PSYCHOMETRIKA—VOL. 84, NO. 2, 333–357
JUNE 2019
https://doi.org/10.1007/s11336-018-9643-8

ESTIMATING THE COGNITIVE DIAGNOSIS $\boldsymbol{Q}$ MATRIX WITH EXPERT KNOWLEDGE: APPLICATION TO THE FRACTION-SUBTRACTION DATASET

STEVEN ANDREW CULPEPPER

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Cognitive diagnosis models (CDMs) are an important psychometric framework for classifying students in terms of attribute and/or skill mastery. The $\boldsymbol{Q}$ matrix, which specifies the required attributes for each item, is central to implementing CDMs. The general unavailability of $\boldsymbol{Q}$ for most content areas and datasets poses a barrier to widespread applications of CDMs, and recent research accordingly developed fully exploratory methods to estimate Q. However, current methods do not always offer clear interpretations of the uncovered skills and existing exploratory methods do not use expert knowledge to estimate Q. We consider Bayesian estimation of $\boldsymbol{Q}$ using a prior based upon expert knowledge using a fully Bayesian formulation for a general diagnostic model. The developed method can be used to validate which of the underlying attributes are predicted by experts and to identify residual attributes that remain unexplained by expert knowledge. We report Monte Carlo evidence about the accuracy of selecting active expert-predictors and present an application using Tatsuoka's fraction-subtraction dataset.

Key words: exploratory cognitive diagnosis models, general diagnostic model, Bayesian, multivariate regression, variable selection, validation, spike–slab priors.

1. Introduction

Cognitive diagnosis models (CDMs) offer a useful psychometric framework for describing how underlying cognitive processes relate to performance on educational tasks. The application of CDMs is predicated upon the availability of knowledge about the underlying structure pertaining to which attributes are needed for each item. The catalogue of skills for each item is referred to as the $J \times K$ $\boldsymbol{Q}$ matrix with element (j, k) denoted by $q_{jk} \in \{0, 1\}$. For instance, Table 1 presents a hypothetical $\boldsymbol{Q}$ matrix with three skills and three items from (Klein, Birenbaum, Standiford, & Tatsuoka 1981). For the first item 3/4–2/4, Table 1 shows that $q_{11} = 0$, $q_{12} = 1$, and $q_{13} = 0$, which indicates that item 1 only requires subtraction of the numerators to obtain the correct response. In contrast, $q_{21} = 1$, $q_{22} = 1$, and $q_{23} = 0$ imply that the first two skills (i.e., finding a common denominator and subtracting the numerators) are needed to solve the second item 8/5–5/6. The third item $4\frac{4}{9}$–$3\frac{5}{6}$ requires all three skills and $q_{31} = 1$, $q_{32} = 1$, and $q_{33} = 1$.

Given the availability of cognitive theory and a $\boldsymbol{Q}$ matrix, CDMs provide a framework to diagnose and classify student i (for $i = 1, \ldots, N$) as a master or non-master of a collection of K binary attributes or skills denoted as $\boldsymbol{\alpha}_i \in \{0, 1\}^K$. Correctly specifying $\boldsymbol{Q}$ is essential for accurate diagnoses (Henson & Templin, 2007; Rupp & Templin, 2008) and prior research primarily relied on expert knowledge to define $\boldsymbol{Q}$. However, cognitive theory may be underdeveloped for experts to form a consensus, so the general unavailability of $\boldsymbol{Q}$ poses a barrier to widespread application of CDMs.

In order to address the fundamental problem of $\boldsymbol{Q}$ matrix specification, researchers developed exploratory CDMs to estimate $\boldsymbol{Q}$ and to infer the underlying cognitive processes (i.e., conjunctive, disjunctive, or compensatory; see de la Torre & Douglas, 2004; Maris, 1999). For instance, recent

Correspondence should be made to Steven Andrew Culpepper, Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL61820, USA. Email: sculpepp@illinois.edu

TABLE 1.
Hypothetical $\boldsymbol{Q}$ matrix cataloguing skills required by three fraction-subtraction items.

	Item	Find common denominator	Subtract numerator	Subtract whole number
1.	$\frac{3}{4}-\frac{2}{4}$	0	1	0
2.	$\frac{8}{5}-\frac{5}{6}$	1	1	0
3.	$4\frac{4}{9}-3\frac{5}{6}$	1	1	1

research supported broader applications of CDMs by: (1) proving model parameter identifiability conditions (Chen, Liu, Xu, & Ying, 2015; Xu, 2017; Xu & Shang, 2018); (2) validating the plausibility of a pre-specified expert $\boldsymbol{Q}$ matrix (Chiu, 2013; de la Torre, 2008; de la Torre & Chiu, 2016); (3) estimating a subset of $\boldsymbol{Q}$ (DeCarlo, 2012; Templin & Henson, 2006); and (4) developing exploratory methods that estimate $\boldsymbol{Q}$ as opposed to the confirmatory approach of using a pre-specified expert $\boldsymbol{Q}$ (Chen et al., 2015; Chen, Culpepper, Chen, & Douglas, 2018; Chung, 2014; Culpepper & Chen, 2018; Liu, Xu, & Ying, 2012, 2013; Xu & Shang 2018).

Expert knowledge may be available regarding item characteristics (e.g., a provisional $\boldsymbol{Q}$ matrix), and there are several reasons to use prior knowledge to estimate $\boldsymbol{Q}$. First, it may be that ignoring expert knowledge when estimating $\boldsymbol{Q}$ is sub-optimal. That is, there are, in general, 2^{JK} possible $\boldsymbol{Q}$ matrices and searching for the optimal matrix is computationally challenging for larger J or K. For instance, our application dataset includes $J = 20$ items and we found evidence for $K = 7$ attributes. In this case, searching for the maximum likelihood estimator (MLE) requires checking approximately 1.39×10^{42} different $\boldsymbol{Q}$ matrices. Clearly, estimating $\boldsymbol{Q}$ is formidable and alternatives are needed. The direction we pursue is to use Bayesian inference to stochastically search for $\boldsymbol{Q}$. Furthermore, using expert knowledge might help to narrow the search further (e.g., see methods of Liu et al., 2012). We demonstrate our methods accurately recover $\boldsymbol{Q}$ for a general diagnostic model in a fraction of the time required to find the MLE.

Second, incorporating expert knowledge in the statistical model may enhance interpretation of uncovered attributes. Similar to factor analysis, exploratory CDMs do not always offer a clear interpretation of the uncovered attributes. Furthermore, a clear interpretation of $\boldsymbol{Q}$ is necessary for explaining model results to practitioners or offering test developers guidance to create new items. Third, explicitly modeling the relationship between $\boldsymbol{Q}$ and expert knowledge may assist with cognitive theory development. That is, using an exploratory method with expert knowledge may help to identify *residual*, or unexplained, attributes that are neither predicted by cognitive theory nor pre-specified in a provisional $\boldsymbol{Q}$ matrix. In such cases, exploratory CDM results can be shared with experts and subsequent conversations may shed light on ways to update cognitive theories about the attributes students use to succeed on tasks.

