Chapter 21
A Sparse Latent Class Model for Polytomous Attributes in Cognitive Diagnostic Assessments

Siqi He, Steven Andrew Culpepper, and Jeff Douglas

Abstract Diagnostic models (DMs) have been widely applied to binary response data. However, in the field of educational and psychological measurement, a wealth of ordinal data are collected to measure latent structures where the traditional binary attributes may not adequately describe the complex response patterns. Considering that, we propose an extension of the sparse latent class model (SLCM) with ordinal attributes, with the purpose of fully exploring the relationships between attributes and response patterns. Furthermore, we discuss the strict and generic identifiability conditions for the ordinal SLCMs. We apply the model to the Short Dark Triad data and revisit the underlying personality structure. Evidence supports that SLCMs have better model fit to this real data than the exploratory factor models. We also confirm the efficiency of a Gibbs algorithm in recovering the empirical item parameters via a Monte Carlo simulation study. This study discusses a way of constructing DMs with ordinal attributes which helps broaden its applicability to personality assessment.

21.1 Introduction

Cognitive diagnostic assessments, aimed at providing fine-grained information about respondents' mastery of latent attributes, have gained increasing research attention in recent decades. The considerable expansion of cognitive diagnostic models (CDMs) has heightened the need for an inclusive and comprehensive modeling approach, where the sparse latent class models (SLCM; Chen et al. 2020) have served this purpose to fit most existing CDMs in an exploratory fashion.

The SLCM was originally proposed with binary attributes, including the deterministic input, noisy, "and" gate model (DINA; De La Torre 2009; Junker & Sijtsma 2001); the deterministic input, noisy, "or" gate model (DINO; J. L. Templin & Henson, 2006); the reduced non-compensatory reparameterized unified model

S. He · S. A. Culpepper · J. Douglas (✉)
University of Illinois at Urbana-Champaign, Champaign, IL, USA
e-mail: siqihe2@illinois.edu; sculpepp@illinois.edu; jeffdoug@illinois.edu

(NC-RUM; DiBello et al. 1995; Templin et al. 2010) and the compensatory-RUM model (C-RUM; Hagenaars 1993; Maris 1999); and the family of general diagnostic models (De La Torre 2011; Henson et al. 2009; von Davier 2005). With binary representations of latent attributes, examinees' attribute patterns are composed of either "mastery" or "non-mastery." However, binary attributes are sometimes not accurate enough to describe the level of mastery as respondents can theoretically possess a specific attribute to different extents. Bolt and Kim (2018) provided empirical evidence that attributes derived from the fraction subtraction test (Tatsuoka 1987) are oversimplified if defined as binary. In addition, previous research also found that allowing attributes to have multiple levels improved the model-data fit (Haberman et al. 2008; von Davier 2018). These justifications support that assuming multiple levels of attributes are sometimes more desirable than binary levels. Therefore, developing CDMs with polytomous attributes would maximize the understanding of response patterns in binary or even polytomous data.

Many existing CDMs have been developed to measure polytomous responses, such as the Ordered Category Attribute Coding DINA model (OCAC-DINA; Karelitz 2004), the reduced reparameterized unified models (R-RUM; Templin 2004), the log-linear cognitive diagnostic model (LCDM ; Templin & Bradshaw(2013), the general diagnostic models (GDM; von Davier 2005), and the pG-DINA model (Chen & Culpepper 2020). Based on whether the interactions between attributes are considered, these models can be further specified as the main-effect cognitive diagnostic models (i.e., the OCAC-DINA, R-RUM, LCDM, and GDM) and the all-effect cognitive diagnostic models (i.e., the pG-DINA). The former involves only the main effects of the required attributes, whereas the latter involves both the main effects and interaction effects. With the between-attribute interaction effects being considered, we are able to discover all types of attribute relationships and how they can affect the observed responses. A fully saturated model is the most general parameterization of the joint attribute distribution, where all the main effects and interaction effects are taken into consideration. However, the challenge is, when the attributes or attribute levels increase, the item parameters increase exponentially. This risk of the overparameterization makes its application restricted to the confirmatory settings. The fact is for confirmatory models, accurate scoring requires the correct specification of item-attribute relationship. Otherwise, misspecification of the item structure could result in inaccurate classification. To this end, an exploratory model has been instrumental in promoting our understanding of the item-attribute structures when they are not prespecified.

With the intention of constructing an exploratory SLCM model, the key concern is to determine whether polytomous attributes are necessary and how many intermediate levels are appropriate. With data-defined polytomous attributes, the levels of attributes and their interpretations can be derived from the model-fitting process. With expert-defined polytomous attributes, the justifications of attribute levels can be provided by experts in the related areas, especially in the area of educational testing and psychopathology. In this study, we take on the first approach to explore the attribute dimensions and levels. Once an adequate model size is determined, we can move to the formal estimation of model parameters.

This chapter is organized in the following sections. The *2. A Sparse Latent Class Model with Polytomous Attributes* section introduces the model parameterization and discusses the identifiability conditions for the model. The *3. Gibbs Sampling* section provides a Gibbs algorithm for Markov chain Monte Carlo estimation with the potential to enforce the identifiability and monotonicity constraints. The *4. An Empirical Application* section illustrates how well the model fits the Short Dark Triad (SD3) data compared to an exploratory item factor model and performs a Monte Carlo simulation study to assess the statistical properties and feasibility of our model. The *5. Discussion* section provides a summary of this study and some potential directions for future research.

21.2 A Sparse Latent Class Model with Polytomous Attributes

21.2.1 Model Configurations

21.2.1.1 Unstructured Mixture Model

Consider a scale that consists of J items and K underlying attributes. We denote the vector of response probabilities for a latent class c on item j as $\boldsymbol{\theta}_{cj} = (\theta_{cj0}, \ldots, \theta_{cj,P-1})'$, where the element θ_{cjp} represents the probability of observing the response category p on item j by the latent class c. Note that the scale can be either dichotomous ($P = 2$) or polytomous ($P > 2$). Given the vector of response probability $\boldsymbol{\theta}_{cj}$, the probability of observing an ordinal response y_j is written as

$$P\left(y_j \mid \boldsymbol{\theta}_{cj}, \boldsymbol{\eta}'\boldsymbol{v} = c\right) = \sum_{p=0}^{P-1} \theta_{cjp}\mathcal{I}\left(y_j = p\right), \tag{21.1}$$

where $y_j \in \{0, \cdots, P-1\}$ and $\mathcal{I}$ is an indicator function. To describe the observed response patterns $\prod_{j=1}^{J} P_j$, we introduce a collection of K ordinal attributes with M levels. In this setting, a total of M^K latent classes can be created. Let the latent class index be $c \in \{0, \ldots, M^K - 1\}$; each latent class has a K-vector of latent ordinal attributes $\boldsymbol{\eta} \in \{0, \ldots, M-1\}^K$ that can be mapped to an integer index c via bijection $\boldsymbol{\eta}'\boldsymbol{v} = c$ with $\boldsymbol{v} = (M^{K-1}, M^{K-2}, \ldots, 1)'$. Next, we assume the membership in class c follows a multinomial distribution with structural parameters $\pi_c \in \{\pi_0, \ldots, \pi_{M^K-1}\}$ where π_c denotes the probability of membership in latent class c and $\sum_{c=0}^{M^K-1} \pi_c = 1$. By integrating out the latent class variable

c, we can write the likelihood of observing a random vector $\boldsymbol{y} = (y_1, \ldots, y_J)$ as

$$p(\boldsymbol{y} \mid \boldsymbol{\Theta}, \boldsymbol{\pi}) = \sum_{c=0}^{M^K - 1} \pi_c \prod_{j=1}^{J} P\left(y_j \mid \boldsymbol{\theta}_{cj}, \boldsymbol{\eta}'\boldsymbol{v} = c\right), \tag{21.2}$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{J \times M^K \times P}$ denotes the response probability array and $\boldsymbol{\pi}$ denotes the structural parameter vector. In a sample of N respondents, we denote the ordinal responses for respondent i as $\boldsymbol{y}_i$, $i = 1, \ldots, N$. The likelihood of observing this sample is

$$p(Y \mid \boldsymbol{\Theta}, \boldsymbol{\pi}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i \mid \boldsymbol{\Theta}, \boldsymbol{\pi}), \tag{21.3}$$

where Y denoted the $N \times J$ response matrix.

In addition, a cumulative link function $\Psi(\cdot)$ is proposed to define the ordinal responses (Culpepper 2019). Specifically, we compute the probability of response category p by taking the difference in two adjacent cumulative probabilities. The response probability for latent class c on item j is written as

$$\theta_{cjp} = \Psi\left(\tau_{j,p+1} - \mu_{cj}\right) - \Psi\left(\tau_{j,p} - \mu_{cj}\right) \tag{21.4}$$

$$\Psi\left(\tau_{j,p} - \mu_{cj}\right) = P\left(y_j \leq p \mid \mu_{cj}, \tau_{j,p}\right), \tag{21.5}$$

where $\boldsymbol{\tau} \in \{\tau_{j,0}, \ldots, \tau_{j,P}\}$. We define $\tau_{j,0} = -\infty$, $\tau_{j,P} = \infty$ which result in $\Psi\left(\tau_{j,0} - \mu_{cj}\right) = 0$, $\Psi\left(\tau_{j,P} - \mu_{cj}\right) = 1$. Here, μ_{cj} is the latent class mean parameter discussed in the next section. $\Psi(\tau_{j,p} - \mu_{cj})$ denotes the cumulative probability of a response at level p or less.

The assumption of local independence implies the response distribution Y given B and $\boldsymbol{\pi}$ can be presented as

$$p(Y \mid B, \boldsymbol{\pi}) = \prod_{j=1}^{J} p\left(Y_j \mid \boldsymbol{\beta}_j, \boldsymbol{\pi}\right) = \prod_{i=1}^{N} \sum_{c=0}^{M^K - 1} \pi_c \prod_{j=1}^{J} p(y_{ij} \mid \boldsymbol{\beta}_j, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c), \tag{21.6}$$

where y_{ij} refers to individual i's response on item j and $\boldsymbol{\alpha}_i$ denotes the attribute profile vector for individual i.

