# Bayesian Transition Diagnostic Classification Models with Pòlya-gamma Augmentation

Joseph Resch, Samuel Baugh, Hao Duan, James Tang, and Minjeong Jeon

May 31, 2024

## 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.



**Keyword:** Diagnostic classification models; longitudinal models; transition analysis; intervention effects; MCMC; Gibbs sampling; Pòlya-gamma Augmentation

# 1 Introduction

The evaluation of intervention effects is an important objective of educational research. One way educational systems can improve is through the systematic implementation of products, policies, practices, and programs that research has identified as successful. The evaluation of educational interventions, however, is not trivial; the answer is usually not a simple "yes, it works", or "no, it does not work". Frequently, the answer is something closer to "it works well in Context X and is best supported by Y and Z". Therefore, methodologies for evaluating interventions should allow for more complex narratives by accounting for contextual effects. The overall objective is to meaningfully evaluate educational interventions while accounting for ever-present contextual effects.

We propose a Bayesian hierarchical regression formulation to extend on the multi-group transition diagnostic classification model (Madison & Bradshaw, 2018a). In essence, the approach presented here can be thought of in terms of two inter-related components. The first is a binomial logistic model which relates the latent classes, i.e., the log-linear cognitive diagnosis model (Henson et al., 2009), which are binary variables indicating whether a student possesses a required attribute at a particular time, to the item responses or not. The second is a multinomial logistic model relating provided covariates to the progression of attributes over time. The probability of the transition diagnostic classification model (Madison & Bradshaw, 2018b) is thus extended to a full regression formulation. Notably, with the transition model, the efficacy of an intervention can be assessed in terms of its effect on the probability of a student gaining an attribute.

Proposed model inference proceeds over three sets of parameters; the response parameters which link attributes to response probabilities (denoted as $\boldsymbol{\beta}$), the latent attribute vectors 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 fit the Bayesian model, Pòlya-gamma augmentation introduced by Polson et al. (2013), is used for both the response and transition logistic models. This procedure is described in Section 3. The R package **cdmfits** is developed, which allows for the user to fit the proposed models with user-provided data, Q matrices, and design matrices for the transition model. Full validation results with empirical and simulated data sets are displayed in Sections 4 and 5.

# 2 Review of Methods

In this section, we first provide conceptual and technical background on diagnostic classification models (DCMs) and longitudinal DCMs. We then motivate the proposed methodological developments through an empirical illustration of evaluating an instructional intervention in mathematics education (Madison & Bradshaw, 2018a).

## 2.1 Diagnostic Classification Models

Diagnostic classification models, or DCMs (Rupp et al., 2010), can be most simply described as siblings to item response theory (IRT) models, with the primary distinguishing feature being that DCMs have categorical latent traits and IRT models have continuous latent traits. While this description is oversimplified, it does point to the source of essentially all differences in these two psychometric models. That is, because DCMs model categorical latent traits, commonly referred to as attributes, they support classification-based inferences and interpretations. 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. In educational settings, these classifications can be used to indicate students' strengths, weaknesses, understandings, and misunderstandings with respect to targeted concepts within a particular domain. Because of the diagnostic and actionable feedback, DCMs are well-suited to support formative assessment efforts aimed at assessing fine-grained attributes and understandings, such as Common Core State Standards and Next Generation Science Standards.

Statistically, DCMs are confirmatory and constrained latent class models (Lazarsfeld, 1955). Latent class models use item responses to group similar respondents into latent classes. In a typical latent class analysis, the number of latent classes is determined in an exploratory fashion; several models are fit with different numbers of latent classes and the model with the best fit (i.e., most parsimonious) is chosen and subsequently interpreted. The substantive meaning of the detected latent classes is then inferred from the characteristics of its members. Once a number of latent classes has been chosen as a candidate model, the structural and measurement components of the latent class model must be estimated. The structural component parametrizes the proportion of respondents belonging to each class, while the measurement component parametrizes how each

latent class responds to items.

DCMs are confirmatory latent class models in two senses: first, the latent classes in a DCM are specified a priori as the patterns of mastery corresponding to the measured attributes. For a test measuring $K$ dichotomous attributes, each respondent is probabilistically classified into one of $2^K$ patterns of attribute mastery, termed attribute profiles. For example, a respondent may be classified into the attribute profile $[0, 1, 1, 0]$, with the 1's indicating that they have mastered Attributes 2 and 3, and the 0's indicating that they have not mastered Attributes 1 and 4. In general, attribute profiles are denoted by $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]$ where $\alpha_k = 0$ indicates non-mastery of Attribute $k$ and $\alpha_k = 1$ indicates mastery of Attribute $k$. Secondly, DCMs are confirmatory latent class models in that the attribute loading structure is specified a priori in an item-by-attribute matrix that indicates which attributes are measured by each item.

DCMs are considered constrained latent class models because they are specified by placing constraints on the item 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. The DCM presented next, the log-linear cognitive diagnosis model, or LCDM (Henson et al., 2009), is a general and flexible model that offers a unified framework through which many previously developed DCMs can be specified by placing constraints on the LCDM parameters.

## 2.2 Log-Linear Cognitive Diagnosis Model

The LCDM uses a log-linear framework to parametrize the relationship between respondent attribute mastery and item response probabilities. In this log-linear framework, other popular DCMs, like the deterministic-inputs, noisy-and-gate (DINA; Haertel, 1989) model, are obtained by constraining certain LCDM parameters to be zero. The LCDM item parameters include intercepts, main effects, and interactions terms, which are interpreted similarly to a reference coded analysis of variance model. The intercept represents the log-odds of a correct response for the reference group: respondents who have mastered none of the required attributes. The attribute main effects represent the increase in log-odds for respondents who have mastered only that respective attribute. Finally, the interaction terms represent the change in log-odds for respondents who have mastered multiple attributes. To obtain the DINA model, main effects are constrained to be zero, and only the intercept and the highest-level interaction term are estimated.

The generality of the LCDM is important because it permits a wider range of tests to be modeled with a DCM and allows for the examination of attribute hierarchies and learning progressions (Templin & Bradshaw, 2014). Constrained DCMs, like the DINA model, are only useful in rare cases where they accurately represent the data. Constrained DCMs make very strict assumptions about the way items function, and the literature has yet to provide an example where a severely constrained model fits better than a more general DCM for every item on an empirical test. When these constraints are misspecified, respondents are misclassified and results can be misleading. And without a general DCM upon which to compare, there is no way to statistically assess the appropriateness of such constraints. Additionally, constrained DCMs do not permit the examination of the attribute hierarchies and potential learning progressions because the measurement model is too limited. The use of the LCDM is critical as the generality of its measurement model permits the best opportunity to capture the underlying item response generating process without unnecessarily risking model misspecification, thereby maximizing the accuracy of model-based inferences. Lastly, it is important to conduct research in the most generalizable way, so that methodological results may extend to and be applied in a variety of educational research contexts.

## 2.3 Transition Diagnostic Classification Model

Traditionally, classical test theory (CTT) and item response theory approaches have been used to model student growth over time. The CTT and IRT frameworks provide measures of student growth on a continuous scale in the form of gain scores or ability gain scores, respectively. To accommodate longitudinal data in a DCM framework, a latent transition model (Collins & Wugalter, 1992) framework was 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 transition diagnostic classification model (TDCM) 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 $y_i \in \{0,1\}^{J\times T}$ for respondent $i$ is given by:

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

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 (2018a). The multi-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 multi-group TDCM is given by

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

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 multi-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 vs. 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 multi-group TDCM presents a promising methodology for evaluating intervention effects with interpretive benefits over other longitudinal psychometric modeling options, it has the limitation in that it is a complex model 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, or more than two time points This is a significant limitation because experimental designs for evaluating intervention effects often include more than two measurement occasions and more than two groups. To provide a more flexible framework for evaluating intervention effects in a DCM framework, a modified TDCM with a practical estimation approach is necessary. In the next section, we provide details of the extended modeling framework we propose and how it can be used in the evaluation of intervention effects.

# 3 Extension to the TDCM

We expand upon the multi-group TDCM framework of Madison & Bradshaw (2018a) by adjusting the model formulation and offering a feasible estimation framework for model inference without `Mplus` (Muthén & Muthén, 2017) as a necessity. Here, we utilize an updated regression-based formulation to incorporate additional covariates outside group-level treatment intervention effects. For Bayesian likelihood inference, a data augmentation framework using the Pòlya-gamma distribution introduced by Polson et al. (2013) is used for the logit link function to allow Markov chain Monte Carlo (MCMC) Gibbs sampling. Pòlya-gamma data augmentation has received increasing attention in Statistics and related fields for efficient Gibbs sampling of logistic models. In Psychometrics, for example, 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 & Culpepper (2022) presented the implementation of Pòlya-gamma data augmentation for a related class of LCDM where the $Q$ matrix is inferred.

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 in exact. The proposed framework's primary contribution to the TDCM can be found in the chosen parameterization for the latent transition model. Stepping away 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 multivariate 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 and 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.

- 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 paper, $Q$ 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 $t$. We use $\mathcal{A}$ 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, we use $\mathbf{v}^{-1}(c)$ 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 $\boldsymbol{\rho}_{ik} \equiv [\boldsymbol{\alpha}_i^{(1)'}\boldsymbol{e}_k, \ldots, \boldsymbol{\alpha}_i^{(T)'}\boldsymbol{e}_k] \in \{0,1\}^T$ 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 attributes in $\mathbf{v}^{-1}(l)$ or $\mathbf{v}^{-1}(l)$ contains any attribute that $\mathbf{v}^{-1}(c)$ does not have. Otherwise, $\delta_{cl}^{(j)} = 1$. We assume there is no column of zero in $\Delta^{(j)}$. That is, $\Delta^{(j)} \in \{0,1\}^{2^K \times 2^{\sum_{k=1}^K q_{jk}}}$, and $\Delta$ 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 multivariate logistic regressions to model latent attribute transition types. $\Gamma$ 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:

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

Using $\Delta$ 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,

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

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 occurrences ($0 \to 0$, $0 \to 1$, $1 \to 0$, 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 point to the next, thus motivating the use of multivariate logistic models. We define the conditional probability to be,

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

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 multivariate logistic regressions such that

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

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 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. It is worth noting here that $\Gamma$ may be set up so that the full multivariate regression is abbreviated to not model certain response levels in order to reduce computation. This may be practically sensible as we may care about growth, i.e., $0 \to 1$ transition, and regression, i.e., $1 \to 0$ transition, more than the two unchanging transition types.

