# AN EXPLORATORY DIAGNOSTIC MODEL FOR ORDINAL RESPONSES WITH BINARY ATTRIBUTES: IDENTIFIABILITY AND ESTIMATION

## STEVEN ANDREW CULPEPPER

### UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Diagnostic models (DMs) provide researchers and practitioners with tools to classify respondents into substantively relevant classes. DMs are widely applied to binary response data; however, binary response models are not applicable to the wealth of ordinal data collected by educational, psychological, and behavioral researchers. Prior research developed confirmatory ordinal DMs that require expert knowledge to specify the underlying structure. This paper introduces an exploratory DM for ordinal data. In particular, we present an exploratory ordinal DM, which uses a cumulative probit link along with Bayesian variable selection techniques to uncover the latent structure. Furthermore, we discuss new identifiability conditions for structured multinomial mixture models with binary attributes. We provide evidence of accurate parameter recovery in a Monte Carlo simulation study across moderate to large sample sizes. We apply the model to twelve items from the public-use, Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999 approaches to learning and self-description questionnaire and report evidence to support a three-attribute solution with eight classes to describe the latent structure underlying the teacher and parent ratings. In short, the developed methodology contributes to the development of ordinal DMs and broadens their applicability to address theoretical and substantive issues more generally across the social sciences.

Key words: multivariate ordinal data, cognitive diagnosis, latent class, Bayesian.

## 1. Introduction

Latent structure models (e.g., see von Davier, 2009), which are also referred to as restricted latent class models or diagnostic models (DMs) in measurement and psychometric research, are frequently used to classify respondents into substantively relevant classes. DMs assume that observed binary (e.g., correct/incorrect on educational test items) or ordinal (e.g., rating scales or partial credit) responses relate to a collection of discrete latent variables or attributes. DMs are popular methods for providing a fine-grained classification in terms of substantively important latent attributes, such as educational skills or psychological states.

Prior research developed both confirmatory (e.g., see de la Torre, 2011; Henson, Templin, & Willse, 2009; von Davier, 2008) and exploratory (e.g., see Y. Chen, Liu, Xu, & Ying, 2015; Y. Chen, Culpepper, Chen, & Douglas, 2018; Culpepper, 2019; Culpepper & Chen, 2018; Xu, 2017; Xu & Shang, 2018) DMs for binary data. Binary DMs are used in educational research to classify students into skill-based attribute profiles using tests of fraction-subtraction (de la Torre & Douglas, 2004,2008; DeCarlo, 2011), spatial rotation (Culpepper, 2015), and geometric sequences (Shute, Hansen, & Almond, 2008). Recent applications also used binary DMs to model learning (Y. Chen, Culpepper, Wang, & Douglas, 2018; Kaya & Leite, 2017; Li, Cohen, Bottge, & Templin, 2016; Madison & Bradshaw, 2018; Wang, Yang, Culpepper, & Douglas, 2017) and detect skill acquisition (Ye, Fellouris, Culpepper, & Douglas, 2016). Additional research applied binary DMs to items measuring pathological gambling (Templin & Henson, 2006) and social anxiety (Y. Chen et al., 2015).

DMs are widely applied to binary response data; however, binary response models are not applicable to the wealth of ordinal data collected by educational, psychological, and behavioral researchers. For instance, large-scale testing programs, such as the National Assessment of Educational Progress, include ordinal response variables to evaluate students' performance on essays or constructed response tasks. In these cases, students may receive partial credit in the form of an ordinal score rather than a binary correct or incorrect. Furthermore, there is a longstanding tradition in the social sciences for researchers to measure constructs with ordinal responses. For instance, large-scale assessment background questionnaires include items with rating scales to measure individual differences in opportunity to learn, educational aspirations, and academic experiences. Each of the aforementioned constructs provides a context for students' dispositions and learning environment, and fine-grained classifications with a general DM could provide educators and practitioners with information to identify students in need of interventions or to describe patterns or profiles of human development as measured by rating scales. Clearly, research on ordinal DMs provides opportunities to advance diagnostic research in education and, more generally, in the social sciences.

In this study, we offer a general, exploratory DM for ordinal response data to broaden the applicability of these methods in the social sciences. Specifically, prior research developed ordinal DMs (J. Chen & de la Torre, 2013,2018; Haberman, von Davier, & Lee, 2008; Karelitz, 2004; R. Liu & Jiang, 2018; Ma & de la Torre, 2016,2019, Templin, 2004; von Davier, 2008) and this paper offers three contributions to the literature. First, we discuss identifiability of structured mixture models for multivariate multinomial response data. Fang, Liu, and Ying (2019) established sufficient conditions for the identifiability of a structured multinomial mixture model for polytomous attributes. We consider the case for binary attributes and show less stringent restrictions are needed to establish sufficient conditions for identifiability than for the polytomous attribute case. Second, we propose a fully exploratory ordinal diagnostic model for uncovering the latent structure and inferring the underlying latent processes. The introduction of an exploratory diagnostic model for ordinal data is an important contribution given current ordinal DMs are confirmatory models that are dependent upon correct specification of the latent structure with expert knowledge or substantive theory. As demonstrated in research on binary DMs (e.g., see Henson & Templin, 2007; Rupp & Templin, 2008), misspecification likely leads to inaccurate classifications and biased parameter estimates. Our model extends the confirmatory general ordinal model of J. Chen and de la Torre (2018) to the exploratory setting. Specifically, our model uses a cumulative probit link function and a fully saturated model with main effects and higher-order interactions involving the latent binary attributes. In order to uncover the underlying structure, we adapt Bayesian model selection tools as described for binary DMs (Culpepper, 2019) to infer the relationship between attributes and observed ordinal responses. Third, we present a Bayesian formulation that provides a convenient approach for imposing monotonicity restrictions on the item response functions (IRFs), so that higher attribute levels are associated with non-decreasing observed response probabilities. It is important to note that enforcing monotonicity conditions is not a requirement for identifying model parameters. Instead, monotonicity of latent class response probabilities may improve interpretation of the exploratory solution.

The remainder of this paper includes four sections. The first section introduces the unstructured mixture model for multivariate multinomial response data and presents new results concerning the identifiability of model parameters. The second section introduces a structured mixture model for ordinal data and presents a Bayesian formulation. Moreover, we review issues related to monotonicity conditions and the coding of attributes. The third section reports results from a Monte Carlo simulation study regarding the recovery of model parameters. We provide evidence of accurate parameter recovery for sample sizes as small as 500 observations. The fourth section presents an application of the exploratory ordinal DM to twelve items of the public-use, Early Childhood Longitudinal Study, Kindergarten (ECLS-K) Class of 1998–1999 approaches to learn-

ing and self-description (ALS) questionnaire with $n = 13{,}354$ complete observations. We discuss the implications of the study, offer future research directions, and provide concluding remarks in the final section. We direct readers to the "Appendix" for a description of the Gibbs sampling algorithm and full conditional distributions.

## 2. Mixture Models for Multinomial Response Data

The purpose of this section is to discuss a mixture model for multinomial response data. We first outline a model and then discuss the identifiability of model parameters.

### 2.1. Mixture of Multinomial Response Data

Let $Y_j$ for $j = 1, \ldots, J$ be a random ordinal response with a realization $y_j \in \{0, \ldots, M_j - 1\}$ where $M_j \geq 2$ denotes the number of response options. The random $J$-vector is denoted by $Y = (Y_1, \ldots, Y_J)^\top$, and the observed vector of responses is $y = (y_1, \ldots, y_J)^\top$. The support for $Y$ is defined as $y \in \times_{j=1}^{J} \{0, \ldots, M_j - 1\}$, which implies there are $\prod_{j=1}^{J} M_j$ observed response patterns.

The purpose of exploratory ordinal DMs is to uncover a latent structure involving fewer attribute profile patterns to describe the $\prod_{j=1}^{J} M_j$ observed response patterns. We introduce a collection of $K$ binary attributes to provide a more parsimonious representation of the $\prod_{j=1}^{J} M_j$ observed patterns using $2^K$ latent profile patterns. We refer to the $K$-vector of latent binary attributes as $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K) \in \{0, 1\}^K$.