21.2.1.2 Structured Mixture Model

A challenge with unstructured mixture model is that, as K or M increases, the number of parameters per item M^K grows exponentially. To reduce the number

of parameters, we impose a specific structure on μ_{cj} by representing μ_{cj} as $\mu_{cj} :=$ $\boldsymbol{\alpha}_c' \boldsymbol{\beta}_j$ where $\boldsymbol{\alpha}_c$ is attribute profile design vector and $\boldsymbol{\beta}_j$ is an item parameter vector of the same length as $\boldsymbol{\alpha}_c$. A saturated model is the most general exploratory model which includes all the main- and higher-order interaction terms of predictors. Within the saturated model framework, K attributes with M levels lead to a total number of M^K predictors. We let $\boldsymbol{A} = (\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{M^K-1})$ be a $M^K \times M^K$ design matrix which comprises the attribute profile vector $\boldsymbol{\alpha}$ for each latent class. Moreover, we let $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)$ be a $J \times M^K$ item coefficient matrix. In this way, since we assume a sparse pattern on $\boldsymbol{B}$ explained later, the number of effective parameters is greatly reduced.

We code the attribute profile $\boldsymbol{\eta}$ as the design vector $\boldsymbol{\alpha}$ using a cumulative coding strategy. Specifically, for attribute k, we define

$$\boldsymbol{a}_k = (1, \mathcal{I}(\eta_k \geq 1), \ldots, \mathcal{I}(\eta_k \geq M-1))', \tag{21.7}$$

so that $\boldsymbol{a}_k$ incorporates an intercept for the first element and main effects for the exceeding different attribute levels. Therefore, the attribute design vector for an arbitrary class can be written as

$$\boldsymbol{\alpha} = \left(\boldsymbol{a}_1' \otimes \cdots \otimes \boldsymbol{a}_K'\right)', \tag{21.8}$$

where $\otimes$ is the Kronecker product sign which frames all possible cross-level interactions between K attributes. Below we illustrate how the coding system works with a specific example.

Table 21.1 displays the attribute profile matrix $\boldsymbol{A}$ for a saturated SLCM with $K = 2$ attributes and $M = 4$ levels per attribute, where the first column prints the latent class integer c and the second column prints the latent class label in the way that each digit represents to which level latent classes master the attributes. For instance, latent class "12" in the column name implies the possession of the first attribute to the first level, and the remaining columns in the table refer to the predictors which compose the attribute profile vector $\boldsymbol{\alpha}$ as Eq. 21.8 describes. The design vector $\boldsymbol{\alpha}$ contains intercept component "00"; main-effect components "01," "02," "03," "10," "20," and "30"; and two-way interaction components "11," "12," "13," "21," "22," "23," "31," "32," and "33." For instance, label "11" in the row name corresponds to the cross-attribute effect between the first level of attribute 1 and the first level of attribute 2. For latent class "12" ($\eta_1 = 1$, $\eta_2 = 2$), component "11" ($a_1 = 1$, $a_2 = 1$) is active given the coding rule $\eta_1 \geq 1$ and $\eta_2 \geq 1$. Specifically, for latent class "12" (i.e., $\eta_1 = 1$, $\eta_2 = 2$), the active components refer to "00," "01," "02," "10," "11," and "12" columns.

21.2.1.3 Model Identifiability

Model identifiability issues have received considerable attention in the study of CDMs. Traditionally, parameter constraints derived for model identifiability

Table 21.1 Example design matrix A of latent classes by attribute predictors for $K = 2$, $M = 4$

		α															
c	η	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
0	00	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	01	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	02	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	03	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	10	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	11	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
6	12	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
7	13	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
8	20	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
9	21	1	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0
10	22	1	1	1	0	1	1	1	0	1	1	1	0	0	0	0	0
11	23	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
12	30	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
13	31	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
14	32	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0
15	33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

are often imposed on a design Q-matrix. The Q-matrix specifies item-attribute relationships by showing which attributes each item measures but not showing how attributes interact to affect response probabilities. Conversely, in SLCM, a sparsity matrix $\boldsymbol{\Delta}_{J \times M^K}$ substitutes the role of the traditional Q-matrix but takes the inter-attribute relationships into consideration. The sparsity matrix $\boldsymbol{\Delta}_{J \times M^K}$ depicts the underlying pattern of the item parameter matrix $\boldsymbol{B}$, where J denotes the total amount of items and M^K denotes the total amount of predictors. An element $\delta = 1$ suggests its corresponding β is active, whereas an element $\delta = 0$ suggests its corresponding β is 0.

Definition 2.1 presents a classic notion of likelihood identifiability where a different set of parameter values leads to different values of likelihood. To this end, a model must be identifiable to elicit consistent parameter estimates. Based on the work established by Culpepper (2019) for the SLCM with binary attributes and ordinal responses, and the work by Chen et al. (2020) with binary attributes and dichotomous responses, we propose identifiability conditions for the SLCM with ordinal responses and ordinal attributes.

Definition 2.1 A parameter set $\Omega(\boldsymbol{\pi}, \boldsymbol{B})$ is identifiable, if there are two sets of parameters $(\boldsymbol{\pi}, \boldsymbol{B})$ and $(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}})$ such that $\mathbb{P}(\boldsymbol{\pi}, \boldsymbol{B}) = \mathbb{P}(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{B}})$ implies $\boldsymbol{\pi} = \bar{\boldsymbol{\pi}}$, $\boldsymbol{B} = \bar{\boldsymbol{B}}$.

We define a parameter space $\Omega(\boldsymbol{\pi}, \boldsymbol{B})$ for the latent class proportion parameter $\boldsymbol{\pi}$ and the item coefficient parameter $\boldsymbol{B}$ as

$$\Omega(\boldsymbol{\pi}, \boldsymbol{B}) = \{(\boldsymbol{\pi}, \boldsymbol{B}) : \boldsymbol{\pi} \in \Omega_1(\boldsymbol{\pi}), \boldsymbol{B} \in \Omega_2(\boldsymbol{B})\}, \tag{21.9}$$

where $\Omega_1(\boldsymbol{\pi}) = \{\boldsymbol{\pi} \in \mathbb{R}^{M^K} : \sum_{c=0}^{M^K-1} \pi_c = 1, \pi_c > 0\}$ and $\Omega_2(\boldsymbol{B}) = \{\boldsymbol{B} \in \mathbb{R}^{J \times M^K}\}$.

Conditioned on the attribute profile vector $\boldsymbol{\alpha}$, the joint distribution of $\boldsymbol{Y}$ is a product of multinomial distributions represented by a P^J vector:

$$\mathbb{P}_{\boldsymbol{\alpha}}(\boldsymbol{B}) = \bigotimes_{j=1}^{J} \boldsymbol{\theta}_{cj}, \tag{21.10}$$

where $\bigotimes$ refers to the Kronecker product and $\boldsymbol{\theta}_{cj} = (\theta_{cj0}, \ldots, \theta_{cj,P-1})'$. Further, the marginal distribution of $\boldsymbol{Y}$ over the proportion parameter $\boldsymbol{\pi}$ is

$$\mathbb{P}(\boldsymbol{\pi}, \boldsymbol{B}) = \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} \mathbb{P}_{\boldsymbol{\alpha}}(\boldsymbol{B}). \tag{21.11}$$

21.2.1.4 Monotonicity Constraints

Imposing monotonicity constraints ensures that mastering any irrelevant skills to an item will not increase the endorsing probability. Xu (2017) and Xu and Shang (2018) proposed the monotonicity constraints as follows:

$$\min_{c \in \mathbb{S}_0} \mu_{cj} \geq \mu_{c_0 j}, \tag{21.12}$$

$$\max_{c \in \mathbb{S}_0} \mu_{cj} < \min_{c \in \mathbb{S}_1} \mu_{cj} = \max_{c \in \mathbb{S}_1} \mu_{cj}, \tag{21.13}$$

where c_0 represents the latent class that does not own any relevant attribute and $\mu_{c_0 j}$ denotes its latent class mean parameters. $\mathbb{S}_0$ denotes a set of latent classes that own at least one but not all relevant attributes, and $\mathbb{S}_1$ denotes a set of latent classes that own all the relevant attributes. Given $\mu_{cj} = \boldsymbol{\beta}'_j \boldsymbol{\alpha}_c$, we can derive a lower bound condition L_{cj} for each item parameters β_{cj} that if β_{cj} is lower bounded, the constraints above are satisfied. The derivation details can be found in Chen et al. (2020).

21.2.1.5 Strict Identifiability

Theorem 2.2 *The parameter space* $\Omega(\boldsymbol{\pi}, \boldsymbol{B})$ *is strictly identifiable if conditions (S1) and (S2) are satisfied.*

(S1) When Δ matrix takes the form of $\Delta = \begin{pmatrix} D_1 \\ D_2 \\ D^* \end{pmatrix}$ after row swapping where D_1,

$$D_2 \in \mathbb{D}_s, \mathcal{D}_s = \left\{ D \in \{0, 1\}^{K \times M^K} : D = \begin{bmatrix} 0 & \mathbf{1}'_{M-1} & 0 & \ldots & 0 & \ldots 0 \\ 0 & 0 & \mathbf{1}'_{M-1} & \ldots & 0 & \ldots 0 \\ \vdots & 0 & 0 & \ddots & 0 & \ldots \vdots \\ 0 & 0 & 0 & \ldots & \mathbf{1}'_{M-1} & \ldots 0 \end{bmatrix} \right\}.$$

Note: $\mathbf{1}_{M-1}$ is a vector of 1 of length $M - 1$ which represents the activeness of an attribute on all $M - 1$ levels.

(S2) In D^*, for any attribute $k = 1, 2, \ldots, K$, there exists an item $j > 2K$, such that all the main-effect components regarding this attribute are active $(\delta_{j,k0}, \delta_{j,k1}, \ldots, \delta_{j,k(M-1)}) = 1$.

The proof details are provided in Appendix "Strict Identifiability Proof".

21.2.1.6 Generic Identifiability

In this section, we propose the generic identifiability condition in (G1) and (G2) in Theorem 2.4. The generic condition is less stringent than the strict conditions (S1) and (S2) given in 2.2. Generic identifiability allows part of the model parameters to be non-identifiable such that these exceptional values are of measure zero in the parameter space.

Definition 2.3 A parameter set $\Omega_\Delta(\boldsymbol{\pi}, \boldsymbol{B})$ is generically identifiable if the Lebesgue measure of the unidentifiable space C_Δ with respect to $\Omega_\Delta(\boldsymbol{\pi}, \boldsymbol{B})$ is zero.