The formulation of this section reintroduces $\mathcal{B}$ for the item-attribute LCDM, and proposes the $\Gamma$ regression parameter to model latent transition using multivariate logistic regressions. It is clear from notation that the extended TDCM is capable of relaxing time- and group- measurement invariance. LCDM $\mathcal{B}$ has the option to vary longitudinally, in settings where the same test is not given at each time point. Transition probabilities model by $\Gamma$ 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.

### 3.3 Pòlya-gamma Sampling Procedure

The computational complexity of the transition model's multivariate 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 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 $\Gamma$.

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 $\rho^*_{irk}$ corresponding to latent transition $\rho_{irk}$ 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 predictor such that $\theta = \mathbf{x}^T\boldsymbol{\beta}$ in case of the logistic regression, the left side of Equation 8 becomes 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. An efficient Gibbs sampling algorithm is adopted here, where $\boldsymbol{\beta}$ and augmented data are sequentially sampled in each iteration, as further explained in Section 3.3.2. The integrand of Equation 8 may also be adapted straightforwardly to the multivariate logit link likelihood for the transition model, as seen in the following Section 3.3.3. Efficiency of Gibbs sampling for logistic regression Bayesian inference makes Pòlya–gamma 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 $\Gamma$ loosely follow the Pòlya-gamma framework of Balamuta & Culpepper (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 $1 \leq p \leq P_j$ and $\Gamma$ 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. Specifically, the posterior distributions 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 & Culpepper, 2022). The exact posteriors distributions are defined in Section 3.3.2.

MCMC procedure for the extended TDCM is hierarchical, mirroring the hierarchy of response model, which is dependent 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 $\Gamma$ 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 $m \ll M$. It is shown in simulation, and empirical studies that posterior samples achieve convergence for small $m$ (e.g., $m = 10$ in implementations), as long as $M$ is large enough to account for burn-in.

The following procedure samples three sets of parameters defined in the previous section: $\mathcal{A}, \mathcal{B}$, and $\Gamma$. 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 $\Gamma$ 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 $\Gamma$ should be randomly initialized in advance. The procedure is fully implemented using R and can be accessed through GitHub at `github.com/samjbaugh/BayesianCDM`. 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.

### 3.3.1 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

$$P(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)|Y, \mathcal{A}_i^{(-t)}, \mathcal{B}, \Gamma) = \frac{P\left(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c)|\mathcal{A}_i^{(-t)}, \Gamma\right) \prod_{j:t_j=t} g\left((2y_{ij}-1)\boldsymbol{\delta}_c^{(j)'}\boldsymbol{\beta}_j\right)}{\sum_{c_*} P\left(\boldsymbol{\alpha}_i^{(t)} = \mathbf{v}^{-1}(c_*)|\mathcal{A}_i^{(-t)}, \Gamma\right) \prod_{j:t_j=t} g\left((2y_{ij}-1)\boldsymbol{\delta}_c^{(j)'}\boldsymbol{\beta}_j\right)}, \tag{12}$$

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 $\Gamma$. 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

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

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.

### 3.3.2 Step 2: Sample $\mathcal{B}$

Following $\mathcal{A}$, we sample item-attribute LCDM parameters $\mathcal{B}$ by first sampling Pòlya–gamma auxiliary response variables $y_{jc}^* \sim PG(n_{jc}, \boldsymbol{\delta}_c^{(j)'}\boldsymbol{\beta}_j)$, where $n_{jc} = \sum_{i=1}^{N} \mathbb{1}\left(\alpha_i^{(t)} = \mathbf{v}^{-1}(c)\right)$ for each question $j$ and profile $c$. Then $\beta_{jp}$ 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.

**Time-Constrained $\mathcal{B}$** The general case for sampling $\mathcal{B}$ in Section 3.3.2 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 (2018a) focuses on assuming a constant $Q$ matrix for all time points, thus fixing the $\beta_{jp}$ 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 $n_{jc} = \sum_{i=1}^N\sum_{t=1}^T \mathbb{1}\left(\alpha_i^{(t)} = \mathbf{v}^{-1}(c)\right)$. $\boldsymbol{\Delta}_p^{(j)}$ 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 $\tilde{\boldsymbol{z}}_j = \boldsymbol{z}_j - \Delta^{(j)}_{-p}\boldsymbol{\beta}_{j,-p}$ 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 paper 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.

### 3.3.3 Step 3: Sample $\Gamma$

In the final step, $\Gamma$ parameters for transition multivariate 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

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

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

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 $\Gamma$ 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 $\Gamma$ are iterated $M$ times to produce full posterior samples.

The extended TDCM formulation in Section 3.2 and the Pòlya-gamma sampling algorithm for Bayesian inference in Section 3.3 are validated in various empirical and simulation settings in the following Sections 4 and 5.

# 4 Empirical Validation

| Item # | 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 |

Table 1: $Q$ matrix for empirical data (Bottge et al., 2014, 2015) indicating attribute requirements for each $J = 21$ test item questions. The four attributes indicated are: ratios and proportional relationships (RPR), measurement and data (MD), number systems (NF), and geometry and graphing (GG).

To demonstrate the utility of the proposed extension to TDCM, we apply our method to the post-test/pre-test empirical data for mathematics education (Bottge et al., 2014, 2015). The goal of the data collection was 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 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 are implicitly measured: ratios and proportional relationships (RPR), measurement and data (MD), number systems (NF), and geometry and graphing (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 1. We see here none of the items measures more than one attribute, meaning the item-attribute regression parameter $\mathcal{B}$ contains only main effects and no interaction terms. Furthermore, the model parameter $\mathcal{B}$ is held constant between time points assuming measurement invariance. This follows the comparison of attribute mastery between the freely estimated TDCM and parameter-constrained TDCM fitted to the empirical data of interest, where the two models agreed in attribute classification 97.9% in Madison & Bradshaw (2018b)'s single-group case, and 97.5% of the time in Madison & Bradshaw (2018a)'s multiple-group case. The authors in both instances favored constrained TDCM results for interpretability. To implement this for the extended TDCM, the following results refer to $\mathcal{B}$ posteriors defined in Section 3.3.2.

We remark here that the same data set and $Q$ matrix was used for the single-group TDCM (Madison & Bradshaw, 2018b) and the multi-group extension (Madison & Bradshaw, 2018a), with the sole difference being the consideration of group-level treatment in the multi-group case versus

ignoring it in the single-group case. In the sections that follow, we present the data analysis results using the proposed extended TDCM in the single-group and multiple-group settings, in comparison to standard TDCM results.

## 4.1 Single-Group Setting

In the single-group setting where no group membership are considered as in Madison & Bradshaw (2018b), the effect of the educational EAI treatment is not of interest and thus the transition regressions are simplified as explained in the following $\Gamma$ posteriors. The Gibbs sampling procedure for the extended TDCM defined in Section 3.3 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 of Section 6 in Supplement. For Bayesian inference, the standard deviations of the normal distribution priors on $\mathcal{B}$ and $\Gamma$ 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.

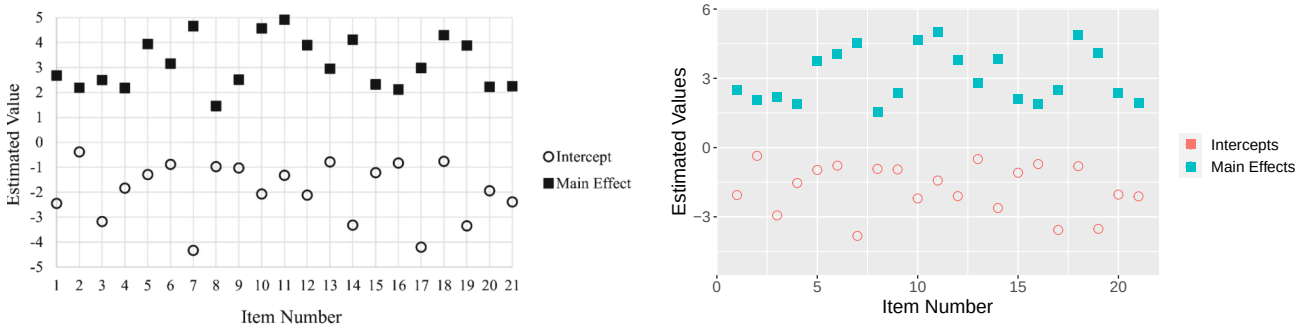


Figure 1: Comparison of $\mathcal{B}$ point estimates between the standard TDCM from Figure 2 of Madison & Bradshaw (2018b) (left) and the extended TDCM (right).

We replicate key results shown by the original study of Madison & Bradshaw (2018b) and present additional results demonstrating extended TDCM's added value. The comparison begins with point estimates for $\mathcal{B}$ in Figure 1. Standard TDCM results on the left are taken directly from Figure 2 of Madison & Bradshaw (2018b), while the extended TDCM point estimates are the taken 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.