Unstructured latent class models are important models in psychometric research (e.g., see Fang et al. 2019; Green, 1951; Hojtink & Molenaar, 1997; McDonald, 1962; Proctor, 1970; Rost, 1988). For multinomial data, the unstructured model includes a $M_j$-vector of category response probabilities for each class and item. To simplify notation, we use the integer representation of the ordinal attributes to refer to the latent classes. That is, let $\boldsymbol{v} = (2^{K-1}, 2^{K-2}, \ldots, 1)^\top$ and note that $\boldsymbol{\alpha}^\top \boldsymbol{v}$ is a bijection that maps the attribute profiles to integers $c \in \{0, \ldots, 2^K - 1\}$. Let the probability of observing a response of $m$ on item $j$ for members with the latent attribute profile of class $c$ is $\theta_{jcm} = P(Y_j = m | \boldsymbol{\alpha}^\top \boldsymbol{v} = c)$ and define the vector of response probabilities for class $c$ as $\boldsymbol{\theta}_{jc} = (\theta_{jc0}, \ldots, \theta_{jc,M_j-1})^\top$. The model for $Y_j$ given membership in class $c$ is a multinomial probability mass function defined as,

$$p(y_j | \boldsymbol{\alpha}^\top \boldsymbol{v} = c, \boldsymbol{\theta}_{jc}) = \sum_{m=0}^{M_j-1} \theta_{jcm} \, \mathcal{I}(y_j = m) \tag{1}$$

where $\mathcal{I}(\cdot)$ is the indicator function. Let $\boldsymbol{\pi}$ be a $2^K$ vector of structural latent class probabilities where element $c$ is defined as $P(\boldsymbol{\alpha}^\top \boldsymbol{v} = c) = \pi_c$. Furthermore, let $\boldsymbol{\Theta} \in \mathbb{R}^{M_1 \times \cdots \times M_J \times 2^K}$ denote the response probabilities for all items and latent classes. We assume the responses in the $J$-vector $Y$ are independent given $\boldsymbol{\alpha}$, which implies the distribution of $Y$ given the item and structural parameters is,

$$p(y | \boldsymbol{\Theta}, \boldsymbol{\pi}) = \sum_{c=0}^{2^K-1} \pi_c \prod_{j=1}^{J} p(y_j | \boldsymbol{\alpha}^\top \boldsymbol{v} = c, \boldsymbol{\theta}_{jc}). \tag{2}$$

Let $Y_i = (Y_{i1}, \ldots, Y_{iJ})^\top$ be a random $J$-vector of ordinal responses for individual $i$ and define the realized values as $y_i = (y_{i1}, \ldots, y_{iJ})^\top$. Let the random $n \times J$ matrix of mixed ordinal responses

be $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^\top$. The likelihood of observing a sample of $i = 1, \ldots, n$ independent observations is,

$$p(\mathbf{y}_1, \ldots, \mathbf{y}_n | \mathbf{\Theta}, \boldsymbol{\pi}) = \prod_{i=1}^n p(\mathbf{y}_i | \mathbf{\Theta}, \boldsymbol{\pi}). \tag{3}$$

An equivalent approach for describing the model in Eq. (3) is to instead consider the implied marginal probabilities for all response patterns (e.g., see Allman, Matias, & Rhodes, 2009; J. Liu, Xu, & Ying, 2013). The distribution of $\mathbf{Y}$ for class $c$ and all response patterns is defined by the $\left(\prod_{j=1}^J M_j\right)$-vector,

$$\mathbb{P}_c = \bigotimes_{j=1}^J \boldsymbol{\theta}_{jc}. \tag{4}$$

which is a Kronecker product of all item response probabilities for class $c$. Let $\mathbf{T} = (\mathbb{P}_0, \ldots, \mathbb{P}_{2^K-1})$ be a $\left(\prod_{j=1}^J M_j\right) \times 2^K$ matrix denoting the model probabilities for all item response patterns and latent classes. The implied marginal probability for the model is a finite mixture of the $2^K$ classes,

$$\mathbb{P} = \mathbf{T}\boldsymbol{\pi} = \sum_{c=0}^{2^K-1} \pi_c \mathbb{P}_c. \tag{5}$$

### 2.2. Identifiability Conditions for Binary Attributes and Multinomial Responses

This subsection discusses conditions for identifying mixtures of multinomial response data. We first introduce a few definitions, examples and prior results and then present an identifiability Theorem for structured multinomial mixture models with binary attributes.

**Definition 1.** *(Identifiability)* The model parameters $(\mathbf{\Theta}, \boldsymbol{\pi}) \in \Omega$ are identifiable if $p(\mathbf{y}_1, \ldots, \mathbf{y}_n | \mathbf{\Theta}, \boldsymbol{\pi}) = p(\mathbf{y}_1, \ldots, \mathbf{y}_n | \mathbf{\Theta}', \boldsymbol{\pi}')$ implies $\mathbf{\Theta} = \mathbf{\Theta}'$ and $\boldsymbol{\pi} = \boldsymbol{\pi}'$.

*Remark 1.* Definition 1 is the classic notion of likelihood identifiability where each unique combination of parameter values corresponds to a different value of the likelihood function.

**Definition 2.** *(Kruskal Rank)* For a matrix $\mathbf{T}$, the Kruskal rank of $\mathbf{T}$, i.e., $\text{rank}_K(\mathbf{T})$, is the largest number $I$ such that every set of $I$ columns of $\mathbf{T}$ are independent.

*Remark 2.* Note that $\text{rank}_K(\mathbf{T}) \leq \text{rank}(\mathbf{T})$. If $\mathbf{T}$ is of full column rank then $\text{rank}_K(\mathbf{T}) = \text{rank}(\mathbf{T})$ (Allman et al., 2009).

**Definition 3.** *(Three-way array)* Let $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]$ be a three-way array where each of the matrices has $R$ columns and the $(i, j, k)$ element is defined as $t_{ijk} = \sum_{r=1}^R t_{1ir} t_{2jr} t_{3kr}$ with $\mathbf{T}$ defined as

$$\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] = \sum_{r=1}^R \boldsymbol{t}_{1r} \otimes \boldsymbol{t}_{2r} \otimes \boldsymbol{t}_{3r} \tag{6}$$

where $\boldsymbol{t}_{ir}$ is column $r$ of $\mathbf{T}_i$ for $i = 1, 2, 3$.

*Example 1.* (Latent class models as a three-way array) We can write the model in Eq. (5) as a three-way array by letting $\mathbf{T}_i$ for $i = 1, 2, 3$ be matrices for three sets of non-overlapping items. Specifically, let $\mathbb{J}_i$ for $i = 1, 2, 3$ denote three mutually exclusive and exhaustive sets of items. The distribution of responses for $j \in \mathbb{J}_i$ for class $c$ is

$$\mathbb{P}_{ic} = \bigotimes_{j \in \mathbb{J}_i} \boldsymbol{\theta}_{jc} \tag{7}$$

where $\mathbb{P}_{ic}$ is a vector of length $\prod_{j \in \mathbb{J}_i} M_j$. Consequently, $\mathbf{T}_i = (\mathbb{P}_{i0}, \ldots, \mathbb{P}_{i,2^K-1})$ and $R = 2^K$ given each $\mathbf{T}_i$ is a $(\prod_{j \in \mathbb{J}_i} M_j) \times 2^K$ matrix of latent class response pattern probabilities for the items in $\mathbb{J}_i$. The three-way array representation of Eq. (5) is,

$$\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] = \sum_{c=0}^{2^K-1} \pi_c \mathbb{P}_{1c} \otimes \mathbb{P}_{2c} \otimes \mathbb{P}_{3c}. \tag{8}$$

**Theorem 1.** (Allman et al., 2009). *Consider the three-way array representation of a multinomial mixture model with $i = 1, \ldots, r$ classes. Suppose all entries of $\boldsymbol{\pi}$ are positive. Then, if*

$$\mathrm{rank}_K(\mathbf{T}_1) + \mathrm{rank}_K(\mathbf{T}_2) + \mathrm{rank}_K(\mathbf{T}_3) \geq 2r + 2$$

*the parameters of the model are uniquely identifiable, up to label swapping.*

*Remark 3.* Note Allman et al.'s (2009) result is based upon Kruskal's (1976, 1977) theorem concerning the uniqueness of three-way arrays. Similar to Y. Chen, Culpepper, and Liang (2018) and Fang et al. (2019), we use Kruskal's (1976, 1977) result to establish an identifiability condition for multinomial response data with binary attributes.

**Definition 4.** *(The $\mathbf{Q}$ matrix)* Let $\mathbf{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_J)^\top$ be a $J \times K$ binary matrix where $q_{jk}$ is the $(j, k)$ element of $\mathbf{Q}$ and $q_{jk} = 1$ if attribute $k$ is required for item $j$ and $q_{jk} = 0$ otherwise.

*Remark 4.* The $\mathbf{Q}$ matrix imposes structure on the response probabilities $\boldsymbol{\Theta}$. For instance, items with $\boldsymbol{q}_j = \boldsymbol{e}_k$ are referred to as simple structure items that load only on attribute $k$. Simple structure items include only two distinct vectors of responses probabilities; $\boldsymbol{\theta}_{j1} = (\theta_{j10}, \ldots, \theta_{j1,M_j-1})^\top$ for $\alpha_k = 1$ and $\boldsymbol{\theta}_{j0} = (\theta_{j00}, \ldots, \theta_{j0,M_j-1})^\top$ for $\alpha_k = 0$. In contrast, items with complex structure that load onto two or more attributes include more than two vectors of response probabilities. Readers are directed to J. Chen and de la Torre (2018) for a discussion of the relation between $\mathbf{Q}$ and various reduced ordinal DMs.

