RESTRICTED LATENT CLASS MODELS FOR NOMINAL RESPONSE DATA: IDENTIFIABILITY AND ESTIMATION

YING LIU AND STEVEN ANDREW CULPEPPER

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Restricted latent class models (RLCMs) provide an important framework for diagnosing and classifying respondents on a collection of multivariate binary responses. Recent research made significant advances in theory for establishing identifiability conditions for RLCMs with binary and polytomous response data. Multiclass data, which are unordered nominal response data, are also widely collected in the social sciences and psychometrics via forced-choice inventories and multiple choice tests. We establish new identifiability conditions for parameters of RLCMs for multiclass data and discuss the implications for substantive applications. The new identifiability conditions are applicable to a wealth of RLCMs for polytomous and nominal response data. We propose a Bayesian framework for inferring model parameters, assess parameter recovery in a Monte Carlo simulation study, and present an application of the model to a real dataset.

Key words: restricted latent class models, nominal response data, cognitive diagnosis model, identifiability, Bayesian.

1. Introduction

Mixture models for multivariate, unordered categorical data, which are also referred to as nominal data, are widely used as a data reduction technique to uncover a partition of latent classes. Nominal response data arises naturally in a diverse collection of fields and associated latent class models have been applied to uncover the structure underlying positional dependence of nucleotides (Dunson & Xing, 2009), surveys responses for political elections (DeYoreo et al., 2017), anuran abundance using calling survey data (Royle & Link, 2005), as well as for multiple imputation of educational (Si & Reiter, 2013) and social science (Murray & Reiter, 2016; Vermunt et al., 2008) surveys. In short, nominal mixture models serve an important role across the physical and social sciences.

Recent psychometric research introduced a class of restricted latent class models (RLCMs) that use a more parsimonious formulation for describing the structure underlying multivariate nominal data (e.g., see Chen & Zhou, 2017, Fang et al., 2019, Templin et al., 2008) than the traditional framework, which we refer to as unrestricted latent class models (ULCMs). For instance, a popular application of nominal RLCMs is to understand how latent classes relate to target and distractor responses on multiple choice tests (e.g., see Bradshaw & Templin, 2014, De La Torre, 2009, DiBello et al., 2015, Ku et al., 2016, Shear & Roussos, 2017, Yigit et al., 2019). In order to distinguish RLCMs and ULCMs we let Θ denote the emission parameters that govern the likelihood that latent classes select different options on J variables with M unordered options for each variable. If C denotes the number of latent classes, the classic ULCM framework includes $J \times M \times C$ parameters to relate latent class membership to observed responses. In contrast, RLCMs impose structure on the elements of Θ by constraining some elements to be equal. Accordingly,

Correspondence should be made to Steven Andrew Culpepper, Department of Statistics, University of Illinois at Urbana-Champaign, Computing Applications Building, Room 152, 605 E. Springfield Ave., Champaign, IL61820, USA. Email: sculpepp@illinois.edu

RLCMs include fewer parameters than ULCMs. Furthermore, as we demonstrate below, RLCMs generally offer a more interpretable framework for understanding the latent structure (i.e., the relationship between the latent classes and observed variables). In fact, the pattern of equal and unequal elements in the RLCM $\boldsymbol{\Theta}$ parameter provides researchers with a guide for interpreting the impact of latent class membership on response probabilities.

Although prior research developed several general models for nominal RLCMs, there are at least two limitations with existing research that limits widespread applicability of these methods for statistical research in education and the social sciences. First, existing methods are primarily confirmatory in nature given that researchers must prespecify the manner by which latent classes relate to observed response probabilities. Specifically, let $\boldsymbol{\Delta}$ be the RLCM parameter with $J \times (M-1) \times C$ binary elements that indicate which elements of $\boldsymbol{\Theta}$ are equal (we formally define the $\boldsymbol{\Delta}$ matrix below with examples). Currently deployed nominal RLCMs must specify every element of $\boldsymbol{\Delta}$, which may be challenging for some research applications. Whereas researchers may be able to correctly articulate the latent structure in $\boldsymbol{\Delta}$ for some applications (e.g., target and distractor responses on some multiple choice tests), the general unavailability of substantive theory would limit widespread application of nominal RLCMs. Second, numerous studies developed RLCMs for multivariate nominal data, yet there has been limited research on conditions that are needed to ensure model parameters are identified. It is important to note that several studies discussed identifiability for ULCMs (e.g., see Allman et al., 2009); however, current results are not specific for RLCMs given that Allman et al. (2009) consider an unrestricted parameter space whereas the parameter space for our RLCM is restricted by the structure of $\boldsymbol{\Delta}$. Consequently, the parameter space falls into a measure zero set with respect to the whole parameter space of ULCMs as discussed in Allman et al. (2009), so identifiability conditions mentioned above for ULCMs cannot be directly applied to our RLCMs. Furthermore, our paper contributes to literature on the identifiability RLCMs. An extensive collection of literatures have delved into local identifiability issues, which aim to ensure the model parameters are identifiable in a neighborhood of the true parameters. McHugh (1956) proposed sufficient conditions to determine the local identifiability condition for latent class model with binary response. Goodman (1974) extended the conditions for latent class models with polytomous response. Huang and Bandeen-Roche (2004) proposed local identifiability conditions for latent class models with covariates. For global identifiability issue, there are numerous papers proposing strict and generic identifiability conditions for binary response data (Chen et al., 2015, 2020; Xu, 2017; Xu & Shang, 2018) and strict identifiability conditions for polytomous response data (Culpepper, 2019; Fang et al., 2019). Additionally, Gu and Dunson (2021) establish strict and generic identifiability conditions for a multiclass, multilayer latent structure model. Gu and Dunson (2021) could be viewed as a more general model than the one we consider as it admits a multilayered, hierarchical structure for attributes. One strength of our paper relative to Gu and Dunson (2021) is that our identifiability conditions provide practitioners with clear guidance for designing nominal response assessments (e.g., forced-choice inventories). Furthermore, our identifiability conditions also provide generic conditions that are applicable to polytomous RLCMs.

Accordingly, the goal of our study is to address the aforementioned shortcomings in the literature. That is, we propose a fully exploratory framework for inferring nominal RLCM parameters and present new theory regarding model identification. The identifiability of model parameters is critical for statistical inference and we also provide researchers with guidance for designing multivariate nominal response studies.

It is also important to distinguish the models we explore in this study in comparison to polytomous latent class models. Specifically, researchers advanced RLCMs for polytomous data for both confirmatory (e.g., see Ma & de la Torre, 2016; 2019) and exploratory methods (Culpepper, 2019; Culpepper & Balamuta, 2021; Jimenez et al., 2023). There are also several studies (Bacci et al., 2014; Bartolucci, 2007; Gnaldi et al., 2020) described latent class models within an item

response theory (IRT) framework with at least three link functions (i.e., graded response, partial credit, and continuation ratio). These prior studies made important contributions and demonstrated how to use link functions for modeling ordered, polytomous response data with latent class models. In contrast, an important innovation of our study is that we deploy the multinomial logistic link function, which is suitable for unordered, nominal responses.

The remainder of this paper includes six sections. The first section provides a general introduction to ULCMs and RLCMs for nominal data and the second section presents new theoretical results concerning the identifiability of RLCMs (please see Appendix for related proofs). The third section outlines a Bayesian formulation for inferring the RLCM parameter posterior distribution. The fourth section reports Monte Carlo results concerning the accuracy of the developed algorithm and the fifth section reports results from an application. The final section discusses the implications of this study and provides concluding remarks.

2. Overview of Mixture Models for Nominal Responses

We consider the setting where multivariate, nominal response data are available such that Y_j (for $j = 1, \ldots, J$) is a random categorical (or nominal) response with a realization $y_j \in \{0, \ldots, M_j - 1\}$ where $M_j \geq 2$ denotes the number of unordered response options. We denote the random J-vector by $\boldsymbol{Y} = (Y_1, \ldots, Y_J)^\top$ and the observed vector of responses as $\boldsymbol{y} = (y_1, \ldots, y_J)^\top$. The support for $\boldsymbol{Y}$ is defined as $\boldsymbol{y} \in \times_{j=1}^{J} \{0, \ldots, M_j - 1\}$, which implies there are $\prod_{j=1}^{J} M_j$ possible observed response patterns. The purpose of this section is discuss the role of mixture models in understanding the multivariate, nominal response patterns. The first subsection reviews existing *unstructured* latent class models (ULCMs) for nominal, unordered response data. ULCMs offer a powerful framework for uncovering substantively meaningful latent classes. However, the results from ULCMs data analyses may not always be easily interpretable as researchers must decipher the meaning of latent classes by comparing many latent class parameters. Accordingly, the second subsection introduces a new general restricted latent class model (RLCM) framework, which has the benefit of directly uncovering the latent structure by providing researchers with a $\boldsymbol{\Delta}$ parameter for more easily interpreting the class labels.

2.1. Unstructured Latent Class Models (ULCMs)

The goal of this section is to review the traditional ULCM framework. Let $c \in \{0, \ldots, C - 1\}$ index the C underlying latent classes. In the case of nominal data, the unstructured model includes a M_j-vector of category response probabilities for each class and item denoted by $\boldsymbol{\theta}_{jc} = (\theta_{jc0}, \ldots, \theta_{jc,M_j-1})^\top$ so that the probability of observing a response of m on item j for members of class c is $\theta_{jcm} = P(Y_j = m|c)$. We define $\boldsymbol{\Theta}_j = (\boldsymbol{\theta}_{j0}, \ldots, \boldsymbol{\theta}_{j,C-1})$ as the $M_j \times C$ matrix of response probabilities by response option and latent class. The goal of ULCMs is to describe the $\prod_{j=1}^{J} M_j$ possible response patterns. ULCMs consider the case where latent classes differ in their chances of responding according to a given response pattern. The probability vector that governs the chance members of class c respond according to one of the $\prod_{j=1}^{J} M_j$ possible response patterns is $\mathbb{P}_c = \bigotimes_{j=1}^{J} \boldsymbol{\theta}_{jc}$ where $\otimes$ denotes a Kronecker product. Let $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{C-1})^\top$ be a C-vector of structural probabilities such that π_c denotes the chance of membership in class c and note that the model implied response pattern probability vector is $\mathbb{P} = \sum_{c=0}^{C-1} \pi_c \mathbb{P}_c$.

2.2. Restricted Latent Class Models (RLCMs)

This subsection introduces a RLCM for nominal data which offers a more interpretable solution by imposing restrictions on the ULCM $\boldsymbol{\theta}_{jc}$ parameters. In particular, the RLCM adapts

the ULCM to describe the $\prod_{j=1}^{J} M_j$ response patterns by reparameterizing both the latent space and parameters. First, the RLCM defines the latent classes using a 2^K binary attribute vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\top \in \{0, 1\}^K$. Therefore, the connection between the number of classes in the ULCM and the RLCM is $C = 2^K$. An advantage of using the binary attribute profile is that researchers can interpret $\alpha_k = 1$ as denoting possession or mastery of attribute k and $\alpha_k = 0$ otherwise. The relationship between the ULCM and RLCM is also apparent when using a bijection between the binary attribute profile $\boldsymbol{\alpha}$ and the integers $c \in \{0, \ldots, 2^K - 1\}$ by defining class $c = \boldsymbol{\alpha}^\top \boldsymbol{v} \in \{0, \ldots, 2^K - 1\}$ with $\boldsymbol{v} = (2^{K-1}, 2^{K-2}, \ldots, 1)^\top$.

Second, the RLCM reparameterizes the elements of $\boldsymbol{\theta}_{jc}$ using the following multinomial logit-link function

$$\theta_{jcm} = \frac{\exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_{jm}\right)}{\sum_{m'=0}^{M_j-1} \exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_{jm'}\right)} \tag{1}$$

where $\boldsymbol{a}_c$ is a design vector for the attribute profile for class c and $\boldsymbol{\beta}_{jm}$ is a P-vector of coefficients for item j and option m (i.e., P depends on the order of the model, $P = 2^K$ if we include main and all interaction-effect terms for latent class). Note that the restriction $\boldsymbol{\beta}_{j0} = \boldsymbol{0}$ is deployed for all j to identify the model. Furthermore, the restriction on $\boldsymbol{\beta}_{j0}$ implies that $y_j = 0$ is the reference response so that $\boldsymbol{\beta}_{jm}$ for $m > 0$ quantifies the impact of the attributes on response values of $y_j = m$ versus $y_j = 0$ on item j. Let the $M_j \times 2^K$ matrix of coefficients for item j be denoted as $\boldsymbol{B}_j = (\boldsymbol{\beta}_{j0}, \ldots, \boldsymbol{\beta}_{j,M_j-1})^\top$.

An important implication of reparameterizing θ_{jcm} with a multinomial logit-link is that the transformed $\boldsymbol{\beta}_{jm}$ parameters provide a more coherent interpretation regarding the process by which the underlying attributes relate to observed responses. For instance, we define the 2^K-vector $\boldsymbol{a}_c$ as including main- and interaction-effect terms for latent class $\boldsymbol{\alpha}^\top \boldsymbol{v} = c$. Consequently, the elements of $\boldsymbol{\beta}_{jm}$ indicate the manner by which the attributes translate into preferences for response option m relative to response option zero.

We next present an example to further illustrate the link between ULCMs and RLCMs and the interpretation of the $\boldsymbol{a}_c$ and $\boldsymbol{\beta}_{jm}$ parameters.

Example 1. Suppose $K = 3$ and $M_j = 3$, so $y_j \in \{0, 1, 2\}$. In this case, the matrix of ULCM parameters is,

$$\boldsymbol{\Theta}_j = \begin{bmatrix} \theta_{j00} & \theta_{j10} & \theta_{j20} & \theta_{j30} & \theta_{j40} & \theta_{j50} & \theta_{j60} & \theta_{j70} \\ \theta_{j01} & \theta_{j11} & \theta_{j21} & \theta_{j31} & \theta_{j41} & \theta_{j51} & \theta_{j61} & \theta_{j71} \\ \theta_{j02} & \theta_{j12} & \theta_{j22} & \theta_{j32} & \theta_{j42} & \theta_{j52} & \theta_{j62} & \theta_{j72} \end{bmatrix} \tag{2}$$

where we note that $\theta_{jc0} = 1 - \sum_{m=1}^{M_j} \theta_{jcm}$ for all $c \in \{0, 1, \ldots, 7\}$. In this setting, the ULCM includes $2 \times 8 = 16$ parameters for each item. Moreover, in order to understand the meaning of the latent classes researchers would need to interpret differences in the $16 \cdot J$ total class probabilities, which may be challenging for even a modest number of items J. The RLCM attempts to address this problem by reparameterizing both the latent classes and item parameters. In the case with $K = 3$, we define the arbitrary design vector $\boldsymbol{a}$ as:

$$\boldsymbol{a}^\top = (1, \alpha_1, \alpha_2, \alpha_3, \alpha_1\alpha_2, \alpha_1\alpha_3, \alpha_2\alpha_3, \alpha_1\alpha_2\alpha_3) \tag{3}$$

so that $\boldsymbol{a}$ includes all main-effect and interaction terms among the attributes and we use $\boldsymbol{a}_c$ to refer to the design vector for attribute profile $\boldsymbol{\alpha}^\top \boldsymbol{v} = c$. The matrix of reparameterized parameters $\boldsymbol{\beta}_j$

for relating $\boldsymbol{\alpha}$ to Y_j is

$$\boldsymbol{B}_j = \begin{bmatrix} \beta_{j00} & \beta_{j10} & \beta_{j20} & \beta_{j30} & \beta_{j40} & \beta_{j50} & \beta_{j60} & \beta_{j70} \\ \beta_{j01} & \beta_{j11} & \beta_{j21} & \beta_{j31} & \beta_{j41} & \beta_{j51} & \beta_{j61} & \beta_{j71} \\ \beta_{j02} & \beta_{j12} & \beta_{j22} & \beta_{j32} & \beta_{j42} & \beta_{j52} & \beta_{j62} & \beta_{j72} \end{bmatrix}. \tag{4}$$

Note we can view $\boldsymbol{a}_c^\top \boldsymbol{\beta}_{jm}$ as the latent response propensity for members of class c to pick option m vs. option 0. Therefore, the definition of $\boldsymbol{a}$ implies that β_{j0m} is an intercept term that corresponds with the latent propensity for the latent class with $\boldsymbol{\alpha} = \boldsymbol{0}$ to select option m vs. 0. Furthermore, the main-effects for α_1, α_2, and α_3 for distinguishing response m from 0 are β_{j1m}, β_{j2m}, and β_{j3m}, respectively. Furthermore, the two-way interaction terms are $\alpha_1\alpha_2$, $\alpha_1\alpha_3$, and $\alpha_2\alpha_3$ with effects β_{j4m}, β_{j5m}, and β_{j6m}, respectively, and the three-way interaction effect is β_{j7m}. In general positive coefficients suggest preference for option m to option 0 and the interactive effects provide researchers with insight regarding the extent to which preferences are determined by a complex interplay of the attributes.