Existing exploratory CDMs do not explicitly incorporate expert knowledge in statistical models beyond using an expert generated $\boldsymbol{Q}$ matrix as a provisional estimate to update or validate. Accordingly, we consider an exploratory CDM framework that directly uses expert knowledge about item features. Certainly, there are many strategies for estimating $\boldsymbol{Q}$ with expert knowledge and in this paper we consider a new prior for $\boldsymbol{Q}$ that incorporates expert knowledge as predictors within a latent, multivariate regression model. As noted in the next section, prior research successfully employed a uniform prior for $\boldsymbol{Q}$ on the space of identified models (Chen et al., 2018), but one possible short-coming is that the uncovered attributes were not easily interpreted. Additionally, the support under the uniform prior may be unnecessarily too large and the availability of expert knowledge could improve the performance of stochastic search algorithms. We examine a method that draws upon the tradition in other exploratory multivariate methods where the latent space underlying a set of variables is projected onto external variables to improve interpretation (e.g., see examples involving multidimensional scaling).

Readers may find similarities between the proposed strategy for selecting features of expert knowledge and $\boldsymbol{Q}$ matrix validation procedures (e.g., de la Torre & Chiu, 2016). Validation procedures first initialize $\boldsymbol{Q}$ with a provisional estimate, possibly from experts, and then use an algorithm to update elements of the provisional $\boldsymbol{Q}$ based upon an index or loss function. Validation procedures essentially predict $\boldsymbol{Q}$ with a provisional matrix, but they are not designed to conduct inference (also, validation procedures may not be consistent as noted by Liu, 2017). Instead, the framework developed in this paper is designed to infer which aspects of expert knowledge are represented in the underlying structure of the data using a fully exploratory CDM framework. That is, we freely estimate $\boldsymbol{Q}$ using a prior that incorporates expert knowledge in the form of a provisional $\boldsymbol{Q}$ matrix. We then use a hierarchical multivariate regression model to relate the provisional $\boldsymbol{Q}$ matrix to the matrix estimated from the underlying structure. Again, the advantage of embedding an expert knowledge validation procedure within an exploratory framework is that we may uncover new, unanticipated (i.e., residual) attributes, which might advance cognitive theory.

The remaining discussion is divided into five sections. In order for the developed methods to be broadly applicable, a flexible measurement model is needed. Prior research on the development of exploratory general diagnostic models (GDMs)[1] uses regularization within the frequentist framework to estimate $\boldsymbol{Q}$ (e.g., see Chen et al., 2015; Xu & Shang, 2018). In the first section, we offer a novel fully Bayesian hierarchical approach for the GDM. The second section introduces the new prior for incorporating expert knowledge in the form of a provisional $\boldsymbol{Q}$ matrix into the exploratory GDM framework. The third section reports results from a Monte Carlo simulation study using the GDM framework to evaluate the relative accuracy of $\boldsymbol{Q}$ matrix estimation, expert knowledge validation, and attribute classification. In particular, the Monte Carlo study generates data with a GDM and assesses the impact of sample size, attribute dependence, and priors on the estimation of model parameters. Note "Appendix A" reports Monte Carlo estimates of bias and expected absolute deviation for the GDM item and structural parameters. Readers are directed to "Appendix B" for additional Monte Carlo simulation study results concerning the impact of the specification of K, attribute dependence, and expert-predictor correlation using a more parsimonious deterministic inputs, noisy "and" gate model (DINA; Culpepper, 2015; Haertel, 1989; Junker & Sijtsma, 2001) generating model. The fourth section reports results from an application of the developed methodology to the fraction-subtraction (FS) data (Tatsuoka, 1984, 1990; Tatsuoka, 2002). Specifically, we use seven of the eight mental operations for the FS items reported in de la Torre and Douglas 2004 as expert-predictors to estimate $\boldsymbol{Q}$. The final section discusses the implications of the developments and results for future research.

2. An Exploratory General Diagnostic Model for Binary Data

Suppose that there are $i = 1, \ldots, N$ individuals and $j = 1, \ldots, J$ items, with $k = 1, \ldots, K$ attributes for each respondent. Let $\boldsymbol{Y}$ be a random $N \times J$ matrix of binary responses with y_{ij} equal to one if individual i correctly responds to item j and zero otherwise. Let row i of $\boldsymbol{Y}$ be $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})$ and denote column j as $\boldsymbol{Y}_j$. Also, the random attribute vector is $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})'$ where α_{ik} takes the value one when individual i possesses attribute k and the value zero otherwise. The 2^K latent class probability vector is $\boldsymbol{\pi}$ such that element $c = 0, \ldots, 2^K - 1$ indicates the probability an individual is a member of class c, i.e., $P(\boldsymbol{\alpha}_i'\boldsymbol{b} = c) = \pi_c$, where $\boldsymbol{b} = (2^{K-1}, \ldots, 1)'$ is a vector used to map $\boldsymbol{\alpha}_i$ to integers between 0 and $2^K - 1$.

[1] Note we use GDM to refer to the general model for binary responses described in the literature by de la Torre (2011), Henson, Templin, and Willse (2009), von Davier (2008).

We let $\boldsymbol{Q}_k$ denote column k of the random $\boldsymbol{Q}$, $\boldsymbol{q}_j$ denotes row j, and q_{jk} indicates the (j, k) element.

CDMs assume that observed responses for individual i are independent when conditioned upon the latent attributes. In general, the likelihood for individual i is,

$$p\left(\boldsymbol{y}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}, \boldsymbol{Q}\right) = \sum_{c=0}^{2^K-1} \pi_c \prod_{j=1}^{J} P(y_{ij} = 1|\boldsymbol{\alpha}_i'\boldsymbol{b} = c, \boldsymbol{\theta}_j, \boldsymbol{q}_j)^{y_{ij}} P(y_{ij} = 0|\boldsymbol{\alpha}_i'\boldsymbol{b} = c, \boldsymbol{\theta}_j, \boldsymbol{q}_j)^{1-y_{ij}} \tag{1}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)'$ is a $J \times 2^K$ matrix of latent class item parameters with the 2^K item j parameters denoted as $\boldsymbol{\theta}_j$. There are many CDMs available to define $\boldsymbol{\Theta}$. For instance, more parsimonious models for $\boldsymbol{\Theta}$ include the noisy inputs, deterministic, "and" gate (NIDA; Maris, 1999), the DINA, and the reduced reparameterized unified model (rRUM; Culpepper & Hudson, 2018; Hartz, 2002). In contrast, more general CDMs for $\boldsymbol{\Theta}$ include the generalized DINA (GDINA; de la Torre, 2011), the log-linear cognitive diagnosis model (LCDM; Henson et al., 2009), the general diagnostic model (GDM; von Davier, 2008), and the restricted latent class model (RLCM; Xu, 2017). Several frameworks exist for modeling $\boldsymbol{\pi}$, as well. That is, the 2^K elements of $\boldsymbol{\pi}$ may be estimated in an unstructured fashion. Alternatively, a higher-order model (de la Torre & Douglas, 2004; Maris, 1999) or a multivariate probit model (Henson et al., 2009; Templin, Henson, Templin, & Roussos, 2008) may be preferred if dependence among attributes is described by a more parsimonious model.