Theorem 2.4 *The parameter space* $\Omega_\Delta(\boldsymbol{\pi}, \boldsymbol{B})$ *is generically identifiable if condition (G1) and (G2) are satisfied.*

(G1) When Δ matrix takes the form of $\Delta = \begin{pmatrix} D_1 \\ D_2 \\ D^* \end{pmatrix}$ after row swapping where D_1,

$$D_2 \in \mathbb{D}_g, \mathbb{D}_g = \left\{ D \in \{0, 1\}^{K \times M^K} : D = \begin{bmatrix} * & \mathbf{1}'_{M-1} & * & \ldots & * & \ldots * \\ * & * & \mathbf{1}'_{M-1} & \ldots & * & \ldots * \\ \vdots & * & * & \ddots & * & \ldots \vdots \\ * & * & * & \ldots & \mathbf{1}'_{M-1} & \ldots * \end{bmatrix} \right\}$$

Note: $\mathbf{1}_{M-1}$ is a vector of 1 of length $M-1$ which represents the activeness of a specific attribute on all levels.

(G2) In D^*, for any attribute $k = 1, 2, \ldots, K$, there exists an item $j > 2K$, such that all the main-effect components regarding this attribute are active $(\delta_{j,k0}, \delta_{j,k1}, \ldots, \delta_{j,k,M-1}) = 1$.

The proof details are provided in Appendix "Generic Identifiability Proof".

21.3   Gibbs Sampling

Following the Bayesian model formulation displayed in Sect. 21.2.1, this section outlines an MCMC approach for the proposed SLCM models. First, we introduce a deterministic relationship between the observed ordinal response Y_{ij} and a continuous augmented latent variable Y_{ij}^* as Eqs. 21.14 and 21.15 show. The augmented variable Y_{ij}^* is generated from a normal distribution conditioned on the latent class mean parameter $\mu_{ij} = \boldsymbol{\alpha}_i' \boldsymbol{\beta}_j$. If Y_{ij}^* falls into the range $[\tau_{jp}, \tau_{j,p+1})$, the random variable Y_{ij} takes the value of p.

$$Y_{ij} = \sum_{p=0}^{P} p \mathcal{I}(\tau_{jp} \le Y_{ij}^* < \tau_{j,p+1}) \tag{21.14}$$

$$Y_{ij}^* | \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j \sim N(\boldsymbol{\alpha}_i' \boldsymbol{\beta}_j, 1) \tag{21.15}$$

We consider a multinomial prior for latent attribute variable $\boldsymbol{\alpha}_i$ as $\boldsymbol{\alpha}_i \sim$ Multinomial$(\boldsymbol{\pi})$. The latent class structural probability $\boldsymbol{\pi}$ follows a conjugate Dirichlet distribution $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{d}_0)$ with hyperparameter $\boldsymbol{d}_0 = \mathbf{1}_{M^K}'$.

In addition, we adopt a spike and slab prior for $\boldsymbol{B}$ as Culpepper (2019) described. For each single β parameter, we formulate the Bayesian model as follows:

$$\beta_{jc} | \delta_{jc} \sim \begin{cases} \mathcal{N}\left(0, \sigma_\beta^2\right) \mathcal{I}(\beta_{jc} > L_{jc}) & \delta_{jc} = 1 \\ \mathcal{I}\left(\beta_{jc} = 0\right) & \delta_{jc} = 0 \end{cases}$$

$$\delta_{jc} | \omega \sim \text{Bernoulli}(\omega)$$

$$\omega \sim \text{Beta}(w_0, w_1),$$

where $(\sigma_\beta^2, w_0, w_1)$ are user-specified hyperparameters and L_{jc} refers to the lower bound for satisfying the monotonicity constraints mentioned in Sect. 21.2.1.4. Noted that the intercepts in $\boldsymbol{\Delta}$ are always set to be active with $\boldsymbol{\delta}_{j0} = 1$.

As a binary variable, we sample δ_{jc} from the following conditional distribution:

$$\delta_{jc} \mid \mathbf{y}_j^*, A, \boldsymbol{\beta}_{j(c)}, \omega, \sigma_\beta^2 \sim \text{Bernoulli}(\tilde{\omega}_{jc}), \tag{21.16}$$

where $A = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{M^K})$ refers to the attribute profile matrix, $\mathbf{y}_j^*$ is the augmented responses vector, and $\boldsymbol{\beta}_{j(c)}$ is the coefficient vector $\boldsymbol{\beta}_j$ that discards the p-th element. Once δ_{jc} is updated, we update $\beta_{j(c)}$ given the full conditional distribution:

$$\beta_{jc} \mid \mathbf{y}_j^*, A, \boldsymbol{\beta}_{j(c)}, \omega, \sigma_\beta^2, \delta_{jc} \sim \mathcal{N}\left(\tilde{\mu}_{jc}, \tilde{\sigma}_c^2\right) [\mathcal{I}(\beta_{jc} > L_{jc})]^{\delta_{jc}} [\mathcal{I}(\beta_{jc} = 0)]^{1-\delta_{jc}}. \tag{21.17}$$

Given Eqs. 21.16 and 21.17, the Bernoulli parameter $\tilde{\omega}_{jc}$ is derived as

$$w_{jc} = \frac{w \Phi(\frac{-L_{jc}}{\sigma_\beta})^{-1} (\frac{\tilde{\sigma}_c}{\sigma_\beta}) \Phi(\frac{\tilde{\mu}_{jc}-L_{jc}}{\tilde{\sigma}_c}) \exp(\frac{\tilde{\mu}_{jc}^2}{2\tilde{\sigma}_c^2})}{w \Phi(\frac{-L_{jc}}{\sigma_\beta})^{-1} (\frac{\tilde{\sigma}_c}{\sigma_\beta}) \Phi(\frac{\tilde{\mu}_{jc}-L_{jc}}{\tilde{\sigma}_c}) \exp(\frac{\tilde{\mu}_{jc}^2}{2\tilde{\sigma}_c^2}) + 1 - w} \tag{21.18}$$

$$\tilde{\mu}_{jc} = \tilde{\sigma}_c^2 A_c' (y_{jc}^* - A_{(c)} \boldsymbol{\beta}_{j(c)}) \tag{21.19}$$

$$\tilde{\sigma}_c^2 = (A_c' A_c + \sigma_\beta^{-2})^{-1} \tag{21.20}$$

where A_c refers to the c-th column in the design matrix A. Note the derivation details can be found in Chen et al. (2020). The full MCMC sampling process is summarized in Algorithm 1, whereas Algorithm 2 presents the detailed sampling steps of the parameter matrix $\boldsymbol{B}$ and $\boldsymbol{\Delta}$.

21.4 An Empirical Application

21.4.1 Short Dark Triad

In the past decade, a great interest has been directed to measure the dark pattern of behaviors, goals, and characters. The Dark Triad (DT; Paulhus & Williams 2002) is one of the most popularly studied personality constructs, which encompasses three substantive dimensions: Machiavellianism, narcissism, and psychopathy. However, different studies have made contrasting conclusions on the construct of the DT (Persson et al. 2019). For instance, Furnham et al. (2013) have argued that psychopathy sometimes subsumes Machiavellianism and narcissism inadvertently. Others have declared that Machiavellianism and psychopathy have the same core and should be deemed as the same measure (Garcia & Rosenberg 2016;

Glenn & Sellbom 2015; McHoskey et al. 1998). Traditionally, assessment of the DT often requires distinct measures for each dimension. To simplify the process of data collection, a measure, namely, Short Dark Triad (SD3; Jones & Paulhus 2014), was created with 27 items selected in a 5-point Likert-type scale (1, "Strongly disagree"; 2,"Slightly disagree"; 3,"Neutral"; 4,"Slightly agree"; and 5, "Strongly agree"). The SD3 items are shown in Appendix "Short Dark Triad" .

In this study, we investigate the latent construct underlying the SD3 scale through exploratory SLCMs. The dataset are available on the Open Psychometrics website (https://openpsychometrics.org/tests/SD3/), where we select a random sample of $N = 5000$ observations. Original item affiliations to the three dimensions can be inferred from the item index in Appendix "Short Dark Triad" ("M" refers to "Machiavellianism"; "N" refers to "narcissism"; "P" refers to "psychopathy").

21.4.2   Model Comparisons

Traditionally, exploratory factor analysis (EFA; Furnham et al. 2014) has been the most popular tool to excavate latent constructs underlying manifest variables in a self-reported questionnaire. Unlike EFA models which treat personality traits as continuous variables, the SLCM allows us to explore the potential for interpreting the personality trait as discrete variables. The purpose of this section is to (1) fit exploratory SLCM with different attributes ($K = 2, 3, 4$) and attribute levels ($M = 2, 3, 4$) and (2) compare the exploratory SLCM to EFA models with ($K = 2, 3, 4$) factors.