The aforementioned example demonstrates the ability of the RLCM to provide researchers with a more clear interpretation of the latent structure (i.e., the relationship between attributes and observed responses). Still, each $\boldsymbol{\beta}_j$ includes many parameters to estimate and interpret. A further refinement we advance to support coherent inferences about the latent structure is to incorporate variable selection methods into the RLCM to infer which of the elements of $\boldsymbol{\beta}_j$ are active (i.e., different from zero) versus inactive (i.e., equal to or near zero). In fact, the pattern of active vs. inactive elements of $\boldsymbol{\beta}_j$ indicates the underlying structure and describes the process by which attributes relate to the observed response Y_j. Accordingly, we introduce a $M_j \times 2^K$ binary matrix $\boldsymbol{\Delta}_j$ in order to indicate which elements of $\boldsymbol{\beta}_j$ are active. Specifically, $\delta_{jpm} = 1$ to denote that β_{jpm} is active (i.e., nonzero) and $\delta_{jpm} = 0$ if $\beta_{jpm} = 0$ (i.e., inactive). Note that we generally always include the intercept and fix $\delta_{j0m} = 1$ for all $m \in \{1, \ldots, M_j - 1\}$.

We next revisit Example 1 to highlight the role of $\boldsymbol{\Delta}_j$ in interpreting the latent structure.

Example 2. Reconsider the case with $M_j = 3$ and $K = 3$. In this case, $\boldsymbol{\Delta}_j$ is generally written as

$$\boldsymbol{\Delta}_j = \begin{bmatrix} \delta_{j00} & \delta_{j10} & \delta_{j20} & \delta_{j30} & \delta_{j40} & \delta_{j50} & \delta_{j60} & \delta_{j70} \\ \delta_{j01} & \delta_{j11} & \delta_{j21} & \delta_{j31} & \delta_{j41} & \delta_{j51} & \delta_{j61} & \delta_{j71} \\ \delta_{j02} & \delta_{j12} & \delta_{j22} & \delta_{j32} & \delta_{j42} & \delta_{j52} & \delta_{j62} & \delta_{j72} \end{bmatrix}. \tag{5}$$

Note that $\delta_{jp0} = 0$ for all $p = 0, \ldots, 2^K - 1$ to identify the model parameters and that terms for the intercepts are generally specified as active so $\delta_{j01} = \delta_{j02} = 1$.

Remark 1. If $\boldsymbol{\Delta}_j = \boldsymbol{1}$ for all $j = 1, \ldots, J$, which implies that all coefficients in $\boldsymbol{B}$ are active, the latent classes have distinct response probabilities, and the RLCM is equivalent to a ULCM in this case. For additional discussion see Example 1 of Chen et al. (2020) for an exposition involving the binary response RLCM.

Note that the pattern of 1's and 0's in $\boldsymbol{\Delta}_j$ convey different types of relationships and structures. The structure of an item is referred to as *simple structure* for attribute k if the response probabilities only differ by levels of α_k.

Definition 1. The structure of $\boldsymbol{\Delta}_j$, which is a slice of $\boldsymbol{\Delta}$ for item j, is referred to as simple structure for attribute k if it satisfies the following structure:

$$
\boldsymbol{\Delta}_j = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & \delta_{jk1} & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 & \delta_{jk,M_j-1} & 0 & \cdots & 0 \end{bmatrix}_{M_j \times P}, \tag{6}
$$

and $\sum_{m=1}^{M_j-1} \delta_{jkm} \geq 1$ where P generally equals 2^K.

Remark 2. Note that for convenience of notation that our identifiability proof below supposes that item j is simple structure for attribute j.

Example 3. Consider $M_1 = M_2 = 3$, $P = 2^K$ and $J = 2$ and note that examples of $\boldsymbol{\Delta}$ matrices that satisfy simple structure according to Definition 1 are:

$$
\boldsymbol{\Delta}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{7}
$$

$$
\boldsymbol{\Delta}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{8}
$$

Items 1 and 2 suppose the item is simple structure such that the probability of selecting the responses only relates to standing on α_1 in $\boldsymbol{\Delta}_1$ and α_2 for $\boldsymbol{\Delta}_2$. $\boldsymbol{\Delta}_1$ indicates that only the main-effect for α_1 differentiates between response option 1 vs. 0 and 2 vs. 0. In contrast, for item 2, $\boldsymbol{\Delta}_2$ represents the case where the main-effect for α_2 is only active for differentiating between response option 2 vs. 0. The associated $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ matrices for the structure parameters $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ are:

$$
\boldsymbol{B}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{101} & \beta_{111} & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{102} & \beta_{112} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{9}
$$

$$
\boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{201} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{202} & 0 & \beta_{222} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{10}
$$

Let $\boldsymbol{\Theta}_j(\boldsymbol{B}_j)$ denote the latent class response probabilities associated with the RLCM $\boldsymbol{B}_j$ matrix. The presence of structure in $\boldsymbol{B}_j$ and $\boldsymbol{\Delta}_j$ implies that elements of $\boldsymbol{\Theta}_j(\boldsymbol{B}_j)$ are restricted to be equal.

Example 4. The rows of $\boldsymbol{\Delta}_1$ in Eq. 7 imply that certain latent classes have a common probability of selecting response 1 vs 0 and 2 vs. 0. That is, latent classes that do not possess the first attribute such that $\boldsymbol{\alpha} = (0, \alpha_2, \alpha_3)$ have common response probabilities for selecting options 0, 1, and 2 of θ_{100}, θ_{101}, and θ_{102}, respectively, whereas classes with the first attribute with $\boldsymbol{\alpha} = (1, \alpha_2, \alpha_3)$ have common response probabilities of θ_{140}, θ_{141}, and θ_{142}. The $\boldsymbol{\Theta}_1(\boldsymbol{B}_1)$ in this case is

$$
\boldsymbol{\Theta}_1(\boldsymbol{B}_1) = \begin{bmatrix} \theta_{100} & \theta_{100} & \theta_{100} & \theta_{100} & \theta_{140} & \theta_{140} & \theta_{140} & \theta_{140} \\ \theta_{101} & \theta_{101} & \theta_{101} & \theta_{101} & \theta_{141} & \theta_{141} & \theta_{141} & \theta_{141} \\ \theta_{102} & \theta_{102} & \theta_{102} & \theta_{102} & \theta_{142} & \theta_{142} & \theta_{142} & \theta_{142} \end{bmatrix}, \tag{11}
$$

where the columns of $\Theta_1(B_1)$ are organized according to the binary-integer bijection and

$$\theta_{10} = (\theta_{100}, \theta_{101}, \theta_{102})^\top = \frac{1}{\sum_{m=0}^2 \exp(\beta_{10m})} (1, \exp(\beta_{101}), \exp(\beta_{102})), \tag{12}$$

$$\theta_{14} = (\theta_{140}, \theta_{141}, \theta_{142})^\top = \frac{1}{\sum_{m=0}^2 \exp(\beta_{10m} + \beta_{11m})} (1, \exp(\beta_{101} + \beta_{111}), \exp(\beta_{102} + \beta_{112})). \tag{13}$$

In contrast, the rows of Δ_2 in Eq. 8 imply a different collection of elements are constrained equal in $\Theta_2(B_2)$. Latent classes that do not possess the second attribute such that $\alpha = (\alpha_1, 0, \alpha_3)$ have common response probabilities for selecting options 0, 1, and 2 of θ_{200}, θ_{201}, and θ_{202}, respectively, whereas classes with the second attribute with $\alpha = (\alpha_1, 1, \alpha_3)$ have common response probabilities of θ_{220}, θ_{221}, and θ_{222}. The $\Theta_2(B_2)$ in this case is

$$\Theta_2(B_2) = \begin{bmatrix} \theta_{200} & \theta_{200} & \theta_{220} & \theta_{220} & \theta_{200} & \theta_{200} & \theta_{220} & \theta_{220} \\ \theta_{201} & \theta_{201} & \theta_{221} & \theta_{221} & \theta_{201} & \theta_{201} & \theta_{221} & \theta_{221} \\ \theta_{202} & \theta_{202} & \theta_{222} & \theta_{222} & \theta_{202} & \theta_{202} & \theta_{222} & \theta_{222} \end{bmatrix}, \tag{14}$$

where the columns of $\Theta_2(B_2)$ are organized according to the binary-integer bijection and

$$\theta_{20} = (\theta_{200}, \theta_{201}, \theta_{202})^\top = \frac{1}{\sum_{m=0}^2 \exp(\beta_{20m})} (1, \exp(\beta_{201}), \exp(\beta_{202})), \tag{15}$$

$$\theta_{22} = (\theta_{220}, \theta_{221}, \theta_{222})^\top = \frac{(1, \exp(\beta_{201}), \exp(\beta_{202} + \beta_{222}))}{1 + \exp(\beta_{201}) + \exp(\beta_{202} + \beta_{222})}. \tag{16}$$

Remark 3. Note that Δ_j can also denote different structures where multiple attributes relate to response variables. For instance, Δ_j might specify the inclusion of interaction terms so that response probabilities are shaped by a more complex relationship of the attributes. Furthermore, we can also draw a connection between the ULCM and RLCM where $\Delta_j = (0, 1_{M_j-1}^\top)^\top 1_P^\top$ corresponds with the ULCM setting with distinct elements in $\Theta_j(B_j)$.

3. Identifiability Issue

3.1. Model Identifiability

As introduced in the previous section, the probability distribution of latent classes is given by $\pi = (\pi_c)^\top \in [0, 1]^{2^K}$ with $\sum \pi_c = 1$. Coefficients array $B = (B_1, \ldots, B_J)$ is a three-dimensional array, where B_j is the j-th slice of B with size $M_j \times P$. Then, we denote the parameter space of (π, B) by

$$\Omega(\pi, B) = \{(\pi, B) : \pi \in \Omega(\pi), B \in \Omega(B)\}, \tag{17}$$

where $\Omega(\pi) = \{\pi \in [0, 1]^{2^K} : \sum_c \pi_c = 1\}$, and $\Omega(B)$ represents the parameter space of the coefficients array B, which could be the whole real space $\mathbb{R}^{J \times P \times \sum_j M_j}$, or a subset of $\mathbb{R}^{J \times P \times \sum_j M_j}$ if constrained by the Δ.

Definition 2. (Strict Identifiability) The parameters $(\boldsymbol{\pi}, \boldsymbol{B}) \in \Omega(\boldsymbol{\pi}, \boldsymbol{B})$ are identifiable if

$$P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\pi}, \boldsymbol{B}) = P(\boldsymbol{Y} = \boldsymbol{y} \mid \bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}}) \Longleftrightarrow (\boldsymbol{\pi}, \boldsymbol{B}) \sim (\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}}),$$

where $(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}})$ is another value from the parameter space $\Omega(\boldsymbol{\pi}, \boldsymbol{B})$ and "$\sim$" means two parameter values are equivalent up to label switching of attributes.

3.2. Generic Identifiability

Generic identifiability, which is a weaker notion of identifiability than Definition 2, was first introduced in Allman et al. (2009). Generic identifiability allows the existence of some exceptional values of parameters for which strict identifiability does not hold, as long as all non-identifiable parameters form a Lebesgue measure zero set within the parameter space. Given that non-identifiable parameters exist in a set of measure zero, one is unlikely to face identifiability problems in performing inference. Thus, generic identifiability is generally sufficient for data analysis purposes.

However, the generic identifiability condition shown in Allman et al. (2009) cannot be applied in this paper. Under the setting of Allman et al. (2009), the parameter space $\Omega(\boldsymbol{B})$ is the whole real space $\mathbb{R}^{J \times P \times \sum_j M_j}$, whereas the parameter space $\Omega(\boldsymbol{B})$ in our RLCM is restricted by the structure of $\boldsymbol{\Delta}$. The dimension of $\Omega(\boldsymbol{B})$ might vary with different $\boldsymbol{\Delta}$ arrays, i.e., the parameter space of $\boldsymbol{B}$ restricted by $\boldsymbol{\Delta}$ might be a measure zero subspace of another parameter space of $\boldsymbol{B}$ restricted by $\tilde{\boldsymbol{\Delta}}$. So, it is important to discuss the generic identifiability issue within a parameter space with a fixed $\boldsymbol{\Delta}$.

Therefore, in order to discuss generic identifiability for our RLCM, we need to define the parameter space $\Omega(\boldsymbol{B})$ by taking into account the sparsity structure due to the $\boldsymbol{\Delta}$ array. Similar to Definition 17, we denote the model parameter space with a given $\boldsymbol{\Delta}$ by

$$\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B}) = \{(\boldsymbol{\pi}, \boldsymbol{B}) : \boldsymbol{\pi} \in \Omega(\boldsymbol{\pi}), \boldsymbol{B} \in \Omega_{\boldsymbol{\Delta}}(\boldsymbol{B})\}. \tag{18}$$

Coefficients in $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{B})$ are active when corresponding elements in $\boldsymbol{\Delta}$ are equal to 1, so the parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{B})$ would be $\mathbb{R}^{|\boldsymbol{\Delta}|}$, where $|\boldsymbol{\Delta}|$ is the total sum of entries of $\boldsymbol{\Delta}$. For generic identifiability, it suffices to consider the parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B})$ with a given sparsity structure $\boldsymbol{\Delta}$.

Let $S_{\boldsymbol{\Delta}}$ denote the set of non-identifiable parameters from $\Omega(\boldsymbol{\pi}, \boldsymbol{B})$:

$$\begin{aligned} S_{\boldsymbol{\Delta}} = \{(\boldsymbol{\pi}, \boldsymbol{B}) : P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\pi}, \boldsymbol{B}) = P(\boldsymbol{Y} = \boldsymbol{y} \mid \bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}}), \\ (\boldsymbol{\pi}, \boldsymbol{B}) \nsim (\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}}), \ (\boldsymbol{\pi}, \boldsymbol{B}) \in \Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B}), \ (\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}}) \in \Omega_{\bar{\boldsymbol{\Delta}}}(\boldsymbol{\pi}, \boldsymbol{B})\}. \end{aligned} \tag{19}$$

Remark 4. The non-identifiable parameters $(\boldsymbol{\pi}, \boldsymbol{B}) \in S_{\boldsymbol{\Delta}}$ could be due to some other parameters $(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}})$ with a different sparsity structure $\bar{\boldsymbol{\Delta}}$.

If the non-identifiable parameter set $S_{\boldsymbol{\Delta}}$ is of measure zero within parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B})$, then we say $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B})$ is a generically identifiable parameter space.

Definition 3. (Generic Identifiability) The parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B})$ is generically identifiable, if the Lebesgue measure of $S_{\boldsymbol{\Delta}}$ with respect to parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B})$ is zero.