The remainder of this section presents a Bayesian model formulation for an exploratory GDM for binary data for the case where there is no expert knowledge available about the underlying structure. We first present the model formulation and then discuss posterior inference.

2.1. Bayesian Formulation for the GDM

Prior research (e.g., see de la Torre, 2011; Henson et al., 2009; von Davier, 2008; Xu, 2017) wrote the GDM as a generalized linear model with latent attribute main-effects and interaction terms. Similar to prior research our formulation for the GDM uses the following item response function (IRF) for individual i on item j,

$$P\left(Y_{ij} = 1 \left|\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j\right.\right) = \Phi\left(\beta_{0,j} + \sum_{k=1}^{K} \beta_{k,j}\alpha_{ik} + \sum_{k>k'}\sum \beta_{kk',j}\alpha_{ik}\alpha_{ik'} + \cdots + \beta_{12\cdots K,j}\prod_{k=1}^{K}\alpha_{ik}\right) \tag{2}$$

where $\Phi(\cdot)$ is a standard normal cumulative distribution function and the linear model includes all main-effects and higher-order interactions of the attributes. To simplify notation, we let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)'$ be a $N \times 2^K$ design matrix where $\boldsymbol{a}_i$ indicates the attribute profile for individual i so that the IRF can be concisely written as $P\left(Y_{ij} = 1 \left|\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j\right.\right) = \Phi\left(\boldsymbol{a}_i'\boldsymbol{\beta}_j\right)$. Also, we index the elements of $\boldsymbol{\beta}_j$ as $\{\beta_{jp}\}_{p=0}^{2^K-1}$.

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 link GDM,

$$Y_{ij} = \mathcal{I}(Y_{ij}^* > 0) \tag{3}$$

$$Y_{ij}^* \left|\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j\right. \sim \mathcal{N}(\boldsymbol{a}_i'\boldsymbol{\beta}_j, 1) \tag{4}$$

where $\mathcal{I}(\cdot)$ is the indicator function specifying a deterministic relationship between the observed binary response and a normally distributed, augmented latent variable Y_{ij}^*.

The next level of the model specifies distributions for $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$. In this paper, we use a categorical prior for $\boldsymbol{\alpha}_i$ where $\boldsymbol{\pi}$ is the 2^K vector of class probabilities (e.g., see Culpepper, 2015), but it is important to note the availability of fully conjugate Bayesian formulations for higher-order (e.g., see Culpepper & Chen, 2018) and multivariate probit (Chen & Culpepper, 2018) models in cases where the attribute structure satisfies a more parsimonious model.

We use a novel prior for β_{jp}, which builds upon literature pertaining to stochastic search variable selection (SSVS; e.g., see O'Hara & Sillanpää 2009). Let $\mathcal{B}_k$ be the set of indices for elements of $\boldsymbol{\beta}_j$ that correspond to the kth attribute. For instance, for $K = 3$, $\mathcal{B}_1 = \{1, 4, 5, 7\}$ for the design vector $\boldsymbol{a} = (1, \alpha_1, \alpha_2, \alpha_3, \alpha_1\alpha_2, \alpha_1\alpha_3, \alpha_2\alpha_3, \alpha_1\alpha_2\alpha_3)'$ where the first element is a one for the intercept. Additionally, let $\tilde{\boldsymbol{q}}_j'$ be a 2^K vector that includes all possible products of the elements of $\boldsymbol{q}_j$. For instance, for $K = 3$, $\boldsymbol{q}_j = (q_{j1}, q_{j2}, q_{j3})'$ and $\tilde{\boldsymbol{q}}_j = (1, q_{j1}, q_{j2}, q_{j3}, q_{j1}q_{j2}, q_{j1}q_{j3}, q_{j2}q_{j3}, q_{j1}q_{j2}q_{j3})'$. The prior for $\boldsymbol{\beta}_j$ given $\boldsymbol{q}_j$ is,

$$p\left(\boldsymbol{\beta}_j|\boldsymbol{q}_j\right) \propto \left[\prod_{p=0}^{2^K-1} \upsilon_{\text{p}}^{-1/2} \exp\left(-\frac{1}{2}\beta_{jp}^2/\upsilon_{\text{p}}\right)\right] \prod_{p=1}^{K} \mathcal{I}\left(\beta_{jp} > 0\right) \tag{5}$$

$$\upsilon_{\text{p}} = \tilde{q}_{jp}/\omega_1 + (1 - \tilde{q}_{jp})/\omega_0 \tag{6}$$

where Eq. 5 is a product of normal densities (i.e., the coefficients are conditionally independent given $\boldsymbol{q}_j$) with the scale parameter defined as a weighted linear combination of constants $1/\omega_0$ and $1/\omega_1$. Also, note we restrict the main-effect for attribute k to be nonnegative to encourage monotonicity, so the class with no attributes has the lowest response probability. The prior for the 2^{K-1} vector of coefficients, $\boldsymbol{\beta}_{jk}$, given $\boldsymbol{q}_j$ includes only those coefficients in Eq. 5 that involve the main-effects and interactions for α_{ik} (i.e., $p \in \mathcal{B}_k$).

The relationship between $\boldsymbol{q}_j$ and the activeness of each β_{jp} is dictated by the weights of the scale parameter, υ_{p}. For instance, a large value for ω_0 corresponds to a normal distribution with a small variance (e.g., a "spike") for inactive coefficients and a smaller value for ω_1 translates to a distribution with a larger variance (i.e., a "slab") for active coefficients. Consequently, β_{jp} is considered active in cases where $\tilde{q}_{jp} = 1$, so the scale parameter is the larger value of $\upsilon_{\text{p}} = 1/\omega_1$. In contrast, $\tilde{q}_{jp} = 0$ denotes that at least one associated q_{jk} is zero, which is consistent with an inactive coefficient that has a smaller prior variance of $1/\omega_0$.