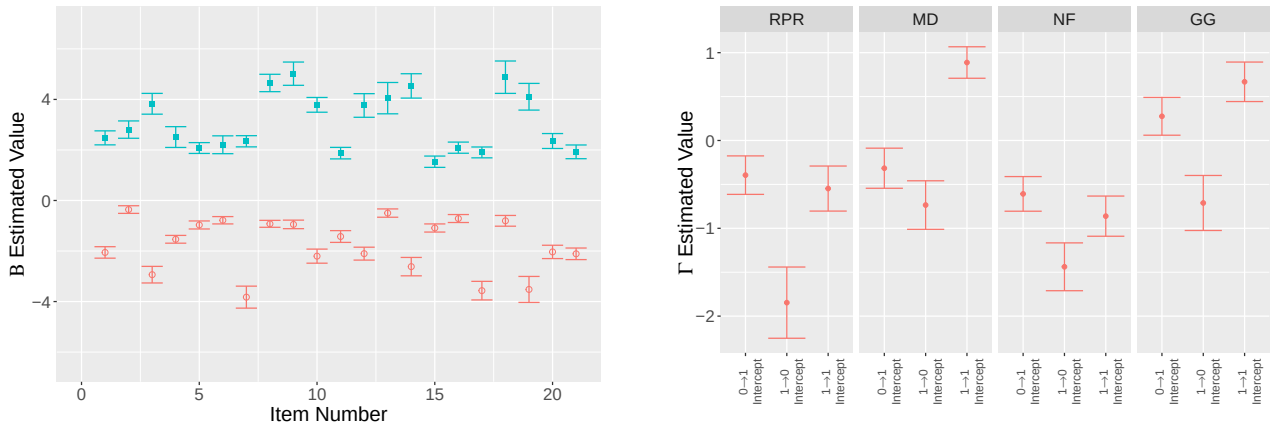


Figure 2: Posterior sample distributions for $\mathcal{B}$ (left) and $\Gamma$ (right) bounded by two standard deviations. For $\mathcal{B}$, each question item contains intercept in red and main effect in blue. The four segments for $\Gamma$ plot correspond to the four attributes: ratios and proportional relationships (RPR), measurement and data (MD), number systems (NF), and geometry and graphing (GG). The indices within each attribute segment denote transition $0 \to 1$ intercept, $1 \to 0$ intercept, and $1 \to 1$ intercept, respectively.

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 $\Gamma$ 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 Section 6. 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 remain roughly constant among all parameters, with $\mathcal{B}$ variability increasing slightly as point estimates depart from zero.

The full posterior distributions of $\Gamma$, shown on the right side of Figure 2, serve as a main appeal of our TDCM extension. For the single-group case, we note the simplified transition regressions can be considered null models in that multivariate regression for each attribute contains only the intercept. The figure shows four segments denoted by the labels on the top showing $\Gamma$ for each of the $K = 4$ attributes. Within each segment, the indices on bottom denote the intercepts for the transition $0 \to 1$, $1 \to 0$, and $1 \to 1$, respectively, with the $0 \to 0$ transition being the baseline level and 0 indicates non-mastery, 1 indicates mastery and $\to$ indicates transition between two time points. The converged estimates here show approximate constant variance. It is also seen that the attribute regress transition $1 \to 0$, the loss of an attribute between the two time points, to be lower in log-odds than other transition types. This makes sense in interpretation given the underlying

behavior that some respondents receive treatment that aid in 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.

| 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) |

Table 2: Comparison of implicit conditional transition probabilities for each $K = 4$ attributes of the single-group empirical study for standard TDCM (left) as shown in Table 9 of Madison & Bradshaw (2018b) and extended TDCM (right) posterior means and standard deviations in parenthesis. The matrices are labeled with 1 being attribute mastery and 0 otherwise.

The conditional transition probability are a direct function of $\Gamma$ posteriors by nature of the logistic link. In the original study (Madison & Bradshaw, 2018b) 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 2, 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 $\Gamma$ posteriors are presented on the right side of Table 2, with point estimates and attached standard deviations in parenthesis. 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 initial time point (e.g., standard deviation given no attribute 1 are both 0.025 in Table 2). Precisely in comparison, the extended TDCM provides the advantage of full posteriors for transition probabilities whereas the standard TDCM only produce point estimates.

| Item # | 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 |

Table 3: Probability 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. Posteriors for latent attribute $\alpha$ 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.

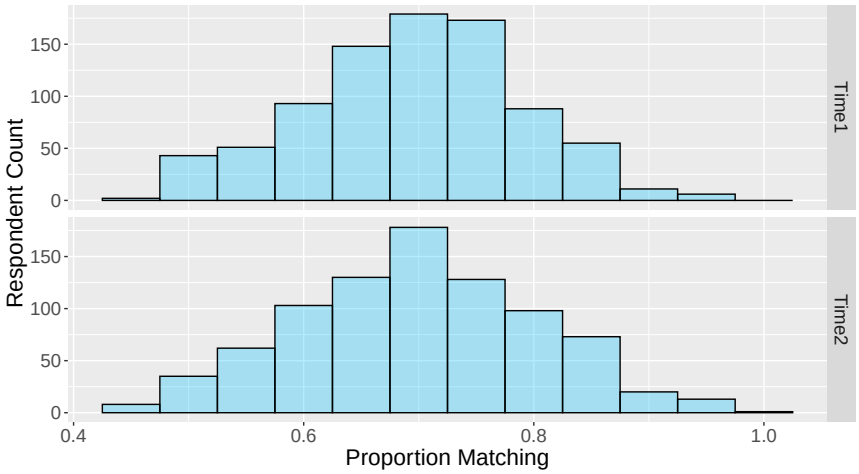


Figure 3: Distribution of probability of extended TDCM fit to match the empirical response data average for 849 respondents averaged over 21 total questions, done for both time points of the study in the single-group setting.

Goodness-of-fit assessment via predictive checks is another advantage that the extended TDCM's transition regressions provide. Although there are no treatment or covariate for single-group case, the null models themselves can be used to perform this predictive check. Here, the fitted LCDM data are simulated and compared with the actual item-attribute data. The one-to-one match provides percentage matching between the fitted data and real data, which can be summarize in that the fitted data matches 70.84% of actual data in the first time point, and 70.62% of actual data in second time point. The matching probabilities can further be broken down by question in Table 3 and by person in Figure 3. We see that the matching probabilities in both sets of results are high, suggesting the fit of the extended TDCM to the data is satisfactory.

## 4.2 Multiple-Group Setting

The procedure shown in the previous Section 4.1 is extended here to the multiple-group case. 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 remaining 456 respondents receives 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 done in our case for computational simplicity, as seen further in detail from the structure of $\Gamma$. 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 $\Gamma$ are provided in Section 6 showing proper convergence.

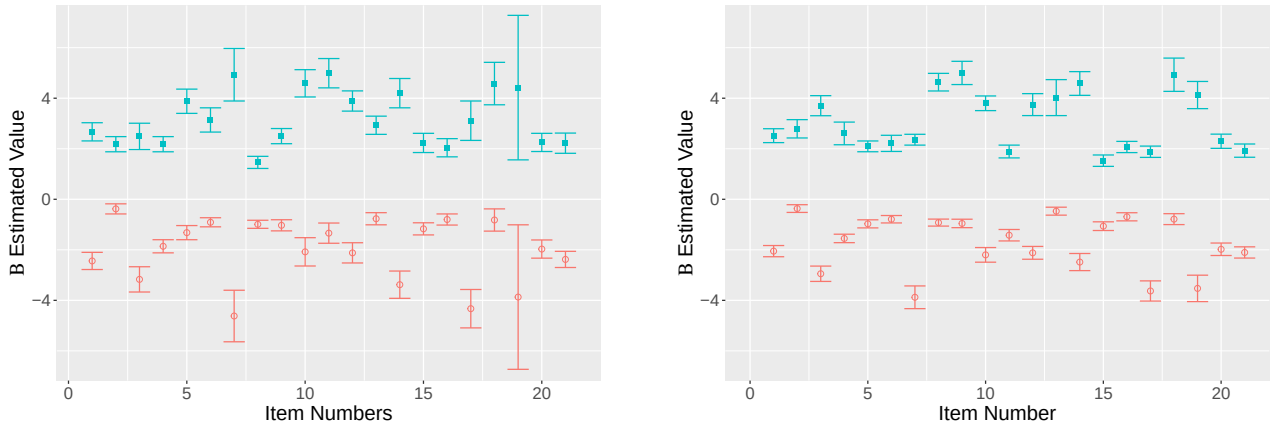


Figure 4: Posterior $\mathcal{B}$ distributions plotted using Table 2 of Madison & Bradshaw (2018a) bounded by two reported standard errors (left) is compared to the extended TDCM bounded by two posterior standard deviations (right). Red indicates the question intercepts, while blue are the main effects.

The resulting estimates for $\mathcal{B}$ are shown in Figure 4 for both the standard multiple-group TDCM (left) and its extension (right). Left side results for the standard TDCM are graphically displayed directed from Table 2 of Madison & Bradshaw (2018a). 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 4 (left) and Figure 2 (left). Intuitively, these results favor the extended TDCM to 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 shows particularly large variability in both $\mathcal{B}$ intercept and main effect. The extended TDCM does not reflect this in contrast.

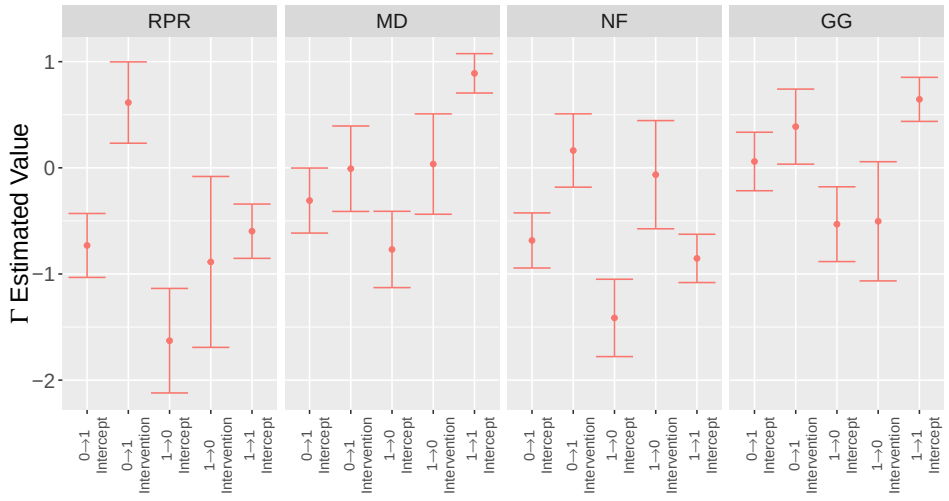


Figure 5: The four segments for $\Gamma$ plot correspond to the four attributes: ratios and proportional relationships (RPR), measurement and data (MD), number systems (NF), and geometry and graphing (GG). The indices within each attribute segment denote transition $0 \to 1$ intercept, $0 \to 1$ intervention, $1 \to 0$ intercept, $1 \to 0$ intervention, and $1 \to 1$ intercept, respectively.