3.3. Identifiability Conditions

In this section, we propose strict and generic identifiability conditions for our RLCM. We start with introducing the form of $\boldsymbol{\Delta}$ as follows.

The sparse 3-dimensional array $\boldsymbol{\Delta}$ takes the form

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}^1 \\ \boldsymbol{\Delta}^2 \\ \boldsymbol{\Delta}' \end{pmatrix}$$

after a permutation of items, where $\boldsymbol{\Delta}^1$ and $\boldsymbol{\Delta}^2$ contain K slices of $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}'$ contains the rest of $J - 2K$ slices. We use $\boldsymbol{\Delta}_j^i$ to denote the j-th slice of $\boldsymbol{\Delta}^i$ for item j.

Theorem 1. *(Strict Identifiability) The parameter space $\Omega(\boldsymbol{\pi}, \boldsymbol{B})$ is strictly identifiable if the following two conditions are satisfied:*

(A1) For $j = 1, \ldots, K$, Δ_j^1 and Δ_j^2 satisfy simple structure shown in Definition 1 and Remark 2;

(A2) For any two classes of subjects, there exists at least one item in $\boldsymbol{\Delta}'$ such that they have different positive response probabilities for some response option.

Remark 5. The $\boldsymbol{\Delta}_j$ shown in Example 3 satisfies the structure in $A1$.

Theorem 2. *(Generic Identifiability) The parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \boldsymbol{B})$ is generically identifiable if the following two conditions are satisfied:*

(B1) For $j = 1, \ldots, K$, Δ_j^1 and Δ_j^2 satisfy the following structure:

$$\boldsymbol{\Delta}_j = \begin{bmatrix} 0 \, 0 \cdots 0 & 0 & 0 \cdots 0 \\ * \, * \cdots * & \delta_{jj1} & * \cdots * \\ \vdots \, \vdots \quad \vdots & \vdots & \vdots \quad \vdots \\ * \, * \cdots * \, \delta_{jj,M_j-1} \, * \cdots * \end{bmatrix}_{M_j \times P}, \tag{20}$$

and $\sum_{m=1}^{M_j-1} \delta_{jjm} \geq 1$, where $*$ can be either 0 or 1, P generally equals 2^K.

(B2) $\boldsymbol{\Delta}'$ satisfies the condition that for every $k = 1, \ldots, K$ there exists a $j > 2K$, such that $\sum_{m=1}^{M_j-1} \delta_{jkm} \geq 1$.

Remark 6. Condition (B2) requires that there is a least one item in the last $J - 2K$ items where attribute k loads onto the main-effect for at least one response option.

4. Bayesian Formulation for the Nominal RLCM

Following the same setting in previous sections, consider a RLCM with N subjects, J items with M_j $(j = 1, \ldots, J)$ unordered response options for each item j, and K skills. We use subscript $i = 1, \ldots, N$ to index subjects, $j = 1, \ldots, J$ to index items, $m = 0, \ldots, M_j - 1$ to index options of each item, and $c = 0, \ldots, 2^K - 1$ to index latent classes. Let $\boldsymbol{\alpha}_i$ denote the

attribute profile of subject i, and Y_{ij} denote the response of subject i to item j. The likelihood of observing a sample of N responses to J items is

$$p\left(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\pi}\right) = \prod_{i=1}^{N} \sum_{\boldsymbol{\alpha}_c \in \{0,1\}^K} \pi_c \prod_{j=1}^{J} \prod_{m=0}^{M_j-1} \left(\frac{\exp\left(\boldsymbol{a}_i^\top \boldsymbol{\beta}_{jm}\right)}{\sum_{m'=0}^{M_j-1} \exp\left(\boldsymbol{a}_i^\top \boldsymbol{\beta}_{jm'}\right)} \right)^{\mathbb{1}(y_{ij}=m)}. \quad (21)$$

The posterior distribution of all parameters for the nominal RLCM is given by

$$p(\boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{\Delta}, \gamma, \sigma_\beta^2, \boldsymbol{\pi} | \boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{B}) p(\boldsymbol{\alpha}|\boldsymbol{\pi}) p(\boldsymbol{\pi}) p(\boldsymbol{B}|\boldsymbol{\Delta}, \sigma_\beta^2) p(\sigma_\beta^2) p(\boldsymbol{\Delta}|\gamma) p(\gamma). \quad (22)$$

Then, we formulate the RLCM Bayesian model as follows.

We outline our Bayesian model and priors. Specifically, we use a categorical likelihood conditioned upon attributes and item parameters,

$$Y_{ij}|\boldsymbol{\alpha}^\top \boldsymbol{v} = c, \boldsymbol{B}_j \sim \text{categorical}\left(\boldsymbol{\theta}_{jc}(\boldsymbol{B}_j)\right). \quad (23)$$

We also use a categorical prior for attributes conditioned upon the latent class probabilities,

$$\boldsymbol{\alpha}_i|\boldsymbol{\pi} \sim \text{categorical}(\boldsymbol{\pi}) \quad (24)$$

and a conjugate Dirichlet prior for the latent class probabilities, $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{d}_0)$ where $\boldsymbol{d}_0$ is a fixed constant vector.

We use a stochastic search variable selection priors for the (j, p, m) elements of $\boldsymbol{B}$ and $\boldsymbol{\Delta}$:

$$\beta_{jpm} \mid \delta_{jpm}, \sigma_\beta^2 \sim \begin{cases} N(0, \sigma_\beta^2) & \delta_{jpm} = 1 \\ N(0, \sigma_\beta^2/D) & \delta_{jpm} = 0 \end{cases}, \quad (25)$$

$$\delta_{jpm} \mid \gamma \sim Bernoulli(\gamma), \quad (26)$$

where $\boldsymbol{B} = (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_J)^\top$ satisfies the generic identifiable condition shown in Theorem 2, and the intercept is always set active with $\delta_{j0m} = 1$. Furthermore, D is a large fixed constant (e.g., we consider $D = 100, 1000$) that is used reduce the variance for the spike distribution for the case with $\delta_{jpm} = 0$. The priors for the hyper-parameters for the coefficients and activeness parameters are:

$$\sigma_\beta^2 \sim IGamma(\alpha_\sigma, \beta_\sigma), \quad (27)$$

$$\gamma \sim Beta(a, b). \quad (28)$$

Here $(\alpha_\sigma, \beta_\sigma, D, a, b, \boldsymbol{d}_0)$ are hyper-parameters.

Model parameters of the nominal RLCM are inferred through applying the Polya-gamma data augmentation approach for multinomial logistic regression (Holmes & Held, 2006; Polson et al., 2013) along with the stochastic search variable selection algorithm (George & McCulloch, 1993) to infer the latent structure. Then, the Gibbs sampling algorithm is implemented from the posterior distribution of model parameters, which is given in Appendix 7. Full sampling algorithm is represented in Algorithm 1. In order to address issues with respect to poor starting values we use a combination of k-means clustering and factor analysis to specify starting values (see the description in Appendix 7).

Algorithm 1

1: Initialize with an identifiable $\mathbf{\Delta}^{(0)}$ array, coefficients array $\boldsymbol{B}^{(0)}$, attribute profiles $\boldsymbol{\alpha}^{(0)}$, attribute categorical probabilities $\boldsymbol{\pi}^{(0)}$, and hyper-parameters $\sigma_\beta^{2(0)}$ and $\omega^{(0)}$.

2: **for** t in $1:T$ **do**

3:      **for** i in $1:N$ **do**

4:         Update $\boldsymbol{\alpha}_i^{(t)}$ to class c sequentially with weight proportional to $(n_{c(i)}+1)p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}^{(t-1)})$.

5:      **end for**

6:      Update $\boldsymbol{\pi}^{(t)} \mid \boldsymbol{\alpha}^{(t)} \sim Dirichlet(\boldsymbol{n}+\boldsymbol{d}_0)$, where $\boldsymbol{n} = (\boldsymbol{n}_0,\ldots,\boldsymbol{n}_{2^K-1}) = (\sum_{i=1}^N \mathcal{I}(\boldsymbol{\alpha}_i^{(t)\top}\boldsymbol{v} = 1),\ldots,\sum_{i=1}^N \mathcal{I}(\boldsymbol{\alpha}_i^{(t)\top}\boldsymbol{v} = 2^K - 1))$.

7:      **for all** j in $1:J$, m in $1:M_j - 1$ **do**

8:         **for** c in $0:2^K - 1$ **do**

9:            Update $w_{jcm}|A, \boldsymbol{\beta}_{jm}^{(t-1)}, \boldsymbol{B}_{j,1:m-1}^{(t)}, \boldsymbol{B}_{j,m+1:M_j-1}^{(t-1)} \sim \mathrm{PG}(n_c, \eta_{jcm})$ (details about n_c and η_{jcm} are given in Appendix 7).

10:         **end for**

11:         **for** p in $0:P-1$ **do**

12:            Update $\delta_{jpm}^{(t)} \mid \beta_{jpm}^{(t-1)}, \gamma^{(t-1)} \sim \mathrm{Bernoulli}\left(\tilde{\gamma}_{jpm}\right)$ for $p > 0$ and set $\delta_{jpm}^{(t)} = 1$ for $p = 0$ ($\tilde{\gamma}_{jpm}$ can be found in Equation B9).

13:            Update $\beta_{jmp}^{(t)}|\boldsymbol{Y}_{1:n,j}, \delta_{jmp}^{(t)}, A, \boldsymbol{\beta}_{j,m,0:p-1}^{(t)}, \boldsymbol{\beta}_{j,m,p+1:P-1}^{(t-1)}, \boldsymbol{B}_{j,1:m-1}^{(t)}, \boldsymbol{B}_{j,m+1:M_j-1}^{(t-1)}, \boldsymbol{\omega}_{jm} \sim \mathcal{N}(\mu_{jmp}, \sigma_{jmp}^2)$ (derivation of the full distribution of β_{jmp} is given in Appendix 7).

14:         **end for**

15:      **end for**

16:      Update $\sigma_\beta^{2(t)} \mid \boldsymbol{B}^{(t)}, \mathbf{\Delta}^{(t)} \sim IGamma(\alpha_\sigma^*, \beta_\sigma^*)$, where $\alpha_\sigma^* = \alpha_\sigma + \frac{1}{2}\sum_{j=1}^J P(M_j - 1)$, $\beta_\sigma^* = \beta_\sigma + \frac{1}{2}\sum_{j=1}^J \sum_{p=0}^{P-1} \sum_{m=1}^{M_j-1} \beta_{jpm}^{2(t)}(D(1 - \delta_{jpm}^{(t)}) + \delta_{jpm}^{(t)})$.

17:      Update $\gamma^{(t)}|\mathbf{\Delta}^{(t)} \sim \mathrm{Beta}\left(a + \sum_{j=1}^J \sum_{p=1}^{P-1} \sum_{m=1}^{M_j-1} \delta_{jpm}^{(t)}, b + \sum_{j=1}^J \sum_{p=1}^{P-1} \sum_{m=1}^{M_j-1} (1 - \delta_{jpm}^{(t)})\right)$.

18: **end for**

5. Monte Carlo Simulation Study

5.1. Settings

In this section, we report results from a Monte Carlo experiment to evaluate the performance of Algorithm 1. We conducted the simulation study under different number of attributes (i.e., $K = 2$ and 3), correlations among the attributes (i.e., $\rho = 0$ and 0.25), and sample size (i.e., $N = 1000, 2000, 5000$ and 10000).

For the $\rho = 0$ case, the attribute profile $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\top$ is generated uniformly from all possible 2^K cases, so the latent class membership probabilities are $\boldsymbol{\pi} = (1/2^K, \ldots, 1/2^K)^\top$. For the $\rho > 0$ case, the dependence among attribute profiles is introduced using the method of Chiu et al. (2009). Suppose $\boldsymbol{Z} = (Z_1, \ldots, Z_K)^\top$ follows a multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ with unit variance and correlation ρ, where $\boldsymbol{\Sigma} = (1 - \rho)\boldsymbol{I}_K + \rho\boldsymbol{1}_K\boldsymbol{1}_K^\top$ and $\boldsymbol{1}_K$ is a column vector of 1 with length K. Then, the attribute profile $\boldsymbol{\alpha}$ is given by $\alpha_k = \mathcal{I}(Z_k \geq \Phi^{-1}(\frac{k}{K+1})), k = 1, \ldots, K$, where Φ is the cumulative distribution function of the standard normal distribution. In this case, the data generating values for $\boldsymbol{\pi}$ are computed from integrals of the multivariate normal distribution (Chen et al., 2015; Culpepper & Balamuta, 2021) .

We assume that there are $J = 18$ items, and $M_j = 4$ unordered options for each item j. For reparameterized latent class variable $\boldsymbol{\alpha}$ shown in Eq. 3, we only include two-way interaction terms among the attributes. Our model does not explicitly contain $\boldsymbol{Q}$ matrices, therefore, we recover

the $\boldsymbol{Q}$ matrices, implied by $\boldsymbol{\Delta}_m$ for each option $m = 1, 2, 3$, using the method shown in Chen et al. (2020). Then, the true $\boldsymbol{\Delta}$ and true $\boldsymbol{Q}$ matrices for each option are shown as follows (columns in $\boldsymbol{\Delta}$ follow the same order as the design vector shown in Eq. 3):

- $\boldsymbol{\Delta}$ cube with $K = 2$

$$
\boldsymbol{\Delta}_{m=1} = \begin{pmatrix} 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \end{pmatrix}, \quad \boldsymbol{\Delta}_{m=2} = \begin{pmatrix} 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \end{pmatrix}, \quad \boldsymbol{\Delta}_{m=3} = \begin{pmatrix} 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \end{pmatrix} \tag{29}
$$

- $\boldsymbol{\Delta}$ cube with $K = 3$

$$
\boldsymbol{\Delta}_{m=1} = \begin{pmatrix} 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \end{pmatrix}, \boldsymbol{\Delta}_{m=2} = \begin{pmatrix} 1&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \end{pmatrix}, \boldsymbol{\Delta}_{m=3} = \begin{pmatrix} 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \end{pmatrix} \tag{30}
$$

- $\boldsymbol{Q}$ matrices with $K = 2$

$$
\boldsymbol{Q}_{m=1} = \begin{pmatrix} 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \end{pmatrix}, \quad \boldsymbol{Q}_{m=2} = \begin{pmatrix} 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 1&0 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 0&1 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \end{pmatrix}, \quad \boldsymbol{Q}_{m=3} = \begin{pmatrix} 1&0 \\ 0&1 \\ 0&1 \\ 1&0 \\ 1&0 \\ 1&0 \\ 1&0 \\ 0&1 \\ 0&1 \\ 0&1 \\ 1&0 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 0&1 \end{pmatrix} \tag{31}
$$

- Q matrices with $K = 3$

$$
\boldsymbol{Q}_{m=1} = \begin{pmatrix} 0\ 0\ 1 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \end{pmatrix}, \quad
\boldsymbol{Q}_{m=2} = \begin{pmatrix} 0\ 1\ 0 \\ 0\ 1\ 0 \\ 0\ 1\ 0 \\ 0\ 1\ 0 \\ 0\ 0\ 1 \\ 0\ 0\ 1 \\ 0\ 0\ 1 \\ 0\ 0\ 1 \\ 0\ 0\ 1 \\ 1\ 0\ 0 \\ 1\ 0\ 0 \\ 1\ 0\ 0 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 1\ 0\ 0 \\ 0\ 1\ 0 \end{pmatrix}, \quad
\boldsymbol{Q}_{m=3} = \begin{pmatrix} 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 1\ 0 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 0\ 0\ 1 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 0\ 1 \\ 1\ 0\ 0 \\ 0\ 1\ 0 \\ 1\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 0\ 1 \\ 1\ 0\ 0 \\ 0\ 1\ 0 \end{pmatrix}
\tag{32}
$$

Given $\boldsymbol{\Delta}$, we generate coefficients in $\boldsymbol{B}$ according to their prior distribution shown in Eq. 25 using $\sigma_\beta^2 = 1$. For hyper-parameters presented in the previous section, we have $\alpha_\sigma = \beta_\sigma = a = b = 1$, $\boldsymbol{d}_0 = \mathbf{1}_{2^K}$, and $D = 100$. We use a Markov chain of length 20,000 with a 10,000 burn-in period for $K = 2$, and a chain length of 30,000 with a 20,000 burn-in period for $K = 3$.