Note in the definition for $\tilde{\boldsymbol{q}}_j$ that the first element is one for the intercept to indicate that we always consider the intercept active. Also, the intercept is the parameter for the class without attributes, which is generally assumed to have the lowest response probability. A negative-valued location parameter can be added to the prior for β_{j0} to reflect this belief a priori.

In this section, we consider the case where expert knowledge is unavailable and we assume the elements of $\boldsymbol{Q}$ are conditionally independent given a hyper-parameter ν. Specifically, each q_{jk} is conditionally distributed as a Bernoulli, i.e., $q_{jk}|\nu \sim \text{Bernoulli}(\nu)$, and we employ a conjugate Beta prior for ν as $\nu \sim \text{Beta}(a, b)$.

2.2. *Posterior Inference for GDM*

The aforementioned formulation for the GDM offers a conjugate Metropolis-within-Gibbs sampler to update the model parameters. We next outline full conditional distributions for the augmented data, attributes, class probabilities, item parameters, and $\boldsymbol{Q}$.

Similar to prior research on Bayesian IRT methods, the augmented data, Y_{ij}^*, are conditionally a truncated normal distribution. Specifically, the full conditional for the augmented data is $Y_{ij}^*|Y_{ij}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j \sim \mathcal{N}(\boldsymbol{a}_i'\boldsymbol{\beta}_j, 1)\mathcal{I}(Y_{ij}^* > 0)^{Y_{ij}}\mathcal{I}(Y_{ij}^* \leq 0)^{1-Y_{ij}}$.

The full conditional distribution for $\boldsymbol{\alpha}_i$ is multinomial with class probabilities defined as,

$$P\left(\boldsymbol{\alpha}_i'\boldsymbol{b}=c|Y_{i1}^*,\ldots,Y_{iJ}^*,\boldsymbol{B},\boldsymbol{\pi}\right)=\frac{\exp\left(-\frac{1}{2}\sum_{j=1}^J(Y_{ij}^*-\boldsymbol{a}_c\boldsymbol{\beta}_j)^2\right)\pi_c}{\sum_{c=0}^{2^K-1}\exp\left(-\frac{1}{2}\sum_{j=1}^J(Y_{ij}^*-\boldsymbol{a}_c\boldsymbol{\beta}_j)^2\right)\pi_c} \tag{7}$$

where $\boldsymbol{B}=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_J)'$ denotes the $J\times 2^K$ matrix of item parameters and $\boldsymbol{a}_c$ is the design vector associated with class c. The full conditional distribution for $\boldsymbol{\pi}$ is a Dirichlet as noted in Culpepper 2015.

Let $\boldsymbol{Y}_j^*=(Y_{1j}^*,\ldots,Y_{Nj}^*)'$, $\boldsymbol{A}_p$ be the pth column of $\boldsymbol{A}$, $\boldsymbol{A}_{(p)}$ a matrix with all of $\boldsymbol{A}$, but the pth column, and $\boldsymbol{\beta}_{j(p)}$ is the 2^K-1 vector that omits β_{jp}. The full conditional distribution for β_{jp} is

$$\beta_{jp}|\boldsymbol{Y}_j^*,\boldsymbol{\alpha},\boldsymbol{\beta}_{j(p)},\boldsymbol{q}_j\sim\begin{cases}\mathcal{N}(\mu_{jp},\sigma_{jp}^2)\mathcal{I}(\beta_{jp}>0) & 1\le p\le K\\ \mathcal{N}(\mu_{jp},\sigma_{jp}^2), & \text{otherwise}\end{cases} \tag{8}$$
$$\sigma_{jp}^2=\frac{1}{\boldsymbol{A}_p'\boldsymbol{A}_p+v_p}$$
$$\mu_{jp}=\sigma_{jp}^2\boldsymbol{A}_p'\left(\boldsymbol{Y}_j^*-\boldsymbol{A}_{(p)}\boldsymbol{\beta}_{j(p)}\right)$$

where the condition $1\le p\le K$ corresponds to a non-negativity constraint for the main-effects. Note that the intercept β_{j0} corresponds to $p=0$ and is always considered active given the expectation that the class without attributes has the lowest response probability.

Finally, we update q_{jk} with either a Gibbs or a Metropolis–Hastings (MH) step. Our preliminary simulation evidence suggests that the MH sampler outperforms Gibbs in terms of recovery of $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ for the GDM, so we discuss the MH sampler. Suppose $q_{jk}^{(t-1)}$ is the current value of q_{jk} in the chain. The MH sampler proceeds by proposing a candidate defined as $q'=1-q_{jk}^{(t-1)}$, which indicates that we propose to change the value of q_{jk} each iteration with probability one. Let $T(q',q_{jk}^{(t-1)})$ be,

$$T(q',q_{jk}^{(t-1)})=\left(\frac{p(\boldsymbol{\beta}_{jk}|\boldsymbol{q}_j,q_{jk}=0)(1-\nu)}{p(\boldsymbol{\beta}_{jk}|\boldsymbol{q}_j,q_{jk}=1)\nu}\right)^{(q_{jk}^{(t-1)}-q')}. \tag{9}$$

where $\boldsymbol{\beta}_{jk}=\left\{\beta_{jp}:p\in\mathcal{B}_k\right\}$. Therefore, the MH decision rule is,

$$\begin{aligned}q_{jk}^{(t)}&=q',\ \min\left(1,T(q',q_{jk}^{(t-1)})\right)>U\\ q_{jk}^{(t)}&=q_{jk}^{(t-1)},\ \text{otherwise}\end{aligned} \tag{10}$$

where U is a uniform random variable. Note the hyper-parameter ν is conditionally a Beta distribution, such that $\nu|\boldsymbol{Q}\sim\text{Beta}(\sum_{j=1}^J\sum_{k=1}^K q_{jk}+a, JK-\sum_{j=1}^J\sum_{k=1}^K q_{jk}+b)$.

3. Expert Knowledge Model for $\boldsymbol{Q}$

Cognitive diagnosis models describe the underlying process by which students respond to items. The $\boldsymbol{Q}$ matrix is an essential feature of all CDMs, and this section summarizes previous Bayesian research on $\boldsymbol{Q}$ estimation and proposes a new modeling framework to incorporate expert knowledge. Specifically, we first discuss the uniform prior of Chen et al. (2018) and then introduce a statistical framework for selecting which features of expert knowledge actively relate to the underlying $\boldsymbol{Q}$.