The complementing set of attribute transition $\Gamma$ posteriors for the extended TDCM are shown in Figure 5. In the same manner as single-group $\Gamma$ in Figure 10 (right), the multiple-group $\Gamma$ is split into quadrants for each attribute. The indices explicitly show the adaptation of the treatment covariate, where transitions $0 \to 1$ and $1 \to 0$ receive the covariate. As is the goal for the EAI treatment, the treatment effects clearly indicate that the treatment increase the log-odd for the attribute gain transition $0 \to 1$ and decrease log-odd for attribute loss transition $1 \to 0$ for attributes RPR, NF, and GG. The treatment has near-zero effects for attribute NF, and it follows that its intercepts are approximate to those of the single-group in Figure 10 (right). These $\Gamma$ posteriors can be further compared using attribute transition probabilities.

To demonstrate the attribute transition aspect of treatment effect, the authors of the standard

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

Table 4: Transition probabilities for $0 \rightarrow 1$ and $1 \rightarrow 0$ transitions between the control and treatment groups, compared between the standard multiple-group TDCM (Standard) and the extended TDCM (Extended).

TDCM show the difference in transition probability between the treatment and control groups. Attribute probabilities for non-mastery to mastery (transition $0 \rightarrow 1$) and mastery to non-mastery (transition $1 \rightarrow 0$) are displayed for each attribute in Table 5 of Madison & Bradshaw (2018a). The $\Gamma$ 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 $\Gamma$ for whether treatment is received. Table 4 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.

| Item # | 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 |

Table 5: Probability 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. Posteriors for latent attribute $\alpha$ 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.

The predictive check done for the previous single-group example is applied for the multiple-group model case, by questions in Table 5 and by respondents in Figure 6. Here, we see that 70.80% of the simulated predictive data matches actual data at the first time point, and 70.60% matches at the second time point. Predictive evidence here indicates a reasonable fit of the extended TDCM

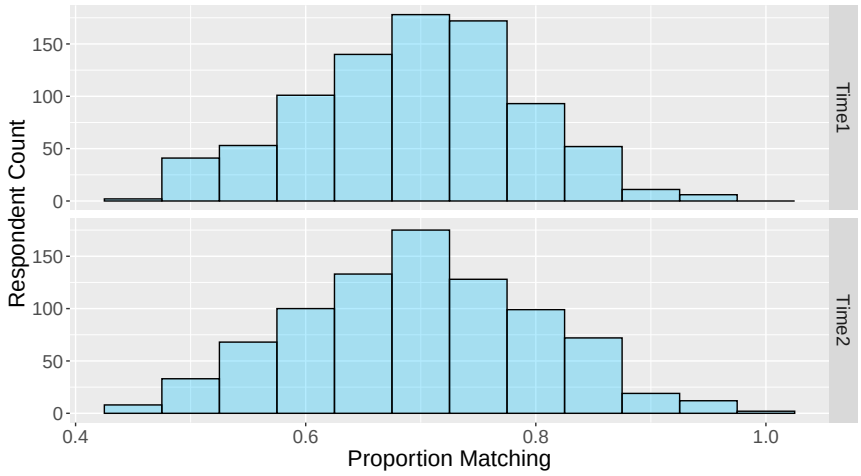


Figure 6: Distribution of probability of extended TDCM fit to match the empirical response data average for 849 respondents averaged over 21 total questions, done for both time points of the study in the multiple-group setting.

with multiple groups to the data, further supporting the utility of the empirically fitted extended TDCM model.

### 4.3 Multiple-Group with Covariates Setting

In an empirical case separate from those found in Madison & Bradshaw (2018b) and Madison & Bradshaw (2018a), two student-level covariates are recorded and taken into account. This translates to additional complexity in the $\Gamma$ structure for the extended TDCM. Given the same exam, 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 whether or not they are a part of the English 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. [mj: Matt, please add additional information about the data. ]

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 $\Gamma$ structure. Intervention treatment and the two covariates are enforced onto the $0 \rightarrow 1$ and $1 \rightarrow 0$ 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

$\Gamma$ are $\sigma_{\text{prior};\beta_{jp}} = 1$ and $\Sigma_{\text{prior};\gamma_{rk}} = 0.5$.

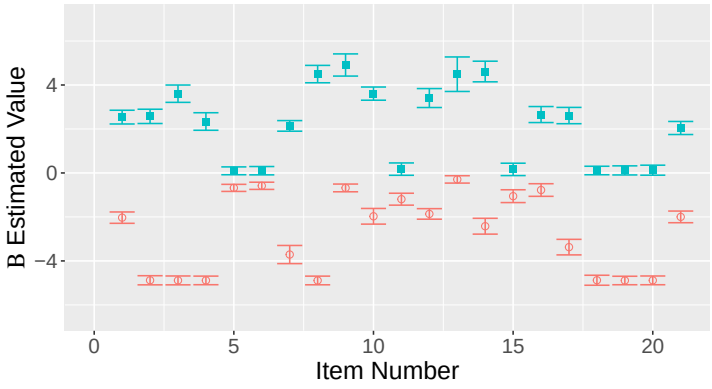


Figure 7: Posterior $\mathcal{B}$ distributions of extended TDCM bounded by two posterior standard deviations (right). Red indicates the question intercepts, while blue are the main effects.

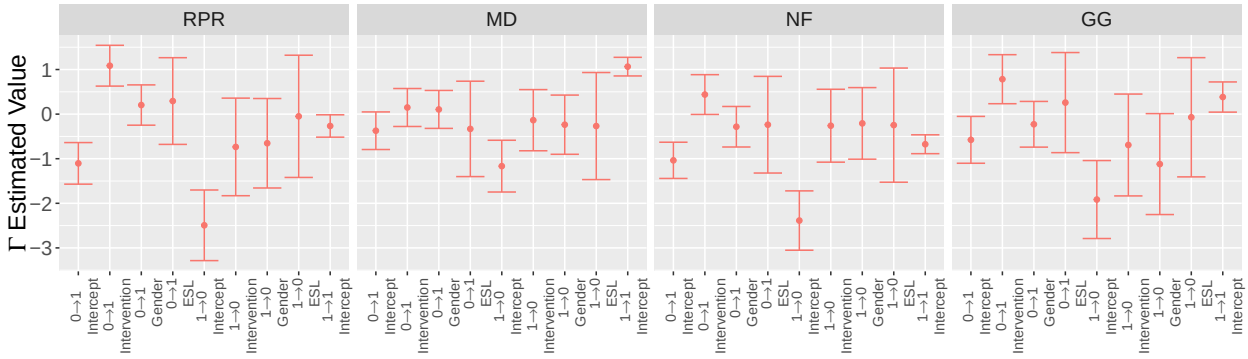


Figure 8: The four segments for $\Gamma$ plot correspond to the four attributes: ratios and proportional relationships (RPR), measurement and data (MD), number systems (NF), and geometry and graphing (GG). The indices within each attribute segment denote transition $0 \to 1$ intercept, $0 \to 1$ intervention, $0 \to 1$ gender , $0 \to 1$ ESL , $1 \to 0$ intercept, $1 \to 0$ intervention, $1 \to 0$ gender , $1 \to 0$ ESL , and $1 \to 1$ intercept, respectively.

Figure 7 show the $\mathcal{B}$ posteriors. Of more interesting is Figure 8, 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 noting that if a student already has an attribute, being in an ESL program does not cause them to lose the attribute, as seen by the $1 \to 0$ ESL effects for each attribute in Figure 8.

| Item # | 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 |

Table 6: Probability 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. Posteriors for latent attribute $\alpha$ 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.

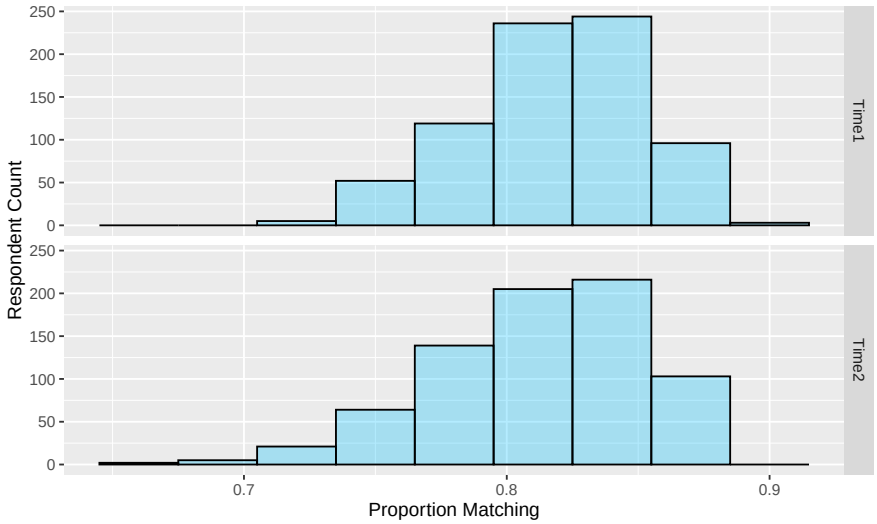


Figure 9: Distribution of probability of extended TDCM fit to match the empirical response data average for 849 respondents averaged over 21 total questions, done for both time points of the study in the multiple-group covariates setting.