5.2. Results

We repeated the simulation study 100 times for each setting. For model performance, we use several metrics to evaluate parameter recovery. Specifically, we report the average element-wise accuracy rate (EAR) for Q by comparing the estimated $\hat{Q}$ and the true Q matrix, where $\hat{Q}$ is recovered by aggregating $\hat{B}$ samples after burn-in period   (Chen et al., 2020) . Note that we transform $\boldsymbol{B}$ to $\boldsymbol{\Theta}$ for every sampled value using Eq. 1 and compute the point estimate $\hat{\boldsymbol{\Theta}}$ as the mean of all sampled $\boldsymbol{\Theta}$ arrays after the burn-in period. We compute the mean absolute deviation (MAD) to assess the accuracy of the estimated latent class response probabilities $\hat{\boldsymbol{\Theta}}$, and report the proportion of attribute profiles that are correctly estimated.

It is important to mention how we address the label-switching problem for the RLCM and ULCM. Similar to latent class models, the exploratory RLCM is identified up to label-switching. However, the RLCM has fewer permutations than the ULCM. For instance, the ULCM as $(2^K)!$ possible arrangements whereas the RLCM has $K! \times 2^K$ arrangements (i.e., there are $K!$ ways to permute the order of attributes and 2^K ways of permuting the attribute levels). Note that for each replication we draw values from the posterior and then compare posterior means of our parameters (e.g., the θ's or β's) with all $K! \times 2^K$ arrangements with the data generating model parameters in order to evaluate parameter recovery. We select the permutations for the ULCM and RLCM that minimizes the difference between the estimates and data generating value. It is important to note that we do not find evidence of label-switching within chain.

Simulation results in Table 1 show a good recovery for model parameters. It suggests that for fixed K, as the sample size gets larger, the MAD of $\hat{\boldsymbol{\Theta}}$ and $\boldsymbol{\pi}$ become smaller and the EARs of $\hat{Q}$ matrices become larger. The EARs of $\hat{Q}$ matrices are higher for smaller K, which is expected given that the number of unknown model parameters that must be estimated increases with larger K. The simulation results also provide evidence that a positive correlation among attributes, represented by $\rho > 0$, results in slightly larger MADs for $\boldsymbol{\Theta}$ in some instances, and this impact is more systematic for $K = 3$. Although $\rho = 0.25$ slightly decreases recovery of $\boldsymbol{\Theta}$, $\boldsymbol{\pi}$, and $\boldsymbol{Q}$, the results in Table 1 show attribute classification accuracy improves by a few percentage points. Overall, the classification accuracy is at acceptable levels and generally exceeds 70% for most scenarios.

We also conduct Monte Carlo experiment for denser $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$ compared with those shown in Eqs. $29 - 32$, true $\boldsymbol{\Delta}$, $\boldsymbol{Q}$ matrices and simulation results are given in Appendix 7. Table 6 shows similar model parameter recovery compared with the results shown in Table 1.

5.3. Unstructured Latent Class Models (ULCMs)

We want to compare our model performance with ULCMs, which assume that there's no latent structure between latent attribute classes and observed response variables. Following the same setting as we represented for the nominal RLCM, the likelihood of observing a sample of N independent responses to J items is

$$p(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\Theta}) = \prod_{j=1}^{J} \prod_{c=0}^{2^K-1} \prod_{m=0}^{M_j-1} \theta_{jc}^{n_{jcm}}, \tag{33}$$

where $n_{jcm} = \sum_{i=1}^{n} = \mathcal{I}(y_{ij} = m)\mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)$. Then, the posterior distribution of all parameters for the nominal ULCM is given by

$$p(\boldsymbol{\alpha}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\Theta})p(\boldsymbol{\Theta})p(\boldsymbol{\alpha}|\boldsymbol{\pi})p(\boldsymbol{\pi}). \tag{34}$$

Below is the Bayesian framework for our nominal ULCM. Given attribute profile $\boldsymbol{\alpha}$ and class-response probability matrix $\boldsymbol{\Theta}$, response data follow a categorical distribution

$$Y_{ij}|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{\theta}_{jc} \sim \text{Categorical}(\boldsymbol{\theta}_{jc}). \tag{35}$$

We use a Dirichlet prior for the class-response probability vectors

$$\boldsymbol{\theta}_{jc} \sim \text{Dirichlet}(\boldsymbol{d}_{M_j}), \tag{36}$$

and a categorical prior for attributes conditioned upon the latent class probabilities,

$$\boldsymbol{\alpha}_i|\boldsymbol{\pi} \sim \text{Categorical}(\boldsymbol{\pi}) \tag{37}$$

with a conjugate Dirichlet prior for the latent class probabilities $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{d}_0)$, where $\boldsymbol{d}_M$ and $\boldsymbol{d}_0$ are constant vectors.

We applied Gibbs sampling algorithm to estimate model parameters via their posterior distributions.

- $\boldsymbol{\theta}_{jc} \mid \boldsymbol{y}_{1:n}, \boldsymbol{\alpha}_{1:n} \sim \text{Dirichlet}(\boldsymbol{n}_{jc} + \boldsymbol{d}_{M_j})$

$$p(\boldsymbol{\theta}_{jc} \mid \boldsymbol{y}_{1:n}, \boldsymbol{\alpha}_{1:n}) \propto p(\boldsymbol{y}_{Ij} \mid \boldsymbol{\alpha}_I, \boldsymbol{\theta}_{jc})p(\boldsymbol{\theta}_{jc})$$
$$\propto \prod_{m=0}^{M_j} \theta_{jcm}^{n_{jcm}} \cdot \prod_{m=0}^{M_j} \theta_{jcm}^{1-1}, \tag{38}$$

where $\boldsymbol{n}_{jc} = (n_{jc0}, \ldots, n_{jc,M_j-1})^\top$ and $I = \left\{i : \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c\right\}$.

- $\boldsymbol{\pi}|\boldsymbol{\alpha} \sim Dirichlet(\boldsymbol{n} + \boldsymbol{d}_0)$

$$p(\boldsymbol{\pi} \mid \boldsymbol{\alpha}_{1:n}) \propto p(\boldsymbol{\alpha}_{1:n} \mid \boldsymbol{\pi})p(\boldsymbol{\pi}) \sim \text{Dirichlet}(\boldsymbol{n} + \boldsymbol{d}_0), \tag{39}$$

where $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{d}_0)$ and $\boldsymbol{n} = (n_0, \cdots, n_{2^K-1})^\top$ represents the frequencies of each attribute pattern $\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, c = 0, \ldots, 2^K - 1$.

- $\boldsymbol{\alpha}_i \mid \boldsymbol{\alpha}_{(i)}, \boldsymbol{y}_{1:n}$ We update $\boldsymbol{\alpha}$ while integrating $\boldsymbol{\pi}$ out

$$\begin{aligned}
p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) &= \int p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N \mid \boldsymbol{\pi})p(\boldsymbol{\pi})d\boldsymbol{\pi} \\
&= \frac{1}{B(\boldsymbol{d}_0)} \int \left(\prod_{c=0}^{2^K-1} \pi_c^{n_c + d_{0,c} - 1} \right) d\boldsymbol{\pi} \\
&= \frac{B(\boldsymbol{n} + \boldsymbol{d}_0)}{B(\boldsymbol{d}_0)}.
\end{aligned} \tag{40}$$

Then, the full conditional distribution for $\boldsymbol{\alpha}_i$ is

$$\begin{aligned}
p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c \mid \boldsymbol{\alpha}_{(i)}) &= \frac{p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)}{p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_N)} \\
&= \frac{n_{c(i)} + 1}{n - 1 + 2^K},
\end{aligned} \tag{41}$$

where $n_{c(i)}$ represents the number of individuals other than i that have attribute profile $\boldsymbol{\alpha}_c$. The full conditional distribution of $\boldsymbol{\alpha}_i$ given $\boldsymbol{y}_{1:n}$ and $\boldsymbol{\alpha}_{(i)}$ is

$$\begin{aligned}
p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c \mid \boldsymbol{\alpha}_{(i)}, \boldsymbol{y}_{1:n}) &\propto p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c \mid \boldsymbol{\alpha}_{(i)})p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{\Theta}) \\
&\propto (n_{c(i)} + 1)p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{\Theta}),
\end{aligned} \tag{42}$$

we update $\boldsymbol{\alpha}_i$ sequentially with weight proportional to $(n_{c(i)} + 1)p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{\Theta})$.

The full Gibbs sampling steps of all parameters are shown in Algorithm 2.

In order to compare the model performance of the ULCM with the RLCM, we generate response data from the RLCM model, and then use both Algorithm 1 and 2 to estimate model parameters in two models. We use same simulation settings as in our RLCM. For hyper-parameters in ULCM, we use $\boldsymbol{d}_{M_j} = \boldsymbol{1}_{M_j}$ for $j = 1, \ldots, J$. Simulation results are shown in Table 1 and provide evidence the RLCM has better parameter recovery. Table 2 reports additional details regarding MADs for $\boldsymbol{\Theta}$ at the item level for the RLCM and ULCM. The results show that the aggregate findings in Table 1 are consistent with item-level performance such that the RLCM has smaller MADs than the ULCM.

Results shown in Tables 1 and 2 indicate that for response data generated via RLCM, Algorithm 1 performs uniformly better than Algorithm 2, which implies that if there is structure in the latent relationship between attributes and observed variables, our RLCM can achieve better recovery for model parameters compared with the ULCM.

TABLE 1.
Summary of simulation performance for RLCM and ULCM.

ρ	K	N	RLCM				ULCM		
			MAD		EAR	Accuracy	MAD		Accuracy
			$\hat{\Theta}$	$\hat{\pi}$	$\hat{Q}$	$\hat{\alpha}$	$\hat{\Theta}$	$\hat{\pi}$	$\hat{\alpha}$
0	2	1000	0.0207	0.0229	0.8660	0.7740	0.0641	0.1107	0.6217
	2	2000	0.0129	0.0140	0.9354	0.7870	0.0369	0.0631	0.7267
	2	5000	0.0116	0.0088	0.9638	0.7919	0.0161	0.0242	0.7688
	2	10000	0.0052	0.0065	0.9949	0.7938	0.0089	0.0122	0.7869
	3	1000	0.0231	0.0175	0.9060	0.6565	0.1256	0.1225	0.3992
	3	2000	0.0214	0.0136	0.9117	0.6677	0.0868	0.0790	0.4948
	3	5000	0.0150	0.0086	0.9262	0.6815	0.0499	0.0441	0.5914
	3	10000	0.0136	0.0064	0.9266	0.6881	0.0313	0.0266	0.6274
0.25	2	1000	0.0203	0.0231	0.8411	0.8030	0.0683	0.0556	0.7554
	2	2000	0.0137	0.0151	0.9192	0.8168	0.0577	0.0418	0.7814
	2	5000	0.0080	0.0079	0.9777	0.8183	0.0238	0.0200	0.8027
	2	10000	0.0056	0.0056	0.9931	0.8208	0.0125	0.0112	0.8144
	3	1000	0.0349	0.0381	0.8620	0.6580	0.1379	0.0756	0.5819
	3	2000	0.0286	0.0290	0.8674	0.6935	0.1101	0.0424	0.6583
	3	5000	0.0225	0.0223	0.8787	0.7208	0.0756	0.0190	0.7125
	3	10000	0.0200	0.0208	0.8810	0.7300	0.0648	0.0128	0.7173

MAD = mean absolute deviation; EAR = element-wise accuracy rate; Accuracy = proportion of attribute profiles that are correctly estimated

TABLE 2.
Summary of mean absolute deviations (MADs) of RLCM and ULCM item response probabilities by item for two selected conditions.

J	$N = 2000, \rho = 0.25$		$N = 5000, \rho = 0$	
	RLCM	ULCM	RLCM	ULCM
1	0.0079	0.1126	0.0023	0.0397
2	0.0150	0.0844	0.0045	0.0281
3	0.0060	0.0866	0.0029	0.0302
4	0.0087	0.0780	0.0026	0.0250
5	0.0068	0.1075	0.0040	0.0369
6	0.0213	0.0833	0.0081	0.0372
7	0.0222	0.0993	0.0044	0.0320
8	0.0363	0.1422	0.0078	0.0506
9	0.0152	0.1385	0.0030	0.0503
10	0.0352	0.1268	0.0062	0.0430
11	0.0153	0.1220	0.0042	0.0453
12	0.0138	0.0971	0.0051	0.0357
13	0.0130	0.1222	0.0046	0.0459
14	0.0197	0.0886	0.0054	0.0325
15	0.0192	0.0906	0.0077	0.0348
16	0.0108	0.1074	0.0044	0.0363
17	0.0100	0.1046	0.0050	0.0385
18	0.0090	0.0872	0.0028	0.0334

Algorithm 2

1: Initialize attribute profiles $\boldsymbol{\alpha}^{(0)}$, attribute categorical probabilities $\boldsymbol{\pi}^{(0)}$, and class-response probability matrix $\boldsymbol{\Theta}^{(0)}$.
2: **for** t in $1 : T$ **do**
3:    **for** i in $1 : N$ **do**
4:       Update $\boldsymbol{\alpha}_i^{(t)}$ to class c sequentially with weight proportional to $(n_{c(i)} + 1) p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^{\top} \boldsymbol{v} = c, \boldsymbol{\Theta}^{(t-1)})$.
5:    **end for**
6:    Update $\boldsymbol{\pi}^{(t)} \mid \boldsymbol{\alpha}^{(t)} \sim Dirichlet(\boldsymbol{n} + \boldsymbol{d}_0)$, where $\boldsymbol{n} = (\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{2^K-1}) = (\sum_{i=1}^{N} \mathcal{I}(\boldsymbol{\alpha}_i^{(t)\top} \boldsymbol{v} = 1), \ldots, \sum_{i=1}^{N} \mathcal{I}(\boldsymbol{\alpha}_i^{(t)\top} \boldsymbol{v} = 2^K - 1))$.
7:    **for all** j in $1 : J, c$ in $0 : 2^K - 1$ **do**
8:       Update $\boldsymbol{\theta}_{jc}^{(t)} \mid \boldsymbol{y}_{1:n}, \boldsymbol{\alpha}_{1:n}^{(t)} \sim \text{Dirichlet}(\boldsymbol{n}_{jc} + \boldsymbol{d}_{M_j})$.
9:    **end for**
10: **end for**

---

6. Applications

6.1. Wagner Preference Inventory

In this section, we apply Algorithm 1 to the dataset in the Wagner Preference Inventory (WAPI II) (Wagner & Wells, 1985) . This data set contains nominal responses to $J = 12$ items, each of which contains $M = 4$ choices. All 13, 502 participants completed the 12 questions, so we have $N = 13, 502$. Table 3 presents the twelve items along with the marginal probability of selecting each response option. The twelve items were originally designed to distinguish preferences along the notion of activities that vary in Left vs. Right brain and logical vs. creative. The proposed two-by-two design included (*a*) Left, logical; (*b*) left, verbal; (*c*) right, manipulative/spatial; and (*d*) right, creative. A separate measure for left and right preference can be obtained by adding (a) and (b) and (c) and (d), respectively. In order to be consistent with   (Wagner & Wells, 1985) , we let $K = 2$ to represent the left-right brain dominance dichotomy. We ran five Markov chains with $K = 2$ for convergence diagnostics of the Markov chain.