3.1. Uniform Prior for $\boldsymbol{Q}$

Chen et al. (2018) extended the Bayesian formulation of the DINA model (e.g., Culpepper, 2015) with the following uniform prior for $\boldsymbol{Q}$,

$$p(\boldsymbol{Q}) \propto \mathcal{I}(\boldsymbol{Q} \in \mathcal{Q}) \tag{11}$$

where $\mathcal{Q}$ is the space of identified DINA $\boldsymbol{Q}$ matrices (e.g., for necessary and sufficient conditions see Gu & Xu, 2018). Chen et al. (2018) describe two strategies for sampling $\boldsymbol{Q}$ from the posterior distribution with the uniform prior in Eq. 11. One approach used a Metropolis–Hastings (MH) within Gibbs sampler to randomly update selected blocks of $\boldsymbol{Q}$ by drawing candidates from $\mathcal{Q}$. A second approach implemented a constrained Gibbs sampler to sequentially update each q_{jk} from a full conditional distribution. Overall, Chen et al. (2018) reported accurate recovery of $\boldsymbol{Q}$ with the uniform prior.

3.2. Content-Expert Prior for $\boldsymbol{Q}$

The uniform prior in Eq. 11 is not without its short-comings. Specifically, the uniform prior is just that it considers all identified $\boldsymbol{Q}$ matrices as equally plausible. An alternative we explore in this subsection incorporates expert knowledge into the prior for $\boldsymbol{Q}$. We consider elements of $\boldsymbol{Q}$ as being independently distributed as Bernoulli random variables, and we include expert knowledge as predictors of elements of $\boldsymbol{Q}$ using a multivariate regression model. That is, expert knowledge about items is treated as "expert-predictors." Let x_{jv} be the value of "expert-predictor" v for item j, $\boldsymbol{x}_j = (1, x_{j1}, \dots, x_{jV})'$ is a vector of predictors for item j, and $\boldsymbol{X}_v = (x_{1v}, \dots, x_{Jv})'$ is expert-predictor v. The $J \times (V+1)$ design matrix is $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_J)' = (\boldsymbol{1}_J, \boldsymbol{X}_1, \dots, \boldsymbol{X}_V)$. Furthermore, let γ_{vk} be a regression coefficient that relates $\boldsymbol{X}_v$ to attribute k, $\boldsymbol{Q}_k$. The $(V+1) \times K$ matrix of multivariate regression coefficients is $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_K)$ with column k defined as $\boldsymbol{\gamma}_k = (\gamma_{0k}, \gamma_{1k}, \dots, \gamma_{Vk})'$.

We relate $\boldsymbol{Q}$ to $\boldsymbol{X}$ in the prior, $p(\boldsymbol{Q} | \boldsymbol{\Gamma})$. The prior for q_{jk} is,

$$q_{jk} | \boldsymbol{\gamma}_k \sim \text{Bernoulli}\left[\Phi\left(\boldsymbol{x}_j' \boldsymbol{\gamma}_k\right)\right] \tag{12}$$

where a probit regression model relates the expert-predictors to q_{jk}. The assumption of conditional independence of the elements of $\boldsymbol{Q}$ given $\boldsymbol{\Gamma}$ implies the prior for $\boldsymbol{Q}$ is,

$$p(\boldsymbol{Q} | \boldsymbol{\Gamma}) = \prod_{j=1}^{J} \prod_{k=1}^{K} \left[\Phi\left(\boldsymbol{x}_j' \boldsymbol{\gamma}_k\right)\right]^{q_{jk}} \left[1 - \Phi\left(\boldsymbol{x}_j' \boldsymbol{\gamma}_k\right)\right]^{1-q_{jk}}. \tag{13}$$

One feature to notice about the prior in Eq. 13 is that the support of $\boldsymbol{Q}$ is not restricted to $\mathcal{Q}$. An implication is that the model parameters are not strictly identified (e.g., see Xu, 2017), so there may be different $\boldsymbol{\Theta}$, $\boldsymbol{\pi}$, and $\boldsymbol{Q}$ that have equal likelihood. In this paper, we consider the Bayesian method, which relies upon the posterior distribution to infer $\boldsymbol{Q}$. Enforcing model identifiability conditions seems most important when using a uniform prior given that the posterior is proportional to the likelihood function and there may be many sets of equally likely parameters. In contrast, the posterior *may* be concentrated around a single mode even if $\boldsymbol{Q} \notin \mathcal{Q}$ with a well-chosen prior. The expert knowledge prior we consider is only uniform in cases where $\boldsymbol{\Gamma} = \mathbf{0}$; i.e., when expert knowledge fails to explain the underlying structure. Otherwise, when $\boldsymbol{\Gamma} \neq 0$ at least one of the expert-predictors is active and the posterior distribution may be concentrated around a single set of parameter values.

3.3. Bayesian Formulation for Expert $\boldsymbol{Q}$ *Model*

We focus on a Bayesian formulation for the expert-predictor prior for $\boldsymbol{Q}$ in Eq. 13. Specifically, we employ the probit data augmentation scheme (e.g., see Albert & Chib, 1993). In short, we introduce a continuous augmented variable q^*_{jk} for each element of $\boldsymbol{Q}$ and define the deterministic relationship $q_{jk} = \mathcal{I}(q^*_{jk} > 0)$. Next, we specify a regression model to relate q^*_{jk} to the vector of expert-predictors, $\boldsymbol{x}_j$, as, $q^*_{jk} | \boldsymbol{\gamma}_k \sim \mathcal{N}(\boldsymbol{x}'_j \boldsymbol{\gamma}_k, 1)$. More concisely, the Bayesian formulation for the prior in Eq. 13 is,

$$p(\boldsymbol{Q}|\boldsymbol{Q}^*) \propto \prod_{j=1}^{J}\prod_{k=1}^{K} \left[\mathcal{I}(q^*_{jk} > 0)\right]^{q_{jk}} \left[\mathcal{I}(q^*_{jk} \leq 0)\right]^{1-q_{jk}} \tag{14}$$

$$\boldsymbol{Q}^* | \boldsymbol{\Gamma} \sim \mathcal{MN}_{J\times K}\left(\boldsymbol{X}\boldsymbol{\Gamma}, \boldsymbol{I}_J, \boldsymbol{I}_K\right) \tag{15}$$

where $\boldsymbol{I}_J$ and $\boldsymbol{I}_K$ are J and K identity matrices and $\boldsymbol{Q}^*|\boldsymbol{\Gamma}$ is distributed as a matrix normal distribution with a mean matrix $\boldsymbol{X\Gamma}$ and independent rows and columns. The model for q^*_{jk} assumes a priori that the columns of $\boldsymbol{Q}^*$ are independent after controlling for $\boldsymbol{X}$. Note that the assumption of independent columns is consistent with model identifiability constraints (see Xu & Shang, 2018) where $\boldsymbol{Q} \in \mathcal{Q}$ include at least two identity matrices, which can be thought of reducing the association between columns.

