PSYCHOMETRIKA—VOL. 87, NO. 3, 903–945
SEPTEMBER 2022
https://doi.org/10.1007/s11336-021-09815-9

EXPLORATORY RESTRICTED LATENT CLASS MODELS WITH MONOTONICITY REQUIREMENTS UNDER PÒLYA–GAMMA DATA AUGMENTATION

JAMES JOSEPH BALAMUTA

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

STEVEN ANDREW CULPEPPER

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Restricted latent class models (RLCMs) provide an important framework for supporting diagnostic research in education and psychology. Recent research proposed fully exploratory methods for inferring the latent structure. However, prior research is limited by the use of restrictive monotonicity condition or prior formulations that are unable to incorporate prior information about the latent structure to validate expert knowledge. We develop new methods that relax existing monotonicity restrictions and provide greater insight about the latent structure. Furthermore, existing Bayesian methods only use a probit link function and we provide a new formulation for using the exploratory RLCM with a logit link function that has an additional advantage of being computationally more efficient for larger sample sizes. We present four new Bayesian formulations that employ different link functions (i.e., the logit using the Pòlya–gamma data augmentation versus the probit) and priors for inducing sparsity in the latent structure. We report Monte Carlo simulation studies to demonstrate accurate parameter recovery. Furthermore, we report results from an application to the Last Series of the Standard Progressive Matrices to illustrate our new methods.

Key words: restricted latent class models, Bayesian, Pòlya–gamma data augmentation.

Restricted latent class models ("RLCMs"), which are also known as either cognitive diagnosis models ("CDMs") in education or latent structure models (von Davier, 2009), are increasingly being used as an assessment framework to support learning interventions (Haertel, 1989, Macready & Dayton, 1977). In contrast with traditional factor analysis methods, which rank individuals along a few continuous traits, RLCMs classify individuals into latent classes, or attribute profiles, that correspond with mastery of a discrete set of skills that are needed for success. More specifically, RLCMs assume that a binary vector of K attributes $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\top \in \{0, 1\}^K$ underlies the observed binary responses to J items, $\boldsymbol{Y} = (Y_1, \ldots, Y_J)^\top \in \{0, 1\}^J$ where the number of dimensions K is smaller than the number of items J. RLCMs offer researchers a framework that is capable of providing educators with detailed information about student mastery in a given subject area with examples including: (1) fraction-subtraction (de la Torre & Douglas, 2004, Tatsuoka, 1984); (2) proportional reasoning Tjoe & de la Torre, 2013; (3) elementary probability theory (Heller & Wickelmaier, 2013); and (4) geometric sequences (Chen & Culpepper, 2020; Shute et al., 2008). Furthermore, the diagnostic modeling paradigm has proved useful for designing interventions that aim to enhance learning (e.g., see the Adaptive Content with Evidence-Based Diagnosis study; Shute et al., 2008). The purpose of our manuscript is to advance exploratory methods to provide researchers with tools for designing and refining diagnostic assessments.

The success of RLCMs as a methodology in education served as a milestone for researchers to apply RLCMs more broadly in the psychological sciences. In fact, researchers applied RLCMs to psychological data involving pathological gambling behaviors (Templin & Henson, 2006), spatial rotation (Chen et al., 2018b; Culpepper, 2015; Wang et al., 2018), and situational judgment tests

Correspondence should be made to James Joseph Balamuta, Departments of Informatics and Statistics, University of Illinois Urbana-Champaign, 725 South Wright Street, Champaign, IL61820, USA. Email: balamut2@illinois.edu

(Sorrel et al., 2016). However, the aforementioned applications of RLCMs are confirmatory in nature and require researchers to prespecify the latent structure a priori. In the cognitive diagnosis context, the $\boldsymbol{Q}$ matrix indicates the latent structure by specifying the attributes related to each item. More specifically, $\boldsymbol{Q}$ is a $J \times K$ binary matrix where the (j, k) element equals one if attribute k loads onto item j (i.e., $q_{jk} = 1$) and zero if having attribute k has no impact on the probability of responding to item j. In order to apply RLCMs, researchers must specify all elements of $\boldsymbol{Q}$. Specifying $\boldsymbol{Q}$ requires detailed theoretical knowledge about the content domain and is challenging in confirmatory studies given the necessary theory for specifying how attributes load onto observed items is typically unavailable.

Defining $\boldsymbol{Q}$ may be easier for educational research domains where the attributes required to succeed on a task are readily known. In contrast, specifying $\boldsymbol{Q}$ is more challenging outside of education where the underlying attributes and their relationship with observed responses may be less clearly understood a priori. The inability to accurately specify $\boldsymbol{Q}$ is a barrier to more widespread application of RLCMs in psychological research. Recent research overcame limitations of confirmatory RLCMs by proposing fully exploratory approaches for estimating the latent structure $\boldsymbol{Q}$ matrix. Specifically, recent research established new theory and methods for identifying (Xu, 2017; Xu & Shang, 2018; Chen et al., 2015, 2020; Culpepper, 2019b) and estimating model parameters (Chen et al., 2015, 2018a; Culpepper & Chen, 2019; Culpepper, 2019a; Xu & Shang, 2018). Exploratory RLCMs have been used to estimate the latent structure for matrix reasoning tasks (Chen et al., 2020), fraction-subtraction (Chen et al., 2015, 2018a; Culpepper, 2019a; Xu & Shang, 2018), English language proficiency (Culpepper & Chen, 2019), clinical assessment of social anxiety disorders (Chen et al., 2015), knowledge of elementary probability (Chen et al., 2020), in addition to ordinal outcomes such as early childhood approaches to learning (Culpepper, 2019b).

Existing exploratory methods provide accurate approaches for inferring the latent structure; yet, there are three significant challenges that must be overcome to encourage widespread application of RLCMs in the educational and psychological sciences. First, new approaches are needed to enforce monotonicity among the latent classes when inferring the $\boldsymbol{Q}$ matrix. In fact, monotonicity conditions are often needed in the exploratory framework to improve interpretation of estimated solutions and to prevent within chain attribute-level swapping where the meaning of $\alpha = 0$ and $\alpha = 1$ changes. Imposing monotonicity conditions fixes the interpretation of attributes by requiring that latent classes with more mastered attributes have a performance level no less than classes with fewer attributes. For instance, for $K = 3$ the monotonicity conditions force response probabilities for the "011" class to be larger than the "000," "010," and "001" classes. Enforcing monotonicity constraints when updating $\boldsymbol{Q}$ conditioned on the item parameters is challenging given it involves integration of a multivariate normal prior for item parameters over the monotone space to obtain a normalizing constant. It is important to note that the sparse latent class model (SLCM; Chen et al., 2020) effectively handles the monotonicity conditions by applying a heuristic, post-processing step for inferring $\boldsymbol{Q}$ from the item parameters. However, one disadvantage of the SLCM is that it is not designed to incorporate prior information about $\boldsymbol{Q}$, which poses a barrier for validating expert knowledge (e.g., see Culpepper, 2019a).

Second, inferring $\boldsymbol{Q}$ can be computationally demanding for Bayesian methods for larger sample sizes of 1000 or more. Access to larger datasets continues to expand and computationally efficient approaches are needed. Consequently, faster algorithms that also impose monotonicity conditions are needed to infer $\boldsymbol{Q}$ for large-scale applications.

We address these challenges by proposing several new Bayesian model formulations. First, we discuss a new hierarchical prior specification that enforces monotonicity conditions while inferring $\boldsymbol{Q}$. In fact, we improve upon the method of Culpepper (2019a), which imposes a restrictive monotonicity constraint that limits the ability of researchers to uncover disjunctive relationships, which may be important to model for understanding attributes in clinical settings (Templin & Henson, 2006). A novel feature of our proposed hierarchical model is that we use a confirmatory deterministic inputs, noisy "and" gate model (DINA) to specify a prior for item parameters conditioned on $\boldsymbol{Q}$. Note that we estimate $\boldsymbol{Q}$ for a general RLCM and the DINA model is used in the hierarchical prior that relates $\boldsymbol{Q}$ to the item parameters. Furthermore, our hierarchical prior provides a natural mechanism for including prior information about $\boldsymbol{Q}$ for validating expert knowledge (e.g., our approach can incorporate the expert prior model described Culpepper (2019a), whereas the Chen et al. (2020) approach cannot). We report Monte Carlo evidence that this hyper-prior provides an accurate framework for selecting active attributes in $\boldsymbol{Q}$. Moreover, the simulation studies suggest that our proposed methods improve upon both the approaches of Culpepper (2019a) and Chen et al. (2020).

Second, we show how using the Pòlya–gamma (PG) data augmentation scheme (Polson et al., 2013a; Windle et al., 2014) for a logistic item response function provides more efficient Bayesian computations for larger sample sizes, n. That is, existing Bayesian methods use a probit-link function that requires the sampling of $n \cdot J$ augmented random variables (i.e., one for each person and item) each step of the Gibbs sampling algorithm. In contrast, we show below that the application of the Pòlya–gamma formulation reduces the number of sampled augmented variables to $2^K \cdot J$ (i.e., one for each class and item). So, the Pòlya–gamma data augmentation approach is likely to be more efficient as long as $n > 2^K$ and the computational cost of sampling a Pòlya–gamma random variable is minimal. In fact, the Pòlya–gamma data augmentation strategy has been studied for the two-parameter logistic item response theory model (Jiang & Templin, 2019) and the confirmatory DINA model (Zhang et al., 2020). We extend the Pòlya–gamma formulation to estimate parameters of a general exploratory diagnostic model. In addition to providing a computationally efficient algorithm for larger sample sizes, we report the first Bayesian formulation for an exploratory version of the well-known log-linear cognitive diagnosis model (LCDM; Henson et al., 2009). Our Monte Carlo simulation results provide evidence that the logit-link approach using the Pòlya–gamma formulation reduces computation time by a half depending on the number of attributes and sample size when compared to methods that use a probit-link (Chen et al., 2020; Culpepper, 2019a).

The remainder of the paper includes five sections. First, we introduce readers to the general restricted latent class framework for multivariate binary data and provide a didactic discussion regarding the distinction between unrestricted latent class models (ULCMs) and RLCMs. The first section also provides readers with intuition for how structure is specified in RLCMs in addition to a discussion of general monotonicity conditions. Second, we discuss a new Bayesian formulation for inferring the latent structure. In particular, we discuss a novel data augmentation approach for the logit-link and provide details about the new hierarchical prior for inferring $\boldsymbol{Q}$ while also enforcing monotonicity conditions. We discuss a logit-link, and the hierarchical prior for $\boldsymbol{Q}$ can also be readily applied to models that use a probit-link as we show in the Monte Carlo simulation study. Note that the second section includes details regarding the specification of the Bayesian model and we report technical details about full conditional distributions for approximating the posterior distribution with Gibbs sampling in "Appendices A, B, and C." The third includes computational time of models with different two link functions (i.e., logit and probit) and three different prior specifications for $\boldsymbol{Q}$. That is, researchers have adopted several different priors for item parameters to infer $\boldsymbol{Q}$ (e.g., the stochastic search variable selection algorithm of Culpepper (2019a) vs. the sparse spike-slab prior of Chen et al. (2020)). We report new Monte Carlo simulation evidence about the performance of these prior formulations as well as the impact of different link functions

(e.g., probit versus logit) for different study design settings. In the fourth section, we report an application to 12 items from the last series of the Standard Progressive Matrix (SPM-LS; see Raven, 1941) to demonstrate the methodology. We discuss the implications of our contributions in the final section, offer directions for future research, and provide concluding remarks.

1. Restricted Latent Class Models for Binary Data

We next discuss the general RLCM framework for multivariate binary response data. First, we introduce the ULCM case and introduce notation and details about the model and likelihood function. Second, we compare ULCMs with RLCMs by describing how the $\boldsymbol{Q}$ matrix imposes structure on latent class response probabilities. Third, we describe details that enable us to use Bayesian variable selection methods to infer the latent structure. Specifically, we outline how to reparameterize latent class response probabilities using a general linear model specification and motivate our novel prior formulation for $\boldsymbol{Q}$ by introducing a $\boldsymbol{\Delta}$ matrix which provides more detailed information about the latent structure and the process by which attributes interact to impact response probabilities. The fourth subsection reviews model identifiability conditions, and the final subsection describes issues related to monotonicity.

1.1. Overview of Unrestricted Latent Class Models (ULCMs)

We first introduce the ULCM framework for binary data. We let i index individuals ($i = 1, \ldots, n$), j index items ($j = 1, \ldots, J$), and consider Y_{ij} as a random binary variable for individual i's response to item j. The observed data are $y_{ij} \in \{0, 1\}$ where $y_{ij} = 1$ indicates an affirmative response, such as a correct answer to a cognitive item, an affirmation of a symptom in clinical settings, or an elicitation of a preference, and $y_{ij} = 0$ otherwise. We let there be K attributes and use k to index attributes ($k = 1, \ldots, K$). Furthermore, the latent attribute profile for individual i is $\boldsymbol{\alpha}_i$, which is a binary vector such that $\boldsymbol{\alpha}_i \in \{0, 1\}^K$. We refer to the J-vector of random responses for individual i as $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^\top$, and let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iJ})^\top$ denote the observed values. In many instances, it is more convenient to use the integer representation for the binary latent class $\boldsymbol{\alpha}_c$ denoted by c. Specifically, we use the bijection between the binary vector and integers $c = \boldsymbol{\alpha}_c^\top \boldsymbol{v} \in \{0, 1, \ldots, 2^K - 1\}$ where $\boldsymbol{v} = (2^{K-1}, 2^{K-2}, \ldots, 1)^\top$. For example, for $K = 2$, $\boldsymbol{v} = (2, 1)^\top$ and the integer representations for attribute profiles $\boldsymbol{\alpha}_0 = (0, 0)^\top$, $\boldsymbol{\alpha}_1 = (0, 1)^\top$, $\boldsymbol{\alpha}_2 = (1, 0)^\top$, and $\boldsymbol{\alpha}_3 = (1, 1)^\top$ are $c = 0$, 1, 2, and 3, respectively.

The ULCM specifies a different item response function (IRF) for each class and is therefore the most general mixture model for binary response data. Let $\boldsymbol{\theta}_j = (\theta_{j0}, \ldots, \theta_{j,2^K-1})^\top$ be a 2^K vector where element c denotes the probability members of class c respond with an affirmative value of one to item j. The probability of an affirmative response for individual i to item j given $\boldsymbol{\alpha}_i$ and $\boldsymbol{\theta}_j$ is

$$\mathrm{P}(Y_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{\theta}_j) = p_{ij} = \sum_{c=0}^{2^K-1} \theta_{jc} \mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c) \tag{1}$$

where $\mathbb{1}(\cdot)$ is the indicator function that equals one if $\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c$ and zero otherwise. Accordingly, Y_{ij} is a Bernoulli random variable with the probability mass function

$$p(y_{ij} | \boldsymbol{\alpha}_i, \boldsymbol{\theta}_j) = \left(p_{ij}\right)^{y_{ij}} \left(1 - p_{ij}\right)^{1 - y_{ij}} \tag{2}$$

and assuming individual i's responses to the J items are conditionally independent given item parameters and $\boldsymbol{\alpha}$ implies the conditional likelihood of observing $\boldsymbol{Y}_i = \boldsymbol{y}_i$ given $\boldsymbol{\alpha}_i$ and item parameters are

$$p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\Theta}) = \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{\theta}_j) \tag{3}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)^\top$ is a $J \times 2^K$ matrix of item parameters.

The attributes are unobserved latent variables, so we must marginalize over them to find the complete-data likelihood. Let $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{2^K-1})^\top$ be a 2^K vector of structural probabilities where element c denotes the probability a given individual belongs to class c, $\pi_c = \mathrm{P}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)$. The likelihood for individual i is obtained by computing a weighted average of conditional likelihoods evaluated at each latent class:

$$p(\boldsymbol{y}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}) = \sum_{c=0}^{2^K-1} \pi_c \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \theta_{jc}) \tag{4}$$

and the likelihood of observing a sample of n independent observations is found by multiplying the likelihoods for all n respondents

$$p(\boldsymbol{y}_{1:n}|\boldsymbol{\Theta}, \boldsymbol{\pi}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}) \tag{5}$$

where $\boldsymbol{y}_{1:n} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\top$ is an $n \times J$ matrix of observed binary responses.

1.2. Restricted Latent Class Models and the $\boldsymbol{Q}$ Matrix

The aforementioned subsection describes mixture models for binary data in a completely unstructured framework. That is, $\boldsymbol{\Theta}$ was introduced in a way that did not impose any structure on its elements. For instance, for $K = 3$ there are 2^K latent classes with a given response probability vector of

$$\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)^\top. \tag{6}$$

where all of the response probabilities are allowed to differ for ULCMs.

Unlike ULCMs, RLCMs allow for the possibility that some elements in the rows of $\boldsymbol{\Theta}$ are equal. The pattern of equal and unequal elements in $\boldsymbol{\theta}_j$ is dictated by the jth row of $\boldsymbol{Q}$, $\boldsymbol{q}_j = (q_{j1}, \ldots, q_{jK})$. For example, consider different types of structure for cases where $K = 3$. If $\boldsymbol{q} = (0, 0, 1)$ only the third attribute impacts item responses. In this case for $\boldsymbol{q}$, the associated vector of item parameters is:

$$\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_0, \theta_1, \theta_0, \theta_1, \theta_0, \theta_1)^\top \tag{7}$$

because the $\boldsymbol{\alpha}_0 = (0, 0, 0)^\top$, $\boldsymbol{\alpha}_2 = (0, 1, 0)^\top$, $\boldsymbol{\alpha}_4 = (1, 0, 0)^\top$, and $\boldsymbol{\alpha}_6 = (1, 1, 0)^\top$ latent classes do not possess attribute three and therefore have the same response probability of θ_0 and the $\boldsymbol{\alpha}_1 = (0, 0, 1)^\top$, $\boldsymbol{\alpha}_3 = (0, 1, 1)^\top$, $\boldsymbol{\alpha}_5 = (1, 0, 1)^\top$, and $\boldsymbol{\alpha}_7 = (1, 1, 1)^\top$ classes all possess

attribute three and have a common response probability of θ_1. We refer to items with a single one in $\boldsymbol{q}_j$ as *simple structure*, given it relates to just one attribute.

An item with $\boldsymbol{q} = (0, 1, 1)$ has *complex structure* as it loads on more than one attribute. In this case, the structure imposed by $\boldsymbol{q}$ includes a more complicated pattern of equalities among elements in the item parameter vector

$$\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_0, \theta_1, \theta_2, \theta_3)^\top. \tag{8}$$

Notice that the classes that are missing both attributes two and three (i.e., $\boldsymbol{\alpha}_0 = (0, 0, 0)^\top$ and $\boldsymbol{\alpha}_4 = (1, 0, 0)^\top$) have the same response probability of θ_0. Similarly, classes with the second attribute and not the third (i.e., $\boldsymbol{\alpha}_2 = (0, 1, 0)^\top$ and $\boldsymbol{\alpha}_6 = (1, 1, 0)^\top$) have the same response probability of θ_2 just as classes with attribute three and not attribute two (i.e., $\boldsymbol{\alpha}_1 = (0, 0, 1)^\top$ and $\boldsymbol{\alpha}_5 = (1, 0, 1)^\top$) have response probabilities of θ_1. Finally, classes that possess both attributes (i.e., $\boldsymbol{\alpha}_3 = (0, 1, 1)^\top$ and $\boldsymbol{\alpha}_7 = (1, 1, 1)^\top$) have a common response probability of θ_3.