**Theorem 2.** $\boldsymbol{\Theta}$ *and $\boldsymbol{\pi}$ are identifiable, up to label swapping, if $\boldsymbol{\pi}$ is positive, each item has distinct item response functions for at least two latent classes, and conditions (C1) and (C2) are satisfied:*

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

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

*Proof.* We show the bound in Theorem 1 is achieved. Let $\mathbf{Q}$ be defined as in (C1) and let $\mathbb{J}_1 = \{1, \ldots, K\}$, $\mathbb{J}_2 = \{K+1, \ldots, 2K\}$, and $\mathbb{J}_3 = \{2K+1, \ldots, J\}$. Fang et al. (2019) showed for the polytomous attribute case that the simple structure assumption and distinct latent class IRFs implies $\mathbf{T}_1$ and $\mathbf{T}_2$ are full column rank, which for the binary attribute case indicates $\mathrm{rank}_K(\mathbf{T}_1) = \mathrm{rank}_K(\mathbf{T}_2) = 2^K$. That is, simple structure for items in $\mathbf{T}_1$ and $\mathbf{T}_2$ implies that

$$\mathbf{T}_1 = \bigotimes_{j=1}^{K} \mathbf{p}_j, \ \ \mathbf{T}_2 = \bigotimes_{j=K+1}^{2K} \mathbf{p}_j \tag{9}$$

where $\mathbf{p}_j = (\boldsymbol{\theta}_{j0}, \boldsymbol{\theta}_{j1})$ is a $M_j \times 2$ matrix of response probabilities. The assumption of distinct probabilities (i.e., $\boldsymbol{\theta}_{j0} \neq \boldsymbol{\theta}_{j1}$) implies $\mathrm{rank}(\mathbf{p}_j) = 2$ for all $j \in \{1, \ldots, 2K\}$ and properties of Kronecker products imply $\mathrm{rank}(\mathbf{T}_1) = \prod_{j=1}^{K} \mathrm{rank}(\mathbf{p}_j) = 2^K$ and, similarly, $\mathrm{rank}(\mathbf{T}_2) = 2^K$.

We next establish that the $\mathrm{rank}_K(\mathbf{T}_3) \geq 2$ for the binary attribute case. In other words, it must be shown that every pair of the $2^K$ columns of $\mathbf{T}_3$ are independent. A proof by contradiction proceeds by assuming columns $c$ and $c'$ are dependent. That is, columns $c$ and $c'$ of $\mathbf{T}_3$ are $\mathbb{P}_{3c}$ and $\mathbb{P}_{3c'}$, respectively, and are dependent if there exists nonzero $u_c$ and $u_{c'}$ such that $u_c \mathbb{P}_{3c} + u_{c'} \mathbb{P}_{3c'} = \mathbf{0}$ for all response patterns for items in $\mathbb{J}_3$. However, if (C2) is satisfied there exists an item $j \in \{2K + 1, \ldots, J\}$ such that $\boldsymbol{\theta}_{jc} \neq \boldsymbol{\theta}_{jc'}$. There is at least one response pattern $\boldsymbol{y}_{3(j)} \in \times_{j' \in \mathbb{J}_3 \setminus j} \{0, \ldots, M_{j'} - 1\}$ with a nonzero probability for items in $\mathbb{J}_3$ that excludes item $j$. Let the value of the distribution function for $\boldsymbol{y}_{3(j)}$ be $p_{3c(j)}$ for class $c$ and $p_{3c'(j)}$ for class $c'$. The $M_j$ elements of $\mathbb{P}_{3c}$ and $\mathbb{P}_{3c'}$ corresponding to the distribution of $\boldsymbol{y}_{3(j)}$ and values of $y_j = 0, \ldots, M_j - 1$ equals $p_{3c(j)} \boldsymbol{\theta}_{jc}$ and $p_{3c'(j)} \boldsymbol{\theta}_{jc'}$, respectively. If $\mathbb{P}_{3c}$ and $\mathbb{P}_{3c'}$ are dependent then $u_c p_{3c(j)} \boldsymbol{\theta}_{jc} + u_{c'} p_{3c'(j)} \boldsymbol{\theta}_{jc'} = \mathbf{0}$ for nonzero $u_c$ and $u_{c'}$. However, $\boldsymbol{\theta}_{jc}$ and $\boldsymbol{\theta}_{jc'}$ are distinct (i.e., independent) and the equality is only achieved if $u_c = u_{c'} = 0$. $c$ and $c'$ were chosen arbitrarily, so $\mathrm{rank}_K(\mathbf{T}_3) \geq 2$. $\qquad \square$

*Remark 5.* Fang et al. (2019) showed that multinomial mixture models with polytomous attributes are identified if $\mathbf{Q}$ include three $\mathbf{I}_K$ matrices for three non-overlapping sets of items. The requirement of three $\mathbf{I}_K$ matrices in $\mathbf{Q}$ is restrictive and Fang et al. (2019) note their sufficient conditions could be relaxed if "additional constraints on parameters are assumed" (p. 25). Theorem 2 shows that the sufficient condition can be reduced to two $\mathbf{I}_K$ matrices if attributes are binary and additional items exist in $\mathbf{Q}'$ that distinguish all of the classes.

## 3. A General Diagnostic Model (GDM) for Ordinal Data

We next discuss two issues related to the GDM for ordinal data. First, we present a structured mixture modeling framework, which uses a cumulative link function for ordinal responses, and discusses monotonicity conditions. Second, we present a fully Bayesian formulation and discuss posterior inference.

### 3.1. Structured Mixture Model with a Cumulative Link Function

The model in Eq. (3) may include many parameters for each item for even modest values of $K$. Consequently, some degree of regularization of the model parameters may improve estimation. There is not a clear approach for regularizing the latent class response probabilities as parameterized in Eq. (3). In this subsection, we describe a reparameterization of $\boldsymbol{\Theta}$ that enables the application of Bayesian variable selection techniques.

One strategy for modeling ordinal responses is to use a cumulative link function (e.g., see Albert & Chib, 1993) for an unstructured mixture model (e.g., see Bao & Hanson, 2015; DeYoreo, Reiter, & Hillygus, 2017; Kottas, Müller, & Quintana, 2005). Specifically, we can rewrite the latent class probabilities for items $j$ $(\theta_{jc0}, \ldots, \theta_{jc,M_j-1})$ using a cumulative link function $\Psi(\cdot)$ and $M_j - 1$ thresholds, $(\tau_{jc1}, \ldots, \tau_{jc,M_j-1})$ where $\tau_{jc0} < \tau_{jc1} < \cdots < \tau_{jc,M_j-1} < \tau_{jcM_j}$ with the endpoints defined as $\tau_{jc0} = -\infty$ and $\tau_{jcM_j} = \infty$. Furthermore, a commonly used parameterization fixes $\tau_{jc1} = 0$ and incorporates a latent class mean parameter for item $j$, $\mu_{jc}$. Therefore, the cumulative probability of a response at level $m$ or less using a cumulative link function is,

$$P(Y_j \leq m \,|\, \boldsymbol{\alpha}' \boldsymbol{v} = c, \tau_{jc,m+1}, \mu_{jc}) = \sum_{m'=0}^{m} \theta_{jcm'} = \Psi\left(\tau_{jc,m+1} - \mu_{jc}\right) \qquad (10)$$

where $\Psi(\cdot)$ is a generic cumulative link function. Accordingly, the probability of response $m$ by members of class $c$ to item $j$ is $\theta_{jcm} = \Psi\left(\tau_{jc,m+1} - \mu_{jc}\right) - \Psi\left(\tau_{jcm} - \mu_{jc}\right)$ where $\Psi\left(\tau_{jc0} - \mu_{jc}\right) = 0$ and $\Psi\left(\tau_{jcM_j} - \mu_{jc}\right) = 1$.

Prior studies estimated the $2^K$ $\mu_{jc}$'s in an unstructured fashion (e.g., see DeYoreo et al., 2017; DeYoreo & Kottas, 2018; Kottas et al., 2005). In contrast, we reparameterize and impose structure on the latent class mean parameters in Eq. (10) by defining $\mu_{jc} = \boldsymbol{a}_c^\top \boldsymbol{\beta}_j$ where $\boldsymbol{\beta}_j$ is a $P$-vector of parameters for item $j$ and $\boldsymbol{a}_c$ is a design vector for class $c$ that uniquely maps $\boldsymbol{\alpha}^\top \boldsymbol{v}$ to a set of $P$ dummy coded variables (the coding for $\boldsymbol{a}_c$ is discussed in the next subsection). Our reparameterized model for the cumulative conditional probability of a response at level $m$ or less is,

$$P(Y_j \le m \left| \boldsymbol{\alpha}^\top \boldsymbol{v} = c, \boldsymbol{\beta}_j, \tau_{jc0}, \dots, \tau_{jcm}\right) = \sum_{m'=0}^{m} \theta_{jcm'} = \Psi\left(\tau_{jc,m+1} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right). \quad (11)$$

Note the model in Eq. (11) implies that the conditional probability an observed response equals $m$ is $\theta_{jcm} = \Psi\left(\tau_{jc,m+1} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right) - \Psi\left(\tau_{jcm} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)$ and the conditional probability mass function for item $j$ under the reparameterized model is,

$$p(y_j \left| \boldsymbol{\alpha}^\top \boldsymbol{v} = c, \boldsymbol{\beta}_j, \tau_0, \dots, \tau_M\right) = \sum_{m=0}^{M_j-1} \left[\Psi\left(\tau_{m+1} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right) - \Psi\left(\tau_m - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j\right)\right] \mathcal{I}(y_j = m). \quad (12)$$

Let $\mathbf{B} = \left(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J\right)^\top$ be a $J \times 2^K$ matrix of regression parameters. Furthermore, let $\boldsymbol{\tau}_j$ denote the $M_j - 1$ thresholds for item $j$ and let $\mathbf{T} = (\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_J)^\top$ denote the thresholds for all $J$ items. Note that we refer to the reparameterized likelihood function for a sample of $n$ independent observations as $p(\mathbf{Y}|\mathbf{B}, \mathbf{T}, \boldsymbol{\pi})$.