Figure 1 shows the plot of maximum proportional scale reduction factor (PSRF) (Brooks & Gelman, 1998) for checking the convergence of Markov chain with multivariate parameters. The approximate convergence is achieved after 5, 000 iterations since the maximum PSRF remains below 1.1 after that. So we ran 100 Markov chains of length 20, 000 (with 10, 000 as burn-in) estimate the parameters and the results are shown in Table 4.

Table 4 implies that participants with attribute profiles $\boldsymbol{\alpha}_i = (0, 1)^{\top}, (1, 1)^{\top}, (0, 0)^{\top}$ and $(1, 0)^{\top}$ prefer option a, b, c and d, respectively. For instance, the choices for item 1 were "a. major in logic", "b. write a letter", "c. fix things at home", and "d. major in art". The estimates for $\boldsymbol{\Theta}$ in Table 4 indicate that members of class 01 were most likely to choose option "a" with an estimated response probability of 0.725. In contrast, members of class 00 had a 0.643 chance of selecting option "c" and respondents in the 10 class chose "d" with a probability equal to 0.693.

We also estimated the latent class probabilities of attribute profiles. Specifically, the proportions of each attribute profile pattern in an increasing order of the bijection $\boldsymbol{\alpha}_c^{\top} \boldsymbol{v}$ are shown in Table 5. The latent classes are nearly equal in size with the most respondents of 0.290 classified with the 11 profile (i.e., Wagner's left-verbal group) and the 00 class having 0.204 proportion of respondents (i.e., Wagner's right-manipulative/spatial).

Also, results shown in Table 4 can be used to evaluate the intended choice design for the items. Most items differentiated among one or two of the underlying latent classes. However, some options did not differentiate the latent classes as intended, such as item 2, 7 and 10. For

TABLE 3.
Wagner preference inventory items, anchors, and response frequencies.

Item	Anchors	%	Item	Anchors	%
1.	a. Major in logic	37	7.	a. Be in charge of computer programming	17
	b. Write a letter	15		b. Study word origins and meaning	41
	c. Fix things at home	23		c. Putter in the yard	15
	d. Major in art	25		d. Invent a new gadget	27
2.	a. Be a movie critic	27	8.	a. Analyze production cost	22
	b. Learn new words	29		b. Describe a new product in words	28
	c. Improve your skills in a game	32		c. Sell a new product on the market	23
	d. Create a new toy	12		d. Draw a picture of a new product	27
3.	a. Improve your strategy in a game	29	9.	a. Explain the logic of a theory	43
	b. Remember people's names	15		b. Be a copy writer for ads	5
	c. Engage in sports	19		c. Work with wood and clay	18
	d. Play an instrument by ear	37		d. Invent a story	34
4.	a. Review a book	27	10.	a. Be a comparison shopper	20
	b. Write for a magazine	27		b. Read about famous men and women	37
	c. Build new shelves at home	21		c. Run a traffic control tower	14
	d. Draw a landscape or seascape	25		d. Mold with clay and putty	29
5.	a. Analyze market trends	23	11.	a. Analyze your budget	23
	b. Remember people's names	33		b. Study literature	25
	c. Do carpentry work	21		c. Visualize and re-arrange furniture	25
	d. Imagine a new play	23		d. Be an artist	27
6.	a. Analyze management practices	27	12.	a. Plan a trip and make a budget	37
	b. Locate words in a dictionary	15		b. Write a novel	30
	c. Put jigsaw puzzles together	31		c. Build a house or shack	18
	d. Paint in oil	27		d. Make crafts your hobby	15

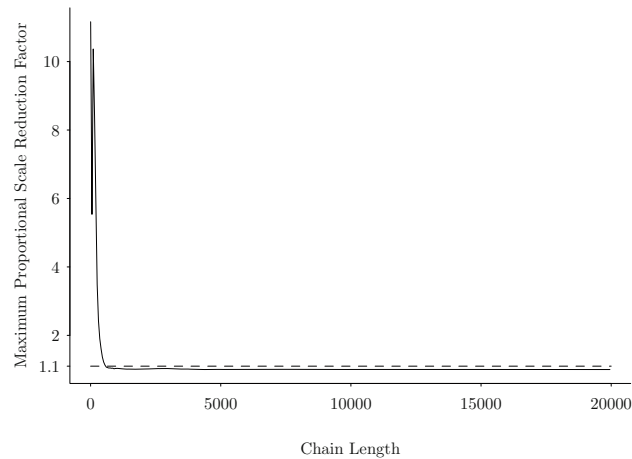a = left/logical; b = left/verbal; c = right/manipulative-spatial; d = right/creative

FIGURE 1.
The maximum PSRF for Wagner Preference Inventory data.

item 2, Wagner originally specified option a as left-logical function and option d as right/creative function. However, according to probabilities represented in Table 4, people with attribute profile $\boldsymbol{\alpha}_i = (0, 1)^\top$ did not strongly prefer option a, and people with attribute profile $\boldsymbol{\alpha}_i = (1, 0)^\top$ did not strongly prefer option d. The choice design for item 2 should be reconsidered.

TABLE 4.
Estimated $\hat{\Theta}$ for Wagner Preference Inventory data.

J	a. Left/Logical Attribute profile patterns				b. Left/Verbal Attribute profile patterns			
	00	01	10	11	00	01	10	11
1	0.214	0.725	0.149	0.327	0.063	0.056	0.083	0.337
2	0.170	0.197	0.287	0.397	0.223	0.260	0.268	0.381
3	0.227	0.491	0.195	0.233	0.170	0.113	0.106	0.197
4	0.095	0.376	0.126	0.399	0.078	0.245	0.220	0.475
5	0.140	0.611	0.060	0.089	0.121	0.156	0.407	0.568
6	0.199	0.613	0.047	0.185	0.125	0.125	0.067	0.265
7	0.129	0.335	0.118	0.092	0.209	0.332	0.416	0.626
8	0.204	0.513	0.041	0.094	0.242	0.166	0.157	0.523
9	0.350	0.805	0.229	0.314	0.041	0.040	0.028	0.102
10	0.220	0.280	0.105	0.185	0.245	0.382	0.244	0.543
11	0.246	0.561	0.020	0.086	0.092	0.206	0.104	0.527
12	0.347	0.617	0.202	0.303	0.024	0.170	0.347	0.586
	c. Right/Manipulative-Spatial				d. Right/Creative			
1	0.643	0.174	0.075	0.122	0.080	0.045	0.693	0.214
2	0.424	0.465	0.258	0.176	0.183	0.078	0.187	0.046
3	0.298	0.206	0.131	0.142	0.305	0.190	0.568	0.428
4	0.614	0.223	0.076	0.039	0.213	0.156	0.578	0.087
5	0.594	0.103	0.182	0.057	0.145	0.130	0.351	0.286
6	0.489	0.224	0.240	0.312	0.187	0.038	0.646	0.238
7	0.325	0.068	0.144	0.093	0.337	0.265	0.322	0.189
8	0.262	0.274	0.152	0.233	0.292	0.047	0.650	0.150
9	0.456	0.023	0.283	0.037	0.153	0.132	0.460	0.547
10	0.191	0.257	0.043	0.080	0.344	0.081	0.608	0.192
11	0.519	0.164	0.146	0.207	0.143	0.069	0.730	0.180
12	0.441	0.156	0.149	0.041	0.188	0.057	0.302	0.070

Note: The estimated probabilities in $\hat{\Theta}$ are averaged over 100 repetitions.

TABLE 5.
Estimated distribution of attributes in Wagner Preference Inventory data.

c	$\boldsymbol{\alpha}_c$	$\hat{\pi}$	Wagner Latent Class Interpretation
0	00	0.204	Right-manipulative/spatial
1	01	0.266	Left-logical
2	10	0.240	Right-creative
3	11	0.290	Left-verbal

7. Discussion

This paper focuses on the identifiability conditions of RLCMs. We proposed strict and generic identifiability conditions based on the unique condition of tensor decomposition shown in Kruskal theorem for the uniqueness of three-way arrays (Kruskal, 1976, 1977) . The established identifiability conditions are applicable to a wealth of models for binary (e.g., Chen et al., 2015; 2020, de

la Torre, 2011), polytomous (e.g., Chen & de la Torre, 2013, Culpepper, 2019, Culpepper & Balamuta, 2021), and nominal response data. Accordingly, the new identifiability results can guide researchers on the design of diagnostic interventions. Then, we developed a Bayesian formulation for the RLCMs where the generic identifiability conditions are taken into consideration. For our simulation study, we apply the Polya-gamma data augmentation for updating coefficients, and compared our algorithm results with ULCMs. Simulation results show that our algorithm can efficiently estimate model parameters, especially when the number of attribute profiles are small. Given latent structures, our model has better performance compared with ULRMs.

In this paper, we assume that the number of attribute profiles, K, is fixed and pre-specified. However, the prior knowledge for K may not be available in practice. Further study may consider K as an unknown parameter that needs to be estimated (e.g., see Chen et al., 2021). Unknown K implies that the dimension of attribute profiles, category response probability array and coefficients array become unavailable, which can be quite challenging for future research.

Acknowledgments

Appendix A Proof of Theorems

Preliminary Results

We start the proof with introducing some notation.

Definition 4. For a matrix M, the Kruskal rank of M is the largest number I such that every set of I columns in M are linearly independent.

Remark 7. Compared with the rank of a matrix M, we have $rank_K(M) \leq rank(M)$. If M has full column rank, then the equality holds.

Consider a tripartition of the set $\mathbb{J} = \{1, 2, \ldots, J\}$ into three disjoint, non-empty subsets $\mathbb{J}_1 = \{1, 2, \ldots, K\}, \mathbb{J}_2 = \{K+1, \ldots, 2K\}$ and $\mathbb{J}_3 = \{2K+1, \ldots, J\}$. Then, the marginal distribution of response Y can be represented as a three-way array T decomposing Y into three parts:

$$
\begin{aligned}
T_{(y^{\mathbb{J}_1}, y^{\mathbb{J}_2}, y^{\mathbb{J}_3})} &= P(Y^{\mathbb{J}_1} = y^{\mathbb{J}_1}, Y^{\mathbb{J}_2} = y^{\mathbb{J}_2}, Y^{\mathbb{J}_3} = y^{\mathbb{J}_3} \mid \pi, B) \\
&= \sum_{\alpha} \pi_{\alpha} P(Y^{\mathbb{J}_1} = y^{\mathbb{J}_1}, Y^{\mathbb{J}_2} = y^{\mathbb{J}_2}, Y^{\mathbb{J}_3} = y^{\mathbb{J}_3} \mid B, \alpha) \\
&= \sum_{\alpha} \pi_{\alpha} P(Y^{\mathbb{J}_1} = y^{\mathbb{J}_1} \mid B, \alpha) P(Y^{\mathbb{J}_2} = y^{\mathbb{J}_2} \mid B, \alpha) P(Y^{\mathbb{J}_3} = y^{\mathbb{J}_3} \mid B, \alpha).
\end{aligned}
\tag{A1}
$$

Let T_1, T_2, T_3 represent distributions of $Y^{\mathbb{J}_1}$, $Y^{\mathbb{J}_2}$, $Y^{\mathbb{J}_3}$ given values of attribute profile $\boldsymbol{\alpha}$. Then, the identifiability is equivalent to the uniqueness of the decomposition of the following tensor (Kruskal, 1977)

$$T = \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} T_{1,\boldsymbol{\alpha}} \otimes T_{2,\boldsymbol{\alpha}} \otimes T_{3,\boldsymbol{\alpha}} = \sum_{\boldsymbol{\alpha}} \tilde{T}_{1,\boldsymbol{\alpha}} \otimes T_{2,\boldsymbol{\alpha}} \otimes T_{3,\boldsymbol{\alpha}}, \tag{A2}$$

where $T_{1,\boldsymbol{\alpha}}$, $T_{2,\boldsymbol{\alpha}}$, $T_{3,\boldsymbol{\alpha}}$ are the $\boldsymbol{\alpha}$-th column of T_1, T_2, T_3, and $\tilde{T}_{1,\boldsymbol{\alpha}} = \pi_{\boldsymbol{\alpha}} T_{1,\boldsymbol{\alpha}}$.
We apply the following theorem shown in Kruskal (1977) for our proof.

Theorem 3. *(Kruskal, 1977) If*

$$rank_K(\tilde{T}_1) + rank_K(T_2) + rank_K(T_3) \geq 2 \cdot 2^K + 2, \tag{A3}$$

then the tensor decomposition of T is unique up to simultaneous permutation and rescaling of the columns.

We have $rank_K(\tilde{T}_1) = rank_K(T_1)$ since $\boldsymbol{\pi}$ has positive entries. Moreover, T_1, T_2 and T_3 are all stochastic matrices with column sum 1, so the decomposition of the tensor T is unique up to permutations of columns if (A3) in Theorem 3 is satisfied, which implies model identifiability. If items have M response levels, we will be able to construct T_1, T_2 and T_3 such that T_1, T_2 are $M^K \times 2^K$, and T_3 is $M^{J-2K} \times 2^K$. So for the first set of K items, the probability matrix T_1 is

$$T_1 = \left[\bigotimes_{j=1}^{K} \boldsymbol{\theta}_{j0}, \ldots, \bigotimes_{j=1}^{K} \boldsymbol{\theta}_{j,2^K-1} \right].$$

We multiply T_1 by a collapsing matrix A that makes $\underset{2^K \times M^K}{A} \underset{M^K \times 2^K}{T_1}$ a square matrix of size $2^K \times 2^K$.

Definition 5. Matrix A_j, with size $2 \times M$, is a collapsing matrix for item j. For the response probability vector θ_{jc}, $A_j \theta_{jc}$ only gives probabilities of selecting a given response level or not in class c.

Following is an example of matrix A_j.

Example 5. Consider $M_j = 3$ and we are collapsing on option 1. Let

$$A_j = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

and the response probability vector $\theta_{jc} = (\theta_{jc,0}, \theta_{jc,1}, \theta_{jc,2})$ for item j in class c. Then, we have

$$A_j \theta_{jc} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 - \theta_{jc,1} - \theta_{jc,2} \\ \theta_{jc,1} \\ \theta_{jc,2} \end{bmatrix} = \begin{bmatrix} 1 - \theta_{jc,1} \\ \theta_{jc,1} \end{bmatrix}.$$

Let $A = \bigotimes_{j=1}^{K} A_j$, where A_j is defined in Definition 5. Compared with T_1, AT_1 is a collapsed version of T_1 which only contains probabilities of selecting a given response level or not by latent class membership. So we can write AT_1 as

$$AT_1 = \bigotimes_{j=1}^{K} A_j \left[\bigotimes_{j=1}^{K} \boldsymbol{\theta}_{j0}, \ldots, \bigotimes_{j=1}^{K} \boldsymbol{\theta}_{j,2^K-1} \right] \tag{A4}$$

$$= \left[\bigotimes_{j=1}^{K} A_j \bigotimes_{j=1}^{K} \boldsymbol{\theta}_{j0}, \ldots, \bigotimes_{j=1}^{K} A_j \bigotimes_{j=1}^{K} \boldsymbol{\theta}_{j,2^K-1} \right] \tag{A5}$$

$$= \left[\bigotimes_{j=1}^{K} (A_j \boldsymbol{\theta}_{j0}), \ldots, \bigotimes_{j=1}^{K} (A_j \boldsymbol{\theta}_{j,2^K-1}) \right]. \tag{A6}$$

Proposition 1. *If Δ^1 follows a simple structure shown in Definition 1 and Remark 2, then we have*

$$AT_1 = \bigotimes_{j=1}^{K} (\boldsymbol{p}_{j0}, \boldsymbol{p}_{j1}) \tag{A7}$$

where $\boldsymbol{p}_{j0} = A_j \boldsymbol{\theta}_{j0}$ and $\boldsymbol{p}_{j1} = A_j \boldsymbol{\theta}_{j,\boldsymbol{e}_j^\top \boldsymbol{v}}$.

