymbol{r^*k}}),$$

where r^* is the transition of attribute k when $\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)$.

Cite this article: Resch, J., Baugh, S., Duan, H., Tang, J., Madison, M. J., Cotterell, M. and Jeon, M., (2025). Bayesian Transition Diagnostic Classification Models with Polya-Gamma Augmentation. *Psychometrika*, 1–32. https://doi.org/10.1017/psy.2025.10031