*3.1.1. Coding Attribute Profiles* We next discuss how to code attribute profile $c$ as the vector of latent predictors, $\boldsymbol{a}_c$, in the cumulative link function. The most general exploratory model is referred to as a saturated model, which includes all main- and higher-order interaction effects terms of the attribute levels. Accordingly, the saturated model includes all effects and the number of design predictors and coefficients is $P = 2^K$. Researchers can in principle estimate exploratory models with fewer parameters (i.e., $P < 2^K$) to estimate more restricted models, but the discussion that follows considers the saturated model.

Table 1 presents an example design matrix for the saturated model with $K = 3$. In this case, the design vector for an arbitrary $\boldsymbol{\alpha}$ is $\boldsymbol{a} = (1, \alpha_3, \alpha_2, \alpha_2\alpha_3, \alpha_1, \alpha_1\alpha_3, \alpha_1\alpha_2, \alpha_1\alpha_2\alpha_3)$. In the saturated model with $K = 3$ the main effects are $\alpha_1$, $\alpha_2$, and $\alpha_3$, the two-way interactions are $\alpha_1\alpha_2$, $\alpha_1\alpha_3$, and $\alpha_2\alpha_3$, the three-way interaction is $\alpha_1\alpha_2\alpha_3$. The second column of Table 1 presents the $2^K = 8$ latent classes, and columns three to ten show the elements of the design vectors $\boldsymbol{a}_c$ for each $c$. For instance, latent class $\boldsymbol{\alpha}_3 = (0, 1, 1)^\top$ corresponds with the design vector $\boldsymbol{a}_3 = (1, 1, 1, 1, 0, 0, 0, 0)^\top$. Finally, Table 1 shows that the first element of $\boldsymbol{a}_c$ is a one for the intercept, which indicates the reference group is the class with $\boldsymbol{\alpha} = (0, 0, 0)^\top$.

It is worth mentioning another reason for studying the design matrix. Specifically, as shown in the first column of Table 1 we order the elements of $\boldsymbol{a}$ according to the integer representation of the binary profile, $\boldsymbol{\alpha}$. Table 1 demonstrates a simple coefficient labeling strategy in the second row of Table 1, which includes attribute profile labels in the order of the integer representation. Table 1 shows that the labels in the second row provide an alternative way to refer to effects denoted by the products of the $a$'s. We see that the alternative main effect labels in Table 1 all have two

TABLE 1.
Example design matrix for $K = 3$ by attribute profile, $\boldsymbol{\alpha}$.

| $c$ | Class $\boldsymbol{\alpha}$ | 1 000 | $\alpha_3$ 001 | $\alpha_2$ 010 | $\alpha_2\alpha_3$ 011 | $\alpha_1$ 100 | $\alpha_1\alpha_3$ 101 | $\alpha_1\alpha_2$ 110 | $\alpha_1\alpha_2\alpha_3$ 111 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 000 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 001 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 010 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 011 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 4 | 100 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 5 | 101 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 6 | 110 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 7 | 111 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

$c$ is the integer representation of $\boldsymbol{\alpha}$.

zeros included. For instance, the main effect for $\alpha_1$ corresponds to the label "100." In contrast, the labels for the two-way interaction effects include one zero. For instance, an alternative label for the interaction involving attributes one and three (i.e., $\alpha_1\alpha_3$) is "101." Furthermore, we see that the intercept label is "000" and the label for the interaction among the highest attribute levels (i.e., the coefficient for $\alpha_1\alpha_2\alpha_3$) is "111." For a general value of $K$, this labeling scheme implies that the intercept label has $K$ zeros, the labels for main effects include $K - 1$ zeros, two-way interaction labels have $K - 2$ zeros, and $K$-way interaction labels are without any zeros.

*3.1.2. Monotonicity Conditions*    We next discuss monotonicity conditions for the latent class IRFs. Specifically, the monotonicity condition we consider is

$$P(Y_j \geq m | \boldsymbol{\alpha}, \boldsymbol{\beta}_j) \geq P(Y_j \geq m | \boldsymbol{\alpha}', \boldsymbol{\beta}_j), \text{ if } \boldsymbol{\alpha} \geq \boldsymbol{\alpha}'. \tag{13}$$

The condition in Eq. (13) states that the probability of responding $m$ or greater on item $j$ for members with attribute profile $\boldsymbol{\alpha}$ is greater for all classes that are element-wise smaller than $\boldsymbol{\alpha}$. It is important to note that the condition in Eq. (13) is satisfied for the ordinal GDM if $\mu_{jc} \geq \mu_{jc'}$ for $\boldsymbol{\alpha}_c \geq \boldsymbol{\alpha}_{c'}$. Consequently, the monotonicity conditions in Eq. (13) can be translated into a lower-bound condition for each coefficient (e.g., see Y. Chen, Culpepper, & Liang, 2018). We let $L_{jp}$ denote the lower bound for $\beta_{jp}$.

### 3.2. Bayesian Model

This subsection outlines a Bayesian formulation for the ordinal latent class model using a cumulative probit link function. We first discuss the model formulation and then discuss posterior inference.

*3.2.1. Model Formulation and Priors*    We use a data augmentation strategy as found in Bayesian item response theory models (e.g., see Albert 1992; Béguin & Glas, 2001; Culpepper, 2016) for the probit cumulative link function. That is, we introduce a deterministic relationship between the random ordinal response $Y_{ij}$ and a continuous augmented latent variable $Y_{ij}^*$ so that $Y_{ij} = m$ whenever $\tau_{jm} < Y_{ij}^* < \tau_{j,m+1}$. Next, we assume the augmented variable $Y_{ij}^*$ is conditionally normally distributed as $Y_{ij}^* | \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j \sim \mathcal{N}(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j, 1)$.

We consider a categorical prior for $\boldsymbol{\alpha}_i$ (e.g., see Culpepper, 2015) conditioned upon the latent class structural probabilities $\boldsymbol{\pi}$, which is the $2^K$ vector of class probabilities. The conditional probability that $\boldsymbol{\alpha}_i$ has the attribute configuration of class $c$ given $\boldsymbol{\pi}$ is,

$$P\left(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c | \boldsymbol{\pi}\right) = \sum_{c'=0}^{2^K-1} \pi_{c'} \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c). \tag{14}$$

Note that we specify conjugate $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{n}_0)$ prior for the structural parameters and implement a uniform prior with $\boldsymbol{n}_0 = \mathbf{1}_{2^K}^\top$.

We employ the stochastic search variable selection prior specification for $\mathbf{B}$ as described in Culpepper (2019). The prior for $\mathbf{B}$ is conditioned on $\mathbf{Q}$. The prior for $\boldsymbol{\beta}_j$ given $\boldsymbol{q}_j$ is,

$$p\left(\boldsymbol{\beta}_j | \boldsymbol{q}_j\right) \propto \prod_{p=0}^{2^K-1} v_p^{-1/2} \exp\left(-\frac{1}{2}\beta_{jp}^2/v_p\right) \mathcal{I}\left(\beta_{jp} > L_{jp}\right) \tag{15}$$

$$v_p = \tilde{q}_{jp}/c_1 + (1 - \tilde{q}_{jp})/c_0. \tag{16}$$

Note as discussed in Culpepper (2019) $\tilde{\boldsymbol{q}}_j$ is a $2^K$ vector that includes all possible products of the elements of $\boldsymbol{q}_j$. For instance, for $K = 2$, $\boldsymbol{q}_j = (q_{j1}, q_{j2})^\top$ and $\tilde{\boldsymbol{q}}_j = (1, q_{j2}, q_{j1}, q_{j1}q_{j2})^\top$. The precision for $\beta_{jp}$ is $c_1$ when $\tilde{q}_{jp} = 1$ and $c_0$ for $\tilde{q}_{jp} = 0$. 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 $c_0 = 500$ in this paper) to reflect a smaller variance for $\beta_{jp}$ and a distribution that is more concentrated near zero. In contrast, $c_1$ is fixed as a smaller value (e.g., we set $c_1 = 1$) to depict a prior distribution for $\beta_{jp}$ that is consistent with an active coefficient.

Under the SSVS formulation, the elements of $\mathbf{Q}$ denote the "activeness" of the coefficients. We assume the elements of $\mathbf{Q}$ are independently distributed given a hyper parameter $\omega$. Specifically, the prior is $q_{jk}|\omega \sim \text{Bernoulli}(\omega)$. Furthermore, $\omega \sim \text{Beta}(a, b)$ is a Beta prior for $\omega$ and we set $a = b = 1$ to employ a uniform prior for $\omega$ in the simulation study and application.

Prior research notes the difficulty associated with estimating the thresholds in mixture models for ordinal data (e.g., see Bao & Hanson, 2015; DeYoreo et al., 2017; DeYoreo & Kottas, 2018; Kottas et al., 2005) and instead recommend treating the thresholds as fixed rather than random. We investigated this issue and conducted several Monte Carlo simulation studies to assess parameter recovery when thresholds are treated as random. In particular, we sampled thresholds using the Metropolis–Hastings approach of Cowles (1996). The results from these numerical experiments supported prior research regarding the difficulty with estimating model parameters when thresholds are random. Accordingly, we follow the recommendation of prior researchers and fix the thresholds. In particular, we fix the thresholds as $(\tau_{jc1}, \ldots, \tau_{jc,M_j-1}) = (0, 2, \ldots, 2(M_j - 2))$. Fixing the thresholds has the added benefit of computational speed by avoiding Metropolis–Hastings updates for thresholds (e.g., see Cowles, 1996).