1.3. Reparameterizing $\boldsymbol{\Theta}$ to Infer the Latent Structure

As noted above, inferring the latent structure can be reduced to the problem of inferring which elements of $\boldsymbol{\Theta}$ are equal. Determining equality among elements in $\boldsymbol{\Theta}$ is more challenging when considering the parameters on their original probability metric. Following the confirmatory model parameterizations of von Davier (2008), Henson et al. (2009), and de la Torre (2011), Chen et al. (2015, 2020) translated the problem of inferring latent structure and equality among elements in $\boldsymbol{\Theta}$ into a variable selection problem by reparameterizing $\boldsymbol{\Theta}$ as a latent, multivariate, generalized linear model. Specifically, element (j, c) of $\boldsymbol{\Theta}$ is reparameterized as

$$\theta_{jc} = \Psi(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j) \tag{9}$$

where $\Psi(\cdot)$ is a general cumulative distribution function (cdf); e.g., Culpepper (2019a) and Chen et al. (2020) used a probit and Henson et al. (2009) and von Davier (2008) used a logit, and de la Torre (2011) used the identity, logit, and log link functions). Furthermore, $\boldsymbol{a}_c$ is a 2^K design vector that corresponds with attribute profile $\boldsymbol{\alpha}_c$, and $\boldsymbol{\beta}_j$ is a 2^K vector of parameters for item j. For instance, for $K = 3$, $\boldsymbol{a}$ and $\boldsymbol{\beta}$ are

$$\boldsymbol{a} = (1, \alpha_3, \alpha_2, \alpha_2\alpha_3, \alpha_1, \alpha_1\alpha_3, \alpha_1\alpha_2, \alpha_1\alpha_2\alpha_3)^\top \tag{10}$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^\top. \tag{11}$$

Equations 10 and 11 show that the reparameterized model can be viewed as a latent analysis of variance model with all main and interaction effects corresponding with the unobserved classes. For instance, the first element of $\boldsymbol{a}$ is one and corresponds with the intercept coefficient in $\boldsymbol{\beta}$ and the fourth element of $\boldsymbol{a}$ is a two-way interaction term involving $\alpha_2\alpha_3$ and the interaction-effect is β_3.

Let $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)^\top$ be a $J \times 2^K$ matrix that includes the reparameterized item parameters. The problem of inferring the latent structure in $\boldsymbol{\Theta}$ corresponds with identifying which elements of $\boldsymbol{B}$ are non-zero and active. In fact, each element of $\boldsymbol{q}$ determines whether multiple elements of $\boldsymbol{\beta}$ are active. For example, for $K = 3$, q_3 determines whether item parameters associated with the third attribute, such as β_1 for α_3, β_3 for $\alpha_2\alpha_3$, β_5 for $\alpha_1\alpha_3$, and β_7 for $\alpha_1\alpha_2\alpha_3$, are active and related to observed response probabilities. Therefore, an item with $\boldsymbol{q} = (0, 0, 1)$ implies that only

the third attribute is active and the $\boldsymbol{\beta}$ vector includes zeros for all coefficients with the exception of the intercept and main effect for α_3

$$\boldsymbol{\beta} = (\beta_0, \beta_1, 0, 0, 0, 0, 0, 0). \tag{12}$$

Moreover, the latent structure represented by an item with attributes two and three active (i.e., $\boldsymbol{q} = (0, 1, 1)$) corresponds with a $\boldsymbol{\beta}$ vector of

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, 0, 0, 0, 0)^\top \tag{13}$$

where only the intercept, main effects for α_2 and α_3, and two-way interaction between α_2 and α_3 are nonzero and active.

Clearly, a given q_{jk} determines the activeness of several parameters in $\boldsymbol{\beta}_j$, so one implication is that $\boldsymbol{Q}$ characterizes which attributes relate to item responses across main- and interaction-effect terms. A limitation of relying on $\boldsymbol{Q}$ to interpret the latent structure is that it does not provide specific details as to the exact underlying process regarding how attributes interact to impact observed responses. An alternative is to provide a more rich description of the latent structure by indicating the activeness of each coefficient in $\boldsymbol{B}$. In fact, Chen et al. (2020) introduced a $\boldsymbol{\Delta}$ matrix to provide more detailed information about the latent structure. Let $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_J)^\top$ be a $J \times 2^K$ binary matrix with row vectors corresponding to each item. That is, for each item the elements of $\boldsymbol{\delta}_j = (\delta_{j0}, \ldots, \delta_{j,2^K-1})$ indicate whether the corresponding elements of $\boldsymbol{\beta}$ are active or inactive. For instance, for $K = 3$,

$$\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7) \tag{14}$$

and each element indicates whether the corresponding element of $\boldsymbol{\beta}$ are active or inactive. For instance, the $\boldsymbol{\delta}$ vector corresponding with the $\boldsymbol{\beta}$ in Eq. 12 is

$$\boldsymbol{\delta} = (1, 1, 0, 0, 0, 0, 0, 0) \tag{15}$$

and the $\boldsymbol{\delta}$ for $\boldsymbol{\beta}$ in Eq. 13 is:

$$\boldsymbol{\delta} = (1, 1, 1, 1, 0, 0, 0, 0). \tag{16}$$

Note it is customary to always include the intercept in the model, so $\delta_{j0} = 1$ for all j.

It is important to note that $\boldsymbol{Q}$ describes the pattern of zeros and ones in $\boldsymbol{\Delta}$ in a similar manner to the active and inactive values of $\boldsymbol{B}$. Therefore, q_{jk} is associated with multiple elements in $\boldsymbol{\delta}_j$. In fact, Chen et al. (2020) note that there is a many-to-one mapping between $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$. We can demonstrate the many-to-one relationship between $\boldsymbol{\delta}$ and $\boldsymbol{q}$ by listing all of the $\boldsymbol{\delta}$'s that are consistent with a given $\boldsymbol{q}$. For instance, $\boldsymbol{q} = (0, 1, 1)$ corresponds with five distinct $\boldsymbol{\delta}$ vectors

$$(1, 0, 0, 1, 0, 0, 0, 0) \tag{17}$$
$$(1, 0, 1, 1, 0, 0, 0, 0) \tag{18}$$
$$(1, 1, 0, 1, 0, 0, 0, 0) \tag{19}$$
$$(1, 1, 1, 0, 0, 0, 0, 0) \tag{20}$$
$$(1, 1, 1, 1, 0, 0, 0, 0). \tag{21}$$

The $\boldsymbol{q} = (0, 1, 1)$ vector implies that both attributes two and three are active in some way and is consistent with the $\boldsymbol{\delta}$ in: (17) where only the interaction effect β_3 for attributes two and three is active; (18) where the both the main effect β_2 for α_2 and the two-way interaction are active; (19) which corresponds to an active main effect β_1 for α_3 and the two-way interaction; (20) where only the main-effects for α_2 and α_3 are active; and (21) where both main-effects and the two-way interaction are active. Notice that a common feature to the five $\boldsymbol{\delta}$'s is that coefficients involving the first attribute are all inactive as specified by $\boldsymbol{q} = (0, 1, 1)$. Furthermore, this example shows the difference in information provided by $\boldsymbol{q}$ and $\boldsymbol{\delta}$. The benefit of considering $\boldsymbol{\delta}$ is that it provides greater detail about exactly how the attributes combine to impact affirmative response probabilities. In contrast, $\boldsymbol{q}$ only indicates if an attribute is needed for an item and does not provide information as to how an attribute relates to an item in terms of main-effects and interactions. In short, both $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ provide researchers with information to interpret the latent structure and a contribution of our paper is that we provide a unifying framework for jointly modeling these parameters.

1.4. Identifiability

Model parameter identifiability is a critical issue for inferring latent structure with RLCMs. Several studies established model parameter identifiability conditions for RLCMs (e.g., see Chen et al., 2015, 2020; Culpepper, 2019b; Gu & Xu 2020; 2021; Xu, 2017; Xu & Shang, 2018). The identifiability conditions established in prior research specify conditions that $\boldsymbol{Q}$ must satisfy in order for the model parameters to be identified. For instance, Theorem 2 of Culpepper (2019b) indicated that the model parameters are strictly identified (i.e., the likelihood is unique for each set of parameter values) if $\boldsymbol{Q}$ is in the identifiable set $\mathcal{Q}$ that satisfies the following conditions:

(C1) The rows of $\boldsymbol{Q}$ can be permuted to the form, $\boldsymbol{Q}^\top = [\boldsymbol{I}_K, \boldsymbol{I}_K, (\boldsymbol{Q}^*)^\top]^\top$ where $\boldsymbol{I}_K$ is a K-dimensional identity matrix and $\boldsymbol{Q}^*$ is a $(J - 2K) \times K$ matrix.

(C2) For any two latent classes c and c', there exists at least one item in $\boldsymbol{Q}^*$, in which $\boldsymbol{\theta}_{jc} \neq \boldsymbol{\theta}_{jc'}$.

Stated differently, C1 states that $\boldsymbol{Q}$ must include two simple structure items for each attribute and C2 indicates that there is at least one item not specified for C1 that distinguishes between all pairs of classes. Additional advances regarding the identification of RLCMs are discussed in Gu and Xu (2021, 2020).

The aforementioned identifiability conditions are stated in terms of the $\boldsymbol{Q}$ matrix and do not directly provide information about identifiability conditions for $\boldsymbol{\Delta}$. Note that Chen et al. (2020) present conditions that $\boldsymbol{\Delta}$ must satisfy to ensure identifiability (e.g., see Theorem 2). Furthermore, Chen et al. (2020) report weaker conditions for identifying model parameters, referred to as generic identifiability. The notion of generic identifiability is more technical and involves topics such as algebraic varieties and measure zero sets, and it essentially states that if certain conditions on $\boldsymbol{\Delta}$ are satisfied, the set of non-identified parameter values resides in a subset of the parameter space and is therefore unlikely to arise in applied computations.

1.5. Monotonicity

Monotonicity conditions are important for ensuring interpretability of the exploratory solution. Xu (2017) defined the monotonicity conditions for a given item of the general diagnostic model as:

$$\min_{c \in \mathbb{S}_0} \theta_{jc} \geq \theta_{j0} \tag{22}$$

$$\max_{c \in \mathbb{S}_0} \theta_{jc} < \min_{c \in \mathbb{S}_1} \theta_{jc} = \max_{c \in \mathbb{S}_1} \theta_{jc} \tag{23}$$

where $\mathbb{S}_0 = \{c : \boldsymbol{\alpha}^\top \boldsymbol{v} = c, \alpha_k < q_{jk} \text{ for } k = 1, \ldots, K\}$ is the set of classes without the required attributes specified by $\boldsymbol{q}_j$ and $\mathbb{S}_1 = \{c : \boldsymbol{\alpha}^\top \boldsymbol{v} = c, \alpha_k \geq q_{jk} \text{ for } k = 1, \ldots, K\}$ is the set of classes with the necessary attributes. In short, Eq. 22 states that the class without any attributes has the lowest response probability and Eq. 23 indicates that: (1) classes without at least one required attribute have a lower response probability than classes with the necessary attributes, and (2) classes that possess all of the necessary attributes have the same response probability.

The monotonicity conditions in Eqs. 22 and 23 are stated in terms of elements of $\boldsymbol{\Theta}$ rather than $\boldsymbol{B}$. Chen et al. (2020) established a monotonicity condition in terms of elements of $\boldsymbol{\beta}_j$ rather than $\boldsymbol{\theta}_j$. In fact, Chen et al. (2020) showed that it is possible to enforce the monotonicity condition by sequentially updating elements of $\boldsymbol{\beta}_j$ subject to a lower bound condition such that $\beta_{jp} > L_{jp}$ where L_{jp} is a function of $\boldsymbol{\beta}_{j(p)}$, which includes all of the elements of $\boldsymbol{\beta}_j$ except element p (see Proposition 1 of Chen et al., 2020 for specific details on the computation of L_{jp}).

As demonstrated in prior research (Culpepper, 2019a; Chen et al., 2020), Bayesian approaches provide a convenient computational framework for inferring the latent structure while enforcing monotonicity conditions; however, there are limitations to existing methods. Chen et al. (2020) report a Gibbs sampling algorithm to estimate $\boldsymbol{\Delta}$ that enforces the monotonicity conditions in Eqs. 22 and 23. However, a drawback is that the algorithm is designed to infer $\boldsymbol{\Delta}$ and not $\boldsymbol{Q}$. Researchers have been traditionally more interested in $\boldsymbol{Q}$ than $\boldsymbol{\Delta}$, and Monte Carlo simulations suggest it is somewhat more difficult to accurately estimate the $J \cdot 2^K$ elements of $\boldsymbol{\Delta}$ in an unstructured fashion than the $J \cdot K$ elements of $\boldsymbol{Q}$. In response, Chen et al. (2020) proposed a post-processing step to obtain a point estimate for $\boldsymbol{Q}$ from the posterior samples of $\boldsymbol{B}$. The method demonstrated accurate performance in the studied Monte Carlo conditions, but the heuristic could fail in cases where items possess a weaker signal for the relationship between attributes and responses (i.e., smaller values for $\boldsymbol{\beta}_j$). Furthermore, the post-processing method does not provide a mechanism for incorporating prior expert knowledge about $\boldsymbol{Q}$. An alternative was proposed by Culpepper (2019a) who directly estimated $\boldsymbol{Q}$. A limitation though of this latter approach is that a more restrictive monotonicity conditions was imposed (i.e., the elements of $\boldsymbol{B}$ were restricted to be positive) in order to approximate the posterior distribution. Whereas the approach taken was accurate for inferring $\boldsymbol{Q}$ and appropriate for main-effects, which must always be positive, the restriction implied the algorithm could not uncover disjunctive relationships (i.e., cases where an interaction effect is negative), which may be present more generally in psychological data (e.g., see Templin & Henson, 2006).

In short, enforcing monotonicity is a critical issue for exploratory RLCMs. The next section presents a solution to address shortcomings of existing methods.

2. Bayesian Formulation

We next introduce a new Bayesian formulation that: (1) enforces latent monotonicity conditions in Eqs. 22 and 23; (2) jointly infers $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$ from the posterior distribution; and (3) uses a logistic IRF with the Pòlya–gamma data augmentation strategy to provide more efficient computations for larger sample sizes. The first subsection outlines the Bayesian model formulation. The second subsection discusses a Gibbs sampling algorithm for approximating the model parameter posterior distribution.

2.1. Bayesian Formulation for the Logit-Link General Diagnostic Model

Figure 1 illustrates a directed acyclic graph of the new Bayesian model and shows the hierarchical structure of the model parameters. We next describe various components of the new model formulation by discussing models for: (1) the observed response Y_{ij} and the augmented response

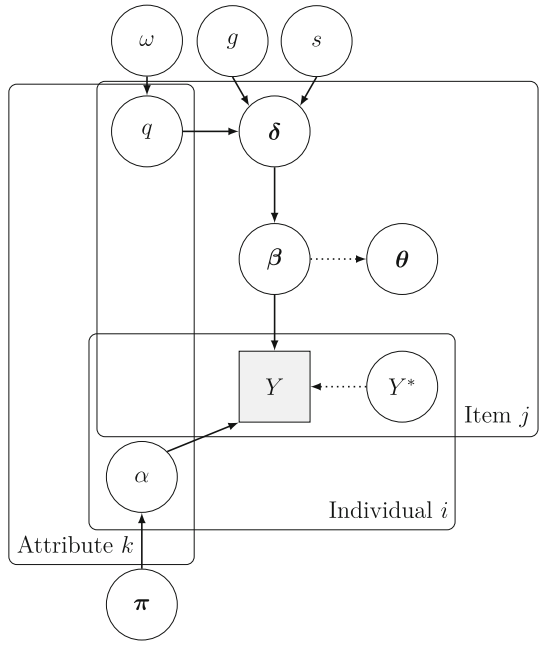

FIGURE 1.
Directed acyclic graph of Bayesian model for the logit-link general diagnostic model. *Note* The square node denotes an observed random variable, circles indicate unknown model parameters, solid arrows denote stochastic relationships, the dashed arrow indicates a deterministic relationship, and the plates with rounded corners indicate parameters that share a common subscript

Y_{ij}^*; (2) the attribute profiles $\boldsymbol{\alpha}_i$ and the latent class probabilities $\boldsymbol{\pi}$; (3) the item parameters $\boldsymbol{\beta}_j$; (4) the item activeness indicators $\boldsymbol{\delta}_j$ and associated hyper-parameters s and g; and (5) the structure vector $\boldsymbol{q}_j$ and its hyper-parameter ω.

2.1.1. Model for Y_{ij} Figure 1 shows that Y_{ij} lies within the intersection of the individual and item plates. Furthermore, the Pòlya–gamma random variable Y_{ij}^* is also specific to each item and individual. We use a logistic cdf for the IRF. The model for Y_{ij} conditioned upon membership in class c and the item parameter vector $\boldsymbol{\beta}_j$ is

$$Y_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j \sim \text{Bernoulli}\left(\prod_{c=0}^{2^K-1} \theta_{jc}^{\mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v}=c)}\right) \tag{24}$$

$$\theta_{jc} = \frac{\exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)}{1+\exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)}. \tag{25}$$

Polson et al. (2013a) showed that the model in Eqs. 24 and 25 can be augmented with the following integral identity

$$\frac{\left(e^\psi\right)^a}{\left(1+e^\psi\right)^b} = 2^{-b} e^{\kappa\psi} \int_0^\infty e^{-w\psi^2/2} p(w) \mathrm{d}w \tag{26}$$

for $b > 0, \kappa = a - b/2$, and w distributed as a Pòlya–gamma random variable (i.e., $w \sim \text{PG}(b, 0)$). We apply the Pòlya–gamma data augmentation for the model in Eqs. 24 and 25 by noting that $\psi = \boldsymbol{a}_c^\top \boldsymbol{\beta}_j$ and specifying the Pòlya–gamma distribution for the augmented data, $Y_{ij}^* \sim \text{PG}(b, 0)$. Polson et al. (2013a) showed that augmenting the likelihood leads to an efficient Gibbs sampling algorithm.

We can see why the Pòlya–gamma approach requires fewer samples by examining the conditional likelihood for a sample of responses to a given item j

$$\begin{aligned} p\left(\boldsymbol{y}_{1:n,j} | \boldsymbol{\alpha}_{1:n}, \boldsymbol{\beta}_j\right) &= \prod_{i=1}^{n} \left[\left(\prod_{c=0}^{2^K-1} \frac{\exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)}{1 + \exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)} \right)^{\mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)} \right]^{y_{ij}} \\ &\quad \left[\left(\prod_{c=0}^{2^K-1} \frac{1}{1 + \exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)} \right)^{\mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)} \right]^{1-y_{ij}} \\ &= \prod_{i=1}^{n} \prod_{c=0}^{2^K-1} \left[\frac{\left(\exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)\right)^{y_{ij}}}{1 + \exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)} \right]^{\mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)} \end{aligned} \tag{27}$$

where $\boldsymbol{y}_{1:n,j} = (y_{1j}, \ldots, y_{nj})^\top$. Recall that $\boldsymbol{\alpha}_i$ equals one of 2^K possible response patterns, which implies that there are groups of terms in the product of Eq. 27 that are equal. We can collapse the conditional likelihood for item j over individuals in the same class to yield

$$p\left(\boldsymbol{y}_{1:n,j} | \boldsymbol{\alpha}_{1:n}, \boldsymbol{\beta}_j\right) = \prod_{c=0}^{2^K-1} \frac{\left(\exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)\right)^{n_{jc}}}{\left(1 + \exp\left(\boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)\right)^{n_c}} \tag{28}$$

where $n_c = \sum_{i=1}^{n} \mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)$ is the number of individuals in class c and $n_{jc} = \sum_{i=1}^{n} \mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c) y_{ij}$ is the number of individuals in class c with a response equal to one on item j. Consequently, we can use the Pòlya–gamma identity in Eq. 26 to augment the likelihood for each class and item rather than each individual and item by defining a $\psi_{jc} = \boldsymbol{a}_c^\top \boldsymbol{\beta}_j$, $a = n_{jc}$, and $b = n_c$ in Eq. 26.

2.1.2. Models for $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$ Figure 1 shows that attributes impact observed responses. Similar to Culpepper (2015), we use a categorical prior for $\boldsymbol{\alpha}_i$ conditioned upon the latent class structural probabilities $\boldsymbol{\pi}$. The conditional probability that $\boldsymbol{\alpha}_i$ has the attribute configuration of class c given $\boldsymbol{\pi}$ is $\text{P}\left(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c | \boldsymbol{\pi}\right) = \pi_c$.

We adopt an unstructured model for the latent class probabilities[1] and specify a semi-conjugate Dirichlet prior

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{n}_0) \tag{29}$$

with prior parameter $\boldsymbol{n}_0$. Note we implement a uniform prior for $\boldsymbol{\pi}$ and set the elements of $\boldsymbol{n}_0$ equal to one (i.e., $\boldsymbol{n}_0 = \mathbf{1}_{2^K}$).