Using a Bayesian approach, we apply a tenfold cross-validation approach to estimate out-of-sample predictive accuracy using within-sample estimates. We choose the k-fold cross-validation approach due to its simplicity compared to the leave-one-out method. Considering that an increasing number of folds help reduce the bias term (Vehtari & Lampinen 2002) caused by data split, we partition the DT3 dataset ($N = 5000$) into *ten* subsets $\{y_k \mid k = 1, \ldots, 10\}$. One fold is used as testing data ($N = 500$), and the remaining folds are used as training data ($N = 4500$). For fold k, the testing and training data are denoted as y_k and $y_{(-k)}$, respectively. For each training data $y_{(-k)}$, we employ the algorithm discussed in Sect. 21.3 to obtain posterior draws of the exploratory SLCM and computed posterior means as the point estimates for $\boldsymbol{B}$ and $\boldsymbol{\Delta}$. Thresholds $\boldsymbol{\tau}$ are fixed to be $\boldsymbol{\tau} = \{-\infty, 0, 2, 4, \infty\}$. We run 10 Markov chain Monte Carlo (MCMC) chains, and for each MCMC chain, a total of 80,000 iterations are generated. Specifically, within each chain, the first 20,000 iterations are discarded as burn-in samples, and the left 60,000 iterations are retained as posterior samples. Finally, one chain is chosen out of the *ten* MCMC chains as it generated the highest marginal likelihood. Note here the

point estimates are computed by taking element-wise means for the item parameter matrix $\boldsymbol{B}$ and the structural parameter vector $\boldsymbol{\pi}$ over posterior distributions. With the posterior distributions estimated from the training data, we then evaluate model fit in the testing data using log point-wise predictive density (lpd) which is defined as

$$
\begin{aligned}
\text{lpd} &= \sum_{k=1}^{10} \sum_{i \in y_k} \log p\left(\boldsymbol{y}_i \mid \boldsymbol{y}_{(-k)}\right) \\
&= \sum_{k=1}^{10} \sum_{i \in y_k} \log \int p\left(\boldsymbol{y}_i \mid \boldsymbol{\Omega}_k\right) p\left(\boldsymbol{\Omega}_k \mid \boldsymbol{y}_{(-k)}\right) d\boldsymbol{\Omega}_k, \quad i \in y_k,
\end{aligned}
\tag{21.21}
$$

where $\boldsymbol{\Omega}_k = \{\boldsymbol{B}_k, \boldsymbol{\pi}_k\}$ represents the item parameters estimated from the training data $\boldsymbol{y}_{(-k)}$. To compute lpd in practice, we evaluate the integration of $\boldsymbol{\Omega}_k$ using MCMC posterior draws, and the log point-wise predictive density for data points in the testing data $\boldsymbol{y}_k$ is written as

$$
\widehat{\text{lpd}}_k = \sum_{i \in y_k} \log \frac{1}{S} \sum_{s=1}^{S} f_m\left(\boldsymbol{y}_i \mid \boldsymbol{B}^{k,s}, \boldsymbol{\pi}\right),
\tag{21.22}
$$

where $\boldsymbol{B}^{k,s}$ are the s_{th} draws from the posterior distributions given the training data $\boldsymbol{y}_{(-k)}$ and i is the index of individuals. The complete log predictive density can be calculated by summing all observations over the 10-folds as $\widehat{\text{lpd}} = \sum_{k=1}^{10} \widehat{\text{lpd}}_k$. Furthermore, the marginal likelihood of response $\boldsymbol{y}_i$ for the s_{th} draws is written as

$$
f_m\left(\boldsymbol{y}_i \mid \boldsymbol{B}^{k,s}, \boldsymbol{\pi}\right) = \sum_{\alpha_i^{k,s}} f_c(\boldsymbol{y}_i \mid \boldsymbol{B}^{k,s}, \boldsymbol{\alpha}_i^{k,s}) g(\boldsymbol{\alpha}_i^{k,s} \mid \boldsymbol{\pi}),
\tag{21.23}
$$

where the latent variable $\alpha_i^{k,s}$ is integrated out with the hyperparameter $\boldsymbol{\pi}$. Since data y_{ij} are independent response data conditioned on the model parameter $\boldsymbol{\beta}_j$ and the attribute pattern $\boldsymbol{\alpha}_i$, we have

$$
f_c\left(\boldsymbol{y}_i \mid \boldsymbol{B}^{k,s}, \boldsymbol{\alpha}_i^{k,s}\right) = \prod_{j=1}^{J} \sum_{p=0}^{P-1} \mathbb{1}(y_{ij} = p) P(y_{ij} = p \mid \boldsymbol{\beta}_j^{k,s}, \boldsymbol{\alpha}_i^{k,s})
\tag{21.24}
$$

For EFA models, we apply the function "MCMCordfactanal" in the R package **MCMCpack** (Martin et al. 2011) to perform Bayesian estimation for posterior inference. We use its default setting of 10,000 burn-in and 20,000 mcmc iterations, and for each training sample $y_{(-k)}$, we take 500 posterior draws to compute predictive accuracy in testing sample y_k. Similar to SLCM,

we calculate the lpd criterion for the EFA models wherein the integration in Eq. (21.22) is obtained via the function "hcubature" in the R package **cubature** (Narasimhan et al. 2020). A higher lpd value indicates a model with higher prediction accuracy. Note that, considering the variety of latent variables assumed by SLCM and EFA, we use the marginal likelihoods f_m instead of the conditional likelihoods f_c in the predictive likelihood function $f_m\left(\boldsymbol{y}_i \mid \boldsymbol{B}^{k,s}, \boldsymbol{\pi}\right)$.

As seen in Table 21.2, inclusion of more factors leads to a superior fit of EFA models for the reason that more nuisance variance is considered. In the two-dimensional case of $K = 2$, the SLCM (lpd ≥ -183595.3) performs better than the EFA (lpd $= -186914.5$) if the attribute is ordinal $M \geq 3$. Furthermore, in the three-/four-dimensional case of $K \geq 3$, the SLCM consistently outperform the EFA regardless of the choice of M. Overall, the lpd of SLCM shows a slight rise when the dimension K or attribute level M increases, suggesting that increasing either the dimension or attribute level could improve the model fit. At this point, the best fitting model is the SLCM with $K = 2$ attributes and $M = 4$ attribute levels (lpd $= -183,443.2$).

Given the information concerning the relative model fit, we estimated an exploratory SLCM with $K = 2$ and $M = 4$ in the same SD3 dataset ($N = 5000$). We ran 10 chains of length 80,000 with a burn-in of 20,000 iterations and keep the chain with the highest marginal likelihood. Figure 21.1 displays the estimated sparse structure of $\boldsymbol{B}$, where we can summarize the two attributes as (1) narcissism and (2) Machiavellianism. The x-axis of Fig. 21.1 presents the predictors, where 01, 02, and 03 refer to the main-effect predictors of Machiavellianism; 10, 20, and 30 refer to the main-effect predictors of narcissism; and sparsity of the matrix is reflected on the fact that loadings manifest on the main-effect predictors. To this end, we can obtain a rough conclusion on the item and attribute relationships. In particular, items M1–M9, P1, P3, P5, P6, and P9 load mostly on Machiavellianism; items N1–N4, N6–N8, P2, and P7 load on narcissism; and items N5, N9, P4, and P8 load equally on the two dimensions.

Table 21.2 Model comparison results in lpd, SD3 ($N = 5000$)

Dimension	EFA		SLCM
$K = 2$	$-186,914.5$	$M = 2$	$-189,320.7$
		$M = 3$	$-183,595.3$
		$M = 4$	$-183,443.2$
$K = 3$	$-185,898.2$	$M = 2$	$-183,929.6$
		$M = 3$	$-183,916.3$
		$M = 4$	$-183,902.9$
$K = 4$	$-184,743.8$	$M = 2$	$-183,577.2$
		$M = 3$	$-183,552.5$
		$M = 4$	$-183,532.1$

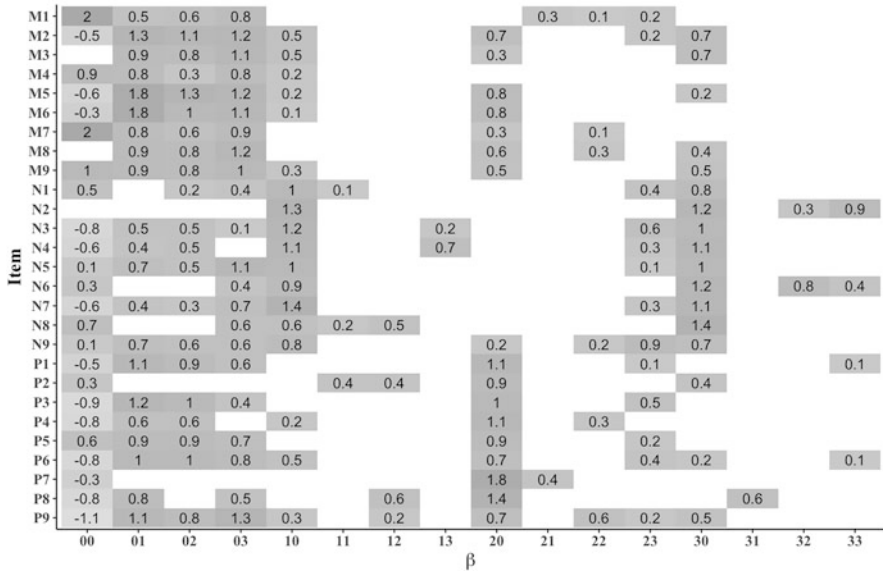| Item	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
M1	2	0.5	0.6	0.8						0.3	0.1	0.2				
M2	-0.5	1.3	1.1	1.2	0.5				0.7		0.2	0.7				
M3		0.9	0.8	1.1	0.5				0.3			0.7				
M4	0.9	0.8	0.3	0.8	0.2											
M5	-0.6	1.8	1.3	1.2	0.2				0.8			0.2				
M6	-0.3	1.8	1	1.1	0.1				0.8							
M7	2	0.8	0.6	0.9					0.3		0.1					
M8		0.9	0.8	1.2					0.6		0.3		0.4			
M9	1	0.9	0.8	1	0.3				0.5				0.5			
N1	0.5		0.2	0.4	1	0.1					0.4	0.8				
N2					1.3							1.2			0.3	0.9
N3	-0.8	0.5	0.5	0.1	1.2		0.2				0.6	1				
N4	-0.6	0.4	0.5		1.1		0.7				0.3	1.1				
N5	0.1	0.7	0.5	1.1	1						0.1	1				
N6	0.3			0.4	0.9							1.2			0.8	0.4
N7	-0.6	0.4	0.3	0.7	1.4						0.3	1.1				
N8	0.7			0.6	0.6	0.2	0.5					1.4				
N9	0.1	0.7	0.6	0.6	0.8				0.2		0.2	0.9	0.7			
P1	-0.5	1.1	0.9	0.6					1.1			0.1				0.1
P2	0.3					0.4	0.4		0.9			0.4				
P3	-0.9	1.2	1	0.4					1			0.5				
P4	-0.8	0.6	0.6		0.2				1.1	0.3						
P5	0.6	0.9	0.9	0.7					0.9			0.2				
P6	-0.8	1	1	0.8	0.5				0.7			0.4	0.2			0.1
P7	-0.3								1.8	0.4						
P8	-0.8	0.8		0.5			0.6		1.4					0.6		
P9	-1.1	1.1	0.8	1.3	0.3		0.2		0.7		0.6	0.2	0.5			