Posterior inference proceeds by sampling each element of $\boldsymbol{Q}$ from the posterior distribution. We update the MH sampler for q_{jk} in Eq. 9 by replacing ν with $\Phi(\boldsymbol{x}'_j\boldsymbol{\gamma}_k)$ as the prior probability $q_{jk} = 1$ in $T\left(q', q^{(t-1)}_{jk}\right)$ as shown below:

$$T(q', q^{(t-1)}_{jk}) = \left(\frac{p(\boldsymbol{\beta}_{jk}|\boldsymbol{q}_j, q_{jk}=0)\left[1-\Phi(\boldsymbol{x}'_j\boldsymbol{\gamma}_k)\right]}{p(\boldsymbol{\beta}_{jk}|\boldsymbol{q}_j, q_{jk}=1)\Phi(\boldsymbol{x}'_j\boldsymbol{\gamma}_k)}\right)^{(q^{(t-1)}_{jk}-q')}. \tag{16}$$

3.4. Expert-Predictor Variable Selection Via a Spike–Slab Prior for γ_{vk}

The multivariate regression coefficients, $\boldsymbol{\Gamma}$, quantify the relationship between $\boldsymbol{Q}$ and expert-predictors, $\boldsymbol{X}$. In general, $\boldsymbol{\Gamma}$ can be freely estimated; however, the nature of the $\boldsymbol{Q}$ matrix suggests $\boldsymbol{\Gamma}$ should be sparse. That is, each column of $\boldsymbol{Q}$ corresponds to a different attribute and it is therefore reasonable to expect that a given expert-predictor relates to at most one column of $\boldsymbol{Q}$. This prior expectation suggests that $\boldsymbol{\Gamma}$ is sparse with a structured pattern. For instance, $(\gamma_{v1}, \ldots, \gamma_{vK})$ are the coefficients for predictor v across the K attributes. We expect that at most one of these γ's is nonzero (or active) and the remaining elements are zero (or inactive).

We consider a spike–slab prior (George & McCulloch, 1993; Ishwaran & Rao, 2005; O'Hara & Sillanpää, 2009) for elements of $\mathbf{\Gamma}$. Let δ_{vk} be one if γ_{vk} is active and $\delta_{vk} = 0$ if $\boldsymbol{X}_v$ is an inactive predictor for attribute k (i.e., $\gamma_{vk} = 0$). The conditional prior for γ_{vk} is a mixture,

$$p(\gamma_{vk}|\delta_{vk}, \omega^2) \propto \delta_{vk} f_1 + (1 - \delta_{vk}) f_0. \tag{17}$$

The "slab" measure, f_1, in Eq. 17 is selected to have a larger variance than the "spike" measure, f_0.

O'Hara and Sillanpää (2009) review several formulations for selecting variables in the Bayesian framework. The classic formulation uses a Dirac delta measure for the "spike" [i.e., $f_0 = \mathcal{I}(\gamma_{vk} = 0)$] and a normal measure for the slab [i.e., $f_1 \propto \exp\left(-\frac{1}{2}\omega^2\gamma_{vk}^2\right)$], which implies $\gamma_{vk} = 0$ and is inactive if $\delta_{vk} = 0$. Note that a conjugate prior for the "slab" precision parameter is $\omega^2 \sim Gamma(a, b)$. In this paper, we consider the SSVS method discussed in the previous section for the GDM and the Monte Carlo results in "Appendix B" employ the classic formulation with the DINA model.

The spike–slab prior in Eq. 17 incorporates sparsity into the estimation of $\mathbf{\Gamma}$, but does not explicitly impose the condition that at most one element in each row of $\mathbf{\Gamma}$ is nonzero. Instead, we can impose structured sparsity by using the following prior for the model selection parameters, $\boldsymbol{\delta}_v = (\delta_{v1}, \ldots, \delta_{vK})$,

$$P(\boldsymbol{\delta}_v = \boldsymbol{d}|\zeta) \propto \mathcal{I}(\boldsymbol{d} \in \{\boldsymbol{e}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_K\}) \prod_{k=1}^{K} \zeta^{d_k}(1-\zeta)^{1-d_k} \tag{18}$$

where $\boldsymbol{e}_k$ is a K vector with a one in element k and zeros elsewhere and $\boldsymbol{e}_0$ is defined as the zero vector $\mathbf{0}_K$. We let $\mathbf{\Delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_V)'$ be a $V \times K$ matrix of the binary predictor selection indicators. Also, a conjugate prior for the hyper-parameter is $\zeta \sim Beta(a, b)$. Note Gibbs sampling steps are available to draw γ_{vk}, δ_{vk}, and ζ from tractable full conditional distributions.

It is important to note the importance of matching the estimated K with the true K when enforcing the structured sparsity condition. Specifically, if too few attributes are estimated, we might find that a larger number of attributes are projected into a smaller space, which would lead to expert-predictors switching between active and inactive on more than one attribute. Certainly, such a finding would serve as evidence that the expert-predictor relates to the underlying structure, but the absence of simple structure would prevent a clear interpretation. One solution is to estimate several GDMs with a range of values of K and then compare models with fit indices such as the Deviance Information Criterion (DIC; Celeux, Forbes, Robert, & Titterington, 2006; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) to assist with selecting the optimal K.

Note the prior in Eq. 18 corresponds to the expert-predictor slope coefficients, but not the multivariate intercepts. We use conjugate independent normal prior distributions for the K intercepts.

3.5. Posterior Inference for Expert-Predictor Selection

The primary goal is to select which expert-predictors relate to uncovered attributes and which attributes are residual, or unexplained by $\boldsymbol{X}$. We infer which expert-predictors are active by summarizing the posterior of δ_{vk}. For instance, let $t = 1, \ldots, T$ index samples from the posterior and define $\bar{\delta}_{vk} = \frac{1}{T}\sum_{t=1}^{T} \delta_{vkt}$. One option is to define estimates as $\hat{\delta}_{vk} = \mathcal{I}\left(\bar{\delta}_{vk} > 0.5\right)$. Similarly, the estimated $\boldsymbol{Q}$ is derived from, $\overline{\boldsymbol{Q}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Q}_t$ and estimates for individual elements can be constructed with the posterior mode as $\widehat{q}_{jk} = \mathcal{I}\left(\bar{q}_{jk} > 0.5\right)$. Another option available with the MVN prior is to use $\boldsymbol{\gamma}_k$ and $\boldsymbol{x}_j$ to predict q_{jk} using the estimated probit model probability. We consider both strategies for estimating $\boldsymbol{Q}$ in the GDM Monte Carlo study discussed in the next section.