*3.2.2. Posterior Inference* The posterior distribution of interest is,

$$p\left(\mathbf{Y}^*, \mathbf{A}, \mathbf{B}, \boldsymbol{\pi}, \mathbf{Q}, \omega | \mathbf{Y}\right) \propto p\left(\mathbf{Y}|\mathbf{Y}^*\right) p\left(\mathbf{Y}^*|\mathbf{A}, \mathbf{B}\right) p(\mathbf{A}|\boldsymbol{\pi}) p(\mathbf{B}|\mathbf{Q}) p(\mathbf{Q}|\omega) p(\boldsymbol{\pi}) p(\omega) \tag{17}$$

where $\mathbf{Y}^*$ is an $n \times J$ random matrix of augmented data and $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$ refers to the latent attributes. We use a Gibbs sampling algorithm to approximate the posterior distribution (see the "Appendix" for additional details).

## 4. Monte Carlo Simulation Study

### 4.1. Overview

This section presents results from a Monte Carlo simulation study to evaluate the accuracy of parameter recovery for different sample sizes. The population model parameters were defined by the values estimated in the application (e.g., see the values for **B** and $\boldsymbol{\pi}$ in Table 5) to ensure the simulated parameters are consistent with applied data. Accordingly, the simulation used $J = 12$, $M_j = 4$, and $K = 3$. We considered four sample sizes of $n = 500, 1000, 1500$, and 2000 and generated 100 replications per condition for a total of 400 simulated datasets. For each replication of the simulation study, we executed a single chain of length 100,000 with a burnin of 20,000 iterations. Additionally, the SSVS parameters were defined as $c_1 = 1$ and $c_0 = 500$.

We measure parameter recovery by computing the expected absolute deviation (EAD) of the estimates from the population values. In particular, we assess the accuracy of item parameter estimates by comparing the estimated and population cumulative IRFs (i.e., see Eq. 11). That is, the estimated cumulative IRF for class $c$ is,

$$\hat{F}_{jcm} = \sum_{m'=0}^{m} \hat{\theta}_{jcm'} \tag{18}$$

where $\hat{\theta}_{jcm'}$ is defined using a probit link and the posterior mean for $\boldsymbol{\beta}_j$. We estimate the expected absolute deviation between the item $j$ across $T$ scans of the posterior using

$$EAD(F_{jcm}) = \frac{1}{T} \sum_{t=1}^{T} \left| \hat{F}_{jcm} - F_{jcm} \right|. \tag{19}$$

For $K = 3$ there are eight parameters for each item, so to simplify the presentation of results we compute an overall average expected absolute deviation for item $j$ and response level $m$ as,

$$EAD(F_{jm}) = \frac{1}{2^K} \sum_{c=0}^{2^K-1} EAD(F_{jcm}). \tag{20}$$

We report EAD for all items and sample sizes for $m = 0, 1, 2$ (note the $m = 3$ case is excluded because $F_{jc3} = 1$).

We also assess accuracy in the estimation of the latent class structural probabilities, $\boldsymbol{\pi}$. That is we estimate $\boldsymbol{\pi}$ by computing the posterior mean and then compute the EAD for each element.

### 4.2. Results

The Monte Carlo simulation results provide evidence that the developed algorithm accurately recovers the item and structural parameters across all sample sizes. Table 2 reports the average EAD for all items by response levels $m = 0, 1, 2$ and sample sizes of $n = 500, 1000, 1500, 2000$. The results in Table 2 show that average EAD for the cumulative IRFs is less than 0.038, 0.027, 0.023, and 0.021 for sample sizes of $n = 500, 1000, 1500, 2000$, respectively. Furthermore, Table 3 reports the expected absolute deviation of the estimate $\boldsymbol{\pi}$ and shows that the latent class probabilities are accurately recovered.

We also recorded the run-time for the developed algorithm. We found evidence that Markov chains with 100,000 iterations required, on average, 146.6 s, 276.0 s, 417.3 s, and 639.4 s for sample sizes of 500, 1000, 1500, and 2000 on a cluster with 2.50 GHz processors.

TABLE 2.
Summary of Monte Carlo simulation study average expected absolute deviation for the cumulative item response functions by response level $m = 0, 1, 2$ and $n = 500, 1000, 1500, 2000$.

| Item | $m = 0$ | | | | $m = 1$ | | | | $m = 2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 500 | 1000 | 1500 | 2000 | 500 | 1000 | 1500 | 2000 | 500 | 1000 | 1500 | 2000 |
| 1 | 0.004 | 0.003 | 0.003 | 0.002 | 0.031 | 0.022 | 0.019 | 0.015 | 0.031 | 0.023 | 0.018 | 0.016 |
| 2 | 0.000 | 0.000 | 0.000 | 0.000 | 0.015 | 0.010 | 0.009 | 0.008 | 0.029 | 0.021 | 0.018 | 0.015 |
| 3 | 0.012 | 0.008 | 0.007 | 0.006 | 0.038 | 0.027 | 0.021 | 0.019 | 0.017 | 0.013 | 0.010 | 0.010 |
| 4 | 0.002 | 0.001 | 0.001 | 0.001 | 0.026 | 0.018 | 0.015 | 0.013 | 0.023 | 0.015 | 0.012 | 0.011 |
| 5 | 0.001 | 0.001 | 0.001 | 0.001 | 0.022 | 0.019 | 0.016 | 0.015 | 0.033 | 0.026 | 0.023 | 0.021 |
| 6 | 0.002 | 0.002 | 0.001 | 0.001 | 0.025 | 0.023 | 0.021 | 0.020 | 0.028 | 0.022 | 0.021 | 0.018 |
| 7 | 0.016 | 0.010 | 0.008 | 0.008 | 0.025 | 0.016 | 0.014 | 0.012 | 0.014 | 0.010 | 0.009 | 0.008 |
| 8 | 0.006 | 0.004 | 0.004 | 0.003 | 0.025 | 0.016 | 0.014 | 0.013 | 0.020 | 0.013 | 0.011 | 0.011 |
| 9 | 0.011 | 0.006 | 0.006 | 0.005 | 0.023 | 0.015 | 0.013 | 0.012 | 0.016 | 0.012 | 0.010 | 0.010 |
| 10 | 0.007 | 0.005 | 0.004 | 0.003 | 0.026 | 0.018 | 0.015 | 0.013 | 0.017 | 0.013 | 0.012 | 0.010 |
| 11 | 0.014 | 0.009 | 0.008 | 0.007 | 0.022 | 0.016 | 0.013 | 0.011 | 0.016 | 0.011 | 0.010 | 0.009 |
| 12 | 0.013 | 0.009 | 0.009 | 0.007 | 0.022 | 0.015 | 0.013 | 0.011 | 0.014 | 0.011 | 0.009 | 0.008 |

TABLE 3.
Summary of Monte Carlo simulation study expected absolute deviation for the latent class structural probabilities $\pi$ by $n = 500, 1000, 1500, 2000$.

| $c$ | Class | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|
| 0 | 000 | 0.015 | 0.008 | 0.009 | 0.006 |
| 1 | 001 | 0.015 | 0.012 | 0.009 | 0.008 |
| 2 | 010 | 0.026 | 0.016 | 0.013 | 0.011 |
| 3 | 011 | 0.016 | 0.012 | 0.009 | 0.007 |
| 4 | 100 | 0.013 | 0.008 | 0.007 | 0.007 |
| 5 | 101 | 0.019 | 0.013 | 0.012 | 0.010 |
| 6 | 110 | 0.012 | 0.009 | 0.008 | 0.007 |
| 7 | 111 | 0.014 | 0.008 | 0.008 | 0.008 |

## 5. Application: Approaches to Learning and Self-Description (ALS) Questionnaire

We next report results from an application of the ordinal DM to the public-use ALS data file, which was collected through the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K; Tourangeau et al., 2015). The public-use ALS dataset included student ratings by parents and teachers on the approaches to learning scale. The dataset includes responses from teachers and parents for a total of 21,409 students of which 13,354 were complete records for the variables studied.

The ALS questionnaire includes twelve items with $M_j = 4$ response categories and scale anchor labels of "0 = never" and "3 = very often." Table 4 reports the item stems and category proportions for parent (i.e., items 1–6) and teacher (i.e., items 7–12) responses. Table 4 shows that proportionally fewer students received a rating of "never" by parents or teachers. The items also demonstrated variability in response proportions. For instance, over 80% of parents responded in the highest two categories for items two ("Show interest in a variety of things?"), five ("Eager to learn new things?"), and six ("Creative in work or in play?"). In contrast, Table 4 shows that teachers tended to report ratings of "1," "2," and "3" more uniformly than parents.

TABLE 4.
Approaches to learning and self-description item stems and category proportions for a random sample of $n = 13,354$ kindergarten students.