[1]Note there are alternatives to the unstructured model. Researchers considered a higher-order structure where the attributes load onto a continuous trait (de la Torre & Douglas, 2004) as well as a multivariate probit model (Henson et al., 2009; Templin et al., 2008) with a tetrachoric correlation matrix (see related Bayesian formulations in Chen & Culpepper, 2020; Culpepper & Chen, 2019).

2.1.3. Models for $\boldsymbol{\beta}_j$ We consider two variable selection priors for $\boldsymbol{\beta}_j$. The first prior is the stochastic search variable selection (SSVS) prior specification for $\boldsymbol{\beta}_j$ (e.g., see Culpepper, 2019a), and the second is a spike-slab mixture distribution with a Dirac delta function (e.g., see Chen et al., 2020).

SSVS Prior The prior for $\boldsymbol{\beta}_j$ conditioned on $\boldsymbol{\delta}_j$ is a truncated multivariate normal distribution

$$\boldsymbol{\beta}_j|\boldsymbol{\delta}_j \sim \mathcal{N}_P\left(\mathbf{0}, \boldsymbol{\Sigma}_j\right) \mathbb{1}(\boldsymbol{\beta}_j \in \mathcal{B}) \tag{30}$$

where $\mathcal{B}$ is the set of coefficients that satisfy the monotonicity condition, $P = 2^K$, the mean vector is zero, and $\boldsymbol{\Sigma}_j$ is the variance-covariance matrix. Note that $\boldsymbol{\Sigma}_j$ is a diagonal matrix with the variances specified as a function of the elements of $\boldsymbol{\delta}_j$,

$$\boldsymbol{\Sigma}_j = \text{diag}(\sigma^2_{j0}, \ldots, \sigma^2_{j,P-1}) \tag{31}$$

$$\sigma^2_{jp} = \delta_{jp}/c_1 + (1 - \delta_{jp})/c_0. \tag{32}$$

The prior for $\boldsymbol{\beta}_j$ in Eq. 30 is a spike-slab prior used to infer active and inactive elements of $\boldsymbol{\beta}_j$ with the SSVS algorithm. That is, the precision for β_{jp} (i.e., the inverse of its variance) is c_1 when $\delta_{jp} = 1$ and c_0 if $\delta_{jp} = 0$. The SSVS approach uses fixed values for the constants c_0 and c_1. Specifically, c_0 is set to a large value (e.g., we set $c_0 = 100/3$ in this paper) to reflect a smaller variance for β_{jp} and a distribution that is more concentrated near zero. Coefficients with smaller variances are therefore interpreted as inactive. In contrast, c_1 is fixed as a smaller value (e.g., we set $c_1 = 1$) to depict a prior distribution for β_{jp} with a larger variance that is consistent with an active coefficient.

Dirac Delta Prior The second prior is a spike-slab prior as described by Chen et al. (2020). The prior formulation in Eq. 30 does not penalize inactive elements of $\boldsymbol{\beta}_j$ exactly to zeros. Instead, the inactive elements will be near zero. In contrast, Chen et al. (2020) used a Dirac delta function for the spike distribution to shrink inactive coefficients exactly to zero. More formally, the conditional spike-slab prior for β_{jp} and $p > 0$ is

$$\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp} \sim \begin{cases} \mathbb{1}(\beta_{jp} = 0) & \delta_{jp} = 0 \\ \mathcal{N}(0, \sigma^2_{jp})\mathbb{1}(\beta_{jp} > L_{jp}) & \delta_{jp} = 1 \end{cases} \tag{33}$$

where σ^2_{jp} is the prior variance for the slab that is fixed to a constant (e.g., see a discussion of our implementation in the description of the simulation study) and L_{jp} is the lower-bound for β_{jp} given $\boldsymbol{\beta}_{j(p)}$ (see Proposition 1 of Chen et al. 2020 for a definition of L_{jp}).

2.1.4. Model for $\boldsymbol{\delta}_j$ and $\boldsymbol{q}_j$ We next describe a novel aspect of our prior specification that enables joint inference of $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$ while enforcing monotonicity constraints. We introduce a stochastic rather than deterministic relationship between elements of $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$. Furthermore, we use a logical "and" gate to describe the relationship between $\boldsymbol{\delta}_j$ and $\boldsymbol{q}_j$. The "and" gate is based on the idea that a given δ_{jp} will be active and equal to one with a higher probability if the associated elements of $\boldsymbol{q}_j$ are active and equal to one. Likewise, δ_{jp} should be zero with high probability to indicate an inactive coefficient when the corresponding elements of $\boldsymbol{q}_j$ are inactive.

We introduce additional notation in order to specify the logical "and" gate function. Let $\boldsymbol{Q}'$ be a $P \times K$ matrix that specifies which elements of $\boldsymbol{q}_j$ and $\boldsymbol{\delta}_j$ are linked. The (p, k) element of $\boldsymbol{Q}'$ is defined as $q'_{pk} = 1$ if q_{jk} is linked with δ_{jp} and zero otherwise. For instance, Table 1 presents the $\boldsymbol{Q}'$ for $K = 3$. Note the rows of Table 1 are in the order of the bijection with the first

TABLE 1.
Relationship between $\boldsymbol{\delta}$, $\boldsymbol{q}$, and $\boldsymbol{q}'$ for $K = 3$

				$\boldsymbol{Q}'$		
p	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\delta}$	q'_1	q'_2	q'_3
0	000	β_0	δ_0	0	0	0
1	001	β_1	δ_1	0	0	1
2	010	β_2	δ_2	0	1	0
3	011	β_3	δ_3	0	1	1
4	100	β_4	δ_4	1	0	0
5	101	β_5	δ_5	1	0	1
6	110	β_6	δ_6	1	1	0
7	111	β_7	δ_7	1	1	1

column including integers from zero to seven and the second column including the associated attribute profiles. The third and fourth columns list item parameters $\boldsymbol{\beta}$ and activeness indicators $\boldsymbol{\delta}$ in the order of the bijection, as well. The fifth through seventh columns of Table 1 include the $\boldsymbol{Q}'$ matrix. For instance, the $p = 1$ row of Table 1 corresponds with the main-effect β_1 for attribute three. In this case, δ_1 is only linked to q_3, so $\boldsymbol{q}'_1 = (0, 0, 1)$. Additionally, the interaction between attribute one and three is represented by the $p = 5$ row with the interaction-effect β_5. The activeness indicator δ_5 is therefore linked with both q_1 and q_3 and noted by $\boldsymbol{q}'_5 = (1, 0, 1)$. Finally, row $p = 7$ is the three-way interaction and δ_7 is related to all elements of $\boldsymbol{q}$ as indicated by $\boldsymbol{q}'_7 = (1, 1, 1)$.

Next, we specify a deterministic "and" gate to denote whether the elements of $\boldsymbol{q}_j$ that are linked to δ_{jp} through $\boldsymbol{q}'_p$ are active. That is, we define η_{jp} as

$$\eta_{jp} = \prod_{k=1}^{K} q_{jk}^{q'_{pk}} \tag{34}$$

where we use the convention that $0^0 = 1$, so that $\eta_{jp} = 1$ if all of the associated q_{jk}'s linked to δ_{jp} are one and zero if at least one associated q_{jk} is not one. We seek a stochastic relationship between $\boldsymbol{\delta}_j$ and $\boldsymbol{q}_j$, so we define $s = \mathrm{P}(\delta_{jp} = 0 | \eta_{jp} = 1)$ as the probability that $\delta_{jp} = 0$ when the associated elements of $\boldsymbol{q}_j$ are active and $g = \mathrm{P}(\delta_{jp} = 1 | \eta_{jp} = 0)$ as the probability that $\delta_{jp} = 1$ when at least one related q_{jk} is inactive. Note this is a DINA model as described in the cognitive diagnosis literature (Haertel, 1989) where s is referred to as the probability of "slipping" and g is the probability of "guessing."

Our confirmatory DINA model prior for relating $\boldsymbol{\delta}_j$ and $\boldsymbol{q}_j$ is

$$\delta_{jp} | \boldsymbol{q}_j, s, g \sim \text{Bernoulli}\left((1 - s)^{\eta_{jp}} g^{1-\eta_{jp}}\right). \tag{35}$$

Note we fix $\delta_{j0} = 1$ for all j and always allow the intercept to be active. We specify a truncated, bivariate, uniform prior for the s and g hyper-parameters

$$p(s, g) \propto \mathbb{1}(0 < g < 1 - s)\mathbb{1}(0 < s < 1). \tag{36}$$

That is, we enforce the monotonicity condition that $g < 1 - s$ (Junker & Sijtsma, 2001), so that $\boldsymbol{q}_j$ and $\boldsymbol{\delta}_j$ are positively related.

2.1.5. Model for $\boldsymbol{q}_j$ The final portion of our model corresponds to a prior for elements of $\boldsymbol{Q}$. Each q_{jk} is binary, so we choose the following Bernoulli distribution with probability ω,

$$q_{jk}|\omega \sim \text{Bernoulli}(\omega)\mathbb{1}(\boldsymbol{Q} \in \mathcal{Q}). \tag{37}$$

It is important to note that the prior for q_{jk} is restricted to the identifiable set of $\boldsymbol{Q}$ matrices, $\mathcal{Q}$ as described above. To complete the model, we use a semi-conjugate Beta prior for ω defined as $\omega \sim \text{beta}(a, b)$.

2.2. Posterior Approximation

The posterior distribution of interest is:

$$p(\boldsymbol{y}^*_{1:n}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{B}, \boldsymbol{\pi}, \boldsymbol{\Delta}, \boldsymbol{Q}, s, g, \omega|\boldsymbol{y}_{1:n}) \tag{38}$$

where $\boldsymbol{y}^*_{1:n}$ is the $n \times J$ matrix of augmented Pòlya–gamma random variables. Note we use a Gibbs sampling Markov chain Monte Carlo (MCMC) algorithm to infer parameters from the posterior distribution. We discuss details for implementing the Gibbs sampling algorithm in "Appendix A." Note we consider two different prior specifications for $\boldsymbol{B}$ and provide details about full conditional distributions for the SSVS approach in "Appendix A," the full conditional distribution for $\boldsymbol{\beta}_j$ using a logit-link in "Appendix B," and the Dirac delta prior in "Appendix C."

3. Monte Carlo Simulation Study

3.1. Overview

We conducted a Monte Carlo simulation to contribute new knowledge about the performance of RLCMs in recovering the latent structure. Specifically, we compare the performance and accuracy of our novel formulation that uses a hierarchical prior involving $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$ with Culpepper (2019a), which we refer to as "probit-Restricted" as it only uses $\boldsymbol{Q}$ and enforces more restrictive monotonicity conditions. Furthermore, we report evidence of model accuracy for four different versions of our algorithm for two different link functions (i.e., probit and logit) and the two prior specifications for $\boldsymbol{B}$ (i.e., SSVS and Dirac delta).

We studied the performance of the algorithms by manipulating three design characteristics: (1) sample size, n; (2) the number of attributes, K; and (3) the correlation structure among attributes, ρ. We considered three sample sizes of $n = 500$, 1000, and 2000 as well as four conditions for $K = 2$, 3, 4, and 5. Attributes were generated using a multivariate normal probit model following Chiu et al. (2009) with a population tetrachoric correlation set to $\rho = 0$, 0.15, or 0.25 with thresholds defined by $\Phi^{-1}(k/(K+1))$ for $k = 1, \ldots, K$. The true elements of $\boldsymbol{\Theta}$ were generated by following the Xu and Shang (2018) approach of computing each element of $\boldsymbol{\Theta}$ as $0.2 + (0.8 - 0.2) \times K'_j/K_j$, where K_j represents the number of attributes as required by item j and K'_j is number of K_j attributes given by a specific class.

Recovery performance for each condition was assessed by comparing the estimated parameter values with the oracle population values used to generate the simulation data. Specifically, the model parameters are identified up to swapping of class labels, so in order to evaluate parameter recovery the estimated values were first oriented according to the oracle values by permuting the

estimated $\widehat{\boldsymbol{\Theta}}$ matrix against the oracle $\boldsymbol{\Theta}$ matrix. Once the permutation order was determined, each parameter of interest was re-oriented under the permutation and then recovery metrics were computed.

Permutation order used was obtained by minimizing the mean absolute deviation ("MAD") difference between the estimated and population IRFs across $t = 1, \ldots, T$ iterations from the posterior. In particular,

$$MAD(\hat{\boldsymbol{\theta}}_{jc}, \boldsymbol{\theta}_{jc}) = \frac{1}{T} \sum_{t=1}^{T} \left| \hat{\theta}_{jc}^{(t)} - \theta_{jc}^{(t)} \right| \tag{39}$$

To condense the result, an overall expected mean average deviation ("EMAD") was taken across the matrix.

$$EMAD(\hat{\boldsymbol{\Theta}}, \boldsymbol{\Theta}) = \frac{1}{2^K J} \sum_{j=0}^{J-1} \sum_{c=0}^{2^K-1} MAD(\hat{\boldsymbol{\theta}}_{jc}, \boldsymbol{\theta}_{jc}) \tag{40}$$

For both the $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ matrix, the element-wise recovery rates were computed. Under the element-wise metric, the proportion of correctly estimated elements within the matrix is calculated by first constructing the posterior mode across $t = 1, \ldots, T$ iterations from the posterior yielding $\overline{\boldsymbol{Q}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Q}^{(t)}$ and $\overline{\boldsymbol{\Delta}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Delta}^{(t)}$. Individual matrix element estimates are then obtained by using the posterior mode under the decision rule of 0.5. In the case of the $\boldsymbol{Q}$ matrix, the posterior estimated element-wise mode is defined as $\hat{q}_{jk} = \mathcal{I}\left(\overline{q}_{jk} > 0.5\right)$. Similarly, for $\boldsymbol{\Delta}$ matrix, the posterior element-wise mode would be $\hat{\delta}_{jp} = \mathcal{I}\left(\overline{\delta}_{jp} > 0.5\right)$. With the posterior estimated modes in hand, the element-wise recovery rate (RR) comparison against the oracle for $\boldsymbol{Q}$ is then computed as

$$RR(\hat{\boldsymbol{Q}}, \boldsymbol{Q}) = \frac{1}{JK} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \mathcal{I}\left(\hat{q}_{jk} = q_{jk}\right). \tag{41}$$

Similarly, for $\boldsymbol{\Delta}$ we have

$$RR(\hat{\boldsymbol{\Delta}}, \boldsymbol{\Delta}) = \frac{1}{J\left(2^K - 1\right)} \sum_{j=0}^{J-1} \sum_{c=1}^{2^K-1} \mathcal{I}\left(\hat{\delta}_{jc} = \delta_{jc}\right). \tag{42}$$

Note, unlike with the $\boldsymbol{Q}$ matrix, the $\boldsymbol{\Delta}$ matrix assumes that the first column containing the intercept will always be active. To fairly assess recovery, we then remove the first column from consideration yielding only $2^K - 1$ columns compared instead of 2^K.

Additional work was done to understand different sampling techniques raised in Windle et al. (2014). Details on how each sampling routine performed are found in "Appendix D." As a result of this work, the underlying sampler for Pòlya–gamma augmented data is taken to be the Saddlepoint approximation approach as it performed the best across varying study conditions in recovery, speed, and stability.

Moreover, in a separate simulation found in "Appendix E," we considered how the estimation methods performed when a disjunctive relationship was present in the data. When the disjunction

was present, the newly derived exploratory variants performed better than the existing probit-Restricted and on par with the SLCM. There was evidence the Delta variants had improved recovery of $\boldsymbol{\Delta}$ in comparison to the SLCM. Thus, if there is suspicion that the data have a disjunctive relationship, it would be advantageous to use the newly derived exploratory Delta and Dirac approaches.

We also considered additional values for dependency between attributes. In particular, we examined additional cases where $\rho = 0.30, 0.35, 0.40, 0.45, 0.50$. Results from this set of simulations are found in "Appendix G." The simulation displays a slight decrease to the existing strong recovery rates as correlations increase. This is inline with what the underlying theory suggests when using an over-parameterized Dirichlet prior for $\boldsymbol{\pi}$ to estimate the attribute dependence. Within the discussion section, we review a potential future methodological approach for addressing attribute dependencies that are described by a factor model.

The simulation was performed using "R" 3.6.2 on a cluster with varying node configurations (R Core Team, 2020). Due to the varying node configurations, the algorithm's run time is more variable. Each algorithm was implemented in "C++" through "armadillo" linear algebra library (Sanderson & Curtin, 2016) and made available in "R" through the combination of "RcppArmadillo" and "Rcpp" packages (Eddelbuettel & Sanderson, 2014; Eddelbuettel & Balamuta, 2018; Eddelbuettel, 2013; Eddelbuettel & François, 2011).

Note we implemented the algorithms with a chain length of 60,000 iterations and 20,000 burn-in iterations. Furthermore, our SSVS algorithm requires researchers to prespecify values for the prior precision parameters, c_0 and c_1, to define active and inactive coefficients. We follow Culpepper (2019a) and use the values $c_0 = 100$ and $c_1 = 1$ for the implementation of the SSVS with a probit-link function and we use $c_0 = 100/3$ and $c_1 = 1/3$ for the logit-link to align with the logistic distribution's variance of $\frac{\pi^2}{3}$. Furthermore, for the Dirac variant of our algorithm we set $c_1 = 2$ for the probit-link and $c_1 = 2/3$ for the logit-link.

3.2. Results

The Monte Carlo simulation results provide ample evidence to the effectiveness of item and structural parameters recovery across all sample sizes. Within Table 2, the average mean absolute deviation (MAD) for the $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ are reported. Specifically, Table 2 shows that average MAD for the probit variants is similar to that of the logit variants when the sample size is large $n = 1000$ and 2000 alongside small $K = 2$ and 3. As the number of attributes increases, the probit variants perform better than the logit variants.

Furthermore, Table 3 contains the accuracy of the posterior element-wise mean for both $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$. In fact, the element-wise recovery rate for $\boldsymbol{Q}$ for the proposed methods exceeds that of the probit-Restricted and SLCM for the majority of studied conditions. However, for the $K = 2$ attributes case and of sample size $N = 500$ and 1000, the proposed Delta and Dirac variants have a lower recovery rate for the $\boldsymbol{\Delta}$ matrix in comparison to the SLCM's recovery rate of 88%. Partially, this is because of the selected spike and slab precision parameters. In particular, we conducted additional simulations and found higher recovery for $\boldsymbol{\Delta}$ when $K = 2$ under a smaller slab precision (i.e., $c_1 < 0.25$). Whereas the recover of $\boldsymbol{\Delta}$ is lower for the proposed methods when $K = 2$, it is important to note that the results in Table 2 suggest that the recovery of $\boldsymbol{\Theta}$ is comparable between methods. As N and K increase, the Delta and Dirac variants considerably improve on both the probit-Restricted and SLCM formulations. Furthermore, the logit variants perform on par with the probit variants when estimating entries in $\boldsymbol{\Delta}$.