The predictive results are derived from sampling of response data using values fitted by the original design matrix to the extended TDCM, just as done in the two empirical examples prior. The results here indicate clear improvement over empirical settings without student-level covariates, such that the fitted data matches 81.81% of the actual data at the first time point, and 81.33% of actual data at the second time point. Table 6 provides detailed matching percentages by question. Furthermore, the histograms composing of matching percentages for individual students are found in Figure 9, where the centers for the left skewed distributions make this improvement over settings without student-level covariates even more apparent. While some of the improvement here could be attributed to the data itself, there is no doubt that the addition of student-level covariates creates

better fitting extended TDCM's.

# 5 Simulation Study

Four simulation settings are used to evaluate the performance of the proposed extended TDCM in this section. These scenarios demonstrate potential for TDCM to model data with additional complexity compared to the empirical data analyzed in Section 4.

The simulations 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). Lastly, 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). The time constrained model is used in each setting such that results may be easily displayed.

We hold constant in all simulation settings such that each setting contains 800 respondents ($N = 800$), 21 questions/items ($J = 21$), and 3 attributes ($K = 3$). A total of 100 data sets, each of which includes item-correct response data $Y$, and respondent profiles $\alpha$, are simulated from fixed $\mathcal{B}$ item-attribute parameters and $\Gamma$ transition regression parameters. We set prior for each parameter with intention to be approximately non-informative, such that $\mathcal{B}$ and $\Gamma$ 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 $\alpha$ 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 $\Gamma$, thus the coverage rate of each simulation's accounting of the true parameter values are report as well.

To supplement the simulations in Section 6, a single simulation's posterior distributions are shown for each of the four settings. Moreover, comparison of computation time for the four settings are displayed in Table 10. [mj: Matt, we may not report this if the computing time is not that

appealing]

## 5.1 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 Supplement (Section 2.1). Intervention here is distributed evenly amongst the total of 800 respondents, with both the control and treatment group 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.

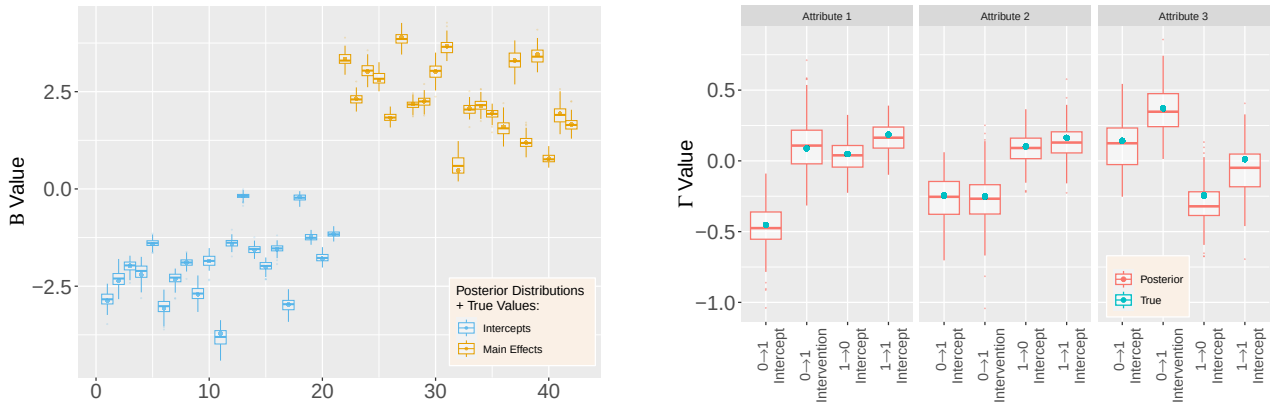


Figure 10: Posterior $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound by two standard deviations for the $K = 3$ simulation setting. The four indices for each attribute denote transition $0 \to 1$ intercept, $0 \to 1$ treatment (single group), $1 \to 0$ intercept, and $1 \to 1$ intercept, respectively.

The resulting posterior point estimates and variation bound of two standard deviations can be seen in Figure 10 for $\mathcal{B}$ (left) and $\Gamma$ (right). For $\mathcal{B}$ indices, the first 21 indicate intercepts and are followed by the 21 main effects. Transition $\Gamma$'s are divided into three sections, corresponding to $K = 3$ multivariate 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 $\Gamma$ to validate our results directly. We see that the coverage of the estimated posteriors $\mathcal{B}$ and $\Gamma$ in Figure 11 indicate high rate of recovery for both sets of parameters, reflecting that almost all of the 100 simulations produced posteriors that covered true parameter values.

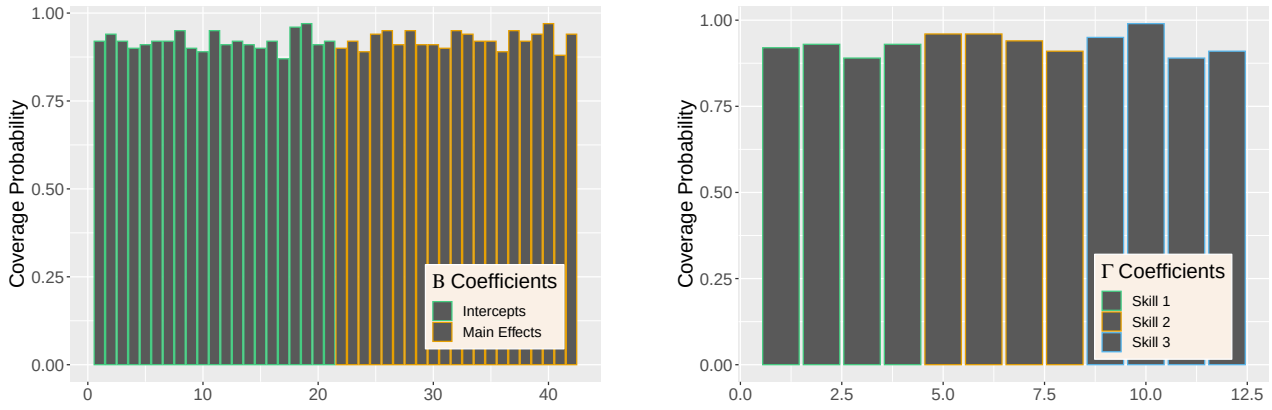


Figure 11: True value coverage rates for 95% credible intervals of $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound for the $K = 3$ simulation setting. The four indices for each attribute denote transition $0 \to 1$ intercept, $0 \to 1$ treatment (single group), $1 \to 0$ intercept, and $1 \to 1$ intercept, respectively.

This simulation setting is closely related to the empirical set-up of Section 4, and serves as the foundational case of these simulation studies. Complexity is then further added to explore how the current efficacy persists.

## 5.2 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 question impacted by multiple attributes, as shown in Supplement (Section 2.1). Recall in Section 3.3.2 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 posterior since having full normal posteriors does not violate monotonicity of the model.

Figure 12 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 $\Gamma$ remain the same as seen in the previous Section 5.1. These results show adequate coverage rate for the added interaction terms in Figure 13.

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 increased number of parameters. In the following, the flexibility of the newly introduced transition regressions

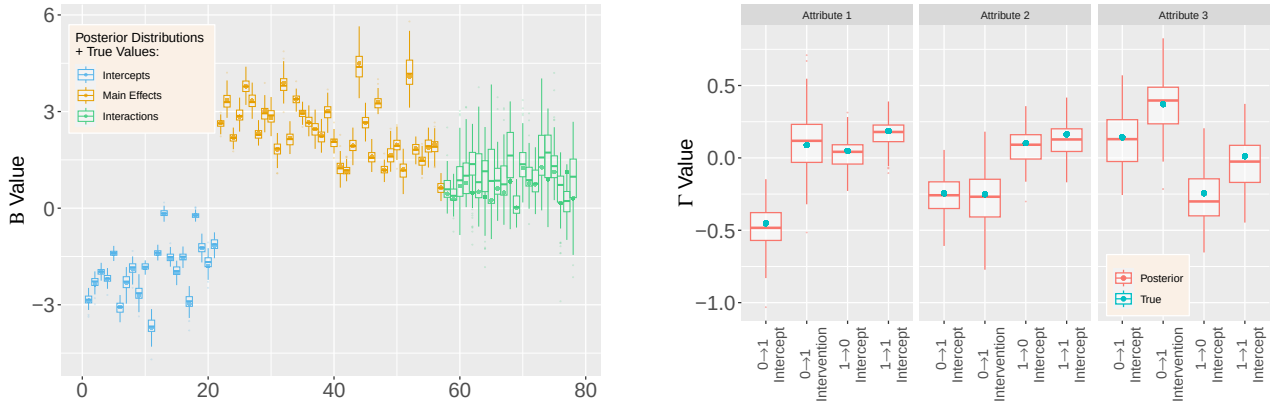


Figure 12: Posterior $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound by two standard deviations for the simulation setting with four covariates.