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 1. Keep working at something until he/she is finished? | 0.017 | 0.255 | 0.388 | 0.341 |
| 2. Show interest in a variety of things? | 0.003 | 0.099 | 0.421 | 0.478 |
| 3. Concentrate on a task and ignore distractions? | 0.044 | 0.452 | 0.348 | 0.157 |
| 4. Help with chores? | 0.018 | 0.313 | 0.377 | 0.292 |
| 5. Eager to learn new things? | 0.003 | 0.079 | 0.360 | 0.559 |
| 6. Creative in work or in play? | 0.005 | 0.156 | 0.405 | 0.433 |
| 7. Keeps belongings organized | 0.044 | 0.309 | 0.407 | 0.240 |
| 8. Shows eagerness to learn new things | 0.020 | 0.214 | 0.378 | 0.388 |
| 9. Works independently. | 0.029 | 0.275 | 0.370 | 0.326 |
| 10. Easily adapts to changes in routine | 0.028 | 0.259 | 0.469 | 0.245 |
| 11. Persists in completing tasks | 0.044 | 0.249 | 0.380 | 0.327 |
| 12. Pays attention well | 0.032 | 0.313 | 0.345 | 0.309 |

The anchor labels were 0 = "never" and 3 = "very often." Students' parents responded to items 1–6 and their teachers responded to items 7–12.

TABLE 5.
Approaches to learning and self-description element-wise means for $\mathbf{Q}$, $\mathbf{B}$, and $\boldsymbol{\pi}$ for a random sample of $n = 13,359$ kindergarten students from the early childhood longitudinal study (ECLS-K) class of 1998–1999.

| Item | $\overline{\mathbf{Q}}$ 1 | 2 | 3 | 1 000 | $\alpha_3$ 001 | $\alpha_2$ 010 | $\alpha_2\alpha_3$ 011 | $\alpha_1$ 100 | $\alpha_1\alpha_3$ 101 | $\alpha_1\alpha_2$ 110 | $\alpha_1\alpha_2\alpha_3$ 111 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 1.00 | 1.37 | 0.76 | 1.10 | −0.22 | 1.51 | 0.12 | 0.65 | −0.47 |
| 2 | 1.00 | 1.00 | 0.00 | 2.50 | 0.04 | 0.75 | −0.02 | 1.77 | −0.00 | 0.00 | 0.00 |
| 3 | 1.00 | 1.00 | 1.00 | 0.82 | 0.64 | 0.76 | −0.15 | 0.86 | 0.26 | 1.19 | −0.41 |
| 4 | 1.00 | 1.00 | 0.00 | 1.94 | 0.01 | 0.47 | 0.00 | 1.32 | −0.00 | 0.03 | 0.00 |
| 5 | 1.00 | 1.00 | 1.00 | 2.17 | 0.47 | 1.12 | −0.25 | 2.10 | −0.19 | −0.02 | 0.04 |
| 6 | 1.00 | 1.00 | 1.00 | 2.08 | 0.45 | 0.92 | −0.35 | 1.90 | −0.37 | 0.00 | 0.28 |
| 7 | 0.00 | 1.00 | 1.00 | 0.48 | 2.33 | 1.33 | 0.24 | 0.01 | 0.00 | 0.00 | 0.01 |
| 8 | 0.01 | 1.00 | 1.00 | 1.32 | 2.37 | 1.15 | 0.27 | 0.04 | 0.08 | −0.01 | −0.02 |
| 9 | 0.00 | 1.00 | 1.00 | 0.84 | 2.78 | 1.29 | 0.50 | 0.05 | 0.02 | −0.02 | −0.02 |
| 10 | 0.00 | 1.00 | 1.00 | 1.21 | 2.01 | 1.09 | 0.12 | 0.01 | 0.00 | −0.00 | 0.00 |
| 11 | 0.00 | 1.00 | 1.00 | 0.62 | 2.87 | 1.34 | 0.61 | 0.02 | 0.00 | 0.00 | −0.00 |
| 12 | 0.00 | 1.00 | 1.00 | 0.58 | 2.64 | 1.10 | 0.80 | 0.03 | −0.01 | 0.01 | −0.01 |
| $\boldsymbol{\pi}$ | | | | 0.07 | 0.12 | 0.15 | 0.18 | 0.07 | 0.19 | 0.08 | 0.14 |

A chain of length of 100,000 was run with a burnin of 20,000. Thresholds were fixed to $\tau_0 = -\infty$, $\tau_1 = 0$, $\tau_2 = 2$, $\tau_3 = 4$, and $\tau_4 = \infty$. $\overline{\mathbf{Q}}$ is the posterior mean of the sampled binary $\mathbf{Q}$ matrices.

### 5.1. Results

We estimated an exploratory ordinal DM solution using $K = 3$ to demonstrate the model and interpret parameter estimates. Note that we ran a single chain of length 100,000 with a burnin of 20,000 iterations. The algorithm required 3087 s to complete 100,000 iterations on a laptop with a 2.20 GHz processor. We implemented the model with the SSVS scale parameters equal to $c_1 = 1$ and $c_0 = 500$.

Table 5 reports ALS model parameter estimates for $\mathbf{Q}$, $\mathbf{B}$, and $\boldsymbol{\pi}$. Specifically, we constructed $\overline{\mathbf{Q}}$ with the posterior element-wise means (i.e., $\bar{q}_{jk} = \frac{1}{T} \sum_{t=1}^{T} q_{jk}^{(t)}$ where $t$ indexes values in the

Markov chain). The results in Table 5 provide evidence of a dense structure for $\mathbf{Q}$ given the absence of simple structure items. The estimated $\mathbf{Q}$ matrix suggests that the latent structure for items is associated with whether ratings were recorded by parents or teachers. In fact, the teacher responses about students approaches to learning (i.e., items seven through twelve) relate to attributes two and three, whereas there are some items where the parent ratings load onto all three attributes.

The values of the coefficients provide insight about how changes in attribute levels relate to observed responses. For instance, consider the coefficients for item twelve ("Pays attention well"). Item 12 relates to the main effects and interaction between attributes two and three. Additionally, the main effects for attributes two and three relate to responses on all but items two and four. There is evidence of several higher-order interaction effects. For instance, $\alpha_2\alpha_3$ relates to item nine and $\alpha_1\alpha_2$ relates to item three. There is also some evidence of disjunctive relationships as indicated by negative interaction effects. In fact, there is evidence of a negative three-way interactions for items one and three.

The IRFs provide another perspective for interpreting the relationship between the attributes and items. For instance, consider the impact of the first attribute on item one by comparing the IRFs for members of classes "000" and "100." Students classified as possessing attribute profile "000" received ratings of (0, 1, 2, 3) from theirs parents with probabilities of (0.09, 0.65, 0.26, 0.00). In contrast, the probabilities for the students in the "100" class had response probabilities of (0.00, 0.19, 0.68, 0.13). A change of $\alpha_1$ from zero to one increases the chance students receive parent ratings of a "2."

Figure 1 presents counter-clock-plots to illustrate the category response probabilities by latent class and item. That is, the "clock" for a given $\boldsymbol{\alpha}$ and item presents the probabilities of observed responses of 0, 1, 2, and 3 in four quadrants in a counter-clockwise order. For instance, Fig. 1 shows that the probability members of classes "011" and "111" of reporting a "3" is nearly one for items eight, nine, eleven, and twelve. Figure 1 also illustrates how increases in latent levels relate to changes in response probabilities. For instance, the probability of reporting a "2" or "3" on items seven through twelve increases as $\alpha_3$ changes from 0 to 1.

The last row of Table 5 reports the posterior mean of the latent class structural parameter, $\boldsymbol{\pi}$. The values of the estimated $\boldsymbol{\pi}$ suggest that the largest three classes were "101," "011," and "010" with class probabilities of $\hat{\pi}_5 = 0.19$, $\hat{\pi}_3 = 0.18$, and $\hat{\pi}_2 = 0.15$, respectively. In contrast, the three smallest classes were "000," "100," and "110" with probabilities equal to $\hat{\pi}_0 = 0.07$, $\hat{\pi}_4 = 0.07$, and $\hat{\pi}_6 = 0.08$.

### 5.2. Model Fit

The previous subsection reported results for a model with two attributes to demonstrate the model in an application. This subsection discusses model fit to understand how well the estimated model describes the observed data.