Having addressed recovery performance, our attention now turns to the computational speed of the algorithms. Reported in Table 4 are the average run times of the algorithm over the considered simulation settings. The logit variant performs over one-half faster than the probit variants for large n. However, as K increases to $K = 5$, the logit speed gains are slowly diluted when compared

TABLE 2.
Summary of exploratory general diagnostic models under Delta and Dirac augmentations ability to estimate $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ measured by the average mean absolute deviation (MAD) across 100 replications for each sample size, N, attributes, K, and attribute dependence, ρ

				Logit				Probit							
				Delta		Dirac		Delta		Dirac		Restricted		SLCM	
N	K	J	ρ	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$
500	2	12	0.00	0.04	0.03	0.04	0.02	0.04	0.03	0.04	0.02	0.04	0.02	0.04	0.02
			0.15	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02
			0.25	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02
500	3	20	0.00	0.05	0.02	0.05	0.02	0.05	0.02	0.05	0.02	0.05	0.02	0.05	0.02
			0.15	0.05	0.02	0.05	0.02	0.05	0.02	0.05	0.02	0.05	0.02	0.05	0.02
			0.25	0.06	0.02	0.05	0.02	0.06	0.03	0.05	0.02	0.05	0.02	0.05	0.02
500	4	20	0.00	0.07	0.02	0.05	0.02	0.06	0.02	0.05	0.02	0.06	0.02	0.06	0.02
			0.15	0.07	0.02	0.06	0.02	0.07	0.02	0.06	0.01	0.06	0.02	0.07	0.02
			0.25	0.07	0.02	0.06	0.02	0.07	0.02	0.06	0.02	0.06	0.02	0.07	0.02
500	5	30	0.00	0.10	0.02	0.08	0.01	0.10	0.02	0.08	0.01	0.09	0.02	0.10	0.02
			0.15	0.10	0.02	0.08	0.02	0.10	0.02	0.08	0.02	0.09	0.02	0.10	0.02
			0.25	0.11	0.02	0.08	0.02	0.10	0.02	0.08	0.02	0.09	0.02	0.11	0.02
1000	2	12	0.00	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02
			0.15	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02
			0.25	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02
1000	3	20	0.00	0.04	0.02	0.03	0.01	0.04	0.02	0.03	0.01	0.04	0.01	0.03	0.01
			0.15	0.04	0.02	0.03	0.01	0.04	0.02	0.03	0.01	0.04	0.01	0.03	0.01
			0.25	0.04	0.02	0.03	0.01	0.05	0.03	0.04	0.02	0.04	0.02	0.04	0.02
1000	4	20	0.00	0.05	0.01	0.04	0.01	0.05	0.01	0.04	0.01	0.05	0.01	0.04	0.01
			0.15	0.05	0.02	0.04	0.01	0.05	0.02	0.04	0.01	0.05	0.01	0.05	0.01
			0.25	0.05	0.02	0.04	0.01	0.05	0.02	0.04	0.01	0.05	0.02	0.05	0.01
1000	5	30	0.00	0.08	0.01	0.05	0.01	0.08	0.01	0.05	0.01	0.07	0.01	0.07	0.01
			0.15	0.08	0.02	0.05	0.01	0.08	0.01	0.05	0.01	0.07	0.01	0.07	0.02
			0.25	0.08	0.02	0.05	0.01	0.08	0.02	0.06	0.01	0.08	0.02	0.08	0.02
2000	2	12	0.00	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
			0.15	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
			0.25	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
2000	3	20	0.00	0.03	0.01	0.02	0.01	0.03	0.01	0.02	0.01	0.03	0.01	0.02	0.01
			0.15	0.03	0.01	0.02	0.01	0.04	0.02	0.02	0.01	0.03	0.01	0.02	0.01
			0.25	0.04	0.02	0.03	0.01	0.04	0.02	0.04	0.02	0.03	0.01	0.04	0.02
2000	4	20	0.00	0.04	0.01	0.03	0.01	0.04	0.01	0.03	0.01	0.04	0.01	0.03	0.01
			0.15	0.04	0.01	0.03	0.01	0.04	0.01	0.03	0.01	0.04	0.01	0.04	0.01
			0.25	0.05	0.02	0.03	0.01	0.05	0.02	0.04	0.02	0.04	0.01	0.04	0.02
2000	5	30	0.00	0.07	0.01	0.03	0.01	0.06	0.01	0.04	0.01	0.06	0.01	0.05	0.01
			0.15	0.07	0.01	0.04	0.01	0.07	0.01	0.04	0.01	0.06	0.01	0.06	0.01
			0.25	0.07	0.01	0.05	0.01	0.07	0.01	0.05	0.01	0.06	0.01	0.07	0.02

against the probit variants. In comparison with the probit-Restricted, all algorithms take double the amount of time due to the calculation of the $\boldsymbol{\Delta}$ matrix.

TABLE 3.
Summary of exploratory general diagnostic models under Delta and Dirac augmentations ability to estimate $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ matrices as measured by element-wise accuracy across sample size, N, attributes, K, and attribute dependence, ρ

				Logit				Probit						
				Delta		Dirac		Delta		Dirac		Restricted	SLCM	
N	K	J	ρ	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$
500	2	12	0.00	0.96	0.80	0.91	0.74	0.95	0.79	0.90	0.72	0.99	1.00	0.91
			0.15	0.96	0.79	0.91	0.73	0.95	0.78	0.89	0.71	0.99	0.99	0.92
			0.25	0.95	0.79	0.90	0.71	0.95	0.78	0.88	0.69	0.99	1.00	0.91
500	3	20	0.00	0.95	0.86	0.97	0.88	0.96	0.87	0.97	0.88	0.94	0.97	0.82
			0.15	0.95	0.86	0.97	0.88	0.96	0.87	0.96	0.87	0.94	0.97	0.82
			0.25	0.95	0.85	0.97	0.87	0.94	0.84	0.96	0.86	0.94	0.97	0.82
500	4	20	0.00	0.91	0.89	0.93	0.90	0.92	0.90	0.93	0.90	0.92	0.93	0.84
			0.15	0.92	0.90	0.93	0.90	0.92	0.90	0.93	0.90	0.91	0.93	0.85
			0.25	0.91	0.90	0.93	0.90	0.92	0.90	0.92	0.90	0.91	0.92	0.84
500	5	30	0.00	0.86	0.93	0.90	0.94	0.87	0.93	0.89	0.94	0.86	0.81	0.86
			0.15	0.85	0.93	0.89	0.94	0.86	0.93	0.88	0.94	0.85	0.80	0.86
			0.25	0.85	0.93	0.88	0.94	0.85	0.93	0.88	0.94	0.85	0.80	0.86
1000	2	12	0.00	0.99	0.82	0.96	0.79	0.98	0.82	0.95	0.77	1.00	1.00	0.92
			0.15	0.99	0.82	0.95	0.78	0.98	0.82	0.93	0.77	1.00	1.00	0.91
			0.25	0.98	0.82	0.95	0.77	0.98	0.81	0.94	0.76	0.99	1.00	0.92
1000	3	20	0.00	0.98	0.89	0.99	0.92	0.98	0.89	0.99	0.92	0.97	0.99	0.83
			0.15	0.98	0.88	0.99	0.92	0.98	0.88	0.99	0.92	0.97	0.99	0.84
			0.25	0.97	0.88	0.99	0.91	0.96	0.86	0.98	0.89	0.97	0.98	0.83
1000	4	20	0.00	0.92	0.89	0.94	0.91	0.93	0.89	0.94	0.91	0.92	0.97	0.85
			0.15	0.93	0.90	0.95	0.91	0.93	0.90	0.95	0.91	0.93	0.96	0.85
			0.25	0.93	0.90	0.94	0.91	0.93	0.90	0.94	0.91	0.93	0.96	0.84
1000	5	30	0.00	0.91	0.94	0.94	0.95	0.91	0.94	0.94	0.95	0.90	0.89	0.87
			0.15	0.90	0.94	0.94	0.95	0.91	0.94	0.94	0.95	0.90	0.88	0.87
			0.25	0.90	0.94	0.94	0.95	0.90	0.94	0.93	0.95	0.89	0.86	0.87
2000	2	12	0.00	0.99	0.92	0.97	0.84	0.99	0.92	0.97	0.84	1.00	1.00	0.91
			0.15	1.00	0.92	0.97	0.85	1.00	0.91	0.97	0.84	1.00	1.00	0.91
			0.25	1.00	0.90	0.97	0.82	1.00	0.89	0.96	0.81	1.00	1.00	0.91
2000	3	20	0.00	1.00	0.90	0.99	0.92	0.99	0.90	0.99	0.92	0.99	1.00	0.84
			0.15	0.99	0.89	0.99	0.92	0.98	0.89	0.99	0.92	0.99	0.99	0.84
			0.25	0.98	0.88	0.98	0.91	0.97	0.87	0.97	0.89	0.98	0.96	0.83
2000	4	20	0.00	0.92	0.89	0.95	0.91	0.92	0.89	0.96	0.92	0.92	0.99	0.85
			0.15	0.93	0.89	0.96	0.92	0.93	0.90	0.96	0.92	0.93	0.97	0.85
			0.25	0.93	0.90	0.96	0.91	0.93	0.90	0.95	0.91	0.93	0.96	0.84
2000	5	30	0.00	0.93	0.94	0.97	0.96	0.94	0.95	0.97	0.96	0.93	0.94	0.87
			0.15	0.93	0.94	0.97	0.95	0.94	0.94	0.96	0.95	0.92	0.92	0.87
			0.25	0.92	0.94	0.95	0.95	0.93	0.94	0.94	0.95	0.91	0.89	0.87

When computing recovery rates for $\boldsymbol{\Delta}$ matrix, the first column is omitted as it is assumed to be always present. In addition, the restricted model only estimates the $\boldsymbol{Q}$ matrix and, thus, we omit $\boldsymbol{\Delta}$ recovery column.

4. Application: Last Series of the Standard Progressive Matrices ("SPM-LS")

4.1. Overview

Within this section, the probit and logit model variants are applied to $n = 499$ responses across $J = 12$ items on the last series of the Standard Progressive Matrices ("SPM-LS," see Raven, 1941). In comparison with other series within the test, the last series represent the most

TABLE 4.
Summary of the average run time (in minutes) to estimate the exploratory models across sample size, n, the number of attributes, K, and attribute dependence, ρ

				Logit		Probit			
N	K	J	ρ	Delta	Dirac	Delta	Dirac	Restricted	SLCM
500	2	12	0.00	0.4	0.5	1.1	0.8	1.0	1.0
			0.15	0.5	0.4	0.8	0.9	1.0	1.0
			0.25	0.6	0.4	0.9	0.9	1.0	0.9
500	3	20	0.00	1.4	1.3	2.4	2.0	2.3	2.3
			0.15	1.4	1.2	1.8	2.3	1.9	2.2
			0.25	1.4	1.2	1.9	1.9	2.3	2.2
500	4	20	0.00	2.7	2.5	3.2	2.8	2.7	3.4
			0.15	2.6	2.3	3.7	2.9	3.4	2.9
			0.25	3.3	2.6	3.0	3.4	3.5	3.2
500	5	30	0.00	10.5	9.1	12.2	8.1	7.6	8.5
			0.15	10.2	8.3	9.7	8.2	7.9	8.1
			0.25	10.2	14.7	9.7	8.1	7.8	8.0
1000	2	12	0.00	0.6	0.7	1.7	1.7	2.0	2.0
			0.15	0.7	0.6	2.0	2.1	1.6	2.0
			0.25	1.2	0.6	1.7	1.7	1.7	1.9
1000	3	20	0.00	1.8	1.8	4.3	4.2	4.4	4.1
			0.15	2.3	2.2	3.5	3.3	4.3	4.1
			0.25	2.3	2.3	3.5	3.2	4.3	3.8
1000	4	20	0.00	3.5	3.3	5.2	4.6	5.1	5.4
			0.15	4.3	4.2	5.2	6.1	5.0	5.9
			0.25	3.5	4.3	4.8	4.7	6.1	5.4
1000	5	30	0.00	15.2	10.1	13.3	13.2	15.3	14.4
			0.15	12.5	10.4	17.3	13.2	15.3	14.9
			0.25	12.3	10.7	14.8	11.9	13.0	13.6
2000	2	12	0.00	1.0	0.9	3.3	3.2	3.2	3.8
			0.15	1.3	1.3	3.4	3.0	3.3	3.3
			0.25	0.9	2.0	3.2	3.9	3.9	3.3
2000	3	20	0.00	2.8	2.4	8.2	6.1	6.6	8.2
			0.15	2.9	2.5	6.6	6.1	6.8	7.9
			0.25	2.5	2.6	8.3	8.2	7.0	7.6
2000	4	20	0.00	5.0	4.4	9.6	9.3	9.6	11.5
			0.15	5.0	4.5	11.7	9.7	9.8	10.3
			0.25	5.0	4.6	10.0	9.4	9.5	11.2
2000	5	30	0.00	16.9	15.4	25.5	28.3	27.5	26.9
			0.15	21.3	14.3	25.0	23.4	22.5	23.8
			0.25	17.0	19.1	21.9	23.4	22.8	25.4

difficult items on the assessment. Each item is structured so that the respondent must observe an incomplete 3×3 matrix and identify the missing element from among 8 possible responses. Identifying the correct answer from among 7 distractors requires the respondent to deduce which response completes the pattern shown in the matrix. As SPM-LS is still actively used as a "general cognitive ability" test, we direct readers to Raven and Raven (2003) for a sample item used on the test.

TABLE 5.
Summary of 20-fold cross-validated deviance on SPM-LS data containing $n = 499$ observations across exploratory general diagnostic model variants

	Logit		Probit			
K	Delta	Dirac	Delta	Dirac	Restricted	SLCM
2	103,651.3	103,963.3	103,457.2	103,797.9	103,911.2	104,886.2
3	102,660.6	103,194.3	**102,459.1**	103,177.6	103,695.0	104,028.3
4	104,457.5	104,982.4	103,833.2	104,366.8	103,908.4	103,624.5
5	104,739.1	105,676.3	104,285.4	105,868.4	104,645.5	105,847.5

Bold value represents the model chosen after applying cross-validation

TABLE 6.
Summary of the average run time (in minutes) for estimating 20-fold cross-validation on SPM-LS data containing either $n = 474$ or $n = 475$ observations across exploratory general diagnostic model variants

	Logit		Probit			
K	Delta	Dirac	Delta	Dirac	Restricted	SLCM
2	0.5	0.6	0.9	0.8	0.6	0.8
3	1.0	1.0	**1.2**	1.1	0.8	1.1
4	1.9	1.6	1.8	1.4	1.0	1.6
5	4.8	4.5	4.0	2.9	2.2	3.3

Bold value indicates the selected model's runtime

4.2. Data

The data analyzed were obtained from Myszkowski and Storme (2018), who gave the assessment to 214 male and 285 female native French undergraduate students aged between 19 and 24 ($\bar{x} = 20.7$, $\sigma_{\bar{x}} = 0.93$). Respondents in the study were encouraged to answer every item even if they were unsure of their answer and faced no time limit. They participated on a voluntary basis without any monetary compensation and were unfamiliar with the assessment.

4.3. Model Selection

We selected a value for K by performing a 20-fold cross-validation across $K = 2, \ldots, 5$. Each model was fit with $n = 474$ or 475 observations, a chain length of 40,000, and a burn-in of 20,000. Table 5 reports the results of the 20-fold cross-validation, which identified $K = 3$ and the probit-Delta variant as the best candidates. Notably, all models selected $K = 3$ and all of the newly proposed methods that utilized a $\boldsymbol{\Delta}$ matrix to infer the latent structure with weaker monotonicity conditions outperformed the probit-Restricted model for $K = 3$. Lastly, Table 6 shows the average runtime in minutes across each fold estimated.

4.4. Results

Following from the model selection phase, we estimated the probit-Delta model fit with $K = 3$ using all $n = 499$ observations. The model was fit with a chain length of 100,000 and a burn-in of 20,000. Fitting the model and summarizing the Markov chains took 2.97 minutes (178 s) on the cluster.

In Table 7, we report the estimated posterior means for $\boldsymbol{Q}$, $\boldsymbol{B}$, and $\boldsymbol{\pi}$. Upon viewing the estimated $\boldsymbol{Q}$ matrix, there is uncertainty on inclusion of each attribute that may be attributed to

TABLE 7.
Summary of element-wise means of $\boldsymbol{Q}$, $\boldsymbol{B}$, and $\boldsymbol{\pi}$ for the SPM-LS data with $n = 499$ using the probit-delta exploratory general diagnostic model variant

	$\overline{\boldsymbol{Q}}$			$\boldsymbol{B}$							
Items	q_1	q_2	q_3	1 000	α_3 001	α_2 010	$\alpha_2\alpha_3$ 011	α_1 100	$\alpha_1\alpha_3$ 101	$\alpha_1\alpha_2$ 110	$\alpha_1\alpha_2\alpha_3$ 111
1	0.29	0.33	0.64	−0.29	1.07	0.11	0.05	0.09	0.05	0.08	0.03
2	0.50	0.33	0.62	0.12	1.44	0.12	0.06	0.33	0.41	0.11	0.19
3	0.30	0.34	0.67	−0.71	1.90	0.09	0.16	0.07	0.19	0.05	0.08
4	0.66	0.44	0.59	−1.33	1.66	0.26	0.08	2.72	0.24	0.68	0.20
5	0.66	0.44	0.59	−1.23	2.03	0.23	0.06	2.60	0.24	0.77	0.23
6	0.55	0.43	0.63	−1.26	1.67	0.23	0.38	0.56	0.51	0.03	−0.11
7	0.55	0.59	0.37	−0.78	0.24	0.96	0.44	0.92	0.08	0.04	−0.09
8	0.54	0.63	0.44	−1.50	0.33	1.55	0.20	0.78	0.34	0.05	−0.28
9	0.39	0.53	0.49	−0.86	0.25	0.20	1.49	0.16	0.11	0.06	0.07
10	0.65	0.41	0.53	−2.14	0.40	0.17	0.24	0.76	0.80	0.19	0.46
11	0.42	0.45	0.34	−1.26	0.07	0.07	0.05	0.09	0.15	0.07	1.71
12	0.47	0.55	0.29	−1.36	0.06	0.15	0.05	0.12	0.06	1.39	0.06
$\boldsymbol{\pi}$				0.10	0.10	0.01	0.13	0.03	0.26	0.05	0.31

A chain of length of 100,000 was run with a burn-in of 20,000. $\overline{\boldsymbol{Q}}$ is the posterior mean of the sampled binary $\boldsymbol{Q}$ matrices.

the tuning of the spike-slab hyperparameters. However, when observing $\boldsymbol{B}$, it is clear the data support a sparse structure. For instance, items 11 and 12 demonstrate conjunctive relationships similar to a DINA model where there is only an intercept and one slope (interaction terms for these items). In particular, item 11 has an estimate of 1.71 for the three-way interaction effect β_7 for $\alpha_1\alpha_2\alpha_3$ and item 12 has an estimate of 1.39 for the two-way interaction β_6 for $\alpha_1\alpha_2$. Note that the other β values for items 11 and 12 are close to zero in value. Comparing the posterior estimates of element-wise posterior means of the $\boldsymbol{\Delta}$ matrix reported in Table 9 (i.e., $\overline{\boldsymbol{\Delta}}$) with the estimated $\boldsymbol{B}$ matrix shows a close reflection of the patterns of active coefficients. Thus, sparsity and latent structure exhibited by the $\boldsymbol{B}$ are more clearly captured by the $\boldsymbol{\Delta}$ matrix in contrast with the $\boldsymbol{Q}$ matrix, which pools information at the item-attribute level from the item-class level of the $\boldsymbol{\Delta}$ matrix. We report the element-wise standard deviations of $\boldsymbol{B}$ within Table 18 found in "Appendix F."

Finally, in the last row of Table 7, we report the posterior mean of the latent class structural parameter, $\boldsymbol{\pi}$. The largest three classes as suggested by estimated values of $\boldsymbol{\pi}$ were "111," "101," and "011" with class probabilities of $\hat{\pi}_7 = 0.31$, $\hat{\pi}_5 = 0.26$, and $\hat{\pi}_3 = 0.13$, respectively. For the lowest group membership, the three smallest classes were "010," "100," and "110" with probabilities equal to $\hat{\pi}_2 = 0.01$, $\hat{\pi}_4 = 0.03$, and $\hat{\pi}_6 = 0.05$. Based on this estimation, the sample suggests that respondents tended to belong to classes with more attributes being present.

Table 8 reports the estimated correlations among the three attributes as well as the marginal probabilities of mastering the three attributes. Let $\hat{r}_{kk'}$ denote the estimated correlation between attribute k and k'. Table 8 shows that the estimated correlations are $\hat{r}_{12} = 0.15$, $\hat{r}_{13} = 0.26$, and $\hat{r}_{23} = 0.20$. Furthermore, the estimated marginal probability of mastering attributes one through three are reported in the last column of Table 8 were 0.65, 0.50, and 0.80, respectively, and the corresponding estimated standard deviations were 0.48, 0.50, and 0.40.