β

Fig. 21.1 Estimated *B* matrix for DT3 data under SLCM with $K = 2$, $M = 4$
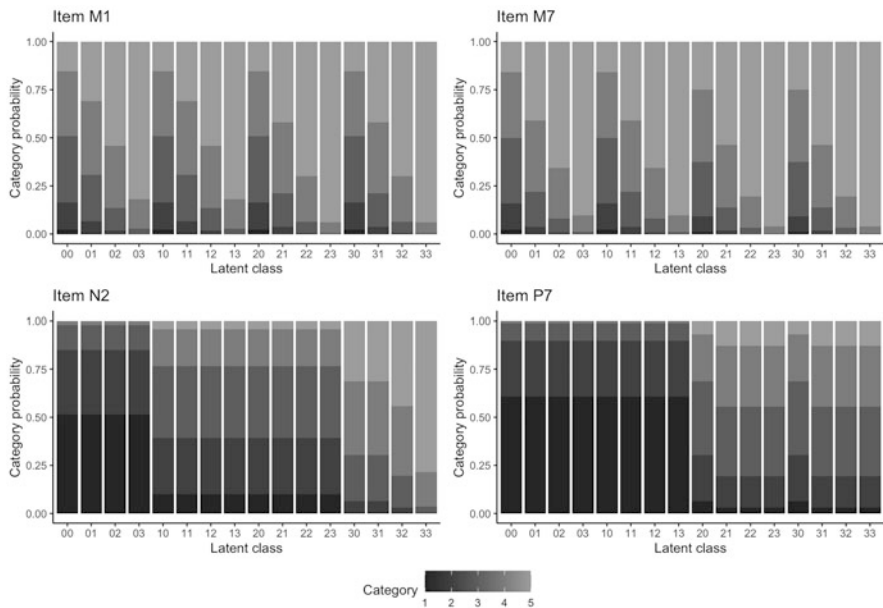
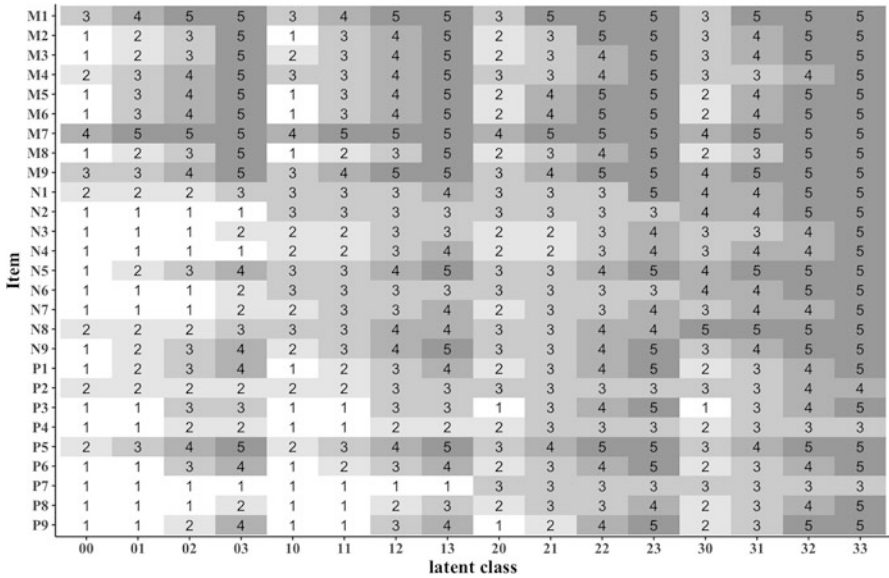**Fig. 21.2** Estimated item category response function by latent class and category probability

Item	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
M1	3	4	5	5	3	4	5	5	3	5	5	5	3	5	5	5
M2	1	2	3	5	1	3	4	5	2	3	5	5	3	4	5	5
M3	1	2	3	5	2	3	4	5	2	3	4	5	3	4	5	5
M4	2	3	4	5	3	3	4	5	3	3	4	5	3	3	4	5
M5	1	3	4	5	1	3	4	5	2	4	5	5	2	4	5	5
M6	1	3	4	5	1	3	4	5	2	4	5	5	2	4	5	5
M7	4	5	5	5	4	5	5	5	4	5	5	5	4	5	5	5
M8	1	2	3	5	1	2	3	5	2	3	4	5	2	3	5	5
M9	3	3	4	5	3	4	5	5	3	4	5	5	4	5	5	5
N1	2	2	2	3	3	3	3	4	3	3	3	5	4	4	5	5
N2	1	1	1	1	3	3	3	3	3	3	3	3	4	4	5	5
N3	1	1	1	2	2	2	3	3	2	2	3	4	3	3	4	5
N4	1	1	1	1	2	2	3	4	2	2	3	4	3	4	4	5
N5	1	2	3	4	3	3	4	5	3	3	4	5	4	5	5	5
N6	1	1	1	2	3	3	3	3	3	3	3	3	4	4	5	5
N7	1	1	1	2	2	3	3	4	2	3	3	4	3	4	4	5
N8	2	2	2	3	3	3	4	4	3	3	4	4	5	5	5	5
N9	1	2	3	4	2	3	4	5	3	4	5	5	3	4	5	5
P1	1	2	3	4	1	2	3	4	2	3	4	5	2	3	4	5
P2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4
P3	1	1	3	3	1	1	3	3	1	3	4	5	1	3	4	5
P4	1	1	2	2	1	1	2	2	2	3	3	3	2	3	3	3
P5	2	3	4	5	2	3	4	5	3	4	5	5	3	4	5	5
P6	1	1	3	4	1	2	3	4	2	3	4	5	2	3	4	5
P7	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3
P8	1	1	1	2	1	1	2	3	2	3	3	4	2	3	4	5
P9	1	1	2	4	1	1	3	4	1	2	4	5	2	3	5	5

latent class

Fig. 21.3 Dominant responses by latent classes and items

As a fully saturated model, interactions can take place between any level of any two attributes. Figure 21.2 are stack barplots that clearly depict how increases in attributes correspond to changes in category response probabilities. The x- and y-axes indicate the latent classes and response probabilities; the stacked bars represent the five response categories. The barplots selectively present the item category response function for items M1, M7, N2, and P7. Item M1 has active coefficients $\beta_{00} = 2.0$, $\beta_{01} = 0.5$, $\beta_{02} = 0.6$, $\beta_{03} = 0.8$, $\beta_{21} = 0.3$, $\beta_{22} = 0.1$, and $\beta_{23} = 0.2$. We can see that all main-effect terms regarding Machiavellianism manifest, while narcissism is active only on the interaction terms. In Fig. 21.3, latent classes 10, 20, and 30 that represents the group are less likely to endorse category 5 (5 = "Strongly agree) compared to other latent classes.

The active coefficients of item N2 are $\beta_{10} = 1.3$, $\beta_{30} = 1.2$, $\beta_{32} = 0.3$, and $\beta_{33} = 0.9$. We can tell that narcissism is more significant than Machiavellianism. In Fig. 21.3, latent classes 01, 02, and 03 which reflects the mastery of three levels on narcissism are most likely to endorse category 1 (5 = "Strongly disagree) compared to other latent classes.

Moreover, Fig. 21.3 shows the dominant response category for each latent class on the 27 items. Given a specific item, the value on the table represents the response category which a latent class has the highest probabilities to endorse over the other response categories.

21.4.3   Monte Carlo Simulation Study

This section presents a simulation study to evaluate the parameter recovery rate for SLCM with different sample sizes. We use the previous estimates of item parameter B and the structural parameter π (see Appendix "Empirical Item Parameters") as the population parameters. Specifically, we have $J = 27$ items and assume the underlying dimension is $K = 2$ and attribute level is $M = 4$. The sample size is set to be $n = 1000, 2000, 3000$, with each sample size condition replicated for 100 times. For each replicated dataset, we run 10 chains where each single chain has a total of 60,000 iterations with a burn-in sample of 20,000 inside. The chain with the highest likelihood is chosen to perform posterior inference and compute the recovery accuracy. Note here we generate different attribute profiles α and responses Y per replication.

The estimation accuracy of B is evaluated in terms of the mean absolute deviation (MAD) for each single β. For each replication, we record the posterior mean of each single parameter in B as the estimates. Next, we compute the absolute deviation between the estimates and the true parameters. Then, we take the average of the absolute deviation over replications. Table 21.3 enunciates the MAD of βs with its activeness into consideration. In specific, the first row "B_Δ" refers to the MAD averaged over the entire matrix B. The second row "$B_{\Delta=1}$" and the third row "$B_{\Delta=0}$" refer to the MAD averaged over the locations where $\delta_s = 1$ and $\delta_s = 0$, respectively. Likewise, we also compute and record the mean absolute deviation (MAD) for each π. Table 21.4 presents the recovery rate of Δ in terms of the proportion of entries that are correctly recovered. The first row "Δ" refers to the proportion of correctly recovered δs over the whole matrix. The second row "$\Delta = 1$" and the third row "$\Delta = 0$" refer to the proportion of 1's and 0's in the population Δ matrix that are correctly recovered.

The result in Table 21.3 shows that the average EAD for B is 0.100, 0.069 and 0.052 for sample sizes of $n = 1000, 2000$, and 3000. The average EAD for π is 0.009, 0.008, and 0.007 corresponding to sample sizes of $n = 1000, 2000$, and 3000. Additionally, Table 21.4 shows the recovery rate for D is 0.883, 0.913,

Table 21.3 Mean absolute deviation (MAD) of B and π

	$n = 1000$	$n = 2000$	$n = 3000$
B_Δ	0.100	0.069	0.052
$B_{\Delta=1}$	0.237	0.166	0.128
$B_{\Delta=0}$	0.046	0.031	0.022
π	0.009	0.008	0.007

Table 21.4 Recovery accuracy of Δ

	$n = 1000$	$n = 2000$	$n = 3000$
Δ	0.883	0.913	0.931
$\Delta = 1$	0.702	0.773	0.814
$\Delta = 0$	0.954	0.968	0.976

and 0.932 for sample sizes of $n = 1000, 2000$, and 3000. As seen, for both of them, the estimation accuracy rises as the sample size increases. Furthermore, in Table 21.3, we observe that the active entries in B have a larger bias than the inactive entries. Similarly, we have at least 0.954 of the inactive entries in D that are correctly estimated as "0" and at least 0.702 of the active entries in D that are correctly estimated as "1." The simulation results support that the model can be mostly recovered by our Gibbs algorithm.