We report posterior predictive probabilities (e.g., see Sinharay, Johnson, & Stern, 2006) to evaluate the fit of the model with $K = 3$. In particular, we assess model fit by assessing how well the model reproduced students' total scores. That is, we used the samples from the posterior to simulate ordinal responses from the model and computed a distribution for each student's total score to calculate the proportion of times total scores generated in the posterior distribution exceeded the observed values. Figure 2 plots the observed total scores against each student's posterior predictive probability (PPP). Figure 2 shows that 50% of students' PPPs equal values between 0.249 and 0.610, which provides some support that the model describes students' total scores. Furthermore, additional evidence of model fit is illustrated by the fact that the PPPs for most students fell between cutoffs of 0.025 and 0.975. The horizontal lines in Fig. 2 separate observations that were less accurately predicted by the model (i.e., PPPs below 0.025 or above
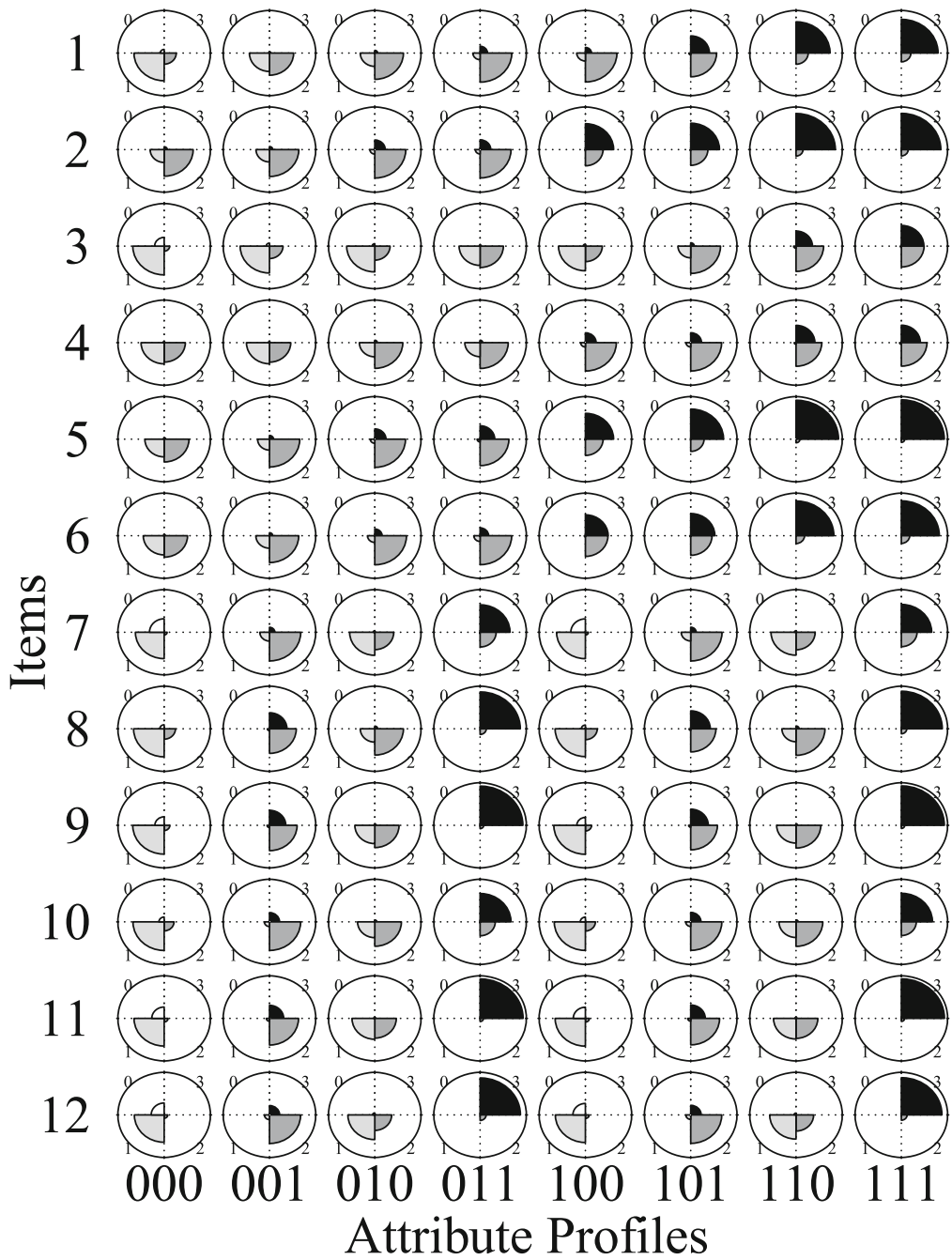
FIGURE 1.
Counter-clock-plots depicting the estimated item response functions by latent class $\boldsymbol{\alpha}$ (x-axis) and item (y-axis) for the ALS data with $M_j = 4$ for $j = 1, \ldots, 12$ and $K = 3$.
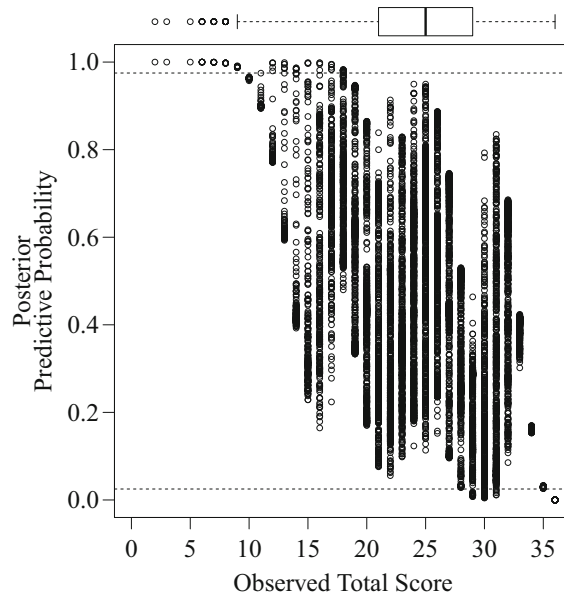
FIGURE 2.

Plot of observed total scores and posterior predictive probabilities (PPPs). *Note* Boxplots are marginal distributions. The dashed, horizontal lines indicate posterior probabilities of 0.025 and 0.975. In the plot, 2.31% of students had total score PPPs less than 0.025 or greater than 0.975.

0.975). In fact, 2.31% of students had extreme PPPs outside of the 0.025 and 0.975 range and could be considered less accurately predicted by the model.

## 6. Discussion

Ordinal DMs are necessary for providing a fine-grained classification on substantively relevant attributes. Ordinal DMs are applicable to address research questions in numerous settings with examples including partial credit responses in education and ratings in survey research. We next discuss the implications of the study, review limitations, offer future research directions, and provide concluding remarks.

We presented a new exploratory diagnostic model for ordinal data that improves upon existing research. Our model used a cumulative probit link along with Bayesian variable selection techniques to infer the relationship between a collection of binary attributes and observed ordinal responses. A potential advantage of the proposed method is that *a priori* knowledge about the latent structure is not needed and our model provides researchers with a framework for inferring the latent structure to inform substantive theory development. It is worth mentioning that although we considered an exploratory approach in this paper the developed framework could be implemented in a confirmatory fashion. That is, the methods could be used in a confirmatory setting by fixing the elements of **Q** to indicate which item parameters are active versus inactive. Furthermore, researchers with partial knowledge about the underlying structure could also fix some $q_{jk}$ and estimate others.

The utility of the proposed framework will likely be judged by the insights the methodology offers applied researchers. Accordingly, our application to the ALS items serves as an example of the type of novel inferences that are available with ordinal DMs. In particular, we uncovered evidence that a three-attribute solution with eight classes describes teacher and parent responses

to the twelve public-use ALS items. The results offered evidence that the latent structure was associated with whether the ratings were from teachers or parents. Additionally, we found evidence that the relationship between observed responses and latent attributes was characterized by main effects and higher-order interactions. We also found that some of the parent item responses loaded on all three attributes. The uncovered structure may suggest that there is a common attribute underlying the approaches to learning and self-description items that both teachers and parents consider in responses about their students. Teacher ratings related to the common attribute and as well as attribute three. One inference could be that the first attribute (which related only to parent ratings) captured learning behaviors that are observed in informal settings outside of schools, whereas the third attribute characterized student feature found in formal academic settings. Future research should apply the developed ordinal DM to additional ALS items to assess the extent to which the uncovered attributes more generally capture the latent structure.

Another contribution relates to the establishment of weaker sufficient conditions for model identifiability. We followed Fang et al.'s (2019) recommendation to relax their sufficient conditions for polytomous response DMs with polytomous attributes. Specifically, Theorem 2 shows that a sufficient condition for identifying ordinal DM parameters with binary attributes is that the **Q** matrix must include simple structure for two items and the latent classes must be distinguished by a least one of the remaining items. It is important to note that we did not explicitly enforce the identifiability conditions when analyzing the ALS items. The algorithm could be restricted to stochastically searching the space of identified **Q** matrices (e.g., see Y. Chen, Culpepper, Chen, & Douglas, 2018). We used the ALS model parameters as to generate data in the Monte Carlo simulation study. The Monte Carlo results provided evidence that, although the estimated **Q** was not strictly identified, the algorithm was successful in recovering model parameters. One implication is that future research may be able to establish even weaker identifiability conditions for ordinal DMs with binary attributes than required by Theorem 2.

There are several additional issues for researchers to consider when applying ordinal DMs. First, researchers may need to reverse-code variables when applying the monotonicity constraints. That is, the monotonicity constraints assume that higher latent attribute levels relate to higher observed responses. A negatively-worded item would need to be reverse-coded to ensure a clear interpretation of the attributes. Alternatively, the identifiability conditions require restrictions on **Q** and the distinctness of some probabilities, so estimation with reverse-coded items could proceed without monotonicity restrictions.

Second, we considered an unstructured model for the latent class probabilities $\boldsymbol{\pi}$. There may be instances where the latent class structure can be approximated with a higher-order factor model (Culpepper & Chen, 2018; de la Torre & Douglas, 2004; Henson et al., 2009) or an underlying multivariate normal distribution with a vector of thresholds and a polychoric correlation matrix (Y. Chen & Culpepper, 2018; Henson et al., 2009; Templin, Henson, Templin, & Roussos, 2008). Future research could establish guidelines for selecting competing structures for $\boldsymbol{\pi}$.