TABLE 8.
Summary of estimated attribute correlations and marginal probabilities for the SPM-LS data with $n = 499$ using the probit-delta exploratory general diagnostic model variant

	Correlations			
Attributes	1	2	3	$P(\alpha_k = 1)$
1	1	0.15	0.26	0.65
2	0.15	1	0.20	0.50
3	0.26	0.20	1	0.80

TABLE 9.
Summary of element-wise posterior means of $\boldsymbol{\Delta}$ for the SPM-LS Data with $n = 499$ using the probit-delta exploratory general diagnostic model variant

	$\overline{\boldsymbol{\Delta}}$							
Items	δ_{000}	δ_{001}	δ_{010}	δ_{011}	δ_{100}	δ_{101}	δ_{110}	δ_{111}
1	1	1.00	0.24	0.17	0.19	0.17	0.18	0.18
2	1	1.00	0.25	0.21	0.58	0.55	0.27	0.34
3	1	1.00	0.22	0.34	0.17	0.36	0.13	0.25
4	1	0.99	0.43	0.33	1.00	0.49	0.64	0.42
5	1	1.00	0.40	0.31	1.00	0.51	0.68	0.44
6	1	1.00	0.43	0.61	0.73	0.66	0.23	0.35
7	1	0.45	0.89	0.56	0.95	0.31	0.33	0.32
8	1	0.55	0.99	0.42	0.85	0.54	0.36	0.46
9	1	0.50	0.42	1.00	0.35	0.27	0.19	0.27
10	1	0.61	0.33	0.43	0.82	0.82	0.39	0.64
11	1	0.17	0.21	0.13	0.23	0.30	0.16	1.00
12	1	0.14	0.37	0.14	0.30	0.14	1.00	0.21

A chain of length of 100,000 was run with a burnin of 20,000. $\overline{\boldsymbol{\Delta}}$ is the posterior mean of the sampled binary $\boldsymbol{\Delta}$ matrices.

5. Discussion

We reported new methodology for inferring structure in restricted latent class models for binary response data. Our Monte Carlo simulation studies provide evidence of accurate parameter recovery as well as improvement to existing Bayesian methods. In this section, we discuss the implications of this work, offer additional research directions, and provide concluding remarks.

We compared several variants of our exploratory RLCM (i.e., two priors for item parameters and two link functions). We offer several important contributions. First, we report methods for enforcing monotoncity in a manner that allows for potentially disjunctive relationships as characterized by negative interaction effects. In fact, we found some evidence of negative interactions in our application, which may explain the improved cross-validated performance of our new algorithms in comparison to previous research. Consequently, our application demonstrates the importance of using the least restrictive monotonicity constraints in practice. Second, we provide evidence that using the Pòlya–gamma formulation yields a significantly faster application of the logit-link function in comparison to the traditional probit-link. Improved computation with a logit-link provides more efficient methods for researchers who are interested in large-scale testing

data, such as the National Assessment of Education Progress ("NAEP"; e.g., see, Culpepper & Park, 2017) or other international assessments.

Furthermore, the developed algorithms have been made available through "R" packages hosted on the Comprehensive R Archival Network ("CRAN") and can be found in supplemental material associated with the paper. By open-sourcing the algorithms accompanying the paper, we hope to lower the barrier of entry to using the newly developed methods in everyday applications and provide a foundation for further research on these methods.

There are opportunities for future research. First, we provided evidence that the latent structure is more difficult to uncover as the number of attributes increases. Future research will consider extensions to deal with cases when there are more than $K = 5$ attributes (i.e., 32 latent classes). Second, researchers should consider Bayesian approaches for inferring the number of attributes, K. In our application, we used k-fold cross-validation to infer K, which was fairly easy with access to high-performance computing facilities, but it could pose a barrier for applied researchers. Bayesian methods may provide a computationally efficient approach for jointly inferring the latent structure and the latent dimensionality. Future research should consider deploying Bayesian methods (e.g., Dirichlet processes) for inferring the number of attributes as described in Fang et al. (2019) and Chen et al. (2021). Third, we used a novel confirmatory DINA model in the hierarchical prior that relates $\boldsymbol{B}$ to $\boldsymbol{Q}$. It is important to note that we could use another CDM for the hierarchical prior and that we picked the DINA as it is the simplest conjunctive model for specifying the relationship between $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$. Conjunctive models are attractive for the purpose of selecting interactions given that all associated attributes must be active for an interaction to be present. More precisely, a δ_{jp} that corresponds with an interaction should be viewed as active only if all associated attributes in $\boldsymbol{q}_j$ are active. In contrast, a disjunctive model such as a Deterministic Input Noisy Output "OR" gate (DINO; Templin & Henson, 2006) would be inappropriate because the activeness of an interaction effect would arise if any associated attribute was active. An alternative to our DINA prior could be to use a more general conjunctive model such as the reduced reparameterized unified model (rRUM; Culpepper & Hudson, 2018, Hartz, 2002) and we recommend future research investigate the relative merits of using a conjunctive prior that includes more parameters.

In closing, throughout our work on exploratory RLCMs, the developed methodology contributes greatly to the customization of exploratory models. By enabling the use of the logit-link function, there now is an alternative to the probit-link function to estimate models. Moreover, the modifications to the existing work on probit models ensure the monotonicity requirements are met.

Acknowledgments

This research was partially supported by National Science Foundation, Division of Methodology, Measurement, and Statistics program grants #1758631 and #1951057. This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign.

Appendix A: Derivation of Metropolis-Within-Gibbs Sampling Steps

The purpose of this section is to derive the MCMC sampling steps. We present full conditional distributions for $\boldsymbol{\alpha}_i, \boldsymbol{\pi}, \boldsymbol{B}, s, g$, and ω and outline Metropolis–Hastings (MH) updates for elements of $\boldsymbol{\Delta}$ and $\boldsymbol{Q}$. The steps of the MCMC sampling algorithm are as follows.

1. For $t = 0$ initialize $\omega^{(0)}$, $s^{(0)}$, $g^{(0)}$, $\boldsymbol{Q}^{(0)}$, $\boldsymbol{\Delta}^{(0)}$, $\boldsymbol{B}^{(0)}$, and $\boldsymbol{\pi}^{(0)}$. Fix $\delta_{j0} = 1$ for all j, so the intercept remains active.
2. For $t = 1, \ldots, T$ repeat the following steps.

   (a) *Update attributes*. For $i = 1, \ldots, n$ sample $\boldsymbol{\alpha}_i^{(t)}$ at iteration t from the categorical full conditional distribution $\boldsymbol{\alpha}_i^{(t)} | \boldsymbol{Y}_i, \boldsymbol{B}^{(t-1)}, \boldsymbol{\alpha}_1^{(t)}, \ldots, \boldsymbol{\alpha}_{i-1}^{(t)}, \boldsymbol{\alpha}_{i+1}^{(t-1)}, \ldots, \boldsymbol{\alpha}_n^{(t-1)}$ where the conditional probability $\boldsymbol{\alpha}_i^{(t)}$ is classified as profile c is

$$\begin{aligned} &\mathrm{P}(\boldsymbol{\alpha}_i^{(t)\top} \boldsymbol{v} = c | \boldsymbol{Y}_i, \boldsymbol{B}^{(t-1)}, \boldsymbol{\alpha}_1^{(t)}, \ldots, \boldsymbol{\alpha}_{i-1}^{(t)}, \boldsymbol{\alpha}_{i+1}^{(t-1)}, \ldots, \boldsymbol{\alpha}_n^{(t-1)}) \\ &\quad = \frac{(n_{ci} + n_{c0}) \prod_{j=1}^{J} \theta_{jc,y_{ij}}^{(t-1)}}{\sum_{c=0}^{2^K-1} (n_{ci} + n_{c0}) \prod_{j=1}^{J} \theta_{jc,y_{ij}}^{(t-1)}} \end{aligned} \tag{A1}$$

   Notice that we integrate $\boldsymbol{\pi}$ from the prior distribution $p(\mathbf{A}, \boldsymbol{\pi}) = p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n | \boldsymbol{\pi})$ $p(\boldsymbol{\pi})$ and instead use the conditional prior distribution $p(\boldsymbol{\alpha}_i | \boldsymbol{\alpha}_1^{(t)}, \ldots, \boldsymbol{\alpha}_{i-1}^{(t)}, \boldsymbol{\alpha}_{i+1}^{(t-1)}, \ldots, \boldsymbol{\alpha}_n^{(t-1)})$ which implies the usual π_c (e.g., see Equation 7 of Culpepper, 2019a) is replaced with $n_{ci} + n_{c0}$ where n_{ci} is the number of respondents other than i that are classified in class c and n_{c0} is the prior Dirichlet parameter (note $n_{c0} = 1$ for a uniform prior).

   (b) *Update* $\boldsymbol{\pi}$. The full conditional distribution for $\boldsymbol{\pi}$ is a Dirichlet distribution defined as:

$$\boldsymbol{\pi} | \boldsymbol{\alpha}_{1:n} \sim \mathrm{Dirichlet}(\boldsymbol{n} + \boldsymbol{n}_0). \tag{A2}$$

   Note $\boldsymbol{n}$ is a 2^K vector with element c defined as $n_c = \sum_{i=1}^{n} \mathbb{1}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c)$. Recall we use a uniform prior for $\boldsymbol{\pi}$ and set $\boldsymbol{n}_0 = \mathbf{1}_{2^K}$.

   (c) For $j = 1, \ldots, J$,

      (a) For $k = 1, \ldots, K$ sample q_{jk} using a MH step. Specifically, the MH acceptance ratio is

$$A(q^*, q_{jk}^{(t-1)}) = \frac{p(\boldsymbol{\delta}_j | s, g, q^*, \boldsymbol{q}_{j(k)}) p(q^* | \boldsymbol{Q}) f(q_{jk}^{(t-1)} | q^*)}{p(\boldsymbol{\delta}_j | s, g, q_{jk}^{(t-1)}, \boldsymbol{q}_{j(k)}) p(q_{jk}^{(t-1)} | \boldsymbol{Q}) f(q^* | q_{jk}^{(t-1)})} \tag{A3}$$

      Note that we use deterministic proposal distributions, so $f(q^* | q_{jk}^{(t-1)}) = f(q_{jk}^{(t-1)} | q^*) = 1$, where the proposed value is defined as $q^* = 1 - q_{jk}^{(t-1)}$. Furthermore, the prior distributions are Bernoulli distributions with the restriction that q^* must yield an identified $\boldsymbol{Q}^*$ otherwise $p(q^* | \boldsymbol{Q}) = 0$. Finally, note that the ratio of $p(\boldsymbol{\delta}_j | s, g, q^*, \boldsymbol{q}_{j(k)})$ to $p(\boldsymbol{\delta}_j | s, g, q_{jk}^{(t-1)}, \boldsymbol{q}_{j(k)})$ simplifies as

$$\frac{p(\boldsymbol{\delta}_j | s, g, q^*)}{p(\boldsymbol{\delta}_j | s, g, q_{jk} = q_{jk}^{(t-1)})}$$

$$= \frac{\prod_{p=1}^{P}\left[(1-s)^{\eta_{jp}^{*}} g^{1-\eta_{jp}^{*}}\right]^{\delta_{jp}}\left[s^{\eta_{jp}^{*}}(1-g)^{1-\eta_{jp}^{*}}\right]^{1-\delta_{jp}}}{\prod_{p=1}^{P}\left[(1-s)^{\eta_{jp}^{(t-1)}} g^{1-\eta_{jp}^{(t-1)}}\right]^{\delta_{jp}}\left[s^{\eta_{jp}^{(t-1)}}(1-g)^{1-\eta_{jp}^{(t-1)}}\right]^{1-\delta_{jp}}}$$

$$= (1-s)^{\mathcal{S}_{11}} g^{\mathcal{S}_{01}} s^{\mathcal{S}_{10}} (1-g)^{\mathcal{S}_{00}} \tag{A4}$$

where η_{jp}^{*} is the "and" gate with the proposed q^{*}, $\eta_{jp}^{(t-1)}$ is the "and" gate from the previous iteration, and

$$\mathcal{S}_{11} = \sum_{p=1}^{P}\left(\eta_{jp}^{*} - \eta_{jp}^{(t-1)}\right)\delta_{jp} q'_{pk} \tag{A5}$$

$$\mathcal{S}_{01} = \sum_{p=1}^{P}\left(\eta_{jp}^{(t-1)} - \eta_{jp}^{*}\right)\delta_{jp} q'_{pk} \tag{A6}$$

$$\mathcal{S}_{10} = \sum_{p=1}^{P}\left(\eta_{jp}^{*} - \eta_{jp}^{(t-1)}\right)(1-\delta_{jp}) q'_{pk} \tag{A7}$$

$$\mathcal{S}_{00} = \sum_{p=1}^{P}\left(\eta_{jp}^{(t-1)} - \eta_{jp}^{*}\right)(1-\delta_{jp}) q'_{pk}. \tag{A8}$$

We accept q^{*} and set $q_{jk}^{(t)} = q^{*}$ if $\ln\left[A(q^{*}, q_{jk}^{(t-1)})\right] > \ln(U)$ for $U \sim$ uniform(0, 1). Otherwise, $q_{jk}^{(t)} = q_{jk}^{(t-1)}$.

(b) For $p = 1, \ldots, 2^{K} - 1$ sample δ_{jp} using a MH step. The MH acceptance ratio is

$$A(\delta^{*}, \delta_{jp}^{(t-1)}) = \frac{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta^{*}) p(\delta^{*}|\boldsymbol{q}_{j}, s, g) f(\delta_{jp}^{(t-1)}|\delta^{*})}{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}^{(t-1)}) p(\delta_{jp}^{(t-1)}|\boldsymbol{q}_{j}, s, g) f(\delta^{*}|\delta_{jp}^{(t-1)})} \tag{A9}$$

where we use a deterministic proposal distribution with $\delta^{*} = 1 - \delta_{jp}^{(t-1)}$. Recall the prior for $\boldsymbol{\beta}_{jp}$ given δ_{jp} and the other item parameters , $\boldsymbol{\beta}_{j(p)}$ is

$$p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}) = \frac{1}{\sqrt{2\pi\sigma_{jp}^{2}}\left(1 - \Phi\left(\frac{L_{jp}}{\sigma_{jp}}\right)\right)} \exp\left(-\frac{1}{2}\frac{\beta_{jp}^{2}}{\sigma_{jp}^{2}}\right) \tag{A10}$$

where L_{jp} is the lower bound for $\boldsymbol{\beta}_{jp}$ given $\boldsymbol{\beta}_{j(p)}$, $\sigma_{jp}^{2} = \delta_{jp}/c_1 + (1-\delta_{jp})/c_0$, and $1 - \Phi\left(\frac{L_{jp}}{\sigma_{jp}}\right)$ is the prior probability $\beta_{jp} > L_{jp}$. The ratio of priors for β_{jp} given the proposal and current state is

$$\frac{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}^{*})}{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}^{(t-1)})} = \sqrt{\frac{\sigma_{jp}^{2,(t-1)}}{\sigma_{jp}^{2*}}}\left(\frac{1 - \Phi\left(\frac{L_{jp}}{\sigma_{jp}^{(t-1)}}\right)}{1 - \Phi\left(\frac{L_{jp}}{\sigma_{jp}^{*}}\right)}\right)$$

$$\exp\left[\frac{1}{2}\left(\frac{1}{\sigma_{jp}^{2,(t-1)}}-\frac{1}{\sigma_{jp}^{2*}}\right)\beta_{jp}^2\right] \tag{A11}$$

where $\sigma_{jp}^{2*}=\delta^*/c_1+(1-\delta^*)/c_0$ and $\sigma_{jp}^{2,(t-1)}=\delta_{jp}^{(t-1)}/c_1+(1-\delta_{jp}^{(t-1)})/c_0$.

$$\begin{aligned}\frac{p(\delta^*|\boldsymbol{q}_j,s,g)}{p(\delta_{jp}^{(t-1)}|\boldsymbol{q}_j,s,g)}&=\frac{\left((1-s)^{\eta_{jp}}g^{1-\eta_{jp}}\right)^{\delta^*}\left(s^{\eta_{jp}}(1-g)^{1-\eta_{jp}}\right)^{1-\delta^*}}{\left((1-s)^{\eta_{jp}}g^{1-\eta_{jp}}\right)^{\delta_{jp}^{(t-1)}}\left(s^{\eta_{jp}}(1-g)^{1-\eta_{jp}}\right)^{1-\delta_{jp}^{(t-1)}}}\\&=\left[\left(\frac{1-s}{s}\right)^{\eta_{jp}}\left(\frac{g}{1-g}\right)^{1-\eta_{jp}}\right]^{\delta^*-\delta_{jp}^{(t-1)}}\end{aligned} \tag{A12}$$

If $\ln\left[A(\delta^*,\delta_{jp}^{(t-1)})\right]>\ln(U)$ for $U\sim$ uniform$(0,1)$ we accept the move to δ^* and set $\delta_{jp}^{(t)}=\delta^*$. Otherwise, $\delta_{jp}^{(t)}=\delta_{jp}^{(t-1)}$.

(c) *Sample augmented data*. For $c=0,\ldots,2^K-1$ sample Y_{jc}^* from the following Pòlya–gamma distribution $Y_{jc}^*\sim PG\left(n_c,\psi_{jc}\right)$ where $\psi_{jc}=\boldsymbol{\alpha}_c^\top\boldsymbol{\beta}_j$.

(d) *Sample item coefficients*. For $p=0,\ldots,2^K-1$ sample β_{jp} from a truncated normal distribution,

$$\beta_{jp}|\boldsymbol{\beta}_{j(p)},\boldsymbol{\alpha}_{1:n},\boldsymbol{Y}_{1:n,j},\boldsymbol{Y}_{1:n,j}^*,\delta_{jp}\sim\mathcal{N}(m_{jp},v_{jp}^2)\mathbb{1}(\beta_{jp}\geq L_{jp}) \tag{A13}$$

where $L_{j0}=-\infty$. The conditional variance and mean are

$$v_{jp}^2=\frac{1}{\boldsymbol{A}_p^\top\boldsymbol{\Omega}_j\boldsymbol{A}_p+1/\sigma_{jp}^2} \tag{A14}$$

$$m_{jp}=v_{jp}^2\boldsymbol{A}_p^\top\boldsymbol{\Omega}_j\tilde{\boldsymbol{z}}_j \tag{A15}$$

where $\boldsymbol{A}_p$ is the pth column of the $2^K\times 2^K$ attribute design matrix $\boldsymbol{A}=(\boldsymbol{a}_0,\ldots,\boldsymbol{a}_{2^K-1})^\top$, $\boldsymbol{\Omega}_j=\text{diag}(y_{j0}^*,\ldots,y_{j,2^K-1}^*)$, $\tilde{\boldsymbol{z}_j}=\boldsymbol{z}_j-\boldsymbol{A}_{(p)}\boldsymbol{\beta}_{j(p)}$, $\boldsymbol{z}_j=(\kappa_{j0}/y_{j0}^*,\ldots,\kappa_{j,2^K-1}/y_{j,2^K-1}^*)^\top$, $\kappa_{jc}=n_{jc}-n_c/2$, and $\boldsymbol{A}_{(p)}$ is the matrix $\boldsymbol{A}$ without column p. See the derivation of the full conditional distribution for β_{jp} in "Appendix B."

(d) *Sample hyper-parameters s and g*. We sample the s and g hyper-parameters from a truncated beta distribution. Specifically, note that the full conditional distribution is

$$\begin{aligned}&p(s,g|\boldsymbol{\Delta},\boldsymbol{Q})\propto p(\boldsymbol{\Delta}|\boldsymbol{Q},s,g)p(s,g)\\&=\left(\prod_{j=1}^{J}\prod_{p=2}^{P}\left[(1-s)^{\eta_{jp}}\,g^{1-\eta_{jp}}\right]^{\delta_{jp}}\left[s^{\eta_{jp}}\,(1-g)^{1-\eta_{jp}}\right]^{1-\delta_{jp}}\right)p(s,g)\\&=s^{a_s}(1-s)^{b_s}g^{a_g}(1-g)^{b_g}\mathbb{1}(0<g<1-s)\mathbb{1}(0<s<1)\end{aligned} \tag{A16}$$

where

$$a_s = \sum_{j=1}^{J}\sum_{p=2}^{P}\eta_{jp}(1-\delta_{jp}),\ b_s = \sum_{j=1}^{J}\sum_{p=2}^{P}\eta_{jp}\delta_{jp}$$
$$a_g = \sum_{j=1}^{J}\sum_{p=2}^{P}(1-\eta_{jp})\delta_{jp},\ b_g = \sum_{j=1}^{J}\sum_{p=2}^{P}(1-\eta_{jp})(1-\delta_{jp}).$$

Note we sample s and g sequentially from the truncated beta distribution as described in Culpepper (2015).