21.4.4 Model Convergence

To evaluate the convergence of the Gibbs sampler, we generate *three* chains for the SLCM with $K = 2$ and $M = 4$ under the most computationally intensive condition $N = 3000$. For each of the latent class mean parameter μ and the structural parameter π, the Gelman-Rubin proportional scale reduction factor (PSRF), also known as $\hat{R}$, is calculated. A $\hat{R}$ value of below 1.2 indicates the acceptable convergence. In our simulation, the maximum $\hat{R}$ is found to be 0.97 for π and 1.06 for μ, with the 80,000 iterations and 20,000 burn-in samples inside. Therefore, we conclude the MCMC chains have reached a steady state.

21.5 Discussion

In this study, we propose a strict and generic model identifiability condition for SLCM with polytomous attributes, which expands the work of SLCM with binary attributes by Chen et al. (2020) and Culpepper (2019). We develop a Gibbs sampling algorithm with the design of enforcing the identifiability and monotonicity constraints. Specifically, with strict identifiability conditions imposed, we notice that the MCMC chains are often trapped and have a slow move forward. A possible explanation is that the strict conditions are too restrictive for the chains to search the right parameter space. Without explicitly enforcing the strict identifiability constraints, the models convergence in 80,000 draws with estimates satisfying the proved generic identifiability conditions. The simulation results demonstrate that the algorithm is efficient in recovering the parameters in different sample sizes.

Overall, our study is innovative in the following aspects. First, we provide a successful case study of applying SLCM to a personality scale. Personality have traditionally been viewed as continuous traits instead of discrete categories, and factor analysis (FA) approach which assumes continuous latent variables is often used in personality measurement. However, with the estimated person scores, it

is always a critical issue to identify cut-offs and classify individuals into different classes. To this end, if variables can be viewed more or less categorical, we can rely on the models with discrete latent variables to provide fine-grained information on the individual differences. Another advantage with discrete latent variables lies in its greatly reduced parameter space, with a potential of facilitating sampling. Overall, our study provides new insights into interpreting personality traits as a discrete measure. Plus, the model also has the potential of being applied to educational settings and contributes to better measurement for educational intervention (Chen & Culpepper 2020).

Second, our study, for the first time, compares SLCM and EFA models from an exploratory perspective. We found the SLCM fit significantly better than the EFA models with a higher prediction accuracy in several configurations. This finding has important implications for promoting applications of CDMs to the areas outside of educational measurement, where another early example is by Cho (2016) who has explored the construct validity of emotional intelligence in situational judgmental tests. In addition, our analysis of the item-attribute structures underlying SD3 supports the previous finding by Persson et al. (2019) that Machiavellianism and psychopathy are subsumable constructs. Moreover, they found the subscale composite scores for the three constructs contain relatively little specific variance, with an implication that reporting the total scores is more appropriate for SD3 than reporting the subscale scores. In our result, most items do not follow a simple structure pattern, which further support this statement that dimensions of SD3 are somewhat inseparable.

Third, from a methodology perspective, our paper addresses the model identifiability concerns of SLCM with polytomous attributes. The strict identifiability condition is way too restrictive in practice. For instance, when the number of attributes is relatively large compared to the items, (e.g., close to half the number of items), enforcing strict identifiability is equivalent to presuming a simple structure on all items. For personality assessment, a simple item structure is often unrealistic to achieve. For this reason, the generic identifiability that loosens some constraints broadens the model applicability.

There are still several recommendations for future study. First, although the MCMC chains successfully converge to the posterior distributions, the Gibbs samplers are still not efficient enough in exploring the parameter space. We have to run several chains and conduct a likelihood selection to find the one with best mixing. The difficulty of mixing could be due to the complexity of the saturated model wherein we have 16 parameters per item. To solve the mixing issue, future work is required to develop more flexible moves in the algorithm that can break local traps or jump between difference spaces.

Second, instead of framing the SLCM in an unstructured way, it is interesting to include a higher-order factor model (Culpepper & Chen 2019; De La Torre & Douglas 2004; Henson et al. 2009) or a multivariate normal distribution with

a vector of thresholds and a polychoric correlation matrix (Chen & Culpepper 2020; Henson et al. 2009; Templin et al. 2008) in the latent class structure. There is also abundant room for future progress in selecting the competing structures for π.

Third, it is also possible to prespecify the number of attributes with a more established approach. For instance, Chen et al. (2021) present a crimp sampling algorithm to jointly infer the number of attribute for DINA model. In our study, the choice of attribute level is limited by the study design. Future work with focus on the selection of attribute level is greatly suggested.

Algorithm 1 Full Gibbs sampling algorithm

Data: $Y_{N \times J}$; π; $\alpha_{1:N}$; $B_{J \times M^K}$; τ; $A_{M^K \times M^K}$; chain length T
Result: Y^*; π; α
for t *in* $(1, \cdots, T)$ **do**
    for j *in* $(1, \cdots, J)$ **do**
        for $c \in (0, \cdots, M^K - 1)$ **do**
            for $y_{ij} \in \{0, \cdots, P - 1\}$ **do**
                $\theta_{jc, y_{ij}} = \Phi(\tau_{y_{ij+1}} - \alpha'_c \beta_j) - \Phi(\tau_{y_{ij}} - \alpha'_c \beta_j);$
            end
        end
    end
    for i *in* $(1, \cdots, N)$ **do**
        Sample α_i from multinomial distribution;
        $P(\alpha_i = \alpha_c \mid \pi, y_i) \propto \dfrac{\pi_c \prod_{j=1}^{J} \theta_{cj, y_{ij}}}{\sum_{c=0}^{M^K - 1} \pi_c \prod_{j=1}^{J} \theta_{cj, y_{ij}}};$
    end
    for $c \in (0, \cdots, M^K - 1)$ **do**
        Sample π from Dirichlet distribution;
        $P(\pi \mid A) \propto \prod_{c=0}^{M^K - 1} \pi_c^{\sum_{i=1}^{N} I(\alpha_i = \alpha_c) + d_{oc}}$
    end
    for j *in* $(1, \cdots, J)$ **do**
        for i *in* $(1, \cdots, N)$ **do**
            Sample y^*_{ij} from truncated normal distribution;
            $P(y^*_{ij} \mid \alpha_i, \beta_j) \propto N(\alpha'_i \beta_j, 1) I(\tau_{j, y_{ij}} < y^*_{ij} < \tau_{j, y_{ij+1}})$
        end
    end
    Sample B and Δ from Algorithm 2
end

Algorithm 2 Full Gibbs sampling algorithm: $\boldsymbol{B}$; $\boldsymbol{\Delta}$

Data: Hyperparameters σ_β, w, w_0, w_1; $\boldsymbol{Y}^0$ from Algorithm 1
Result: $\boldsymbol{B}$,$\boldsymbol{\Delta}$
for t *in* $(1, \cdots, T)$ **do**
    for c *in* $(1, \cdots, M^K - 1)$ **do**
        $\tilde{\sigma}_c^2 = \frac{1}{A_c' A_c + \sigma_\beta^{-2}}$;
        for j *in* $1, \cdots, J$ **do**
            $\tilde{\mu}_{jc} = \tilde{\sigma}_c^2 A_c' (\boldsymbol{Y}_j^0 - A_{(c)}\beta_{j(c)})$;
            $L \leftarrow \max \left\{ \max_{\boldsymbol{\alpha} \in \mathbb{L}_1} -\gamma_{\boldsymbol{\alpha}}', \max_{\boldsymbol{\alpha} \in \mathbb{L}_0} \gamma_{\boldsymbol{\alpha}}' - \gamma_{\boldsymbol{q}_j}' \right\}$;
            if *($L \leq 0$)* and the identifiability condition is satisfied **then**

$$w_{jc} = \frac{w\Phi(\frac{-L}{\sigma_\beta})^{-1}(\frac{\tilde{\sigma}_c^2}{\sigma_\beta^2})^{\frac{1}{2}}\Phi(\frac{\tilde{\mu}_{jc}-L}{\tilde{\sigma}_c})\exp(\frac{\tilde{\mu}_{jc}^2}{2\tilde{\sigma}_c^2})}{w\Phi(\frac{-L}{\sigma_\beta})^{-1}(\frac{\tilde{\sigma}_c^2}{\sigma_\beta^2})^{\frac{1}{2}}\Phi(\frac{\tilde{\mu}_{jc}-L}{\tilde{\sigma}_c})\exp(\frac{\tilde{\mu}_{jc}^2}{2\tilde{\sigma}_c^2})+1-w};$$

                Sample δ_{jc} from Bernoulli(w_{jc})
            end
            if $\delta_{jc} = 1$ **then**
                Sample β_{jc} from a truncated normal distribution;
                $P(\beta_{jc} \mid \tilde{\mu}_{jc}, \tilde{\sigma}_c, \delta_{jc} = 1) \propto N(\tilde{\mu}_{jc}, \tilde{\sigma}_c^2)I(\beta_{jc} > L)$
            else
                $\beta_{jc} = 0$
            end
        end
    end
    Sample w from Beta($\sum_{j,c}(1 - \delta_{jc}) + w_0$, $\sum_{j,c} \delta_{jc} + w_1$)
end

Appendices

Short Dark Triad

See Table 21.5.

Table 21.5 Short dark triad items in the original item affiliation

	Statements
M1	It's not wise to tell your secrets.
M2	I like to use clever manipulation to get my way.
M3	Whatever it takes, you must get the important people on your side.
M4	Avoid direct conflict with others because they may be useful in the future.
M5	It's wise to keep track of information that you can use against people later.
M6	You should wait for the right time to get back at people.
M7	There are things you should hide from other people because they don't need to know.
M8	Make sure your plans benefit you, not others.
M9	Most people can be manipulated.
N1	People see me as a natural leader.
N2	I hate being the center of attention.
N3	Many group activities tend to be dull without me.
N4	I know that I am special because everyone keeps telling me so.
N5	I like to get acquainted with important people.
N6	I feel embarrassed if someone compliments me.
N7	I have been compared to famous people.
N8	I am an average person.
N9	I insist on getting the respect I deserve.
P1	I like to get revenge on authorities.
P2	I avoid dangerous situations.
P3	Payback needs to be quick and nasty.
P4	People often say I'm out of control.
P5	It's true that I can be mean to others.
P6	People who mess with me always regret it.
P7	I have never gotten into trouble with the law.
P8	I enjoy having sex with people I hardly know.
P9	I'll say anything to get what I want.