Third, in our application of the ordinal DM to the ALS data we analyzed the subset of complete cases, which implies that we implicitly assumed incomplete data were ignorable and missing at random. Future research should investigate the plausibility of the ignorability assumption with the restricted-use data and consider modeling the missing data mechanism if the assumption is untenable.

Lastly, the Monte Carlo simulation study was limited to one set of population parameters derived from applied data. Additional applications of the ordinal DM are needed to identify a distribution of parameter values that can be expected in practice. Subsequent Monte Carlo simulation studies are needed to evaluate parameter recovery for an expanded set of population parameter values.

In conclusion, advancing ordinal DMs is critical for establishing tools that yield fine-grained classification of respondents into substantively meaningful latent clusters. The developed method-

ology contributes to research on ordinal DMs. We provided an exploratory methodology that broadens the applicability of ordinal DMs to social science research and allows researchers to refine substantive theory about the latent structure underlying ordinal response data.

## Acknowledgments

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Appendix: Gibbs Sampling Algorithm and Full Conditional Distributions

This section discusses the full conditional distributions used to approximate the posterior distribution of the ordinal diagnostic model parameters with Gibbs sampling. For iteration $t = 1, \ldots, T$ we sample:

1. For $i = 1, \ldots, n$,

   (a) $\boldsymbol{\alpha}_i^{(t)}$ from the multinomial full conditional distribution $\boldsymbol{\alpha}_i^{(t)} | \boldsymbol{Y}_i, \mathbf{B}^{(t-1)}, \boldsymbol{\alpha}_1^{(t)}, \ldots,$ $\boldsymbol{\alpha}_{i-1}^{(t)}, \boldsymbol{\alpha}_{i+1}^{(t-1)}, \ldots, \boldsymbol{\alpha}_n^{(t-1)}$ where the conditional probability $\boldsymbol{\alpha}_i^{(t)}$ is classified as profile $c$ is,

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

   where $\theta_{jc,y_{ij}}^{(t-1)} = \Phi\left(\tau_{jc,y_{ij}+1} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j^{(t-1)}\right) - \Phi\left(\tau_{jc,y_{ij}} - \boldsymbol{a}_c^\top \boldsymbol{\beta}_j^{(t-1)}\right)$. 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 $\pi_c$ (e.g., see Equation 7 of Culpepper, 2019) is replaced with $n_{ci} + n_{c0}$ where $n_{ci}$ is the number of respondents other than $i$ that are classified in class $c$ (e.g., see Jain & Neal, 2004) and $n_{c0}$ is the prior Dirichlet parameter (note $n_{c0} = 1$ for a uniform prior).

   (b) For $j = 1, \ldots, J$ update the latent augmented data from the full conditional distribution

$$
Y_{ij}^{*(t)} | Y_{ij}, \boldsymbol{\alpha}_i^{(t)}, \boldsymbol{\beta}_j^{(t-1)} \sim \mathcal{N}(\boldsymbol{a}_i^{(t)\top} \boldsymbol{\beta}_j^{(t-1)}, 1) \mathcal{I}(\tau_{jc,y_{ij}} < Y_{ij}^{*(t)} < \tau_{jc,y_{ij}+1})
\tag{A2}
$$

where $\tau_{jc,y_{ij}}$ and $\tau_{jc,y_{ij}+1}$ are lower and upper thresholds for the observed value of $Y_{ij}$ for class $c$ and item $j$. Recall we follow previously discussed strategies and fix the thresholds as $\boldsymbol{\tau}_j = (0, 2, \ldots, 2(M_j - 2))^\top$.

2. Update the latent class probabilities (i.e., the mixing weights) as $\boldsymbol{\pi}^{(t)}|\mathbf{A}^{(t)}$ from the Dirichlet full conditional distribution (e.g., see Culpepper, 2015).
3. For $j = 1, \ldots, J$,

   (a) For $k = 1, \ldots, K$ sample $q_{jk}^{(t)}$ from the Bernoulli full conditional distribution $q_{jk}^{(t)}|\boldsymbol{\beta}_j^{(t-1)}, q_{j1}^{(t)}, \ldots, q_{j,k-1}^{(t)}, q_{j,k+1}^{(t-1)}, \ldots, q_{jK}^{(t-1)}, \omega^{(t-1)}$ (e.g., see Culpepper, 2019).

   (b) For $p = 1, \ldots, P$ sample $\beta_{jp}^{(t)}$ from the truncated normal full conditional distribution $\beta_{jp}^{(t)}|Y_{1j}^{*(t)}, \ldots, Y_{nj}^{*(t)}, \mathbf{A}^{(t)}, \beta_{j1}^{(t)}, \ldots, \beta_{j,p-1}^{(t)}, \beta_{j,p+1}^{(t-1)}, \ldots, \beta_{j,P+1}^{(t-1)}, \boldsymbol{q}_j^{(t)}$ (e.g., see Culpepper, 2019).

4. Sample $\omega^{(t)}$ from the Beta full conditional distribution $\omega^{(t)}|\mathbf{Q}^{(t)}$ (e.g., see Culpepper, 2019).

## References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *17*(3), 251–269.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

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

Bao, J., & Hanson, T. E. (2015). Bayesian nonparametric multivariate ordinal regression. *Canadian Journal of Statistics*, *43*(3), 337–357.

Béguin, A. A., & Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*(4), 541–561.

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

Chen, J., & de la Torre, J. (2018). Introducing the general polytomous diagnosis modeling framework. *Frontiers in Psychology*, *9*, 1–9.

Chen, Y., & Culpepper, S. A. (2018). A multivariate probit model for learning trajectories with application to classroom assessment. In *Paper presentation at the international meeting of the psychometric society, New York*.

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

Chen, Y., Culpepper, S. A., & Liang, F. (2018). Beyond the Q-matrix: A general approach to cognitive diagnostic models. In *Paper presentation at the international meeting of the psychometric society, New York*.

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

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, *6*(2), 101–111.

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. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, *81*(4), 1142–1163.

Culpepper, S. A. (2019). Estimating the cognitive diagnosis Q matrix with expert knowledge: Application to the fraction-subtraction dataset. *Psychometrika*, *84*, 333–357. 10.1007/s11336-018-9643-8.

Culpepper, S. A., & Chen, Y. (2018). Development and application of an exploratory reduced reparameterized unified model. *Journal of Educational and Behavioral Statistics*, *44*, 3–24.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*(1), 8–26.

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.

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*(4), 595–624.

DeYoreo, M., & Kottas, A. (2018). Bayesian nonparametric modeling for multivariate ordinal regression. *Journal of Computational and Graphical Statistics*, *27*(1), 71–84.

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

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

Green, B. F. (1951). A general solution for the latent class model of latent structure analysis. *Psychometrika*, *16*(2), 151–166.

Haberman, S. J., von Davier, M., & Lee, Y.-H. (2008). Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions. *ETS Research Report Series*, *2008*(2), 1–25.

Henson, R. A., & Templin, J. (2007). Importance of Q-matrix construction and its effects cognitive diagnosis model results. In *Annual meeting of the national council on measurement in education, Chicago, IL*.

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.

Hojtink, H., & Molenaar, I. W. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika*, *62*(2), 171–189.

Jain, S., & Neal, R. M. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, *13*(1), 158–182.

Karelitz, T. M. (2004). *Ordered category attribute coding framework for cognitive assessments*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement*, *77*(3), 369–388.

Kottas, A., Müller, P., & Quintana, F. (2005). Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphical Statistics*, *14*(3), 610–625.

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

Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, *18*(2), 95–138.

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

Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli*, *19*(5A), 1790–1817.

Liu, R., & Jiang, Z. (2018). Diagnostic classification models for ordinal item responses. *Frontiers in Psychology*, *9*, 1–12.

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

Ma, W., & de la Torre, J. (2019). An empirical Q-matrix validation method for the sequential generalized DINA model. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12156.

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

McDonald, R. P. (1962). A note on the derivation of the general latent class model. *Psychometrika*, *27*(2), 203–206.

Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of guttman scaling. *Psychometrika*, *35*(1), 73–78.

Rost, J. (1988). Rating scale analysis with latent class models. *Psychometrika*, *53*(3), 327–348.

Rupp, A. A., & Templin, J. L. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, *68*(1), 78–96.

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.

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

Templin, J. L. (2004). Generalized linear mixed proficiency models. Unpublished doctoral dissertation, 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.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A., Hagedorn, M., Daly, P., & Najarian, M. (2015). *Early childhood longitudinal study, kindergarten class of 2010–2011 (ECLS-K:2011), user's manual for the ECLS-K:2011 kindergarten data file and electronic codebook, public version (NCES 2015-074)*. Early childhood longitudinal study, kindergarten class of 2010–2011 (ECLS-K:2011), user's manual for the ECLS-K:2011 kindergarten data file and electronic codebook, public version (NCES 2015-074). U.S. Department of Education. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010070. Accessed 19 Apr 2018.

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. (2017). 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.

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.

Ye, S., Fellouris, G., Culpepper, S. A., & Douglas, J. (2016). Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology*, *69*(2), 139–158.