Proof. Simple structure implies that we can then write AT_1 as a block matrix:

$$AT_1 = \left(\boldsymbol{p}_{10} \otimes \left(\bigotimes_{j>1} A_j \boldsymbol{\theta}_{j0} \right), \ldots, \boldsymbol{p}_{10} \otimes \left(\bigotimes_{j>1} A_j \boldsymbol{\theta}_{j,2^{K-1}-1} \right), \right.$$
$$\left. \boldsymbol{p}_{11} \otimes \left(\bigotimes_{j>1} A_j \boldsymbol{\theta}_{j,2^{K-1}} \right), \ldots, \boldsymbol{p}_{11} \otimes \left(\bigotimes_{j>1} A_j \boldsymbol{\theta}_{j,2^K-1} \right) \right)$$
$$= \left(\boldsymbol{p}_{10} \otimes \boldsymbol{T}_{(1)0}, \boldsymbol{p}_{11} \otimes \boldsymbol{T}_{(1)1} \right). \tag{A8}$$

We next show that simple structure of item 1 implies that,

$$\bigotimes_{j>1} A_j \boldsymbol{\Theta}_{j0} = \bigotimes_{j>1} A_j \boldsymbol{\Theta}_{j,2^{K-1}}, \ldots, \bigotimes_{j>1} A_j \boldsymbol{\Theta}_{j,2^{K-1}-1} = \bigotimes_{j>1} A_j \boldsymbol{\Theta}_{j,2^K-1}.$$

Notice that items $j > 1$ are unrelated to attribute one. Let $\boldsymbol{\alpha}_{(1)} = (\alpha_2, \ldots, \alpha_K)^\top$ denote the response pattern on attributes two through K. Simple structure of item 1 implies that classes with $\boldsymbol{\alpha} = (0, \boldsymbol{\alpha}_{(1)})^\top$ and $\boldsymbol{\alpha} = (1, \boldsymbol{\alpha}_{(1)})^\top$ will have identical response probabilities on items $j > 1$. Stated differently, classes $(0, \boldsymbol{\alpha}_{(1)})\boldsymbol{v} = c_0$ and $(1, \boldsymbol{\alpha}_{(1)})\boldsymbol{v} = c_0 + 2^{K-1}$ have equivalent response probabilities on the remaining $j > 1$ items so that $\bigotimes_{j>1} A_j \boldsymbol{\Theta}_{j,c_0} = \bigotimes_{j>1} A_j \boldsymbol{\Theta}_{j,c_0+2^{K-1}}$ for all $c_0 \in \{0, \ldots, 2^{K-1} - 1\}$. Consequently, $\boldsymbol{T}_{(1)0} = \boldsymbol{T}_{(1)1} = \boldsymbol{T}_{(1)}$ and properties of the Kronecker product imply that

$$AT_1 = \left(\boldsymbol{p}_{10} \otimes \boldsymbol{T}_{(1)}, \boldsymbol{p}_{11} \otimes \boldsymbol{T}_{(1)} \right) = \left(\boldsymbol{p}_{10}, \boldsymbol{p}_{11} \right) \otimes \boldsymbol{T}_{(1)} \tag{A9}$$

where

$$T_{(1)} = \left(\bigotimes_{j>1} A_j \Theta_{j0}, \ldots, \bigotimes_{j>1} A_j \Theta_{j,2^{K-1}-1} \right). \tag{A10}$$

Item two is also simple structure and repeating the aforementioned steps on $T_{(1)}$ implies that

$$AT_1 = (p_{10}, p_{11}) \otimes (p_{20}, p_{21}) \otimes T_{(1:2)} \tag{A11}$$

where

$$T_{(1:2)} = \left(\bigotimes_{j>2} A_j \Theta_{j0}, \ldots, \bigotimes_{j>2} A_j \Theta_{j,2^{K-2}-1} \right). \tag{A12}$$

Consequently, simple structure for the remaining $j \in \{3, \ldots, K\}$ items implies that

$$AT_1 = \bigotimes_{j=1}^{K} (p_{j0}, p_{j1}). \tag{A13}$$

$\square$

Remark 8. Note that properties of the Kronecker product and simple structure in T_1 imply that AT_1 has rank 2^K if p_{j0} and p_{j1} are linearly independent for all j. According to Definition 1, there is at least one $\beta_{jjm} \neq 0$ for $m \in \{1, \ldots, M_j - 1\}$, which implies that $p_{j0} \neq p_{j1}$.

Proposition 2. $rank(A^i T_i) = 2^K$ *if and only if* $rank(T_i) = 2^K$, $i = 1, 2$.

Proof. By Sylvester's rank inequality (Matsaglia & Styan, 1974) , we have

$$rank(A^i) + rank(T_i) - 2^K \leq rank(A^i T_i) \leq \min\{rank(A^i), rank(T_i)\}.$$

Given $rank(A_j^i) = 2$, and the property of the rank of a Kronecker product, we have $rank(A^i) = \prod_{j=1}^{K} rank(A_j^i) = 2^K$. Then, we get

$$rank(T_i) \leq rank(A^i T_i) \leq \min\{2^K, rank(T_i)\},$$

which implies that $rank(A^i T_i) = 2^K$ if and only if $rank(T_i) = 2^K$. $\square$

Proof of Theorem 1

By Proposition 2, it suffices to show that $rank(A^1 T_1) = rank(A^2 T_2) = 2^K$ and $rank_K(T_3) \geq 2$, where $A^1 = \bigotimes_{j=1}^{K} A_j^1$ and $A^2 = \bigotimes_{j=1}^{K} A_j^2$ are collapsing matrices introduced in Definition 5. According to condition $(A1)$ in Theorem 1, both Δ^1 and Δ^2 satisfy the simple structure, then by Proposition 1 and remark 8 we have $rank(A^1 T_1) = rank(A^2 T_2) = 2^K$. For T_3, since each element is nonnegative and each column sums to 1, then given condition $(A2)$, for any two different classes c and c', there must exist one item j such that $\theta_{jcm} \neq \theta_{j'm}$, so that $rank_K(T_3) \geq 2$. By Theorem 3, the model is strictly identified.

Proof of Theorem 2

According to the tripartition of items set $\mathbb{J}$, we can decompose $\mathbf{\Delta}$ into $\mathbf{\Delta}^1$, $\mathbf{\Delta}^2$, $\mathbf{\Delta}'$ corresponding to $\mathbb{J}_1$, $\mathbb{J}_2$, $\mathbb{J}_3$, respectively. Similarly, we can decompose parameter space $\Omega(\mathbf{B})$ into three parts, $\Omega_{\mathbf{\Delta}}(\mathbf{B}) = \Omega_{\mathbf{\Delta}^1} \times \Omega_{\mathbf{\Delta}^2} \times \Omega_{\mathbf{\Delta}'}$. Therefore, to prove Theorem 2, it suffices to show that under conditions $(B1)$ and $(B2)$, $rank_K(\mathbf{T}_1) = rank_K(\mathbf{T}_2) = 2^K$ and $rank_K(\mathbf{T}_3) \geq 2$ hold almost everywhere in $\Omega_{\mathbf{\Delta}^1}$, $\Omega_{\mathbf{\Delta}^2}$, $\Omega_{\mathbf{\Delta}'}$, respectively. Then, by Theorem 3, the identifiability holds almost everywhere in $\Omega_{\mathbf{\Delta}}(\mathbf{B})$.

Based on Theorem 3 and Proposition 2, we first show that $rank(\mathbf{A}^i \mathbf{T}_i) = 2^K$ holds almost everywhere in $\Omega_{\mathbf{\Delta}^i}$, $i = 1, 2$, given $\mathbf{\Delta}$ satisfying the structure shown in Theorem 2. Let

$$f_i(\mathbf{B}) = det(\mathbf{A}^i \mathbf{T}_i) : \ \Omega_{\mathbf{\Delta}^i} \to \mathbb{R} \tag{A14}$$

denote the determinant of matrix $\mathbf{A}^i \mathbf{T}_i$, where $\mathbf{\Delta}^i$ satisfies condition $(B1)$.

Proposition 3. $f_i(\mathbf{B})$ *is a real analytic function of* $\mathbf{B}$.

Proof. $f_i(\mathbf{B})$ is a composition function shown as below.

$$f_i(\mathbf{B}) = det(\mathbf{A}^i \mathbf{T}_i) = g(\boldsymbol{\theta}_{1,0}, \ldots, \boldsymbol{\theta}_{K,2^K-1}),$$

where $\boldsymbol{\theta}_{jc} = (\theta_{jc0}, \ldots, \theta_{jc,M_j-1})^\top$ and $\theta_{jcm} = \dfrac{\exp(\boldsymbol{\alpha}_c^\top \boldsymbol{\beta}_{jm})}{\sum_{m'=0}^{M_j-1} \exp(\boldsymbol{\alpha}_c^\top \boldsymbol{\beta}_{jm'})} = \dfrac{1}{1 + \sum_{m' \neq m} \exp(\boldsymbol{\alpha}_c^\top \boldsymbol{\beta}_{jm'} - \boldsymbol{\alpha}_c^\top \boldsymbol{\beta}_{jm})}$.

θ_{jcm} is an analytic function because exponential functions are positive analytic functions, and $g(\boldsymbol{\theta}_{1,0}, \ldots, \boldsymbol{\theta}_{K,2^K-1})$ is also a real analytic function of $(\boldsymbol{\theta}_{1,0}, \ldots, \boldsymbol{\theta}_{K,2^K-1})$ given that it is a polynomial. Therefore, we know that $f_i(\mathbf{B})$ is a real analytic function of $\mathbf{B}$, given the fact that the composition of real analytic functions is a real analytic function. $\square$

Next we introduce the following lemma, which shows that the zero set of a real analytic function has Lebesgue measure zero if the function is not constantly equal to zero.

Lemma 1. *(Mityagin, 2020; Dang, 2015)* *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a real analytic function on* $\mathbb{R}^d$. *If* f *is not identically zero, then its zero set* $\{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) = 0\}$ *has Lebesgue measure zero.*

Proposition 4. *If* $\mathbf{\Delta}^i$ *satisfies the structure shown in Theorem 2, then there exists some* $\mathbf{B} \in \Omega_{\mathbf{\Delta}^i}$, *such that* $f_i(\mathbf{B}) \neq 0$, $i = 1, 2$.

Proof. As shown in condition $B1$ of Theorem 2, assume that for $j = 1, \ldots, K$, Δ_j^1 and Δ_j^2 satisfy the following structure:

$$\mathbf{\Delta}_j = \begin{bmatrix} 0\ 0\ \cdots\ 0 & 0 & 0\ \cdots\ 0 \\ 1\ 0\ \cdots\ 0\ \delta_{jj1} = 1 & 0\ \cdots\ 0 \\ 1\ 0\ \cdots\ 0 & 0 & 0\ \cdots\ 0 \\ \vdots\ \vdots\quad \vdots & \vdots & \vdots\quad \vdots \\ 1\ 0\ \cdots\ 0 & 0 & 0\ \cdots\ 0 \end{bmatrix}_{M_j \times P}. \tag{A15}$$

Then, we have

$$\boldsymbol{p}_{j0} = \left(1 - \frac{\exp\left(\beta_{j01}\right)}{\exp\left(\beta_{j01}\right) + h_j}, \frac{\exp\left(\beta_{j01}\right)}{\exp\left(\beta_{j01}\right) + h_j}\right)^{\top},$$

$$\boldsymbol{p}_{j1} = \left(1 - \frac{\exp\left(\beta_{j01} + \beta_{jj1}\right)}{\exp\left(\beta_{j01} + \beta_{jj1}\right) + h_j}, \frac{\exp\left(\beta_{j01} + \beta_{jj1}\right)}{\exp\left(\beta_{j01} + \beta_{jj1}\right) + h_j}\right)^{\top},$$

where

$$h_j = \sum_{m' \neq 1} \exp\left(\beta_{j0m'}\right).$$

As shown in Definition 1, Δ_j^1 and Δ_j^2 satisfy simple structure, and $\boldsymbol{p}_{j0}$ and $\boldsymbol{p}_{j1}$ are linearly independent for $j = 1, \ldots, K$ due to $\beta_{jk1} \neq 0$. Therefore, according to Proposition 1 and Remark 8, we have $rank(\boldsymbol{A}^i \boldsymbol{T}_i) = \prod_{j=1}^{K} rank(\boldsymbol{p}_{j0}, \boldsymbol{p}_{j1}) = 2^K$, which implies that $f_i(\boldsymbol{B}) \neq 0$, $i = 1, 2$. $\qquad\square$

Let S_i denote the zero set of function $f_i(\boldsymbol{B})$:

$$S_i = \{\boldsymbol{B} \in \Omega_{\boldsymbol{\Delta}^i} : f_i(\boldsymbol{B}) = det(\boldsymbol{A}^i \boldsymbol{T}_i) = 0\},$$

then by Lemma 1 we can conclude that S_i is a measure zero set with respect to $\Omega_{\boldsymbol{\Delta}^i}$ provided $\boldsymbol{\Delta}^i$ satisfies condition $(B1)$.

For condition $(B2)$, we need to show that $rank_K(\boldsymbol{T}_3) \geq 2$ holds almost everywhere in $\boldsymbol{\Omega}_{\boldsymbol{\Delta}'}$. Note $rank_K(\boldsymbol{T}_3) \geq 2$ implies that for any two different attribute profiles $\boldsymbol{\alpha}_c, \boldsymbol{\alpha}_{c'}$, there always exist one $j^* > 2K$, such that $\theta_{j^*cm} \neq \theta_{j^*c'm}$ for some choice m with $0 < m < M_{j^*}$. Note that the $\boldsymbol{\alpha}_c$-th and $\boldsymbol{\alpha}_{c'}$-th columns of $\boldsymbol{T}_3$ are $\boldsymbol{T}_{3,\boldsymbol{\alpha}_c} = \bigotimes_{j>2K} \boldsymbol{\theta}_{j\boldsymbol{\alpha}_c}$ and $\boldsymbol{T}_{3,\boldsymbol{\alpha}_{c'}} = \bigotimes_{j>2K} \boldsymbol{\theta}_{j\boldsymbol{\alpha}_{c'}}$. Then, under condition $B2$, we have $\boldsymbol{\theta}_{j^*\boldsymbol{\alpha}_c} \neq \boldsymbol{\theta}_{j^*\boldsymbol{\alpha}_{c'}}$, which implies $\boldsymbol{T}_{3,\boldsymbol{\alpha}_c} \neq \boldsymbol{T}_{3,\boldsymbol{\alpha}_{c'}}$. Therefore, $\boldsymbol{T}_{3,\boldsymbol{\alpha}_c} = \boldsymbol{T}_{3,\boldsymbol{\alpha}_{c'}}$ holds only when $\beta_{j^*mk} = 0$ for some k and choice m with $\delta_{j^*mk} = 1$, which are of Lebesgue measure zero within $\boldsymbol{\Omega}_{\boldsymbol{\Delta}'}$. That proves $rank_K(\boldsymbol{T}_3) \geq 2$ almost everywhere within $\boldsymbol{\Omega}_{\boldsymbol{\Delta}'}$.

Therefore, the inequality A3 shown in Theorem 3 holds almost everywhere in $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{B})$.

Appendix B Posterior Inference

The goal of this section is to describe our strategy for inferring the nominal RLCM parameters. We first discuss the conditional likelihood and apply the Polya-gamma identity to augment the likelihood function. An important feature to note is that we collapse the conditional likelihood, which has the advantage of requiring fewer draws of Polya-gamma augmented variables. Then, we show the derivation of full conditional distributions for a Gibbs sampling algorithm.