(e) *Sample hyper-parameter* ω. The hyper-parameter for $\boldsymbol{Q}$ is conditionally distributed as a beta distribution. Specifically, we sample ω from

$$\omega|\boldsymbol{Q} \sim \text{Beta}\left(\sum_{j=1}^{J}\sum_{k=1}^{K}q_{jk}+a, \sum_{j=1}^{J}\sum_{k=1}^{K}(1-q_{jk})+b\right). \tag{A17}$$

Note we use a uniform prior and set $a = b = 1$.

Appendix B: Derivation of Full Conditional Distribution for β_{jp} Under the Logit-Link

We next derive the full conditional distribution for the item parameters $\boldsymbol{\beta}_j$. Let $\boldsymbol{y}_{1:n,j} = (y_{1j},\ldots,y_{nj})^\top$ be the observed responses for item j and $\boldsymbol{y}_j^* = (y_{j0}^*,\ldots,y_{j,2^K-1}^*)^\top$ the 2^K vector of augmented data for item j.

In order to derive the full conditional distribution for $\boldsymbol{\beta}_j$, we first substitute the Pòlya–gamma identity in Eq. 26 into the collapsed conditional likelihood in Eq. 28 to yield,

$$\begin{aligned} p\left(\boldsymbol{y}_{1:n,j}|\boldsymbol{\alpha}_{1:n},\boldsymbol{\beta}_j\right) &= \prod_{c=0}^{2^K-1}\frac{\left(\exp\left(\psi_{jc}\right)\right)^{n_{jc}}}{\left(1+\exp\left(\psi_{jc}\right)\right)^{n_c}} \\ &= \prod_{c=0}^{2^K-1}2^{-n_c}\exp\left(\kappa_{jc}\psi_{jc}\right)\int_0^\infty \exp\left(-\frac{1}{2}y_{jc}^*\psi_{jc}^2\right)p(y_{jc}^*)\mathrm{d}y_{jc}^*. \end{aligned} \tag{B1}$$

where $\kappa_{jc} = \left(n_{jc}-\frac{n_c}{2}\right)$. Therefore, the joint conditional distribution for $\boldsymbol{y}_{1:n,j}$ and $\boldsymbol{y}_j^*$ is proportional in terms of ψ_{jc} to

$$p\left(\boldsymbol{y}_{1:n,j},\boldsymbol{y}_j^*|\boldsymbol{\alpha}_{1:n},\boldsymbol{\beta}_j\right) \propto \prod_{c=0}^{2^K-1}2^{-n_c}\exp\left(\kappa_{jc}\psi_{jc}\right)\exp\left(-\frac{1}{2}y_{jc}^*\psi_{jc}^2\right). \tag{B2}$$

Recall that completing the square under the transformation of: $x^2 - bx = \left(x-\frac{b}{2}\right)^2 + (b)^2$. Therefore, the product can be rewritten as

$$\prod_{c=0}^{2^K-1}\exp\left(\kappa_{jc}\psi_{jc}-\frac{1}{2}y_{jc}^*\psi_{jc}^2\right) = \prod_{c=0}^{2^K-1}\exp\left(-\frac{1}{2}y_{jc}^*\left(\psi_{jc}^2-\frac{2\kappa_{jc}}{y_{jc}^*}\psi_{jc}\right)\right)$$

$$= \prod_{c=0}^{2^K-1} \exp\left(-\frac{1}{2} y_{jc}^* \left(\left(\psi_{jc} - \frac{\kappa_{jc}}{y_{jc}^*}\right)^2 - \left(\frac{2\kappa_{jc}}{y_{jc}^*}\right)^2\right)\right)$$
$$= \prod_{c=0}^{2^K-1} \exp\left(-\frac{1}{2} y_{jc}^* \left(\psi_{jc} - \frac{\kappa_{jc}}{y_{jc}^*}\right)^2 + 2\frac{\kappa_{jc}^2}{y_{jc}^*}\right). \tag{B3}$$

Retaining the terms that are proportional to $\boldsymbol{\beta}_j$ yields

$$\propto \prod_{c=0}^{2^k-1} \exp\left(-\frac{1}{2} y_{jc}^* \left(\psi_{jc} - \frac{\kappa_{jc}}{y_{jc}^*}\right)^2\right) = \exp\left(-\frac{1}{2} \sum_{c=0}^{2^k-1} \left[y_{jc}^* \left(\psi_{jc} - \frac{\kappa_{jc}}{y_{jc}^*}\right)^2\right]\right)$$
$$= \exp\left(-\frac{1}{2} \sum_{c=0}^{2^k-1} \left[y_{jc}^* \left(\boldsymbol{\alpha}_c^\top \boldsymbol{\beta}_j - \frac{\kappa_{jc}}{y_{jc}^*}\right)^2\right]\right)$$
$$= \exp\left(-\frac{1}{2} \sum_{c=0}^{2^k-1} \left[y_{jc}^* \left(\frac{\kappa_{jc}}{y_{jc}^*} - \boldsymbol{\alpha}_c^\top \boldsymbol{\beta}_j\right)^2\right]\right)$$
$$= \exp\left(-\frac{1}{2}\left(\boldsymbol{z}_j - \boldsymbol{A}\boldsymbol{\beta}_j\right)^T \boldsymbol{\Omega}_j \left(\boldsymbol{z}_j - \boldsymbol{A}\boldsymbol{\beta}_j\right)\right) \tag{B4}$$

where $\boldsymbol{A} = (\boldsymbol{a}_0, \ldots, \boldsymbol{a}_{2^K-1})^\top$ is a $2^K \times 2^K$ matrix with the latent class design vectors, $\boldsymbol{z}_j$ is a 2^K vector with element c defined as κ_{jc}/y_{jc}^*, and $\boldsymbol{\Omega}_j = \text{diag}(y_{j0}^*, \ldots, y_{j,2^K-1}^*)$.
Let $\boldsymbol{\kappa}_j = (\kappa_{j0}, \ldots, \kappa_{j,2^K-1})$. Multiplying Eq. B4 by the prior for $\boldsymbol{\beta}_j$ in Eq. 30 implies the full conditional distribution for $\boldsymbol{\beta}_j$ is

$$\boldsymbol{\beta}_j | \boldsymbol{\alpha}_{1:n}, \boldsymbol{y}_j^*, \boldsymbol{y}_{1:n,j}, \boldsymbol{\delta}_j \sim \mathcal{N}_P\left(\boldsymbol{m}_j, \boldsymbol{V}_j\right) \mathbb{1}(\boldsymbol{\beta}_j \in \mathcal{B}) \tag{B5}$$
$$\boldsymbol{V}_j = \left(\boldsymbol{A}^\top \boldsymbol{\Omega}_j \boldsymbol{A} + \boldsymbol{\Sigma}_j^{-1}\right)^{-1} \tag{B6}$$
$$\boldsymbol{m}_j = \boldsymbol{V}_j \boldsymbol{A}^\top \boldsymbol{\kappa}_j. \tag{B7}$$

We update the elements of $\boldsymbol{\beta}_j$ sequentially in order to enforce monotonicity conditions. Specifically, the term in the exponential of Eq. B4 times the prior for β_{jp} equals

$$-\frac{1}{2}\left(\boldsymbol{z}_j - \boldsymbol{A}\boldsymbol{\beta}_j\right)^T \boldsymbol{\Omega}_j \left(\boldsymbol{z}_j - \boldsymbol{A}\boldsymbol{\beta}_j\right) - \beta_{jp}^2 \sigma_{jp}^2$$
$$= -\frac{1}{2}\left[\left(\tilde{\boldsymbol{z}}_j - A_p \beta_{jp}\right)^T \boldsymbol{\Omega}_j \left(\tilde{\boldsymbol{z}}_j - A_p \beta_{jp}\right) + \beta_{jp}^2 \sigma_{jp}^2\right] \tag{B8}$$

where $\tilde{\boldsymbol{z}}_j = \boldsymbol{z}_j - \boldsymbol{A}_{(p)}\boldsymbol{\beta}_{j(p)}$, $\boldsymbol{A}_{(p)}$ is the matrix $\boldsymbol{A}$ without column p, and $\boldsymbol{\beta}_{j(p)}$ is $\boldsymbol{\beta}_j$ without element p. Completing the square implies that the full conditional distribution for β_{jp} is a truncated normal with $\beta_{jp} > L_{jp}$ with

$$v_{jp}^2 = \frac{1}{\boldsymbol{A}_p^\top \boldsymbol{\Omega}_j \boldsymbol{A}_p + 1/\sigma_{jp}^2} \tag{B9}$$
$$m_{jp} = v_{jp}^2 \boldsymbol{A}_p^\top \boldsymbol{\Omega}_j \tilde{\boldsymbol{z}}_j. \tag{B10}$$

Note $L_{j0} = -\infty$ so the full conditional distribution for intercepts are unrestricted normal distributions.

Appendix C: Implementation of a Dirac Delta Spike Distribution

The purpose of this section is to provide details on updating item parameters under a spike-slab prior for $\boldsymbol{B}$ using a mixture of a normal distribution for the slab and a spike distribution with a point mass at zero. Special care is needed for updating δ_{jp} and β_{jp} in the case where the spike distribution is a point mass. We next report steps for updating δ_{jp} and β_{jp}.

5.1. Updating δ_{jp}

We update δ_{jp} after integrating β_{jp} from the posterior distribution. We let the posterior distribution after marginalizing over β_{jp} be

$$p(\boldsymbol{y}^*_{1:n}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{B}_{(j,p)}, \boldsymbol{\Delta}, \boldsymbol{Q}, \boldsymbol{\pi}, \omega, s, g|\boldsymbol{y}_{1:n}) = \int p(\boldsymbol{y}^*_{1:n}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{B}, \boldsymbol{\Delta}, \boldsymbol{Q}, \boldsymbol{\pi}, \omega, s, g|\boldsymbol{y}_{1:n})\mathrm{d}\beta_{jp} \tag{C1}$$

where $\boldsymbol{B}_{(j,p)}$ is all of $\boldsymbol{B}$ except β_{jp}.
We next construct a Metropolis–Hastings step using the marginalized posterior distribution to update δ_{jp}. Specifically, as described above, we use a deterministic proposal distribution and define the candidate as $\delta^* = 1 - \delta^{(t-1)}_{jp}$. The acceptance ratio is:

$$A(\delta^*, \delta^{(t-1)}_{jp}) = \frac{p(\boldsymbol{y}^*_{1:n}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{B}_{(j,p)}, \boldsymbol{\Delta}_{(j,p)}, \delta_{jp} = \delta^*, \boldsymbol{Q}, \boldsymbol{\pi}, \omega, s, g|\boldsymbol{y}_{1:n})}{p(\boldsymbol{y}^*_{1:n}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{B}_{(j,p)}, \boldsymbol{\Delta}_{(j,p)}, \delta_{jp} = \delta^{(t-1)}_{jp}, \boldsymbol{Q}, \boldsymbol{\pi}, \omega, s, g|\boldsymbol{y}_{1:n})} \tag{C2}$$

where $\boldsymbol{\Delta}_{(j,p)}$ excludes δ_{jp} and the ratio of proposal distributions is excluded as it equals one. The challenge is integrating β_{jp} from the posterior. Our framework allows sampling from the full conditional distributions and we can therefore find an analytic solution to each integral. In fact, derivations in Appendix D of Chen et al. (2020) can be adapted to show that the acceptance probability can be written as

$$A(\delta^*, \delta^{(t-1)}_{jp}) = \left[\frac{\Phi\left(\frac{-L_{jp}}{\sigma_{jp}}\right)^{-1} \frac{v_{jp}}{\sigma_{jp}} \exp\left(\frac{1}{2}\frac{m^2_{jp}}{v^2_{jp}}\right) \Phi\left(\frac{m_{jp}-L_{jp}}{v_{jp}}\right) \mathrm{P}(\delta_{jp} = 1|\boldsymbol{q}_j, s, g)}{\mathrm{P}(\delta_{jp} = 0|\boldsymbol{q}_j, s, g)} \right]^{\delta^*_{jp} - \delta^{(t-1)}_{jp}} \tag{C3}$$

where L_{jp} is the lower-bound for β_{jp}, σ^2_{jp} is the prior variance for β_{jp} for the case where $\delta_{jp} = 1$, and, as defined in Eqs. B9 and B10, v^2_{jp} is the conditional variance and m_{jp} is the conditional mean of β_{jp}. Rearranging terms and plugging in the prior for δ_{jp} yield

$$A(\delta^*, \delta^{(t-1)}_{jp}) = \left[\left(\frac{v_{jp}}{\sigma_{jp}}\right) \frac{\Phi\left(\frac{m_{jp}-L_{jp}}{v_{jp}}\right)}{\Phi\left(\frac{-L_{jp}}{\sigma_{jp}}\right)} \exp\left(\frac{1}{2}\frac{m^2_{jp}}{v^2_{jp}}\right) \left(\frac{1-s}{s}\right)^{\eta_{jp}} \left(\frac{g}{1-g}\right)^{1-\eta_{jp}} \right]^{\delta^*_{jp} - \delta^{(t-1)}_{jp}} \tag{C4}$$

Chen et al. (2020) indicated that $\delta_{jp}^{(t)} = 0$ is not possible in cases where $L_{jp} > 0$. Therefore, $\delta_{jp}^{(t-1)}$ can only be updated when $L_{jp} \leq 0$. Therefore, the update rule for δ_{jp} is if $L_{jp} \leq 0$ and $\ln\left[A(\delta^*, \delta_{jp}^{(t-1)})\right] > \ln(U)$ for $U \sim \text{uniform}(0, 1)$ we accept the move to δ^* and set $\delta_{jp}^{(t)} = \delta^*$. Otherwise, $\delta_{jp}^{(t)} = \delta_{jp}^{(t-1)}$.

5.2. *Updating* β_{jp}

A Gibbs step is available for sampling β_{jp} from its full conditional distribution. Specifically, the full conditional distribution is

$$\beta_{jp}|\boldsymbol{y}_{1:n,j}, \boldsymbol{y}^*_{1:n,j}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{\beta}_{j(p)}, \delta_{jp} \sim \begin{cases} \mathbb{1}(\beta_{jp} = 0) & \delta_{jp} = 0 \\ \mathcal{N}(m_{jp}, v_{jp}^2)\mathbb{1}(\beta_{jp} > L_{jp}) & \delta_{jp} = 1 \end{cases}. \tag{C5}$$

Appendix D: Evaluating Routines for Generating Pòlya–gamma Random Variables

We next present a closer look at how different logit variants perform across the techniques for sampling a random variables from a Pòlya–gamma distribution. The implementation of Pòlya–gamma samplers follows from the techniques presented in Windle et al. (2014). Principally, these methods are given as sum of gammas ("G"), Devroye method ("DM"), hybrid ("H"), normal approximation ("N"), and saddlepoint approximation ("SP"). Omitted from this list is the alternate sampler given the routine was deactivated pending review in the "BayesLogit" package (Polson, Scott, & Windle, 2013b).

With this being said, the hybrid sampler consists of a mixture of the four other methods at various sample size n cutoffs. In particular, the sum of gammas method is used for $n \in (1, 13)\backslash\{1, 2\}$, Devroye method for $n = 1, 2$, Saddlepoint method when $3 \leq n < 170$, and Normal Approximation for $n \geq 170$. If the alternate sampler was enabled, then it would be used to sample when $n \in (1, 13)\backslash\{2\}$ and cause the sum of gammas to non-integer values between $0 < n < 1$.

In Table 10, we present the element-wise recovery rates for the $\boldsymbol{Q}$ matrix under different sampling routines. Each of the samplers returns a similar conclusion with recovery accuracy being nearly identical across sample sizes and K attributes. As the correlation increases, there is a slightly worse recovery across the board. Similar conclusions are able to be made when observing the average MAD entries for $\boldsymbol{\Theta}$ in Table 11 (Table 12).

Reported with Table 13 are the run times associated with each sampler. The normal approximation is the quickest to be estimated with times ranging between 18 s when $n = 500$ and $K = 2$ are small to 14 minutes when $n = 2000$ and $K = 5$ are large. However, the stability of the normal approximation is problematic as shown in Table 14 as indicated by the finding that not all models returned estimable results. Non-estimable results were returned as a "Not a Number" (NaN). From Table 14, the instability was striking in extent when $K = 5$ and $\rho = 0.15, 0.25$, especially under the logit-Delta formulation. As the hybrid sampler relies on the normal approximation, it also falls into similar stability issues due to its reliance on the sampler when $n > 170$.

Appendix E: Disjunction Simulation Study

Within this appendix, we apply the novel Bayesian formulations to 100 simulated disjunctive data sets. Unlike in the prior simulations, which simulated data from a tetrachoric population

TABLE 10.
Overview of how each Pòlya–gamma random variable sampling routine performed in the respective Logit Delta and Dirac model variants under $\boldsymbol{Q}$

			Logit-Delta					Logit-Dirac				
N	K	ρ	DM	G	H	N	SP	DM	G	H	N	SP
500	2	0.00	0.960	0.960	0.960	0.960	0.956	0.925	0.922	0.921	0.925	0.924
		0.15	0.950	0.948	0.950	0.949	0.949	0.902	0.901	0.903	0.903	0.902
		0.25	0.957	0.957	0.956	0.951	0.957	0.902	0.901	0.902	0.902	0.900
500	3	0.00	0.953	0.952	0.953	0.901	0.952	0.970	0.969	0.970	0.963	0.971
		0.15	0.950	0.952	0.948	0.872	0.951	0.968	0.968	0.968	0.963	0.968
		0.25	0.955	0.950	0.951	0.842	0.952	0.967	0.968	0.967	0.953	0.967
500	4	0.00	0.910	0.910	0.910	0.749	0.910	0.928	0.928	0.928	0.919	0.928
		0.15	0.910	0.910	0.909	0.761	0.910	0.925	0.925	0.923	0.917	0.925
		0.25	0.912	0.912	0.912	0.747	0.913	0.928	0.928	0.928	0.923	0.928
500	5	0.00	0.865	0.865	0.866	0.599	0.865	0.899	0.901	0.898	0.757	0.900
		0.15	0.857	0.857	0.860	0.599	0.858	0.888	0.888	0.890	0.765	0.891
		0.25	0.854	0.854	0.853	0.599	0.853	0.879	0.879	0.879	0.766	0.881
1000	2	0.00	0.990	0.990	0.990	0.990	0.990	0.955	0.954	0.955	0.951	0.955
		0.15	0.989	0.988	0.989	0.986	0.989	0.950	0.951	0.952	0.950	0.952
		0.25	0.980	0.982	0.980	0.979	0.981	0.935	0.935	0.935	0.933	0.935
1000	3	0.00	0.979	0.979	0.973	0.971	0.979	0.990	0.990	0.991	0.991	0.990
		0.15	0.978	0.975	0.969	0.935	0.978	0.988	0.987	0.977	0.983	0.989
		0.25	0.967	0.962	0.962	0.868	0.966	0.975	0.977	0.978	0.965	0.981
1000	4	0.00	0.922	0.922	0.916	0.865	0.923	0.935	0.938	0.939	0.931	0.939
		0.15	0.926	0.928	0.924	0.869	0.926	0.947	0.948	0.949	0.948	0.946
		0.25	0.928	0.926	0.926	0.854	0.928	0.946	0.945	0.946	0.938	0.945
1000	5	0.00	0.904	0.904	0.904	0.598	0.904	0.941	0.941	0.941	0.901	0.941
		0.15	0.901	0.901	0.901	0.604	0.901	0.936	0.935	0.935	0.908	0.935
		0.25	0.899	0.899	0.899	0.623	0.898	0.931	0.930	0.931	0.914	0.931
2000	2	0.00	0.996	0.996	0.996	0.996	0.996	0.973	0.968	0.969	0.969	0.974
		0.15	0.993	0.993	0.992	0.991	0.992	0.965	0.965	0.962	0.964	0.965
		0.25	0.994	0.994	0.995	0.990	0.995	0.971	0.972	0.969	0.972	0.971
2000	3	0.00	0.987	0.994	0.994	0.994	0.988	0.996	0.993	0.995	0.991	0.992
		0.15	0.985	0.986	0.983	0.980	0.986	0.989	0.992	0.989	0.985	0.986
		0.25	0.973	0.986	0.983	0.895	0.983	0.979	0.986	0.981	0.990	0.973
2000	4	0.00	0.923	0.923	0.926	0.895	0.925	0.958	0.959	0.956	0.950	0.959
		0.15	0.930	0.932	0.923	0.874	0.932	0.970	0.967	0.965	0.949	0.961
		0.25	0.930	0.932	0.927	0.865	0.929	0.954	0.958	0.954	0.950	0.959
2000	5	0.00	0.932	0.934	0.935	0.625	0.931	0.967	0.970	0.968	0.961	0.971
		0.15	0.929	0.929	0.920	0.665	0.929	0.967	0.967	0.961	0.932	0.964
		0.25	0.920	0.920	0.923	0.678	0.923	0.950	0.954	0.952	0.927	0.956

Closer to 1 indicates higher element-wise recovery. "G" is sum of gammas, "DM" Devroye method, "H" is hybrid, "N" normal approximation, and "SP" saddlepoint approximation. Results based on the number of numerically stable replications as reported in Table 14.