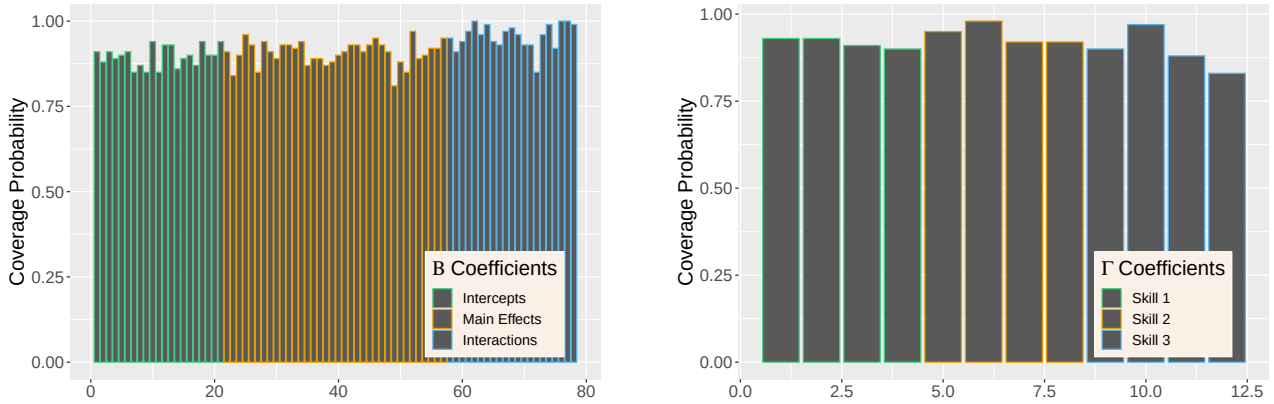


Figure 13: True value coverage rates for 95% credible intervals of $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound for the $K = 3$ with additional covariates simulation setting. The first five estimates correspond to $0 \to 1$ transition (intercept plus four covariate effects), followed by the next five for $1 \to 0$, and the intercept for $1 \to 1$.

is updated.

## 5.3 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 latent transition analysis. 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 $\Gamma$ 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 $0 \to 1$ transitions for computational simplicity and ease of display.

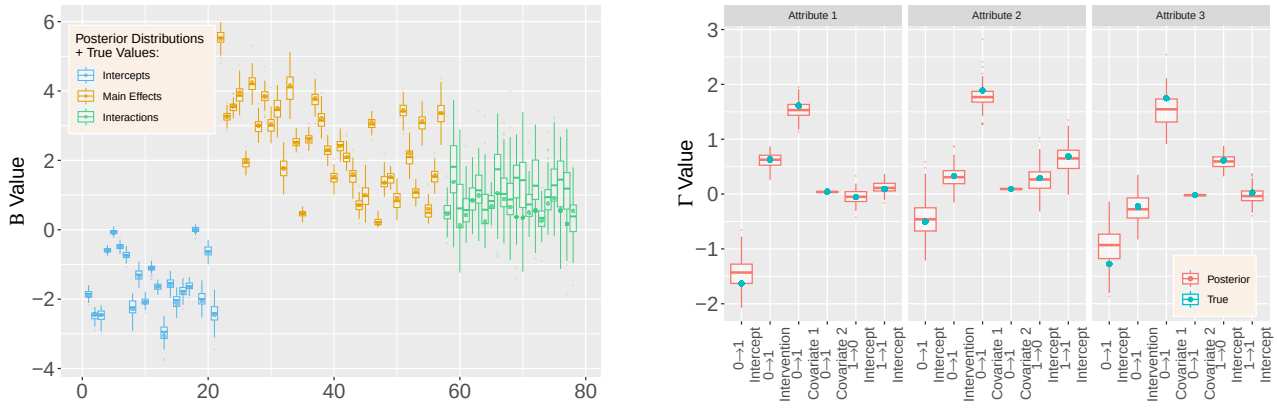


Figure 14: Posterior $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound by two standard deviations for the simulation setting with four covariates.

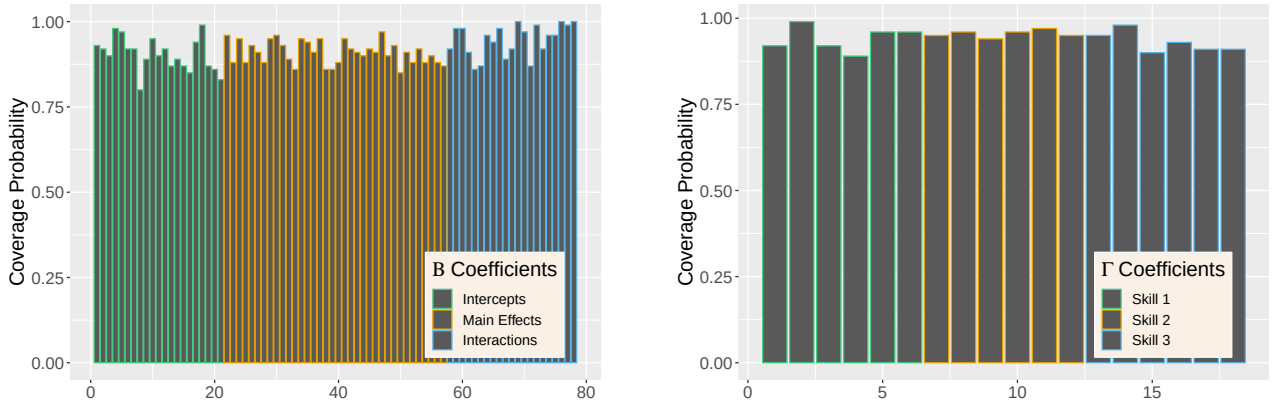


Figure 15: True value coverage rates for 95% credible intervals of $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound for the $K = 3$ with additional covariates simulation setting. The first five estimates correspond to $0 \to 1$ transition (intercept plus four covariate effects), followed by the next five for $1 \to 0$, and the intercept for $1 \to 1$.

Figure 14 shows the estimated posteriors with the true simulation values for our two sets of parameters. Indices for $\mathcal{B}$ remain the same, whereas $\Gamma$ are organized such that for each attribute, the indices correspond to the $0 \to 1$ transition intercept, intervention effect, uniform distribution

covariate effect, normal distribution covariate effect, followed by the intercepts for $1 \to 0$, and $1 \to 1$. The coverage rates for both $\mathcal{B}$ and $\Gamma$ remain favorable and are not noticeably different from those of the previous simulation settings.

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

### 5.4 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 invention treatment covariate. Note that the previously discussed empirical studies (Madison & Bradshaw, 2018a, 2018b) in Section 4 were based on two time points only. Thus, this simulation demonstrates the capacity of the extended TDCM that goes beyond the standard TDCM applications.

| 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 |

Table 7: The $2^T$ types of transitions (denoted $r$) for each attribute for $T = 3$, modeled by multivariate logistic regressions. The eight different transition are the possible ways that a respondent can have for a certain attribute over the three time points.

Here, transition multivariate 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 for the multivariate regression doubles from the previous simulation and empirical settings, where $T = 2$. That is, for each of the $K = 3$ attributes, Table 7 explicitly lays out 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 $0 \to 1 \to 0$, $0 \to 1 \to 1$, $1 \to 0 \to 0$, and $1 \to 0 \to 1$ (corresponding

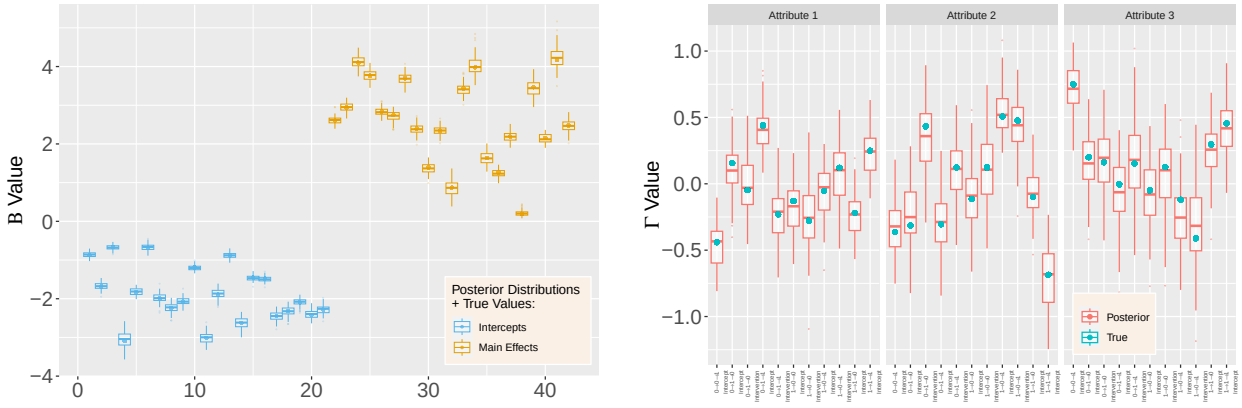


Figure 16: Posterior $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound by two standard deviations for the simulation setting with time-varying $\mathcal{B}$ structure.

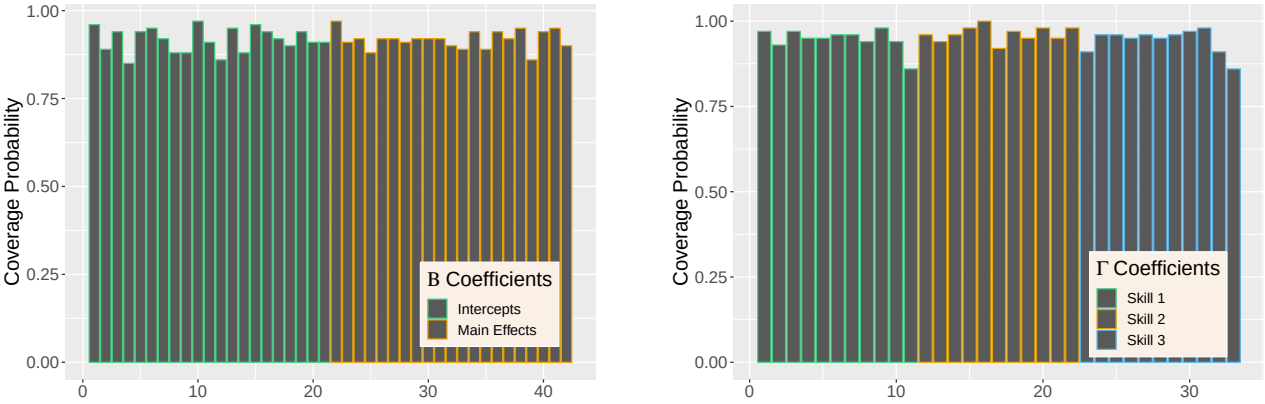


Figure 17: True value coverage rates for 95% credible intervals of $\mathcal{B}$ (left) and $\Gamma$ (right) distributions bound for the $K = 3$ simulation setting. The four indices for each attribute denote transition $0 \to 1$ intercept, $0 \to 1$ treatment (single group), $1 \to 0$ intercept, and $1 \to 1$ intercept, respectively.

to $r = \{3, 4, 5, 6\}$) in Table 7 while the first transition type $0 \to 0 \to 0$ serves as the baseline constraint. See that this setup is consistent with applying covariates to transition $0 \to 1$ and $1 \to 0$ for $T = 2$.