Conditional Likelihood and Polya Gamma Data Augmentation

Let $i = 1, \ldots, n$ index individuals so that y_{ij} denotes the observed response for individual i to item j and $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iJ})^{\top}$ is a J-vector of responses for individual i. Let $\boldsymbol{B}$ denote all of

the item parameters. The conditional distribution of a sample of n responses is,

$$p(\boldsymbol{y}_{1:n}|\boldsymbol{\alpha}_{1:n}, \boldsymbol{B}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{B}) = \prod_{i=1}^{n}\prod_{j=1}^{J}\prod_{m=0}^{M_j-1}\left(\frac{\exp\left(\boldsymbol{a}_i^\top\boldsymbol{\beta}_{jm}\right)}{\sum_{m'=0}^{M_j-1}\exp\left(\boldsymbol{a}_i^\top\boldsymbol{\beta}_{jm'}\right)}\right)^{\mathbb{1}(y_{ij}=m)}. \quad \text{(B1)}$$

where $\boldsymbol{y}_i$ is the response vector for individual i, $\boldsymbol{\alpha}_i$ is individual i's attribute profile, $\boldsymbol{y}_{1:n} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\top$ is a $n \times J$ matrix of responses, and $\boldsymbol{\alpha}_{1:n} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)^\top$ denotes all attribute profiles. An important feature to note about the conditional likelihood is that we can aggregate terms in the product for individuals who reside in the same class and select the same response options on items. That is, after switching the order of the i and j products and substituting $\boldsymbol{a}_i = \boldsymbol{a}_{\boldsymbol{\alpha}_i^\top\boldsymbol{v}}$ we can aggregate Eq. B1 as

$$p(\boldsymbol{y}_{1:n}|\boldsymbol{\alpha}_{1:n}, \boldsymbol{B}) = \prod_{j=1}^{J}\prod_{c=0}^{2^K-1}\frac{\prod_{m=0}^{M_j-1}\left[\exp\left(\boldsymbol{a}_c^\top\boldsymbol{\beta}_{jm}\right)\right]^{n_{jcm}}}{\left[\sum_{m'=0}^{M_j-1}\exp\left(\boldsymbol{a}_c^\top\boldsymbol{\beta}_{jm'}\right)\right]^{n_c}} \quad \text{(B2)}$$

where $n_{jcm} = \sum_{i=1}^{n}\mathbb{1}(y_{ij}=m)\mathbb{1}(\boldsymbol{\alpha}_i^\top\boldsymbol{v}=c)$ indicates the number of individuals within class c that select option m on item j and $n_c = \sum_{i=1}^{n}\mathbb{1}(\boldsymbol{\alpha}_i^\top\boldsymbol{v}=c)$ is the number of individuals residing in class c.

The form of Eq. B2 enables us to adopt the Polya-Gamma data augmentation strategy for models involving logistic functions (Polson et al., 2013). In particular, Polson et al. (2013) reported the following identity relating the logistic function with an integral for a Polya-Gamma (PG) random variable,

$$\frac{(e^\psi)^a}{(1+e^\psi)^b} = 2^{-b}e^{\kappa\psi}\int_0^\infty e^{-w\psi^2/2}p(w)dw, \ a \in \mathbb{R}, \ b > 0 \quad \text{(B3)}$$

where $\kappa = a - b/2$ and $w \sim \text{PG}(b, 0)$. Equation B3 provides a data augmentation strategy for the random variable ψ that is conjugate with normal priors. For instance, see Polson et al. (2013) for a summary of a Gibbs sampling algorithm to infer the posterior distribution of logistic regression model parameters using Markov chain Monte Carlo (MCMC). Additionally, the PG strategy was also used for Bayesian estimation of the two parameter logistic item response theory model (Jiang & Templin, 2019) and binary diagnostic models (Balamuta & Culpepper, 2022).

We apply the PG identity in Eq. B3 by rewriting the portion of the conditional likelihood in Eq. B2 that corresponds with item j in a logistic format. Specifically, as noted by Holmes and Held and Polson et al., the full conditional distribution for $\boldsymbol{\beta}_{jm}$ can be written as,

$$p(\boldsymbol{\beta}_{jm}|\boldsymbol{y}_{1:n,j}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{B}_{j(m)}) = p(\boldsymbol{\beta}_{jm})p(\boldsymbol{y}_{1:n,j}|\boldsymbol{B}_j)$$

$$\propto p(\boldsymbol{\beta}_{jm})\prod_{c=0}^{2^K-1}\left[\frac{\exp\left(\eta_{jcm}\right)}{1+\exp\left(\eta_{jcm}\right)}\right]^{n_{jcm}}\left[\frac{1}{1+\exp\left(\eta_{ijm}\right)}\right]^{n_c-n_{jcm}}. \quad \text{(B4)}$$

where $\mathbf{y}_{1:n,j} = (y_{1j}, \ldots, y_{nj})$, $\mathbf{B}_{j(m)} = (\boldsymbol{\beta}_{j0}, \ldots, \boldsymbol{\beta}_{j,m-1}, \boldsymbol{\beta}_{j,m+1}, \ldots, \boldsymbol{\beta}_{j,M_j-1})$ excludes coefficients for response m on j, and

$$\eta_{jcm} = \mathbf{a}_c^\top \boldsymbol{\beta}_{jm} - C_{jcm} \tag{B5}$$

$$C_{jcm} = \ln \left(\sum_{m' \neq m} \exp(\mathbf{a}_c^\top \boldsymbol{\beta}_{jm'}) \right). \tag{B6}$$

Full Conditional Distributions

δ_{jpm} *and* γ The full conditional distribution for δ_{jpm} is

$$p(\delta_{jpm}|\beta_{jpm}, \gamma, \sigma_\beta^2) \propto p(\beta_{jpm}|\delta_{jpm}, \sigma_\beta^2) p(\delta_{jpm}|\gamma) \tag{B7}$$

which is

$$\delta_{jpm}|\beta_{jpm}, \gamma \sim \text{Bernoulli}\left(\tilde{\gamma}_{jpm}\right) \tag{B8}$$

where

$$\tilde{\gamma}_{jpm} = \frac{\gamma p(\beta_{jpm}|\delta_{jpm} = 1, \sigma_\beta^2)}{\gamma p(\beta_{jpm}|\delta_{jpm} = 1, \sigma_\beta^2) + (1 - \gamma) p(\beta_{jpm}|\delta_{jpm} = 0, \sigma_\beta^2)} \tag{B9}$$

The full conditional distribution for γ is

$$p(\gamma|\boldsymbol{\Delta}) \propto p(\boldsymbol{\Delta}|\gamma) p(\gamma) = \left(\prod_{j=1}^{J} \prod_{p=1}^{P-1} \prod_{m=1}^{M_j-1} p(\delta_{jpm}|\gamma) \right) p(\gamma) \tag{B10}$$

so

$$\gamma|\boldsymbol{\Delta} \sim \text{Beta}\left(\sum_{j=1}^{J} \sum_{p=1}^{P-1} \sum_{m=1}^{M_j-1} \delta_{jpm} + a, \; J(P-1)(M-1) - \sum_{j=1}^{J} \sum_{p=1}^{P-1} \sum_{m=1}^{M_j-1} \delta_{jpm} + b \right) \tag{B11}$$

σ_β^2 The full conditional distribution for σ_β^2 is

$$\sigma_\beta^2 \mid \mathbf{B}, \boldsymbol{\Delta} \sim IGamma\left(\alpha_\sigma + \frac{1}{2} \sum_{j=1}^{J} P(M_j - 1), \beta_\sigma + \frac{1}{2} \sum_{j=1}^{J} \sum_{p=0}^{P-1} \sum_{m=1}^{M_j-1} \beta_{jpm}^2 (D(1 - \delta_{jpm}) + \delta_{jpm}) \right) \tag{B12}$$

$\boldsymbol{\alpha}_i$ *and* $\boldsymbol{\pi}$ We update $\boldsymbol{\alpha}$ while integrating $\boldsymbol{\pi}$ out

$$
\begin{aligned}
p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) &= \int p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}) d\boldsymbol{\pi} \\
&= \frac{1}{B(\boldsymbol{d}_0)} \int \left(\prod_{c=0}^{2^K-1} \pi_c^{n_c+d_{0,c}-1} \right) \mathrm{d}\boldsymbol{\pi} \\
&= \frac{B(\boldsymbol{n}+\boldsymbol{d}_0)}{B(\boldsymbol{d}_0)}.
\end{aligned}
\tag{B13}
$$

Then, the full conditional prior distribution for $\boldsymbol{\alpha}_i$ is

$$
\begin{aligned}
p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c \mid \boldsymbol{\alpha}_{(i)}) &= \frac{p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)}{p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_N)} \\
&= \frac{n_{c(i)}+1}{n-1+2^K},
\end{aligned}
\tag{B14}
$$

where $n_{c(i)}$ represents the number of individuals other than i that have attribute profile $\boldsymbol{\alpha}_c$. Full conditional distribution of $\boldsymbol{\alpha}_i$ given $\boldsymbol{y}_{1:n}$ and $\boldsymbol{\alpha}_{(i)}$ is

$$
\begin{aligned}
p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c \mid \boldsymbol{\alpha}_{(i)}, \boldsymbol{y}_{1:n}, \boldsymbol{B}^{(t-1)}) &\propto p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c \mid \boldsymbol{\alpha}_{(i)}) p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}^{(t-1)}) \\
&\propto (n_{c(i)}+1) p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}^{(t-1)}),
\end{aligned}
\tag{B15}
$$

we update $\boldsymbol{\alpha}_i$ sequentially with weight proportional to $(n_{c(i)}+1) p(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}^{(t-1)})$. For $\boldsymbol{\pi}$, we have

$$
\boldsymbol{\pi} \mid \boldsymbol{\alpha}_{1:n} \sim Dirichlet(\boldsymbol{n}+\boldsymbol{d}_0),
\tag{B16}
$$

where $\boldsymbol{n} = (\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{2^K-1})$.

Item Parameters, $\boldsymbol{\beta}_{jm}$, *and Augmented Parameters,* w_{jcm} Applying the PG identity in Eq. B3 to Eq. B4 yields,

$$
\begin{aligned}
p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{y}_{1:n,j}, \boldsymbol{A}, \boldsymbol{B}_{j(m)}, \boldsymbol{\delta}_{jm}, \boldsymbol{w}_{jm}) &\propto p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm}) \prod_{c=0}^{2^K-1} \frac{\left[\exp(\eta_{jcm}) \right]^{n_{jcm}}}{\left[1 + \exp(\eta_{jcm}) \right]^{n_c}} \\
&\propto p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm}) \prod_{c=0}^{2^K-1} \exp\left(\tilde{y}_{jcm}\eta_{jcm} - \frac{w_{jcm}\eta_{jcm}^2}{2} \right),
\end{aligned}
\tag{B17}
$$

where $\tilde{y}_{jcm} = n_{jcm} - n_c/2$ and w_{jm} is a 2^K vector with element c defined as a PG random variable w_{jcm} with full conditional distribution of $w_{jcm}|A, B_j \sim \text{PG}(n_c, \eta_{jcm})$. We see,

$$p(\boldsymbol{\beta}_{jm}|\boldsymbol{y}_{1:n,j}, A, B_{j(m)}, \boldsymbol{\delta}_{jm}, \boldsymbol{w}_{jm}) \propto p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm}) \prod_{c=0}^{2^K-1} \exp\left\{ -\frac{\omega_{jcm}}{2} \left(\frac{\tilde{y}_{jcm}}{\omega_{jcm}} - \eta_{jcm} \right)^2 \right\}$$

$$= p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm}) \prod_{c=0}^{2^K-1} \exp\left\{ -\frac{\omega_{jcm}}{2} \left(z_{jcm} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_{jm} \right)^2 \right\}$$

$$= p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm}) \exp\left\{ -\frac{1}{2} \left(\boldsymbol{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} \right)^\top \boldsymbol{\Omega}_{jm} \left(\boldsymbol{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} \right) \right\} \tag{B18}$$

where $\mathbf{A}$ is a $2^K \times P$ design matrix and

$$z_{jcm} = \frac{\tilde{y}_{jcm}}{\omega_{jcm}} + C_{jcm} \tag{B19}$$

$$\boldsymbol{z}_{jm} = (z_{0jm}, \ldots, z_{2^K-1, jm})^\top. \tag{B20}$$

Since the prior distribution of $\boldsymbol{\beta}_{jm}$ given $\boldsymbol{\delta}_{jm}$ is

$$\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{jm}), \tag{B21}$$

where $\boldsymbol{\Sigma}_{jm} = \sigma_\beta^2 \text{diag}(v_{jm0}, \ldots, v_{jm,2^K-1})$ and $v_{jmp} = \delta_{jmp} + (1 - \delta_{jmp})/D$. Then, adding the prior term from the exponent for $\boldsymbol{\beta}_{jm}$ yields the posterior distribution

$$p(\boldsymbol{\beta}_{jm}|\boldsymbol{y}_{1:n,j}, A, B_{j(m)}, \boldsymbol{\delta}_{jm}, \boldsymbol{w}_{jm}) \propto p(\boldsymbol{\beta}_{jm} \mid \boldsymbol{\delta}_{jm}) \prod_{c=0}^{2^K-1} \exp\left\{ -\frac{\omega_{jcm}}{2} \left(\frac{\tilde{y}_{jcm}}{\omega_{jcm}} - \eta_{jcm} \right)^2 \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\beta}_{jm}^\top \boldsymbol{\Sigma}_{jm}^{-1} \boldsymbol{\beta}_{jm} + \left(\boldsymbol{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} \right)^\top \boldsymbol{\Omega}_{jm} \left(\boldsymbol{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} \right) \right) \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\beta}_{jm} - \boldsymbol{\mu}_{jm} \right)^\top \mathbf{V}_{jm} \left(\boldsymbol{\beta}_{jm} - \boldsymbol{\mu}_{jm} \right) \right\}. \tag{B22}$$

Therefore, the full conditional distribution of $\boldsymbol{\beta}_{jm}$ is

$$\boldsymbol{\beta}_{jm} \mid \boldsymbol{Y}_{1:n,j}, A, B_{j(m)}, \boldsymbol{\delta}_{jm}, \boldsymbol{w}_{jm} \sim \mathcal{N}_{2^K} \left(\boldsymbol{\mu}_{jm}, \mathbf{V}_{jm} \right) \tag{B23}$$

$$\mathbf{V}_{jm} = A^\top \boldsymbol{\Omega}_{jm} A + \boldsymbol{\Sigma}_{jm}^{-1} \tag{B24}$$

$$\boldsymbol{\mu}_{jm} = \mathbf{V}_{jm}^{-1} A^\top \boldsymbol{\Omega}_{jm} \boldsymbol{z}_{jm} = \mathbf{V}_{jm}^{-1} A^\top \left(\tilde{\boldsymbol{y}}_{jm} + \boldsymbol{\Omega}_{jm} \boldsymbol{C}_{jm} \right) \tag{B25}$$

where $\boldsymbol{\Sigma}_{jm} = \sigma_\beta^2 \text{diag}(v_{jm0}, \ldots, v_{jmP})$ and $v_{jmp} = \delta_{jmp} + (1 - \delta_{jmp})/D$. Note we use the second equality in Eq. B25 to avoid numerical issues associated with dividing by a ω_{jcm} that is zero.