Empirical Item Parameters

See Table 21.6.

Table 21.6 DT3 element-wise means for B and π for a random sample of $N = 5000$ respondents

	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
M1	2.0	0.5	0.6	0.8						0.3	0.1	0.2				
M2	−0.5	1.3	1.1	1.2	0.5				0.7			0.2	0.7			
M3	0.0	0.9	0.8	1.1	0.5				0.3				0.7			
M4	0.9	0.8	0.3	0.8	0.2											
M5	−0.6	1.8	1.3	1.2	0.2				0.8				0.2			
M6	−0.3	1.8	1.0	1.1	0.1				0.8							
M7	2.0	0.8	0.6	0.9					0.3							
M8	0.0	0.9	0.8	1.2					0.6		0.1		0.4			
M9	1.0	0.9	0.8	1.0	0.3				0.5		0.3		0.5			
N1	0.5		0.2	0.4	1.0	0.1						0.4	0.8			
N2	0.0				1.3								1.2		0.3	0.9
N3	−0.8	0.5	0.5	0.1	1.2			0.2				0.6	1.0			
N4	−0.6	0.4	0.5		1.1			0.7				0.3	1.1			
N5	0.1	0.7	0.5	1.1	1.0							0.1	1.0			
N6	0.3			0.4	0.9								1.2		0.8	0.4
N7	−0.6	0.4	0.3	0.7	1.4	0.2						0.3	1.1			
N8	0.7			0.6	0.6		0.5						1.4			
N9	0.1	0.7	0.6	0.6	0.8				0.2		0.2	0.9	0.7			
P1	−0.5	1.1	0.9	0.6					1.1			0.1				0.1
P2	0.3					0.4	0.4		0.9				0.4			

(continued)

Table 21.6 (continued)

	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
P3	−0.9	1.2	1.0	0.4					1.0			0.5				
P4	−0.8	0.6	0.6		0.2				1.1		0.3					
P5	0.6	0.9	0.9	0.7					0.9			0.2				
P6	−0.8	1.0	1.0	0.8	0.5				0.7			0.4	0.2			0.1
P7	−0.3								1.8	0.4						
P8	−0.8	0.8		0.5			0.6		1.4					0.6		
P9	−1.1	1.1	0.8	1.3	0.3		0.2		0.7		0.6	0.2	0.5			
π	0.03	0.03	0.06	0.03	0.03	0.09	0.13	0.06	0.07	0.15	0.09	0.03	0.03	0.09	0.06	0.02

Strict Identifiability Proof

The proof is mainly based on Kruskal's theorem (Kruskal 1976, 1977) and the tripartition strategy proposed by Allman et al. (2009). We first introduce the probability matrix $H(\Delta, B)$ and its Kruskal rank in Definitions 1 and 2.

Definition 1 The class-response matrix $H(\Delta, B)$ is a matrix of size $M^K \times P^J$, wherein the rows denote attribute patterns and the columns denote response patterns. An arbitrary element (α_c, y) in $H(\Delta, B)$ presents the probability of observing a response pattern y from the latent class with attribute profile α_c:

$$H_{c,j}(\Delta, B) = P(Y = y \mid \beta_j, \alpha_c) = \prod_{j=1}^{J} \sum_{p=0}^{P-1} \theta_{cjp} \mathcal{I}(p = y_j), \qquad (21.25)$$

Definition 2 (Kruskal Rank) The Kruskal rank of matrix H is the largest number j such that every set of j columns in H is independent. If H has full row rank, the Kruskal rank of H is its row rank.

Theorem 3 (Allman et al. 2009) *Consider a general latent class model with r classes and $\mathcal{J}$ items, where $J \geq 3$. Suppose all entries of π are positive. If there exists a tripartition of the item set $\mathcal{J} = 1, 2, \ldots, J$ that divides $\mathcal{J}$ into three disjoint, nonempty subsets $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$ such that the Kruskal ranks of the three class-response matrices H_1, H_2, and H_3 satisfy*

$$I_1 + I_2 + I_3 \geq 2r + 2,$$

then the parameters of the model are uniquely determined, up to label switching.

To prove the model parameters are uniquely determined, we need to find three subsets of items in SLCMs that satisfy Theorem 3. Suppose we have three disjoint, nonempty subsets $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$, the marginal probability of response Y can be reframed as a three-way tensor T of dimension $P^{|\mathcal{J}_1|} \times P^{|\mathcal{J}_2|} \times P^{|\mathcal{J}_3|}$. Specifically, the (y_1, y_2, y_3)-th element in T is the marginal probability of the products of the three subsets items:

$$
\begin{aligned}
T_{(y_1, y_2, y_3)} &= P\left(Y_{\mathcal{J}_1} = y_1, Y_{\mathcal{J}_2} = y_2, Y_{\mathcal{J}_3} = y_3 \mid B, \pi\right) \\
&= \sum_c \pi_c P\left(Y_{\mathcal{J}_1} = y_1, Y_{\mathcal{J}_2} = y_2, Y_{\mathcal{J}_3} = y_3 \mid B, \alpha_c\right) \\
&= \sum_c \pi_c P\left(Y_{\mathcal{J}_1} = y_1 \mid B_1, \alpha_c\right) P\left(Y_{\mathcal{J}_2} = y_2 \mid B_2, \alpha_c\right) \\
&\quad \times P\left(Y_{\mathcal{J}_3} = y_3 \mid B_3, \alpha_c\right).
\end{aligned}
\qquad (21.26)
$$

In other words, tensor T can be decomposed as an outer product of three vectors:

$$T = \sum_c \pi_c H_{\alpha_c}(D_1, B_1) \otimes H_{\alpha_c}(D_2, B_2) \otimes H_{\alpha_c}(D_3, B_3),$$

where $H_{\alpha_c}(D_i, B_i)$ represents a row vector of M^K in the class-probability matrix $H(D_i, B_i)$ of size $M^K \times P^{\mathcal{J}_I}$.

Kruskal (1976, 1977) and Allman et al. (2009) state that if the sum of the Kruskal ranks of $H(D_1, B_1)$, $H(D_2, B_2)$, and $H(D_3, B_3)$ is greater or equal to $2M^K + 2$, the tensor decomposition is unique up to the row rescaling and label switching. Now we will give the proof by showing the existence of three item subsets $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$. If the corresponding item-attribute matrices D_1, D_2, D_3 satisfy the structure of D_s, D_s, $and D^*$ in Theorem 2.2, the Kruskal rank sum of their class-probability matrix H_1, H_2, and H_3 satisfies the minimum requirement $2M^K + 2$, and the model parameters are uniquely determined.

Proof 4 shows if D_1 and D_2 take the form of D_s in Theorem 2.2, the class-response matrices $H(D_1, B_1)$ and $H(D_2, B_2)$ have a full Kruskal row rank of M^K.

Proof 4 D_s is the defined item structure matrix of dimension $K \times M^K$ in Theorem 2.2, wherein the k-th item loads on all of the levels in attribute k, i.e., $\delta_k = (\delta_{k,0}, \ldots, \delta_{k,M-1}) = \mathbf{1}$, and the corresponding item parameters $\beta_{k,m} \neq 0$ for $m \in \{0, \ldots, M-1\}$. The class-response matrix $H(D_s, B_s)$ of dimension $M^K \times P^K$ can be written as the Kronecker product of K sub-matrices H_k:

$$H(D_s, B_s) := \bigotimes_{k=1}^{K} H_k$$

$$= \bigotimes_{k=1}^{K} \begin{bmatrix} \Psi(\tau_1 - \mu_{k,0}) & \Psi(\tau_2 - \mu_{k,0}) - \Psi(\tau_1 - \mu_{k,0}) & \ldots & 1 - \Psi(\tau_{(P-1)} - \mu_{k,0}) \\ \Psi(\tau_1 - \mu_{k,1}) & \Psi(\tau_2 - \mu_{k,1}) - \Psi(\tau_1 - \mu_{k,1}) & \ldots & 1 - \Psi(\tau_{(P-1)} - \mu_{k,1}) \\ \Psi(\tau_1 - \mu_{k,2}) & \Psi(\tau_2 - \mu_{k,2}) - \Psi(\tau_1 - \mu_{k,2}) & \ldots & 1 - \Psi(\tau_{(P-1)} - \mu_{k,2}) \\ \Psi(\tau_1 - \mu_{k,M-1}) & \Psi(\tau_2 - \mu_{k,M-1}) - \Psi(\tau_1 - \mu_{k,M-1}) & \ldots & 1 - \Psi(\tau_{(P-1)} - \mu_{k,M-1}) \end{bmatrix},$$

$$(21.27)$$

where H_k can be viewed as the attribute-category matrix of dimension $M \times P$ for the k-th item in D_s. In H_k, the rows indicate the attribute levels and columns indicate the response categories.

Given the item parameters are all nonzero $\beta_{k,m} \neq 0$, the latent class mean parameter $\mu_{k,m}$ is different from each other given $\mu_{k,0} = \beta_{k,0}$, $\mu_{k,1} = \beta_{k,0} + \beta_{k,1}, \mu_{k,2} = \beta_{k,0} + \beta_{k,1} + \beta_{k,2}$, and $\mu_{k,M-1} = \beta_{k,0} + \beta_{k,1} + \cdots + \beta_{k,M-1}$. Then, the rows in matrix H_k are not linearly dependent so that H_k is of full row Kruskal rank, namely, $\text{rank}(H_k) = M$. For each item k in D_s, we have $\text{rank}(H_k) = M$. According to the property of Kronecker products, $H(D_s, B_s) = \bigotimes_{k=1}^{K} H_k$ is also full Kruskal row rank with $\text{rank}(H(D_s, B_s)) = M^K$.

The following Proof 5 shows if D_3 takes the form of D^* in Theorem 2.2, the class-response matrix $H(D_3, B_3)$ has Kruskal row rank of 2.

Proof 5 Condition (S2) in Theorem 2.2 ensures the main-effect components of each attribute to be nonzero in at least one item in D^* so that D^* can distinguish every pair of latent classes. Specifically, there must exist an item j_0 in D^* that any two latent class c_1 and c_2 has different response probability matrix, i.e., $\Theta_{j_0,c_1} \neq \Theta_{j_0,c_2}$. Therefore, there must exist two rows in $H(D^*, B^*)$ that are independent of each other, implying that the Kruskal rank of $H(D^*, B^*)$ is at least 2.