For $\Gamma$, 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 $\gamma_k$ for each attribute. These 11 parameters for each attribute are explicitly labeled as indices in the $\Gamma$ posteriors on the right side of Figure 16. The structure of $\Gamma$ here is indexed following the order of Table 7. We see that overall the estimate intervals show good coverage.

In summary of the simulations studies of this section, we have extended the settings used by Madison & Bradshaw (2018a) and Madison & Bradshaw (2018b) for the standard TDCM and shown the flexible capabilities of the extended model formulation in Section 3.2 as well as the efficacy of the Gibbs sampling framework for Bayesian inference via Pòlya-gamma data augmentation of Section 3.3. The increased complexity does not decrease the ability for the extended TDCM to produce desired results. These simulations, along with the empirical studies, have shown the extended TDCM to be a valuable contribution to longitudinal DCMs.

# 6 Discussion

This study introduced a more flexible extension to the recent Transition Diagnostic Classification Model (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 multivariate 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 transition parameter $\Gamma$. 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. 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 $\Gamma$ and provide measure on 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`, the extended TDCM offers a Gibbs sampling framework made efficient by Pòlya-gamma data augmentation. Model validation implementations are laid out in the R package **cdmfits**. 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 set-

tings 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.

In the current study, we assumed measurement invariance for $\mathcal{B}$ across time in all analysis. This was set for the comparisons with previous work based on the same data set. 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.

On another note, the Gibbs sampling framework of the extended TDCM does alleviate 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 increases drastically and may hinder extended TDCM's computation performance. The flexible structure and inference method of the extended TDCM provides plentiful opportunities for future works in longitudinal DCMs.

In summary, 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 eased computational requirements for longitudinal (prettest/posttest) design in educational research. We are hopeful that the method and software package we provided will assist educators in assessing growth over time, and thereby better supporting students' learning and improvements.

# References


Balamuta, J. J., & Culpepper, S. A. (2022). Exploratory restricted latent class models with monotonicity requirements under pólya–gamma data augmentation. *Psychometrika*, *87*(3), 1–43.

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.

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.

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

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*(4), 301–321.

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*, 191–210.

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.

Lazarsfeld, P. F. (1955). Recent developments in latent structure analysis. *Sociometry*, *18*(4), 391–403.

Madison, M. J., & Bradshaw, L. P. (2018a). Evaluating intervention effects in a diagnostic classification model framework. *Journal of Educational Measurement*, *55*(1), 32–51.

Madison, M. J., & Bradshaw, L. P. (2018b). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*, 963–990.

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

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.

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

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, *79*, 317–339.

Zhang, Z., Zhang, J., Lu, J., & Tao, J. (2020). Bayesian estimation of the dina model with pólya-gamma gibbs sampling. *Frontiers in Psychology*, *11*, 384.

# Appendix

## A. Derivation of Posterior Distributions in MCMC Procedure

The notation in the appendix is the same as stated in 3.1. Let $g(\rho^*_{irk})$ be the probability density function of a Pòlya-Gamma distribution with parameter $(1, 0)$ for all $i, r, k$. $h(y^*_{jc})$ 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 $\Gamma$ are listed as the following:

$$\begin{aligned} P(Y|\mathcal{A},\mathcal{B}) &= \prod_{i=1}^{N}\prod_{j=1}^{J} P(Y_{ij} = y_{ij}|\boldsymbol{\alpha}_i^{(t_j)}, \boldsymbol{\beta}_j) \\ &= \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{c=1}^{2^K} \left(\frac{\exp\left(y_{ij}\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}{1+\exp\left(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\right)}\right)^{\mathbb{1}\left(\boldsymbol{\alpha}_i^{(t_j)}=\mathbf{v}^{-1}(c)\right)}, \end{aligned} \tag{24}$$

$$\begin{aligned} P(\mathcal{A}|\Gamma) &= \prod_{i=1}^{N}\prod_{k=1}^{K} P(\boldsymbol{\rho}_{ik}|\boldsymbol{\gamma}_{rk}) \\ &= \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T} \left(\frac{\exp\psi_{irk}}{\sum_{r^*=1}^{2^T}\exp\psi_{ir^*k}}\right)^{\mathbb{1}\left(\boldsymbol{\rho}_{ik}=\boldsymbol{\nu}^{-1}(r)\right)} \\ &= \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T} \left(\frac{\exp\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk}}{\sum_{r^*=1}^{2^T}\exp\boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k}}\right)^{\mathbb{1}\left(\boldsymbol{\rho}_{ik}=\boldsymbol{\nu}^{-1}(r)\right)}, \end{aligned} \tag{25}$$

$$P(\mathcal{B}) = \prod_{j=1}^{J}\prod_{p=1}^{2^{\sum \mathbf{q_j}}} \frac{1}{(1-\Phi(\frac{L_{jp}}{\sigma^2_{\text{prior};\beta_{jp}}}))\sqrt{2\pi\sigma^2_{\text{prior};\beta_{jp}}}} \exp\left(-\frac{\beta^2_{jp}}{2\sigma^2_{\text{prior};\beta_{jp}}}\right)\mathbb{1}(\beta_{jp} \geq L_{jp}), \tag{26}$$

$$P(\Gamma) = \prod_{k=1}^{K}\prod_{r=1}^{2^T}(2\pi)^{-\frac{M_{rk}}{2}} \det(\Sigma_{\text{prior};\boldsymbol{\gamma}_{rk}})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\gamma}'_{rk}\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}}\boldsymbol{\gamma}_{rk}\right). \tag{27}$$

Also, we have the probability density functions of Pòlya-Gamma augmentation auxiliary vari-

ables as the following:

$$
\begin{aligned}
P(\boldsymbol{\rho}^*|\Gamma) &= \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T} P(\rho^*_{irk}|\boldsymbol{\gamma}_{rk}) \\
&= \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T} \cosh\left(\frac{\eta_{irk}}{2}\right)\exp\left(-\frac{\eta^2_{irk}\rho^*_{irk}}{2}\right)g(\rho^*_{irk}) \qquad (28) \\
P(Y^*|\mathcal{B}) &= \prod_{j=1}^{J}\prod_{c=1}^{2^K} P(y^*_{jc}|\boldsymbol{\beta}_j) \\
&= \prod_{j=1}^{J}\prod_{c=1}^{2^K} \cosh^{n_{jc}}\left(\frac{\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j}{2}\right)\exp\left(-\frac{(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)^2 y^*_{jc}}{2}\right)h(y^*_{jc}) \qquad (29)
\end{aligned}
$$

The joint likelihood function is:

$$
\begin{aligned}
& P(Y, Y^*, \mathcal{A}, \mathcal{B}, \Gamma, \boldsymbol{\rho}^*) \qquad (30)\\
&= P(Y|\mathcal{A}, \mathcal{B}) P(\mathcal{A}|\Gamma) P(\mathcal{B}) P(\Gamma) P(\boldsymbol{\rho}^*|\Gamma) P(Y^*|\mathcal{B}) \\
&= \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) \\
&\times \prod_{j=1}^{J}\prod_{c=1}^{2^K}\cosh^{n_{jc}}\left(\frac{\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j}{2}\right)\exp\left(-\frac{(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)^2 y^*_{jc}}{2}\right)h(y^*_{jc}) \\
&= \prod_{j=1}^{J}\prod_{c=1}^{2^K}\exp\left(\sum_{i=1}^{N}y_{ij}\mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)}=\mathbf{v}^{-1}(c))\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j\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\boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k}} \\
&\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) \\
&\times \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{r=1}^{2^T}\cosh\left(\frac{\eta_{irk}}{2}\right)\exp\left(-\frac{\eta_{irk}^2\rho^*_{irk}}{2}\right)g(\rho^*_{irk}) \\
&\times \prod_{j=1}^{J}\prod_{c=1}^{2^K}2^{-n_{jc}}\exp\left(-\frac{n_{jc}\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j+(\boldsymbol{\delta}_c^{(j)\prime}\boldsymbol{\beta}_j)^2 y^*_{jc}}{2}\right)h(y^*_{jc})
\end{aligned}
$$

(31)

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 posterior distribution of $\beta_{jp}$ is

$$P(\beta_{jp}|\cdot) \propto [\prod_{c=1}^{2^K} \exp\,(\sum_{i=1}^{N} y_{ij}\,\mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c))\delta_{cp}^{(j)}\beta_{jp})] \exp\,(-\frac{\beta_{jp}^2}{2\sigma^2_{\text{prior};\beta_{jp}}})\mathbb{1}(\beta_{jp} \geq L_{jp})$$
$$\times \prod_{c=1}^{2^K} \exp\,(-\frac{n_{jc}\delta_{cp}^{(j)}\beta_{jp} + (\sum_{p=1}^{2^{\sum \mathbf{q_j}}} \delta_{cp}^{(j)}\beta_{jp})^2 y^*_{jc}}{2}). \tag{32}$$