4. GDM Monte Carlo Simulation Study

4.1. Overview

A Monte Carlo study was conducted to assess the accuracy of GDM parameter recovery. We examined the impact of different priors, sample sizes, and the attribute structure on the accuracy of $\boldsymbol{Q}$ matrix estimation and expert-predictor validation. First, we compared the performance of the MVN prior in Eq. 12 versus a prior that does not relate $\boldsymbol{Q}$ to expert-predictors (i.e., $q_{jk}|\nu \sim \text{Bernoulli}(\nu)$). Second, we considered sample sizes of $N = 500$ and 1500 and we expect accuracy to improve with N. Third, attribute dependence was manipulated by generating attributes using the multivariate normal probit model described by Chiu, Douglas, and Li (2009) with the population tetrachoric correlation taking values of $\rho = 0$ or 0.5. The implied structural probabilities for the latent classes (000, 001, 010, 011, 100, 101, 110, 111) were (0.094, 0.031, 0.094, 0.031, 0.281, 0.094, 0.281, 0.094) for $\rho = 0$ and (0.181, 0.009, 0.051, 0.009, 0.259, 0.051, 0.259, 0.181) for $\rho = 0.5$.

We assessed model accuracy by examining recovery of $\boldsymbol{Q}$, model selection accuracy (i.e., $\boldsymbol{\Delta}$), in addition to item parameters (i.e., the latent class response probabilities $\boldsymbol{\Theta}$), latent structure $\boldsymbol{\pi}$, as well as classification accuracy for each $\boldsymbol{\alpha}_i$. We assess accuracy of estimating $\boldsymbol{Q}$ by documenting the matrix- and element-wise recovery rates $\boldsymbol{Q}$. Matrix-wise recovery rates are based upon whether the estimated $\boldsymbol{Q}$ is exactly recovered in all elements. For replication r the measure of matrix-wise accuracy is $\mathcal{I}(\widehat{\boldsymbol{Q}}^{(r)} = \boldsymbol{Q})$ and the recovery rate is $\frac{1}{R}\sum_{r=1}^{R}\mathcal{I}(\widehat{\boldsymbol{Q}}^{(r)} = \boldsymbol{Q})$ where R is the number of replications. In contrast, element-wise recovery rates are defined as the proportion of correctly estimated elements in $\boldsymbol{Q}$ and the accuracy rate is $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathcal{I}(\hat{q}_{jk}^{(r)} = q_{jk})$.

We considered two methods for defining estimators of q_{jk}. The first method uses the element-wise posterior mode of $\boldsymbol{Q}$ by defining $\hat{q}_{jk} = \mathcal{I}\left(\frac{1}{T}\sum_{t=1}^{T} q_{jk}^{(t)} > 0.5\right)$. The second approach, which is only available for the MVN prior, predicts q_{jk} with the expert-predictors. That is, the second method constructs estimates as $\hat{q}_{jk} = \mathcal{I}\left(\Phi\left(\boldsymbol{x}_j'\hat{\boldsymbol{\gamma}}_k\right) > 0.5\right)$.

We also assess model selection accuracy. Specifically, we set $K = 3$ for the true $\boldsymbol{Q}$ matrix (see "Appendix A") and randomly generated an additional three columns, which do not relate to the latent structure. Therefore, the expert-predictor matrix, $\boldsymbol{X}$, included a column of ones, three columns for the true $\boldsymbol{Q}$ matrix, and three randomly generated columns that did not relate to the underlying structure. Model selection accuracy was evaluated by computing the proportion of times the true expert-predictors were selected and the spurious predictors were not selected.

We estimated the bias and expected absolute deviation of $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$ in the Monte Carlo simulation. We followed Xu and Shang (2018) and defined elements of the true $\boldsymbol{\Theta}$ as $0.2 + (0.8 - 0.2) \times K_j'/K_j$, where K_j is the number of attributes required by item j and K_j' is the number of the K_j attributes a given class possesses. For instance, if $K = 3$ and $\boldsymbol{q}_j = (1, 1, 0)$, the $\boldsymbol{\alpha} = (0, 0, 0)'$ and $\boldsymbol{\alpha} = (0, 0, 1)'$ classes have correct response probabilities of 0.2, the $\boldsymbol{\alpha} = (0, 1, 0)'$, $\boldsymbol{\alpha} = (0, 1, 1)'$, $\boldsymbol{\alpha} = (1, 0, 0)'$, and $\boldsymbol{\alpha} = (1, 0, 1)'$ classes have a 0.5 chance, and the $\boldsymbol{\alpha} = (1, 1, 0)'$ and $\boldsymbol{\alpha} = (1, 1, 1)'$ classes have the highest correct response probability of 0.8.

We generated 100 replications of the four conditions (i.e., two sample sizes and two attribute structures) to estimate model accuracy. For each replication, we independently estimated a fully saturated, exploratory GDM with three attributes using the MVN prior and a model without the MVN prior using chain lengths of 80,000 and a burnin of 20,000. We wrote the MCMC algorithm in C++ and found that running a chain length of 80,000 with a sample size of 500 and $K = 3$ required approximately 587 s on a cluster with 2.50 GHz processor compute nodes. Note we fixed the "spike–slab" precision parameters across all conditions. Specifically, we estimated the models by fixing the "spike" precision parameters to $\omega_0 = 100$ for the main-effects and $\omega_0 = 500$ for the interaction-effects. In preliminary simulations, we found evidence of improved recovery when using different values for the "slab" scale parameter ω_1, for main- and interaction-effects. We set $\omega_1 = 0.1$ for the main-effects and $\omega_1 = 10$ for all interaction-effects. Finally, we fixed the precision parameter for the intercepts to 100 with a prior mean of -1 to reflect the prior belief that the $\boldsymbol{\alpha} = (0, 0, 0)'$ class has the lowest response probability.

4.2. Results

We next report Monte Carlo evidence concerning the accuracy of the proposed exploratory GDM. Specifically, we discuss the results for the accuracy of estimation for $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$, in addition to classification accuracy for $\boldsymbol{\alpha}$. Readers are directed to "Appendix A" for Monte Carlo results pertaining to the bias and mean absolute deviation for the Bayesian estimators of $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$.