With Proofs 4 and 5, we have $\text{rank}(H_1) + \text{rank}(H_2) + \text{rank}(H_3) \geq 2M^K + 2$.

Generic Identifiability Proof

In the context of SLCMs, the item structure matrix Δ is a sparse matrix, so the real parameter space should be of dimension less than $J \times M^K$. To be differentiated from Eq. 21.9, we denote the parameter space with a sparsity structure Δ as

$$\Omega_{\Delta}(\pi, B) = \{(\pi, B) : \pi \in \Omega_1(\pi), B \in \Omega_{\Delta}^*(B)\}, \tag{21.28}$$

where $\Omega_{\Delta}^*(B)$ presents the parameter space for B which have nonzero entry at position β when the corresponding $\delta = 1$. Then, we define the unidentifiable parameter set C_{Δ} as

$$C_{\Delta} = \{(\pi, B) : \mathbb{P}(\pi, B) = \mathbb{P}(\tilde{\pi}, \tilde{B}), \quad (\pi, B) \not\sim (\bar{\pi}, \bar{B}),$$
$$(\pi, B) \in \Omega_{\Delta}(\pi, B), \quad (\tilde{\pi}, \tilde{B}) \in \Omega_{\tilde{\Delta}}(\pi, B)\}. \tag{21.29}$$

As Definition 2.3 stated, $\Omega_{\Delta}(\pi, B)$ is a generically identifiable parameter space if the unidentifiable set C_{Δ} is of measure zero within $\Omega_{\Delta}(\pi, B)$.

Similar to the strict identifiability proof, we will use the tripartition strategy to find three item subsets $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$ that generate a tensor decomposition of D_1, D_2, and D_3. We proceed to show if D_1, D_2, and D_3 satisfy the structure of D_g, D_g, $and D^*$ in Theorem 2.4, the Kruskal rank sum of the corresponding class-probability matrices H_1, H_2, and H_3 satisfies the minimum requirement $2M^K + 2$.

Proof 6 For H_1 and H_2, we use Theorem 7 to show that $\text{rank}(H_1) = M^K$ and $\text{rank}(H_2) = M^K$ hold almost everywhere in Ω_{D_1} and Ω_{D_2}, respectively. Different from the Theorem 4 in Chen et al. (2020), we perform a transpose multiplication to the response-class matrix $H(D_g, B_g)$ so that it can be transformed into a square matrix $H(D_g, B_g)' H(D_g, B_g)$ which has an accessible determinant function. Given Proof 10, we show $G_D(B) \to \mathbb{R}$ is a real analytical function of B, and then we know $\lambda_{\Omega_D}(A)$ has the Lebesgue measure zero. By Theorem 7, we can infer that $H(D_g, B_g)$ is a full row rank matrix. Therefore, if $D_1 \in \mathbb{D}_g$ and $D_2 \in \mathbb{D}_g$, we have $\text{rank}(H_1) + \text{rank}(H_2) = 2M^K$ holds almost everywhere in $\Omega_{D_1} \bigotimes \Omega_{D_2}$.

Theorem 7 *Given $D \in \mathbb{D}_g$, the corresponding class-response matrix $H(D, B)$ is of full rank except some values of B from a measure zero set with respect to Ω_D, i.e.,*

$$\lambda_{\Omega_D}\{B \in \Omega_D : det[H(D, B)'H(D, B) = 0]\} = 0,$$

where $\lambda_{\Omega_D}(A)$ denotes the Lebesgue measure of set A with respect to Ω_D.

Proposition 8 *If $f : \mathbb{R}^n \to \mathbb{R}$ is a real analytic function which is not identically zero, then the set $\{x : f(x) = 0\}$ has Lebesgue measure zero.*

Remark 9 $G_D(B) = det[H(\Delta, B)'H(\Delta, B)] : \Omega_D \to \mathbb{R}$ is a real analytic function of B.

Proof 10 $G_D(B)$ is a composition function:

$$G_D(B) = det[H(\Delta, B)'H(\Delta, B)] = h[(\theta_{\alpha_0}, \ldots, \theta_{\alpha_{M^K-1}})'(\theta_{\alpha_0}, \ldots, \theta_{\alpha_{M^K-1}})]$$

where $h(\theta) : [0, 1]^{K \times M^K}$ denotes a polynomial function and θ_{α_c} represents the probability vector for the latent class α_c, which can be further written as the difference of two CDFs. A polynomial function is a real analytic function. Since the CDF is an integral of a real analytic function, the composition of real analytic functions (difference between two CDFs) is still a real analytic function. Furthermore, $h(\theta)$ is also a real analytic function of B. $G_D(B)$, as a determinant of $h(\theta)'h(\theta)$, is also a real analytic function.

Proof 11 For H_3, if D_3 takes the form of D^* in [S2], we can infer that there must exist an item j_0 in D_3 that for any two latent classes c_1 and c_2, we have $\mu_{j_0,c_1} \neq \mu_{j_0,c_2}$. Then we have at least two rows in matrix $H(D_3, B_3)$ to be independent of each other, implying that the Kruskal rank of $H(D_3, B_3)$ is at least 2. The exceptional case could exist when $\beta_{k,m} = 0$ holds for some k and m, which has Lebesgue measure zero with respect to Ω_{Δ^*}. Consequently, $\text{rank}(H_3) \geq 2$ holds almost everywhere in Ω_{Δ^*}

With Proofs 6 and 5, we have $\text{rank}(H_1) + \text{rank}(H_2) + \text{rank}(H_3) \geq 2M^K + 2$ holds almost everywhere in $\Omega_\Delta(\pi, B)$.

References

Allman, E. S., Matias, C., Rhodes, J. A., et al. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics, 37*(6), 3099–3132. https://doi.org/10.1214/09-AOS689

Bolt, D. M., & Kim, J.-S. (2018). Parameter invariance and skill attribute continuity in the DINA model. *Journal of Educational Measurement, 55*(2), 264–280. https://doi.org/10.1111/jedm.12175

Chen, Y., Culpepper, S., & Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika, 85*(1), 121–153. https://doi.org/10.1007/s11336-019-09693-2

Chen, Y., & Culpepper, S. A. (2020). A multivariate probit model for learning trajectories: A fine-grained evaluation of an educational intervention. *Applied Psychological Measurement, 44*(7–8), 515–530. https://doi.org/10.1177/0146621620920928

Chen, Y., Liu, Y., Culpepper, S. A., & Chen, Y. (2021). Inferring the number of attributes for the exploratory DINA model. *Psychometrika, 86*(1), 30–64. https://doi.org/10.1007/s11336-021-09750-9

Cho, S. H. (2016). *An application of diagnostic modeling to a situational judgment test assessing emotional intelligence* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Culpepper, S. A. (2019). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation. *Psychometrika, 84*(4), 921–940. https://doi.org/10.1007/s11336-019-09683-4

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. https://doi.org/10.3102/1076998618791306

De La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115–130. https://doi.org/10.3102/1076998607309474

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

De La Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333–353. https://doi.org/10.1007/BF02295640

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In R. L. B. P. D. Nichols & S. F. Chipman (Eds.), *Cognitively diagnostic assessment* (pp. 361–390). Routledge. https://doi.org/10.4324/9780203052969

Furnham, A., Richards, S., Rangel, L., & Jones, D. N. (2014). Measuring malevolence: Quantitative issues surrounding the dark triad of personality. *Personality and Individual Differences, 67*, 114–121. https://doi.org/10.1016/j.paid.2014.02.001

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass, 7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Garcia, D., & Rosenberg, P. (2016). *The dark cube: dark and light character profiles* (Vol. 4). PeerJ Inc. https://doi.org/10.7717/peerj.1675

Glenn, A. L., & Sellbom, M. (2015). Theoretical and empirical concerns regarding the dark triad as a construct. *Journal of Personality Disorders, 29*(3), 360–377. https://doi.org/10.1521/pedi_2014_28_162

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), i–25. https://doi.org/10.1002/j.2333-8504.2008.tb02131

Hagenaars, J. A. (1993). *Loglinear models with latent variables* (No. 94). Sage Publications, Inc. https://dx.doi.org/10.4135/9781412984850

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

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3) a brief measure of dark personality traits. *Assessment, 21*(1), 28–41. https://doi.org/10.1177/1073191113514105

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. https://doi.org/10.1177/01466210122032064

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

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. https://doi.org/10.1007/BF02293554

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. https://doi.org/10.1016/0024-3795(77)90069-6

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*(2), 187–212. https://psycnet.apa.org/doi/10.1007/BF02294535

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov chain monte carlo in R. *Journal of Statistical Software, 42*(9), 1–21. https://doi.org/10.18637/jss.v042.i09

McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology, 74*(1), 192. https://doi.org/10.1037/0022-3514.74.1.192

Narasimhan, B., Johnson, S. G., Hahn, T., Bouvier, A., & Kiêu, K. (2020). cubature: Adaptive multivariate integration over hypercubes [Computer software manual]. (R package version 2.0.4.1)

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Persson, B. N., Kajonius, P. J., & Garcia, D. (2019). Revisiting the structure of the short dark triad. *Assessment, 26*(1), 3–16. https://doi.org/10.1177/1073191117701192

Tatsuoka, K. K. (1987). Toward an integration of item-response theory and cognitive error diagnosis. In *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9780203056899

Templin, J. (2004). *Generalized linear mixed proficiency models for cognitive diagnosis.* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification, 30*(2), 251–275. http://dx.doi.org/10.1007/s00357-013-9129-4

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

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287–305. https://doi.org/10.1037/1082-989x.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*(7), 559–574. https://doi.org/10.1177/0146621607300286

Vehtari, A., & Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation, 14*(10), 2439–2468. https://doi.org/10.1162/08997660260293292

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series, 2005*(2), i–35. https://doi.org/10.1348/000711007x193957

von Davier, M. (2018). Diagnosing diagnostic models: From von neumann's elephant to model equivalencies and network psychometrics. *Measurement: Interdisciplinary Research and Perspectives, 16*(1), 59–70. https://doi.org/10.1080/15366367.2018.1436827

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics, 45*(2), 675–707. http://www.jstor.org/stable/44245820

Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association, 113*(523), 1284–1295. https://doi.org/10.1080/01621459.2017.1340889