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

$$\sigma^2_{\text{post};\beta_{jp}} = (\sigma^{-2}_{\text{prior};\beta_{jp}} + \sum_{c=1}^{2^K} {\delta_{cp}^{(j)}}^2 y^*_{jc})^{-1} = (\sigma^{-2}_{\text{prior};\beta_{jp}} + {\boldsymbol{\Delta}_p^{(j)}}' \text{diag}\left([y^*_{jc}]_{c=1}^{2^K}\right) \boldsymbol{\Delta}_p^{(j)})^{-1} \tag{33}$$

$$\mu_{\text{post};\beta_{jp}} = \sigma^2_{\text{post};\beta_{jp}} \sum_{c=1}^{2^K} [(\sum_{i=1}^{N} y_{ij}\,\mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c))\delta_{cp}^{(j)}) - \frac{1}{2} n_{jc}\delta_{cp}^{(j)} - \sum_{p^* \neq p} y^*_{jc}\delta_{cp^*}^{(j)}\beta_{jp^*}\delta_{cp}^{(j)}]$$
$$= \sigma^2_{\text{post};\beta_{jp}} {\boldsymbol{\Delta}_p^{(j)}}' \text{diag}(y^*_{jc})\tilde{\boldsymbol{z}}_j \tag{34}$$

$$\boldsymbol{\Delta}_p^{(j)} = [\delta_{1p}^{(j)}, \delta_{2p}^{(j)}, \ldots, \delta_{2^K p}^{(j)}] \tag{35}$$

$$\tilde{\boldsymbol{z}}_j = \boldsymbol{z}_j - \boldsymbol{\delta}_{-p}^{(j)}\boldsymbol{\beta}_{j,-p} \tag{36}$$

$$\boldsymbol{z}_j = [\frac{\kappa_{j1}}{y^*_{j1}}, \frac{\kappa_{j2}}{y^*_{j2}}, \ldots, \frac{\kappa_{j2^K}}{y^*_{j2^K}}] \tag{37}$$

$$\kappa_{jc} = -\frac{1}{2} n_{jc} + \sum_{i=1}^{N} y_{ij}\,\mathbb{1}(\boldsymbol{\alpha}_i^{(t_j)} = \mathbf{v}^{-1}(c)). \tag{38}$$

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

$$P(\boldsymbol{\gamma}_{rk}|\cdot) \propto \prod_{i=1}^{N} \frac{\exp\,(\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk}\mathbb{1}(\rho_{ik} = \mathbf{v}^{-1}(r)))}{\sum_{r^*=1}^{2^T} \exp\,(\boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k})} \times \prod_{i=1}^{N} \cosh\,(\frac{\eta_{irk}}{2}) \exp\,(-\frac{\eta^2_{irk}\rho^*_{irk}}{2})$$
$$\times \exp\,(-\frac{1}{2}\boldsymbol{\gamma}_{rk}\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}}\boldsymbol{\gamma}_{rk})$$
$$\propto \prod_{i=1}^{N} \exp\,(\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk}\mathbb{1}(\rho_{ik} = \mathbf{v}^{-1}(r)) - \frac{\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk}}{2} - \frac{(\boldsymbol{x}'_{irk}\boldsymbol{\gamma}_{rk} - \log\sum_{r \neq r^*} \exp \boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k})^2\rho^*_{irk}}{2})$$
$$\times \exp\,(-\frac{1}{2}\boldsymbol{\gamma}_{rk}\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}}\boldsymbol{\gamma}_{rk}). \tag{39}$$

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

$$\Sigma_{\text{post};\boldsymbol{\gamma}_{rk}} = (\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}} + \sum_{i=1}^{N} \boldsymbol{x}_{irk}\boldsymbol{x}'_{irk}\rho^*_{irk})^{-1} = (\Sigma^{-1}_{\text{prior};\boldsymbol{\gamma}_{rk}} + X'_{rk}\text{diag}\left([\rho^*_{irk}]_{i=1}^{N}\right) X_{rk})^{-1} \tag{40}$$

$$\begin{aligned}\mu_{\text{post};\boldsymbol{\gamma}_{rk}} &= \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}}[\sum_{i=1}^{N} \boldsymbol{x}_{irk}(\rho_{irk} - \frac{1}{2} - \rho^*_{irk}\log\sum_{r\neq r^*}\exp \boldsymbol{x}'_{ir^*k}\boldsymbol{\gamma}_{r^*k})] \\ &= \Sigma_{\text{post};\boldsymbol{\gamma}_{rk}} X'_{rk}(\boldsymbol{\zeta}_{rk} - \text{diag}\left([\rho^*_{irk}]_{i=1}^{N}\right)\boldsymbol{c}_{rk}) \end{aligned} \tag{41}$$

$$\boldsymbol{\zeta}_{rk} = [\rho^*_{1rk} - 1/2, \ldots, \rho^*_{Nrk} - \text{“}1/2] \tag{42}$$

$$\mathbf{c}_{rk} = [\log(\sum_{r_*\neq r}\exp\psi_{1r_*k}), \ldots, \log(\sum_{r_*\neq r}\exp\psi_{Nr_*k})], \tag{43}$$

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

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

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

# Supplemental Material

# 1 Empirical Study: Traceplots

## 1.1 Empirical Data Analysis - Single-Group

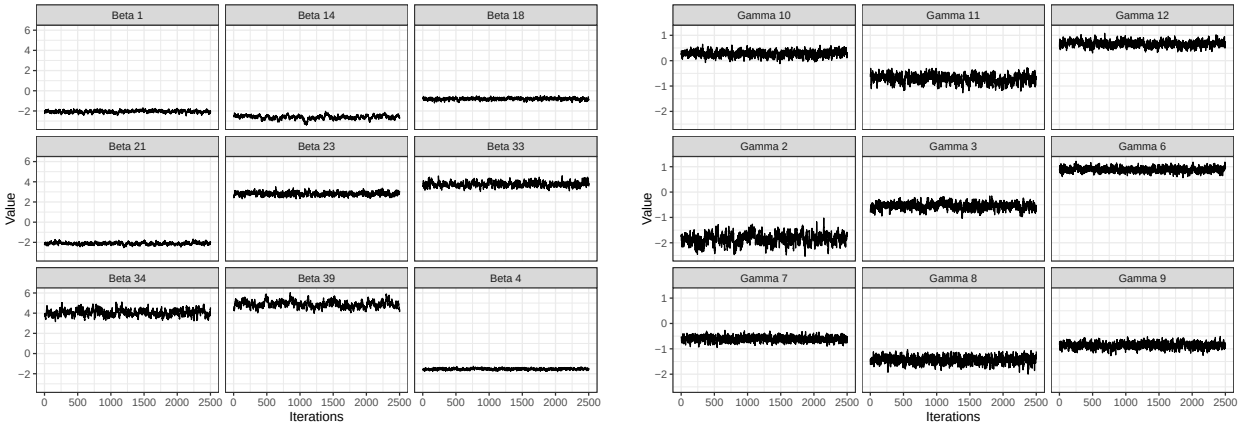


Figure 18: Trace plots of nine random parameters from $\mathcal{B}$ (left) and $\Gamma$ (right) showing convergence for the single-group empirical example in Section 4.1.

## 1.2 Empirical Data Analysis - Multiple-Group

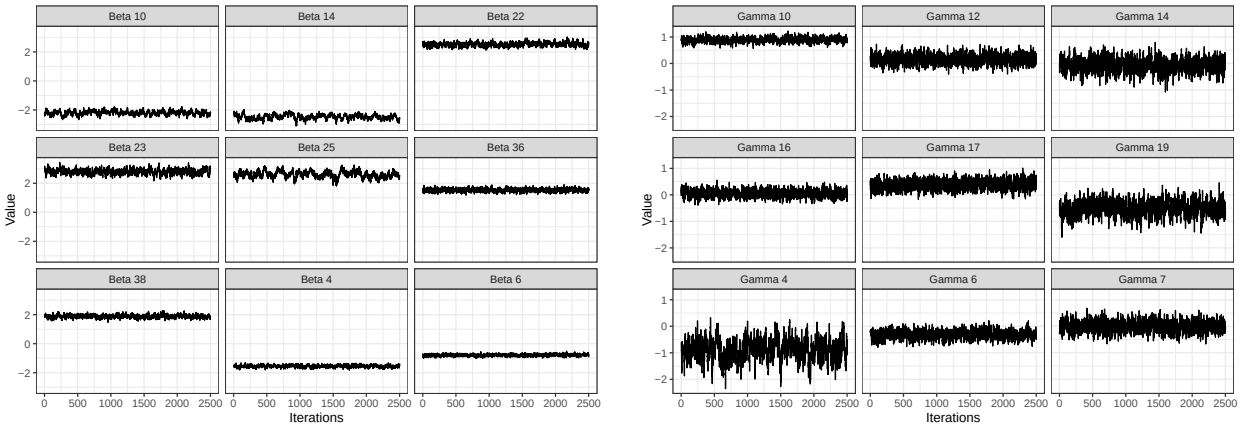


Figure 19: Trace plots of nine random parameters from $\mathcal{B}$ (left) and $\Gamma$ (right) showing convergence for the multiple-group empirical example in Section 4.2.

# 2 Simulation Study

## 2.1 Q matrices

| $J$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| k=1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| k=2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| k=3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 8: $Q$ matrix without interactions for $K = 3$ simulation data indicating attribute requirements for each $J = 21$ test questions, used in Section 5.1 and Section 5.4

| $J$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| k=1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| k=2 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| k=3 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |

Table 9: $Q$ matrix with interactions for $K = 3$ simulation data indicating attribute requirements for each $J = 21$ test questions, used in Section 5.2 and Section 5.3

## 2.2 Simulation Computation Times

| | Setting 1 | Setting 2 | Setting 3 | Setting 4 |
|---|---|---|---|---|
| Seconds | 2988.39 | 3093.11 | 3159.47 | 4717.73 |
| Minutes | 49.81 | 51.55 | 52.66 | 78.63 |

Table 10: Time elapsed for the completion of a single simulation for each simulation setting shown in in Section 5, where 100 simulations are paralleled and summarized. Computations are performed on the Apple silicon M1 process.