TABLE 11.
Summary of the average mean absolute deviation (MAD) of $\boldsymbol{\Theta}$ estimated by the Logit Delta and Dirac variants under different sampling routines for Pòlya–gamma distribution

				Logit-Delta					Logit-Dirac				
N	K	J	ρ	DM	G	H	N	SP	DM	G	H	N	SP
500	2	12	0.00	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
			0.15	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
			0.25	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
500	3	20	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
			0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
			0.25	0.05	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05
500	4	20	0.00	0.07	0.07	0.07	0.06	0.07	0.05	0.05	0.05	0.06	0.05
			0.15	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06
			0.25	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06
500	5	30	0.00	0.10	0.10	0.10	NaN	0.10	0.08	0.08	0.08	0.08	0.08
			0.15	0.10	0.11	0.10	NaN	0.10	0.08	0.08	0.08	0.08	0.08
			0.25	0.11	0.11	0.11	NaN	0.11	0.08	0.08	0.08	0.08	0.08
1000	2	12	0.00	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
			0.15	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
			0.25	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
1000	3	20	0.00	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
			0.15	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
			0.25	0.04	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.04	0.03
1000	4	20	0.00	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04
			0.15	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04
			0.25	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04
1000	5	30	0.00	0.08	0.08	0.08	NaN	0.08	0.05	0.05	0.05	0.05	0.05
			0.15	0.08	0.09	0.09	NaN	0.08	0.05	0.05	0.05	0.05	0.05
			0.25	0.08	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.05	0.05
2000	2	12	0.00	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
			0.15	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
			0.25	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
2000	3	20	0.00	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
			0.15	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
			0.25	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.02	0.03
2000	4	20	0.00	0.04	0.04	0.04	0.04	0.04	0.03	0.02	0.02	0.03	0.02
			0.15	0.04	0.04	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.03
			0.25	0.05	0.04	0.05	0.04	0.04	0.04	0.03	0.04	0.03	0.03
2000	5	30	0.00	0.07	0.06	0.06	0.06	0.07	0.04	0.03	0.04	0.03	0.03
			0.15	0.06	0.07	0.07	0.06	0.06	0.04	0.04	0.04	0.04	0.04
			0.25	0.07	0.07	0.07	0.06	0.07	0.05	0.04	0.05	0.04	0.04

Values closer to 0 indicates higher recovery. "G" is sum of gammas, "DM" Devroye method, "H" is hybrid, "N" normal approximation, and "SP" saddlepoint approximation. Results based on the number of numerically stable replications as reported in Table 14. "NaN" denotes 0 replications were completed successfully.

TABLE 12.
Comparison across each Pòlya–gamma random variable sampling routine performed in the respective Logit Delta and Dirac model variants under $\boldsymbol{\Delta}$

			Logit-Delta					Logit-Dirac				
N	K	ρ	DM	G	H	N	SP	DM	G	H	N	SP
500	2	0.00	0.792	0.792	0.793	0.792	0.794	0.743	0.743	0.740	0.743	0.744
		0.15	0.782	0.783	0.781	0.780	0.783	0.719	0.718	0.719	0.718	0.718
		0.25	0.789	0.789	0.790	0.785	0.788	0.716	0.716	0.717	0.716	0.715
500	3	0.00	0.860	0.861	0.860	0.819	0.860	0.884	0.883	0.884	0.880	0.884
		0.15	0.857	0.860	0.853	0.788	0.861	0.878	0.878	0.879	0.871	0.878
		0.25	0.861	0.855	0.857	0.740	0.860	0.873	0.875	0.872	0.855	0.872
500	4	0.00	0.893	0.893	0.892	0.827	0.893	0.897	0.897	0.897	0.896	0.898
		0.15	0.894	0.894	0.894	0.851	0.894	0.899	0.899	0.899	0.897	0.899
		0.25	0.895	0.895	0.895	0.829	0.895	0.899	0.899	0.899	0.897	0.899
500	5	0.00	0.929	0.929	0.929	0.828	0.928	0.940	0.940	0.940	0.908	0.939
		0.15	0.928	0.928	0.929	0.802	0.928	0.939	0.938	0.938	0.881	0.938
		0.25	0.928	0.928	0.928	0.806	0.928	0.937	0.937	0.936	0.877	0.937
1000	2	0.00	0.822	0.822	0.821	0.823	0.822	0.784	0.783	0.784	0.777	0.784
		0.15	0.822	0.821	0.822	0.819	0.822	0.780	0.780	0.780	0.775	0.780
		0.25	0.816	0.816	0.816	0.816	0.816	0.764	0.764	0.764	0.761	0.765
1000	3	0.00	0.887	0.886	0.876	0.876	0.886	0.922	0.922	0.922	0.922	0.922
		0.15	0.877	0.875	0.864	0.837	0.879	0.921	0.917	0.914	0.913	0.922
		0.25	0.868	0.858	0.860	0.778	0.866	0.898	0.898	0.901	0.886	0.904
1000	4	0.00	0.893	0.893	0.892	0.874	0.893	0.907	0.907	0.907	0.906	0.907
		0.15	0.895	0.896	0.895	0.873	0.895	0.909	0.909	0.909	0.910	0.909
		0.25	0.896	0.897	0.896	0.868	0.897	0.908	0.907	0.908	0.906	0.908
1000	5	0.00	0.936	0.937	0.937	0.794	0.936	0.952	0.952	0.952	0.937	0.952
		0.15	0.937	0.937	0.937	0.823	0.937	0.949	0.949	0.949	0.937	0.949
		0.25	0.937	0.937	0.937	0.836	0.937	0.947	0.948	0.948	0.937	0.948
2000	2	0.00	0.916	0.917	0.916	0.914	0.916	0.844	0.843	0.839	0.838	0.844
		0.15	0.906	0.905	0.904	0.903	0.905	0.832	0.830	0.824	0.827	0.830
		0.25	0.894	0.894	0.893	0.885	0.891	0.818	0.818	0.816	0.818	0.818
2000	3	0.00	0.895	0.903	0.903	0.903	0.894	0.927	0.922	0.926	0.919	0.922
		0.15	0.884	0.886	0.880	0.884	0.885	0.919	0.924	0.919	0.914	0.915
		0.25	0.873	0.890	0.887	0.790	0.887	0.905	0.911	0.907	0.920	0.896
2000	4	0.00	0.893	0.893	0.894	0.883	0.893	0.917	0.918	0.919	0.916	0.918
		0.15	0.896	0.896	0.894	0.875	0.896	0.919	0.919	0.918	0.916	0.916
		0.25	0.895	0.896	0.893	0.842	0.895	0.912	0.913	0.913	0.907	0.914
2000	5	0.00	0.943	0.943	0.944	0.792	0.943	0.955	0.956	0.956	0.954	0.957
		0.15	0.941	0.941	0.940	0.838	0.941	0.953	0.954	0.952	0.940	0.953
		0.25	0.941	0.941	0.941	0.827	0.941	0.949	0.949	0.949	0.946	0.950

Closer to 1 indicates higher element-wise recovery. "G" is sum of gammas, "DM" Devroye method, "H" is hybrid, "N" normal approximation, and "SP" saddlepoint approximation. Results based on the number of numerically stable replications as reported in Table 14.

TABLE 13.
Summary of the average run time in minutes of the Delta and Dirac Logit variants under different sampling routines for Pòlya–gamma distribution

				Logit-Delta					Logit-Dirac				
N	K	J	ρ	DM	G	H	N	SP	DM	G	H	N	SP
500	2	12	0.00	3.3	4.8	0.4	0.3	0.6	3.8	4.3	0.5	0.3	1.1
			0.15	3.3	5.0	0.5	0.4	0.6	3.8	4.6	0.4	0.3	0.5
			0.25	3.3	4.6	0.5	0.3	0.5	3.2	4.4	0.5	0.3	0.5
500	3	20	0.00	5.4	16.0	1.5	1.1	1.7	5.9	14.7	1.5	1.1	1.4
			0.15	5.9	15.9	1.6	1.1	1.7	6.1	15.3	1.6	0.9	1.3
			0.25	5.5	14.6	1.7	2.2	3.1	5.6	16.0	1.7	1.1	1.4
500	4	20	0.00	7.1	29.5	5.2	1.7	3.3	6.7	30.6	4.9	2.0	2.6
			0.15	6.8	31.4	4.2	1.7	3.3	6.6	30.0	4.8	1.6	2.6
			0.25	7.9	30.2	5.0	2.1	3.3	6.5	39.3	4.8	1.6	2.5
500	5	30	0.00	15.5	101.6	18.8	9.1	12.5	17.3	91.9	18.3	6.5	9.1
			0.15	17.4	96.0	16.1	7.6	10.2	14.9	97.3	15.1	7.2	9.3
			0.25	15.3	101.3	19.3	7.5	12.4	13.8	93.3	15.6	6.6	9.3
1000	2	12	0.00	6.4	4.7	0.6	0.5	0.8	6.8	4.6	0.7	0.6	0.7
			0.15	6.5	4.6	0.6	0.6	0.7	6.2	4.8	0.6	0.6	0.8
			0.25	6.1	5.0	0.7	0.5	0.8	6.5	5.2	0.7	0.6	0.7
1000	3	20	0.00	12.6	14.8	2.1	1.7	2.3	11.2	15.6	1.7	1.3	2.3
			0.15	12.3	15.8	2.2	1.4	2.3	11.3	20.3	1.8	1.6	1.9
			0.25	12.2	16.7	2.3	1.3	1.9	11.6	15.8	2.2	1.3	1.8
1000	4	20	0.00	12.1	33.1	4.1	3.1	4.3	15.5	33.3	4.4	3.0	3.4
			0.15	12.6	30.2	4.1	3.1	4.3	12.4	33.7	5.0	2.9	4.2
			0.25	11.7	30.2	5.1	3.1	3.6	11.4	30.3	5.1	2.9	4.2
1000	5	30	0.00	26.1	97.8	16.9	9.7	12.6	23.4	100.1	15.9	10.3	13.7
			0.15	26.4	110.4	20.7	9.9	15.3	24.0	98.4	16.1	8.6	13.8
			0.25	23.5	106.7	20.5	11.9	15.3	23.0	106.8	19.3	8.6	11.5
2000	2	12	0.00	14.0	5.4	1.2	0.9	1.3	12.1	5.8	1.1	1.1	1.1
			0.15	12.1	6.2	0.9	2.3	1.3	12.2	5.2	1.1	1.1	1.1
			0.25	12.8	6.8	1.1	1.1	1.3	12.9	4.9	1.1	1.0	1.1
2000	3	20	0.00	23.0	20.3	2.5	2.3	3.4	22.8	17.0	3.0	2.8	2.9
			0.15	21.6	16.6	3.1	2.8	2.8	22.1	17.3	2.5	2.3	2.8
			0.25	21.2	18.1	2.6	2.8	3.4	21.1	17.2	2.5	2.8	2.9
2000	4	20	0.00	29.2	34.4	6.3	5.0	5.2	24.0	30.8	6.1	4.8	11.2
			0.15	30.1	34.8	5.1	4.2	6.3	25.0	31.9	5.1	4.8	4.8
			0.25	24.7	34.5	5.2	5.2	6.4	24.2	33.0	5.2	4.0	4.9
2000	5	30	0.00	48.2	115.3	19.8	17.2	21.3	52.1	101.5	22.5	12.9	19.3
			0.15	45.5	97.2	20.6	14.0	17.4	57.5	104.5	18.5	15.7	16.4
			0.25	45.5	102.0	19.4	14.6	17.2	51.0	98.7	19.2	13.0	19.2

"G" is sum of gammas, "DM" Devroye method, "H" is hybrid, "N" normal approximation, and "SP" saddlepoint approximation. Results based on all 100 iterations regardless of stability.

TABLE 14.
Summary of Pòlya–gamma distribution random variable sampler stability

				Logit-Delta					Logit-Dirac				
N	K	J	ρ	DM	H	N	SP	G	DM	H	N	SP	G
500	2	12	0.00	100	100	100	100	100	100	99	100	100	100
			0.15	100	100	99	100	100	100	100	100	100	100
			0.25	100	100	98	100	100	100	100	100	100	100
500	3	20	0.00	100	100	79	100	100	100	100	98	100	100
			0.15	100	100	75	100	100	100	100	99	100	100
			0.25	100	100	63	100	100	100	100	97	100	100
500	4	20	0.00	100	100	29	100	100	100	100	96	100	100
			0.15	100	100	35	100	100	100	100	97	100	100
			0.25	100	100	34	100	100	100	100	97	100	100
500	5	30	0.00	100	100	0	100	100	100	100	36	100	100
			0.15	100	100	0	100	100	100	100	40	100	100
			0.25	100	100	0	100	100	100	100	41	100	100
1000	2	12	0.00	100	100	100	100	100	100	100	98	100	100
			0.15	100	100	99	100	100	100	100	98	100	100
			0.25	100	100	99	100	100	100	100	99	100	100
1000	3	20	0.00	100	99	96	100	100	100	99	99	100	100
			0.15	100	99	88	100	100	100	98	100	100	100
			0.25	100	99	72	100	100	100	100	97	100	100
1000	4	20	0.00	100	98	68	100	100	100	100	95	100	100
			0.15	100	99	75	100	100	100	100	99	100	100
			0.25	100	100	75	100	100	100	100	97	100	100
1000	5	30	0.00	100	100	0	100	100	100	100	83	100	100
			0.15	100	100	0	100	100	100	100	88	100	100
			0.25	100	100	3	100	100	100	100	93	100	100
2000	2	12	0.00	100	100	100	100	100	100	99	98	100	100
			0.15	100	100	99	100	100	100	99	99	100	100
			0.25	100	100	98	100	100	100	99	100	100	100
2000	3	20	0.00	100	100	100	100	100	100	99	97	100	100
			0.15	100	99	94	100	100	100	100	99	100	100
			0.25	100	99	74	100	100	100	99	99	100	100
2000	4	20	0.00	100	99	88	100	100	100	99	97	100	100
			0.15	100	97	82	100	100	100	100	98	100	100
			0.25	100	99	76	100	100	100	100	97	100	100
2000	5	30	0.00	100	99	2	100	100	100	99	97	100	100
			0.15	100	96	14	100	100	100	99	86	100	100
			0.25	100	100	16	100	100	100	100	89	100	100

Samplers with less than 100 replications provided problematic output under the given condition. “G” is sum of gammas, “DM” Devroye method, “H” is hybrid, “N” normal approximation, and “SP” saddlepoint approximation

TABLE 15.
Summary of exploratory general diagnostic models under Delta and Dirac augmentations ability to estimate $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ measured by the average mean absolute deviation (MAD) across 100 replications for each sample size, N, attributes, $K = 3$, and attribute dependence, $\rho = 0$

	Logit				Probit							
	Delta		Dirac		Delta		Dirac		Restricted		SLCM	
N	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$
500	0.09	0.03	0.09	0.04	0.09	0.04	0.10	0.04	0.14	0.08	0.10	0.04
1000	0.08	0.04	0.08	0.04	0.09	0.04	0.09	0.05	0.14	0.09	0.09	0.04
2000	0.07	0.03	0.08	0.04	0.09	0.05	0.09	0.04	0.13	0.08	0.09	0.05

correlation according to Chiu et al. (2009), we define a $\boldsymbol{B}$ and $\boldsymbol{Q}$ matrix with $J = 21$ items and $K = 3$ attributes that exhibits the desired disjunctive behavior under $\rho = 0$.

$$\boldsymbol{B} = \begin{bmatrix} -0.84 & 1.68 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.84 & 1.68 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.84 & 1.68 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.84 & 0 & 1.68 & 0 & 0 & 0 & 0 & 0 \\ -0.84 & 0 & 1.68 & 0 & 0 & 0 & 0 & 0 \\ -0.84 & 0 & 1.68 & 0 & 0 & 0 & 0 & 0 \\ -0.84 & 0 & 0 & 0 & 1.68 & 0 & 0 & 0 \\ -0.84 & 0 & 0 & 0 & 1.68 & 0 & 0 & 0 \\ -0.84 & 0 & 0 & 0 & 1.68 & 0 & 0 & 0 \\ -0.84 & 1.68 & 1.68 & -1.68 & 0 & 0 & 0 & 0 \\ -0.84 & 1.68 & 1.68 & -1.68 & 0 & 0 & 0 & 0 \\ -0.84 & 1.68 & 1.68 & -1.68 & 0 & 0 & 0 & 0 \\ -0.84 & 1.68 & 1.68 & -1.68 & 0 & 0 & 0 & 0 \\ -0.84 & 1.68 & 0 & 0 & 1.68 & -1.68 & 0 & 0 \\ -0.84 & 1.68 & 0 & 0 & 1.68 & -1.68 & 0 & 0 \\ -0.84 & 1.68 & 0 & 0 & 1.68 & -1.68 & 0 & 0 \\ -0.84 & 1.68 & 0 & 0 & 1.68 & -1.68 & 0 & 0 \\ -0.84 & 0 & 1.68 & 0 & 1.68 & 0 & -1.68 & 0 \\ -0.84 & 0 & 1.68 & 0 & 1.68 & 0 & -1.68 & 0 \\ -0.84 & 0 & 1.68 & 0 & 1.68 & 0 & -1.68 & 0 \\ -0.84 & 0 & 1.68 & 0 & 1.68 & 0 & -1.68 & 0 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \tag{E1}$$

Found in Table 16 are the average MAD entries for $\boldsymbol{\Theta}$ matrix. Within Table 15, we present the element-wise recovery rate (RR) for the $\boldsymbol{Q}$ matrix and $\boldsymbol{\Delta}$ matrix under different sample sizes as defined by Eqs. 41 and 42. As the number of subjects increases, there is improved recovery of elements in $\boldsymbol{\Theta}$ marked by lower MAD values. Most notably, the recovery rate for probit-Restricted are lower in the presence of a disjunctive relationship in comparison to the newly developed methodology and SLCM. Meanwhile, the SLCM method performs on par in terms of $\boldsymbol{\Theta}$, $\boldsymbol{Q}$, $\boldsymbol{\pi}$ recovery; though, the recovery with respect to $\boldsymbol{\Delta}$ is slightly lower (Table 16). Finally, Table 17 provides an overview of runtime.

TABLE 16.
Summary of exploratory general diagnostic models to estimate a disjunction population under Delta and Dirac augmentations ability to estimate $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ matrices as measured by element-wise accuracy across sample size, N, attributes, $K = 3$, and attribute dependence, $\rho = 0$

	Logit				Probit			
	Delta		Dirac		Delta		Dirac	
N	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$
500	0.92	0.88	0.88	0.83	0.90	0.86	0.84	0.80
1000	0.91	0.89	0.86	0.82	0.90	0.87	0.82	0.79
2000	0.95	0.92	0.86	0.80	0.86	0.84	0.86	0.80

TABLE 16.
continued

	Probit		
	Restricted	SLCM	
N	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$
500	0.79	0.89	0.81
1000	0.79	0.89	0.81
2000	0.80	0.88	0.81

The column for $\boldsymbol{\Delta}$ matrix is omitted for the Restricted model as the method only estimates the $\boldsymbol{Q}$ matrix.