Similarly, if we instead sample one coefficient at a time, we need the full conditional distribution of β_{jmp} as follows:

$$
\begin{aligned}
p(\beta_{jmp}|\mathbf{y}_{1:n,j}, \mathbf{A}, \mathbf{B}_{j(m)}, \mathbf{q}_j, \mathbf{w}_{jm}) &\propto p(\beta_{jmp}) \prod_{c=0}^{2^K-1} \exp\left\{ -\frac{\omega_{jcm}}{2} \left(\frac{\tilde{y}_{jcm}}{\omega_{jcm}} - \eta_{jcm} \right)^2 \right\} \\
&= p(\beta_{jmp}) \exp\left\{ -\frac{1}{2} \left(\mathbf{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} \right)^\top \boldsymbol{\Omega}_{jm} \left(\mathbf{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} \right) \right\} \\
&\propto \exp\left\{ -\frac{\beta_{jmp}^2}{2\boldsymbol{\Sigma}_{jmp}} \right\} \exp\left\{ -\frac{1}{2} \left(\tilde{\mathbf{z}}_{jm} - \mathbf{A}_p \beta_{jmp} \right)^\top \boldsymbol{\Omega}_{jm} \left(\tilde{\mathbf{z}}_{jm} - \mathbf{A}_p \beta_{jmp} \right) \right\} \\
&\propto \left(\boldsymbol{\beta}_{jm} - \boldsymbol{\mu}_{jm} \right)^\top \mathbf{V}_{jm} \left(\boldsymbol{\beta}_{jm} - \boldsymbol{\mu}_{jm} \right). \quad (B26)
\end{aligned}
$$

Therefore, the full conditional distribution for β_{jmp} given $\mathbf{Y}_{1:n,j}$, $\mathbf{A}$, $\boldsymbol{\beta}_{jm(p)}$, and $\mathbf{B}_{j(m)}$ is

$$
\beta_{jmp} \mid \mathbf{Y}_{1:n,j}, \delta_{jmp}, \mathbf{A}, \boldsymbol{\beta}_{jm(p)}, \mathbf{B}_{j(m)} \sim \mathcal{N}(\mu_{jmp}, \sigma_{jmp}^2), \quad (B27)
$$

$$
\sigma_{jmp}^2 = \frac{1}{\mathbf{A}_p^\top \boldsymbol{\Omega}_{jm} \mathbf{A}_p + 1/\sigma_\beta^2 v_{jmp}}, \quad (B28)
$$

$$
\mu_{jmp} = \sigma_{jmp}^2 \mathbf{A}_p^\top \boldsymbol{\Omega}_{jm} \tilde{\mathbf{z}}_{jm} = \sigma_{jmp}^2 \mathbf{A}_p^\top \left(\tilde{\mathbf{y}}_{jm} + \boldsymbol{\Omega}_{jm} \mathbf{C}_{jm} - \boldsymbol{\Omega}_{jm} \mathbf{A}_{(p)} \boldsymbol{\beta}_{jm(p)} \right), \quad (B29)
$$

$$
\tilde{\mathbf{z}}_{jm} = \mathbf{z}_{jm} - \mathbf{A}_{(p)} \boldsymbol{\beta}_{jm(p)} = \mathbf{z}_{jm} - \mathbf{A}\boldsymbol{\beta}_{jm} + \mathbf{A}_p \beta_{jmp}, \quad (B30)
$$

where $\mathbf{A}_p$ is column p of $\mathbf{A}$ and $\mathbf{A}_{(p)}$ excludes column p of $\mathbf{A}$. Note computation of the conditional mean and variance requires, $\mathbf{A}_p^\top \boldsymbol{\Omega}_{jm} \mathbf{A}_p = \sum_{c=0}^{2^K-1} \mathbf{A}_{pc} \omega_{jcm}$.

Appendix C Starting Values of $\mathbf{B}$

In this appendix, we will show the starting value generation steps of coefficients in $\mathbf{B}$.

1. Perform a $k-$means clustering (MacQueen, 1967) on the observed responses.

   (a) First, define the binary response array $\mathbf{Y}^b$ with size $N \times J \times M$, where $Y_{ijm}^b = \mathbf{I}\left\{Y_{ij} = m\right\}$. Then, reshape the three dimensional array $\mathbf{Y}^b$ into a matrix $\mathbf{Y}^*$ with size $N \times JM$, through combining columns of every slice of $\mathbf{Y}^b$.

   (b) Partition the $JM-$dimensional vectors corresponding to the N respondents $\mathbf{Y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_N^*)^\top$ into 2^K distinct groups with $n_{c'}$ respondents per group.

   (c) Initialize the category response probabilities described in Eq. 1 such that $\theta_{j'c'} \in \boldsymbol{\Theta}_{JM \times 2^K}$ is the j'-th element of the cluster center for group c'.

2. Assuming K factors, perform an exploratory factor analysis (EFA) on the slices of observed responses array $\mathbf{Y}_m^b$, $m = 1, \ldots, M-1$.

   (a) Generate factor scores for the $i = 1, \ldots, N$ respondents across the $j = 1, \ldots, J$ items and $k = 1, \ldots, K$ attributes, $\tilde{\theta}_{ik}^m$.

   (b) Compute within-cluster averages of the factor scores $\tilde{\theta}_{c'k}^m = \frac{1}{n_{c'}} \sum_{i^{c'}=1}^{n_{c'}} \tilde{\theta}_{i^{c'}k}^m$ for each of the 2^K groups.

TABLE 6.
Summary of simulation performance for RLCM with dense $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$

ρ	K	N	MAD		EAR
			$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{Q}}$
0	2	1000	0.0214	0.0151	0.8840
	2	2000	0.0179	0.0108	0.9234
	2	5000	0.0120	0.0067	0.9531
	2	10000	0.0077	0.0048	0.9733
	3	1000	0.1097	0.0204	0.8513
	3	2000	0.0135	0.0152	0.8981
	3	5000	0.0086	0.0062	0.9345
	3	10000	0.0069	0.0046	0.9474
0.25	2	1000	0.0229	0.0204	0.8593
	2	2000	0.0170	0.0152	0.9134
	2	5000	0.0084	0.0062	0.9630
	2	10000	0.0061	0.0046	0.9783
	3	1000	0.0256	0.0111	0.8299
	3	2000	0.0195	0.0089	0.8749
	3	5000	0.0155	0.0085	0.9099
	3	10000	0.0120	0.0069	0.9231

MAD = mean absolute deviation; EAR = element-wise accuracy rate

(c) Dichotomize the within-cluster factor score averages into pseudo-attributes as $\tilde{\alpha}^m_{c'k} = I\left(\tilde{\theta}^m_{c'k} > 0\right)$.

3. Define the pseudo-attribute profiles $\tilde{\boldsymbol{\alpha}}^m_{c'} = (\tilde{\alpha}^m_{c'1}, \ldots, \tilde{\alpha}^m_{c'K})^\top$ in terms of the binary-integer bijection $(\tilde{\boldsymbol{\alpha}}^m_{c'})^\top \boldsymbol{v} = c$.

4. Based on the bijection integers in step 3, swap the label of latent class of the initialized category response probabilities matrix got in step 1.

5. For $m = 1, 2, \ldots, M - 1$, initialize the m-th slice of $\boldsymbol{B}$ as follows:

(a) Define the category response probability as $Logit^{-1}(\theta_{jcm}) = \boldsymbol{\alpha}_c^\top \boldsymbol{\beta}^{(0)}_{jm} \in \boldsymbol{M}^{(0)}_{\alpha\beta:J\times 2^K}$.

(b) Calculate $\boldsymbol{B}^{(0)}_m = \boldsymbol{M}^{(0)}_{\alpha\beta}\boldsymbol{A}(\boldsymbol{A}^\top\boldsymbol{A})^{-1}$, where $\boldsymbol{A}$ is the $2^K \times P$ design matrix.

Appendix D Simulation Results for Dense $\boldsymbol{\Delta}$

The unknown denser true $\boldsymbol{\Delta}$ and true $\boldsymbol{Q}$ matrices for each option are shown as follows (columns in $\boldsymbol{\Delta}$ follow the same order as the design vector shown in Eq. 3):

- $\boldsymbol{\Delta}$ cube with $K = 2$

$$\Delta_{m=1} = \begin{pmatrix} 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&1&1 \\ 1&1&1&1 \\ 1&1&1&1 \end{pmatrix}, \quad \Delta_{m=2} = \begin{pmatrix} 1&0&1&0 \\ 1&1&0&0 \\ 1&1&1&1 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&1&1&1 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&1&1 \\ 1&0&1&0 \\ 1&1&1&1 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \end{pmatrix}, \quad \Delta_{m=3} = \begin{pmatrix} 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&1&1&1 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&1&1 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&1&1&1 \end{pmatrix} \tag{B31}$$

- Δ cube with $K = 3$

$$\Delta_{m=1} = \begin{pmatrix} 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&1&0&0&1 \\ 1&1&1&0&1&0&0 \\ 1&1&0&1&0&1&0 \\ 1&0&1&1&0&0&1 \\ 1&1&1&0&1&0&0 \\ 1&1&0&1&0&1&0 \\ 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1 \end{pmatrix}, \Delta_{m=2} = \begin{pmatrix} 1&1&1&0&1&0&0 \\ 1&1&1&0&1&0&0 \\ 1&1&1&1&1&1&1 \\ 1&0&1&0&0&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&1&1&1&1&1 \\ 1&0&0&1&0&0&0 \\ 1&0&1&1&0&0&1 \\ 1&0&1&1&0&0&1 \\ 1&1&1&1&1&1&1 \\ 1&1&0&0&0&0&0 \\ 1&1&1&1&1&1&1 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&1&0&1&0 \\ 1&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \end{pmatrix}, \Delta_{m=3} = \begin{pmatrix} 1&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0 \\ 1&1&0&1&0&1&0 \\ 1&0&0&1&0&0&0 \\ 1&1&1&0&1&0&0 \\ 1&0&1&1&0&0&1 \\ 1&1&0&1&0&1&0 \\ 1&0&0&1&0&0&0 \\ 1&1&1&1&1&1&1 \\ 1&1&0&0&0&0&0 \\ 1&1&1&0&1&0&0 \\ 1&1&1&1&1&1&1 \\ 1&0&0&1&0&0&0 \\ 1&0&1&1&0&0&1 \\ 1&1&0&0&0&0&0 \\ 1&1&1&1&1&1&1 \end{pmatrix} \tag{B32}$$

- Q matrices with $K = 2$

$$Q_{m=1} = \begin{pmatrix} 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&1 \\ 1&1 \\ 1&1 \end{pmatrix}, \quad Q_{m=2} = \begin{pmatrix} 0&1 \\ 1&0 \\ 1&1 \\ 1&0 \\ 1&0 \\ 1&1 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&1 \\ 0&1 \\ 1&1 \\ 0&1 \\ 1&0 \\ 1&0 \\ 0&1 \\ 1&0 \\ 0&1 \end{pmatrix}, \quad Q_{m=3} = \begin{pmatrix} 1&0 \\ 0&1 \\ 0&1 \\ 1&0 \\ 1&0 \\ 1&0 \\ 1&0 \\ 0&1 \\ 0&1 \\ 0&1 \\ 1&1 \\ 1&0 \\ 0&1 \\ 1&1 \\ 0&1 \\ 1&0 \\ 0&1 \\ 1&1 \end{pmatrix} \tag{B33}$$

- Q matrices with $K = 3$

$$Q_{m=1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Q_{m=2} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad Q_{m=3} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \tag{B34}$$

Simulation results are shown in Table 6.

References

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics, 37*, 3099–3132.

Bacci, S., Bartolucci, F., & Gnaldi, M. (2014). A class of multidimensional latent class IRT models for ordinal polytomous item responses. *Communications in Statistics-Theory and Methods, 43*(4), 787–800.

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

Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika, 72*, 141–157.

Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: A psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika, 79*(3), 403–425.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*(4), 434–455.

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement, 37*(6), 419–437.

Chen, J., & Zhou, H. (2017). Test designs and modeling under the general nominal diagnosis model framework. *PLoS ONE, 12*(6), e0180016.

Chen, Y., Culpepper, S., & Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika, 85*, 121–153.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association, 110*(510), 850–866.

Chen, Y., Liu, Y., Culpepper, S. A., & Chen, Y. (2021). Inferring the number of attributes for the exploratory DINA model. *Psychometrika, 86*(1), 30–64.

Chiu, C.-Y., Douglas, J., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika, 74*, 633–665.

Culpepper, S. A. (2019). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation. *Psychometrika, 84*(4), 921–940.

Culpepper, S. A., & Balamuta, J. J. (2021). Inferring latent structure in polytomous data with a higher-order diagnostic model. *Multivariate Behavioral Research* 1–19.

Dang, N. V. (2015). Complex powers of analytic functions and meromorphic renormalization in QFT. arXiv preprint arXiv:1503.00995 .

De La Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement, 33*(3), 163–183.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179–199.

DeYoreo, M., Reiter, J. P., & Hillygus, D. S. (2017). Bayesian mixture models with focused clustering for mixed ordinal and nominal data. *Bayesian Analysis, 12*(3), 679–703.

DiBello, L. V., Henson, R. A., & Stout, W. F. (2015). A family of generalized diagnostic classification models for multiple choice option-based scoring. *Applied Psychological Measurement, 39*(1), 62–79.

Dunson, D. B., & Xing, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *Journal of the American Statistical Association, 104*(487), 1042–1051.

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika, 84*(1), 19–40.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*(423), 881–889.

Gnaldi, M., Bacci, S., Kunze, T., & Greiff, S. (2020). Students' complex problem solving profiles. *Psychometrika, 85*, 469–501.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*(2), 215–231.

Gu, Y., & Dunson, D. B. (2021). Bayesian pyramids: Identifiable multilayer discrete latent structure models for discrete data. arXiv preprint arXiv:2101.10373 .

Holmes, C. C., & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis, 1*(1), 145–168.

Huang, G.-H., & Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika, 69*(1), 5–32.

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.

Jimenez, A., Balamuta, J. J., & Culpepper, S. A. (2023). A sequential exploratory diagnostic model using a Pólya-gamma data augmentation strategy. *British Journal of Mathematical and Statistical Psychology*.

Kruskal, J. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications, 18*, 95–138.

Kruskal, J. B. (1976). More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika, 41*(3), 281–293.

Kuo, B.-C., Chen, C.-H., Yang, C.-W., & Mok, M. M. C. (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items. *Educational Psychology, 36*(6), 1115–1133.

Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology, 69*(3), 253–275.

Ma, W., & de la Torre, J. (2019). Category-level model selection for the sequential G-DINA model. *Journal of Educational and Behavioral Statistics, 44*(1), 45–77.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).

Matsaglia, G., & Styan, G. P. H. (1974). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra, 2*(3), 269–292.

McHugh, R. B. (1956). Efficient estimation and local identification in latent class analysis. *Psychometrika, 21*(4), 331–347.

Mityagin, B. S. (2020). The zero set of a real analytic function. *Mathematical Notes, 107*, 529–530.

Murray, J. S., & Reiter, J. P. (2016). Multiple imputation of missing categorical and continuous values via Bayesian mixture models with local dependence. *Journal of the American Statistical Association, 111*(516), 1466–1479.

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.

Royle, J. A., & Link, W. A. (2005). A general class of multinomial mixture models for anuran calling survey data. *Ecology, 86*(9), 2505–2512.

Shear, B. R., & Roussos, L. A. (2017). Validating a distractor-driven geometry test using a generalized diagnostic classification model. In *Understanding and investigating response processes in validation research* (pp. 277–304). Springer.

Si, Y., & Reiter, J. P. (2013). Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *Journal of Educational and Behavioral Statistics, 38*(5), 499–521.

Templin, J., Henson, R. A., Rupp, E., Jang, A. A., & Ahmed, M. (2008). Cognitive diagnosis models for nominal response data. New York, NY: In *Annual Meeting of the National Council on Measurement in Education*.

Vermunt, J. K., Van Ginkel, J. R., Van der Ark, L. A., & Sijtsma, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociological Methodology, 38*(1), 369–397.

Wagner, R. F., & Wells, K. A. (1985). A refined neurobehavioral inventory of hemispheric preference. *Journal of Clinical Psychology, 41*(5), 671–676.

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *Annals of Statistics, 45*(2), 675–707.

Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association, 113*(523), 1284–1295.

Yigit, H. D., Sorrel, M. A., & de la Torre, J. (2019). Computerized adaptive testing for cognitively based multiple-choice data. *Applied Psychological Measurement, 43*(5), 388–401.