TABLE 17.
Summary of the average run time (in minutes) to estimate a disjunction population using exploratory models across sample size, n, with $K = 3$ attributes, and an attribute dependence $\rho = 0$

	Logit		Probit			
N	Delta	Dirac	Delta	Dirac	Restricted	SLCM
500	1.8	1.7	2.5	2.4	1.9	1.9
1000	2.4	2.3	4.5	4.4	3.5	3.6
2000	3.6	3.7	8.5	8.3	7.0	8.2

Appendix F: Additional Estimation Details on the SPM-LS Application

Due to space limitations within the main manuscript, readers may be interested in understanding the standard deviation of the element-wise estimates for the $\boldsymbol{B}$ matrix. Inside of Table 18, we provide the element-wise standard deviation. Most of the active standard deviations are small with the exception of two elements $\beta_{4,6}$ and $\beta_{5,6}$. In these cases, the standard deviations are around 0.7 which is comparable in magnitude to the size of the actual coefficient. The corresponding elements of $\boldsymbol{\Delta}$ convey the probability the coefficients are active to be around 0.6.

TABLE 18.
Summary of element-wise standard deviation of $\boldsymbol{B}$ for the SPM-LS data with $n = 499$ using the probit-delta exploratory general diagnostic model variant

	$\boldsymbol{B}$							
Items	000	001	010	011	100	101	110	111
1	0.16	0.20	0.11	0.10	0.09	0.10	0.11	0.12
2	0.17	0.30	0.15	0.17	0.28	0.55	0.19	0.42
3	0.20	0.30	0.10	0.23	0.07	0.26	0.08	0.19
4	0.25	0.43	0.32	0.28	0.57	0.56	0.74	0.55
5	0.26	0.38	0.29	0.22	0.61	0.58	0.76	0.58
6	0.24	0.34	0.26	0.41	0.42	0.51	0.15	0.30
7	0.19	0.25	0.49	0.54	0.34	0.25	0.25	0.31
8	0.29	0.32	0.47	0.41	0.44	0.47	0.27	0.42
9	0.18	0.23	0.22	0.36	0.16	0.17	0.12	0.22
10	0.35	0.38	0.21	0.34	0.50	0.56	0.31	0.46
11	0.15	0.06	0.10	0.08	0.09	0.19	0.11	0.27
12	0.14	0.06	0.16	0.09	0.13	0.09	0.29	0.17
π	0.02	0.03	0.01	0.04	0.02	0.05	0.02	0.03

A chain of length of 100,000 was run with a burn-in of 20,000.

Appendix G: Impact of Larger Attribute Dependence on Parameter Recovery

This subsection includes extended Monte Carlo simulation results regarding parameter recovery for cases with larger values for the attribute correlation, ρ. We examined attribute dependency when $\rho = 0.3, 0.4, 0.5$. Table 19 reports information about recovery of $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ for the various methods, Table 20 reports details concerning the recovery of $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$, and Table 21 reports the average run time of the algorithms in minutes. Overall, the results from the extended simulation provide evidence that the methods, which use an over-parameterized Dirichlet prior for attribute correlations, are accurate for larger attribute correlations. Note that the use of an over-specified model for attribute dependence does lead to slightly larger MADs for $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ and smaller recovery rates for $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ as ρ increases. Furthermore, note that parameter recovery improves as N increases.

TABLE 19.
Summary of exploratory general diagnostic models under Delta and Dirac augmentations ability to estimate $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ measured by the average mean absolute deviation (MAD) across 100 replications for each sample size, N, attributes, K, and attribute dependence, ρ

				Logit				Probit							
				Delta		Dirac		Delta		Dirac		Restricted		SLCM	
N	K	J	ρ	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$	$\hat{\boldsymbol{\Theta}}$	$\hat{\boldsymbol{\pi}}$
500	2	12	0.3	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02
			0.4	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02
			0.5	0.04	0.02	0.04	0.02	0.04	0.02	0.05	0.02	0.04	0.02	0.04	0.02
500	3	20	0.3	0.06	0.02	0.05	0.02	0.06	0.03	0.05	0.02	0.05	0.02	0.05	0.02
			0.4	0.06	0.03	0.05	0.02	0.07	0.03	0.06	0.02	0.05	0.02	0.05	0.02
			0.5	0.08	0.04	0.06	0.03	0.09	0.06	0.07	0.03	0.05	0.02	0.05	0.02
500	4	20	0.3	0.07	0.02	0.06	0.02	0.07	0.02	0.06	0.02	0.06	0.02	0.07	0.02
			0.4	0.08	0.02	0.06	0.02	0.07	0.02	0.07	0.02	0.07	0.02	0.08	0.02
			0.5	0.08	0.03	0.07	0.02	0.08	0.03	0.08	0.02	0.07	0.02	0.08	0.03
500	5	30	0.3	0.11	0.02	0.09	0.02	0.10	0.02	0.09	0.02	0.10	0.02	0.11	0.03
			0.4	0.11	0.03	0.09	0.02	0.11	0.02	0.09	0.02	0.10	0.02	0.12	0.03
			0.5	0.11	0.03	0.10	0.02	0.11	0.03	0.10	0.02	0.10	0.03	0.12	0.03
1000	2	12	0.3	0.03	0.02	0.03	0.01	0.03	0.02	0.03	0.01	0.03	0.02	0.03	0.01
			0.4	0.03	0.02	0.03	0.01	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02
			0.5	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02	0.03	0.02
1000	3	20	0.3	0.05	0.03	0.04	0.02	0.06	0.03	0.04	0.02	0.04	0.02	0.04	0.02
			0.4	0.07	0.04	0.05	0.02	0.07	0.05	0.05	0.03	0.04	0.02	0.05	0.03
			0.5	0.08	0.06	0.06	0.03	0.09	0.06	0.06	0.04	0.04	0.01	0.06	0.04
1000	4	20	0.3	0.05	0.02	0.04	0.01	0.06	0.02	0.04	0.01	0.05	0.01	0.05	0.02
			0.4	0.06	0.02	0.05	0.01	0.06	0.02	0.05	0.02	0.05	0.02	0.06	0.02
			0.5	0.07	0.03	0.06	0.02	0.08	0.03	0.07	0.03	0.05	0.02	0.07	0.03
1000	5	30	0.3	0.09	0.02	0.06	0.01	0.08	0.02	0.06	0.01	0.08	0.02	0.08	0.02
			0.4	0.09	0.02	0.06	0.01	0.09	0.02	0.07	0.01	0.08	0.02	0.10	0.02
			0.5	0.09	0.02	0.07	0.01	0.09	0.02	0.08	0.02	0.08	0.02	0.10	0.02
2000	2	12	0.3	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
			0.4	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.02
			0.5	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.03	0.03
2000	3	20	0.3	0.05	0.02	0.04	0.02	0.06	0.03	0.04	0.02	0.03	0.01	0.05	0.03
			0.4	0.07	0.04	0.05	0.03	0.08	0.05	0.06	0.04	0.03	0.01	0.06	0.04
			0.5	0.09	0.06	0.07	0.05	0.10	0.08	0.07	0.05	0.03	0.01	0.08	0.07
2000	4	20	0.3	0.05	0.02	0.04	0.02	0.05	0.02	0.05	0.02	0.04	0.01	0.05	0.02
			0.4	0.07	0.03	0.05	0.03	0.08	0.03	0.06	0.03	0.04	0.01	0.07	0.03
			0.5	0.09	0.04	0.07	0.03	0.10	0.04	0.08	0.04	0.05	0.01	0.08	0.04
2000	5	30	0.3	0.07	0.01	0.05	0.01	0.07	0.02	0.06	0.02	0.06	0.01	0.08	0.02
			0.4	0.08	0.02	0.07	0.02	0.08	0.02	0.07	0.02	0.07	0.01	0.09	0.02
			0.5	0.09	0.02	0.09	0.02	0.12	0.03	0.10	0.03	0.07	0.02	0.10	0.03

TABLE 20.
Summary of exploratory general diagnostic models under Delta and Dirac augmentations ability to estimate $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ matrices as measured by element-wise accuracy across sample size, N, attributes, K, and high attribute dependence, ρ

				Logit				Probit						
				Delta		Dirac		Delta		Dirac		Restricted	SLCM	
N	K	J	ρ	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{Q}}$	$\overline{\boldsymbol{\Delta}}$
500	2	12	0.3	0.95	0.79	0.90	0.71	0.94	0.77	0.87	0.69	0.99	0.99	0.91
			0.4	0.95	0.78	0.88	0.70	0.93	0.77	0.85	0.67	0.99	0.99	0.91
			0.5	0.94	0.77	0.87	0.68	0.92	0.76	0.85	0.65	0.99	0.99	0.90
500	3	20	0.3	0.95	0.86	0.97	0.87	0.95	0.85	0.96	0.86	0.94	0.96	0.82
			0.4	0.94	0.84	0.95	0.84	0.93	0.82	0.95	0.83	0.94	0.95	0.82
			0.5	0.91	0.81	0.94	0.80	0.90	0.76	0.92	0.76	0.93	0.95	0.82
500	4	20	0.3	0.92	0.90	0.93	0.90	0.92	0.90	0.93	0.90	0.92	0.92	0.84
			0.4	0.91	0.90	0.93	0.90	0.91	0.90	0.92	0.89	0.91	0.90	0.83
			0.5	0.91	0.90	0.92	0.89	0.90	0.89	0.91	0.89	0.91	0.90	0.83
500	5	30	0.3	0.84	0.93	0.87	0.93	0.85	0.93	0.87	0.93	0.85	0.78	0.86
			0.4	0.83	0.92	0.86	0.93	0.84	0.93	0.85	0.93	0.84	0.78	0.86
			0.5	0.82	0.92	0.83	0.93	0.82	0.92	0.83	0.93	0.83	0.77	0.86
1000	2	12	0.3	0.98	0.82	0.94	0.76	0.98	0.81	0.92	0.74	1.00	1.00	0.91
			0.4	0.98	0.82	0.94	0.76	0.97	0.81	0.92	0.74	1.00	1.00	0.91
			0.5	0.98	0.81	0.93	0.75	0.97	0.80	0.91	0.73	0.99	0.99	0.91
1000	3	20	0.3	0.96	0.86	0.97	0.89	0.95	0.85	0.97	0.89	0.97	0.98	0.83
			0.4	0.93	0.83	0.96	0.86	0.93	0.83	0.93	0.82	0.96	0.96	0.82
			0.5	0.91	0.80	0.94	0.81	0.89	0.77	0.91	0.75	0.96	0.92	0.81
1000	4	20	0.3	0.93	0.90	0.94	0.91	0.93	0.90	0.94	0.90	0.93	0.95	0.84
			0.4	0.92	0.90	0.94	0.90	0.92	0.90	0.94	0.90	0.93	0.93	0.84
			0.5	0.92	0.90	0.94	0.90	0.91	0.89	0.92	0.88	0.92	0.91	0.84
1000	5	30	0.3	0.90	0.94	0.93	0.95	0.90	0.94	0.93	0.95	0.89	0.86	0.87
			0.4	0.89	0.94	0.92	0.94	0.89	0.94	0.91	0.94	0.88	0.84	0.86
			0.5	0.88	0.93	0.90	0.94	0.87	0.93	0.89	0.94	0.87	0.83	0.86
2000	2	12	0.3	0.99	0.90	0.97	0.82	0.99	0.89	0.96	0.81	1.00	1.00	0.92
			0.4	0.99	0.87	0.96	0.80	0.99	0.86	0.95	0.79	1.00	0.99	0.91
			0.5	0.99	0.84	0.96	0.79	0.99	0.83	0.95	0.78	1.00	0.97	0.89
2000	3	20	0.3	0.96	0.86	0.97	0.88	0.95	0.84	0.96	0.87	0.98	0.94	0.81
			0.4	0.93	0.81	0.95	0.86	0.92	0.80	0.92	0.81	0.98	0.91	0.80
			0.5	0.90	0.78	0.91	0.78	0.88	0.75	0.90	0.76	0.98	0.87	0.79
2000	4	20	0.3	0.93	0.90	0.95	0.91	0.93	0.90	0.94	0.90	0.93	0.94	0.84
			0.4	0.91	0.89	0.95	0.90	0.91	0.89	0.94	0.90	0.93	0.90	0.83
			0.5	0.89	0.89	0.92	0.89	0.89	0.88	0.92	0.88	0.93	0.89	0.84
2000	5	30	0.3	0.92	0.94	0.95	0.95	0.93	0.94	0.93	0.94	0.91	0.87	0.86
			0.4	0.91	0.94	0.92	0.94	0.91	0.94	0.91	0.94	0.91	0.85	0.86
			0.5	0.90	0.94	0.89	0.93	0.88	0.93	0.88	0.93	0.90	0.83	0.86

TABLE 21.
Summary of the average run time (in minutes) to estimate the exploratory models across sample size, n, the number of attributes, K, and high attribute dependence, ρ

				Logit		Probit			
N	K	J	ρ	Delta	Dirac	Delta	Dirac	Restricted	SLCM
500	2	12	0.3	0.6	0.6	0.9	0.9	1.0	0.9
			0.4	0.5	0.5	1.1	0.9	1.0	0.9
			0.5	0.4	0.5	0.9	1.1	0.9	0.9
500	3	20	0.3	1.4	1.7	2.4	2.0	1.8	1.9
			0.4	1.4	1.7	2.0	1.9	2.2	1.8
			0.5	1.7	1.7	2.4	1.9	1.8	1.9
500	4	20	0.3	2.7	3.2	3.7	2.8	2.7	2.8
			0.4	2.7	2.6	3.0	2.8	2.7	2.8
			0.5	2.7	2.6	3.7	3.5	2.7	2.7
500	5	30	0.3	12.3	10.9	10.0	9.5	7.6	8.0
			0.4	12.3	9.4	9.8	9.5	7.5	8.0
			0.5	12.2	9.4	11.3	8.6	7.6	8.0
1000	2	12	0.3	0.8	0.8	2.0	1.7	1.6	1.7
			0.4	0.7	0.7	1.7	2.0	1.7	1.6
			0.5	0.6	0.7	1.7	1.7	2.0	1.7
1000	3	20	0.3	2.3	2.2	3.5	3.5	3.4	3.6
			0.4	1.9	2.2	4.3	3.5	4.2	3.3
			0.5	2.3	1.9	3.6	4.2	4.2	3.6
1000	4	20	0.3	4.3	3.5	5.1	5.0	4.9	5.1
			0.4	3.7	3.4	6.4	5.2	4.9	5.0
			0.5	4.4	3.5	5.1	5.1	4.9	5.0
1000	5	30	0.3	12.8	13.6	17.4	12.7	15.3	13.2
			0.4	15.1	13.6	14.4	12.9	15.3	12.8
			0.5	15.2	13.6	15.1	12.8	15.3	12.8
2000	2	12	0.3	1.1	1.1	3.9	3.4	3.4	3.4
			0.4	1.1	1.3	3.3	3.3	3.2	3.4
			0.5	1.1	1.3	3.3	3.3	3.2	3.2
2000	3	20	0.3	2.9	3.4	8.1	6.7	6.8	7.1
			0.4	3.4	2.9	6.7	6.7	8.0	6.5
			0.5	3.4	2.9	8.1	6.8	7.0	7.0
2000	4	20	0.3	6.3	6.1	9.7	9.4	9.2	9.7
			0.4	6.4	6.1	9.8	9.7	9.3	9.7
			0.5	6.4	5.1	9.4	9.3	9.2	9.7
2000	5	30	0.3	17.6	19.2	29.5	27.6	27.4	23.5
			0.4	17.6	19.3	29.5	23.4	27.5	22.6
			0.5	20.7	16.6	29.5	23.0	23.3	23.2

References

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

Chen, Y., & Culpepper, S. A. (2020). A multivariate probit model for learning trajectories: A fine-grained evaluation of an educational intervention. Applied Psychological Measurement.

Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. (2018). Bayesian estimation of the DINA Q. *Psychometrika, 83*(1), 89–108.

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

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. (2020). Estimation of K and Q matrix in restricted latent class models. In *International Meeting of the Psychometric Society, Virtual*.
Chen, Y., Liu, Y., Culpepper, S. A., & Chen, Y. (2021). Inferring the number of attributes for the exploratory Dina model. *Psychometrika*. https://doi.org/10.1007/s11336-021-09750-9
Chiu, C. Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika, 74*(4), 633–665.
Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics, 40*(5), 454–476.
Culpepper, S. A. (2019a). Estimating the cognitive diagnosis $\boldsymbol{Q}$ matrix with expert knowledge: Application to the fraction–subtraction dataset. *Psychometrika, 84*(2), 333–357.
Culpepper, S. A. (2019b). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation. *Psychometrika, 84*(4), 921–940.
Culpepper, S. A., & Chen, Y. (2019). Development and application of an exploratory reduced reparameterized unified model. *Journal of Educational and Behavioral Statistics, 44*(1), 3–24.
Culpepper, S. A., & Hudson, A. (2018). An improved strategy for Bayesian estimation of the reduced reparameterized unified model. *Applied Psychological Measurement*. https://doi.org/10.1177/0146621617707511
Culpepper, S. A., & Park, T. (2017). Bayesian estimation of multivariate latent regression models: Gauss versus Laplace. *Journal of Educational and Behavioral Statistics, 42*(5), 591–616.
de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179–199.
de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333–353.
Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. Springer. ISBN 978-1-4614-6867-7. https://doi.org/10.1007/978-1-4614-6868-4.
Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A brief introduction to Rcpp. *The American Statistician, 72*(1), 28–36. https://doi.org/10.1080/00031305.2017.1375990
Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software, 40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08.
Eddelbuettel, D., & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis, 71*, 1054–1063. https://doi.org/10.1016/j.csda.2013.02.005
Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. Psychometrika.
Gu, Y., & Xu, G. (2020). Partial identifiability of restricted latent class models. *Annals of Statistics, 48*(4), 2082–2107.
Gu, Y., & Xu, G. (2021). Sufficient and necessary conditions for the identifiability of the Q-matrix. *Statistica Sinica, 31*, 449–472. https://doi.org/10.5705/ss.202018.0410
Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*(4), 301–321.
Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics, 42*, 49–56.
Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191–210.
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.
Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258–272.
Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics, 2*(2), 99–120. https://doi.org/10.3102/10769986002002099
Myszkowski, N., & Storme, M. (2018). A snapshot of g? Binary and polytomous item-response theory investigations of the last series of the standard progressive matrices (SPM-LS). *Intelligence, 68*, 109–116. https://doi.org/10.1016/j.intell.2018.03.010
Polson, N. G., Scott, J. G., & Windle, J. (2013a). Bayesian inference for logistic models using Pólya-gamma latent variables. *Journal of the American Statistical Association, 108*(504), 1339–1349. https://doi.org/10.1080/01621459.2013.829001
Polson, N. G., Scott, J. G., & Windle, J. (2013b). Bayesian inference for logistic models using P, ólya-gamma latent variables. Retrieved from arXiv:1205.0310 (Most recent version: Feb. 2013).
R Core Team. (2020). R: A language and environment for statistical computing, [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/.
Raven, J. & Raven, J. (2003). Raven progressive matrices. In: R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Springer US. https://doi.org/10.1007/978-1-4615-0153-4_11.
Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology, 19*(1), 137–150. https://doi.org/10.1111/j.2044-8341.1941.tb00316.x
Sanderson, C., & Curtin, R. (2016). Armadillo: A template-based C++ library for linear algebra. *Journal of Open Source Software, 1*(2), 26. https://doi.org/10.21105/joss.00026

Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it-or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education, 18*(4), 289–316.

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods, 19*(3), 506–532.

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Computer-Based Education Research Laboratory: University of Illinois at Urbana-Champaign.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287.

Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement, 32*, 559–574.

Tjoe, H., & de la Torre, J. (2013). Designing cognitively-based proportional reasoning problems as an application of modern psychological measurement models. *Journal of Mathematics Education, 6*(2), 17–26.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology, 61*(2), 287–307.

von Davier, M. (2009). Some notes on the reinvention of latent structure models as diagnostic classification models. *Measurement: Interdisciplinary Research and Perspectives, 7*, 67–74.

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

Windle, J., Polson, N. G., & Scott, J. G. (2014). Sampling pólya-gamma random variates: Alternate and approximate techniques. arXiv preprint arXiv:1405.0506.

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.

Zhang, Z., Zhang, J., Lu, J., & Tao, J. (2020). Bayesian estimation of the DINA model with P, ólya-gamma Gibbs sampling. *Frontiers in Psychology, 11*, 384. https://doi.org/10.3389/fpsyg.2020.00384

Manuscript Received: 23 DEC 2020
Final Version Received: 17 SEP 2021
Published Online Date: 13 JAN 2022