Table 2 reports results concerning the accuracy of the exploratory GDM algorithm to recover $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ across sample size, attribute dependence, and prior distributions for $\boldsymbol{Q}$. A first observation is that the element-wise recovery rates of $\boldsymbol{Q}$ exceed 0.968 for all conditions, priors, and methods for estimating $\boldsymbol{Q}$. Table 2 also shows that recovery of $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ improves with larger N. For instance, the matrix-wise recovery rate for the Method 1 estimator, which is defined as $\hat{q}_{jk} = \mathcal{I}\left(\frac{1}{T}\sum_{t=1}^{T} q_{jk}^{(t)} > 0.5\right)$, increases from 0.25 for $N = 500$ to 0.85 for $N = 1500$. Additionally, attribute dependence has an effect on recovery, but to a smaller extent than N. For example, for $N = 500$ the matrix-wise recovery of $\boldsymbol{Q}$ using the MVN prior decreases from 0.25 for $\rho = 0$ to 0.19 for $\rho = 0.5$ and for $N = 1500$ recovery rates decrease from 0.85 to 0.72.

Table 2 provides evidence that Method 2 for estimating elements of $\boldsymbol{Q}$, which is defined as $\hat{q}_{jk} = \mathcal{I}\left(\Phi\left(\boldsymbol{x}'_j\hat{\boldsymbol{\gamma}}_k\right) > 0.5\right)$, outperforms using the element-wise posterior mode. The improvement from using Method 2 relative to Method 1 was greatest for $N = 500$ given increases from 0.25 to 0.98 for $\rho = 0$ and 0.19 to 0.92 for $\rho = 0.5$. For $N = 1500$, accuracy rates increased from 0.85 to 0.99 for $\rho = 0$ and 0.72 to 0.92 for $\rho = 0.5$. Furthermore, Table 2 shows that Method 1 performs similarly for both the expert MVN prior and the prior that does not use expert information (i.e., the last two columns where a priori $P\left(q_{jk} = 1|\nu\right) = \nu$).

The results in Table 2 provide evidence of accurate selection of expert-predictors in the multivariate regression. Specifically, the probability of selecting the true underlying predictors and of excluding the non-related predictors for the $N = 500$ case was 0.58 and 0.41 for $\rho = 0$ and

TABLE 2.
Summary of GDM expert validation and $\boldsymbol{Q}$ matrix estimation accuracy for $K = 3$ and $J = 20$ across sample size, N, attribute dependence, ρ, and priors.

N	ρ	Expert prior, $P(q_{jk} = 1\|\boldsymbol{\gamma}_k) = \Phi(\boldsymbol{x}'_j\boldsymbol{\gamma}_k)$					No expert prior	
		Method 1		Method 2				
		$\hat{q}_{jk} = \mathcal{I}(\frac{1}{T}\sum_{t=1}^{T} q_{jk}^{(t)} > 0.5)$		$\hat{q}_{jk} = \mathcal{I}\left(\Phi(\boldsymbol{x}'_j\hat{\boldsymbol{\gamma}}_k\right) > 0.5)$			$P\left(q_{jk} = 1\|\nu\right) = \nu$	
		Matrix	Element	Matrix	Element	$\widehat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$	Matrix	Element
500	0.0	0.250	0.978	0.980	0.999	0.580	0.300	0.980
1500	0.0	0.850	0.997	0.990	0.999	0.990	0.830	0.997
500	0.5	0.190	0.968	0.920	0.994	0.410	0.140	0.968
1500	0.5	0.720	0.988	0.920	0.988	0.900	0.770	0.994

Method 1 estimates q_{jk} with the posterior mode; Method 2 predicts elements of $\boldsymbol{Q}$ with the MVN prior regression coefficients; Method 1 is used for the No Expert Prior; $\widehat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$ indicates the empirical probability of correctly selecting the expert-predictors; Matrix = matrix-wise accuracy; Element = element-wise accuracy; $\hat{\gamma}_k$ is the posterior mean. Results are based upon 100 replications.

0.5, respectively. The chance of selecting the true model increased with N to 0.99 when $\rho = 0$ and 0.9 when $\rho = 0.5$.

A comparison of the GDM Monte Carlo results with those of the DINA reported in "Appendix B" suggest that expert-predictor validation is more challenging for the GDM when $N = 500$ and $K = 3$. Although selecting the true expert-predictors is less accurate for the GDM than DINA, the results in Table 2 suggest that element-wise $\boldsymbol{Q}$ matrix recovery is accurate even though the expert-predictors are not exactly selected.

Finally, we estimated individual attribute profiles with the posterior mode and found evidence the exploratory GDM algorithm accurately classified individual attribute profiles. Specifically, for $N = 500$ the median vector-wise correct classification rates for attributes were 78.4% and 80.0% for $\rho = 0$ and 0.5, respectively. The vector-wise classification rates increased slightly for $N = 1500$ to 79.3% and 82.0% for $\rho = 0$ and 0.5. The element-wise classification accuracy of attributes for $\rho = 0$ and 0.5 was 91.9% and 92.9% for $N = 500$ and 92.2% and 93.7% for $N = 1500$. Additionally, the Monte Carlo evidence for the GDM suggested that the attribute classification accuracy was similar with and without the expert prior, which suggests that the expert prior did not significantly improve accuracy in the studied Monte Carlo conditions.

5. Application to Tatsuoka's Fraction-Subtraction (FS) Data

5.1. Overview

We apply the expert $\boldsymbol{Q}$ model to Tatsuoka's FS dataset. The FS dataset includes responses of $N = 536$ middle school students to $J = 20$ items. More specifically, Tatsuoka (1990) defined the following eight attributes for the FS items (see Table 3):

(I) Convert a whole number to fraction,
(II) Separate a whole number from fraction,
(III) Simplify before subtraction,
(IV) Find a common denominator,
(V) Borrow from the whole number part,
(VI) Column borrow to subtract the 2nd numerator from the 1st,
(VII) Subtract numerators,
(VIII) Reduce answers to simplest form.

The purpose of the application is to infer which expert-predictors describe the structure in the FS dataset. As observed in Table 3, expert-predictor (VII) is nonzero for all but one item, so this variable is dropped from the analysis. We consider the remaining $V = 7$ expert-predictors as $\boldsymbol{X}$ and use exploratory CDMs with the expert $\boldsymbol{Q}$ prior.

"Appendix B" provides Monte Carlo simulation evidence that over-specifying K impacts the accuracy of expert-predictor selection. Accordingly, we estimated the GDM with $K = 5, 6, 7,$ and 8 to ascertain the number of underlying attributes and assessed relative fit of the models using the marginal Deviance Information Criterion (DIC; Celeux et al., 2006; Spiegelhalter et al., 2002). Note that we also compare the fit of the exploratory GDMs with results from the more parsimonious DINA model using $K = 5$.

The number of item parameters for a fully saturated GDM is $J \times 2^K$. For the FS data with $K = 5, 6, 7,$ and 8 the saturated models have 640, 1280, 2560, and 5120 parameters, respectively. Clearly, the number of parameters for all of the saturated models for the FS data exceeds the sample size and this is a scenario where regularization such as the SSVS algorithm is needed to force many of the coefficients toward zero. In addition to using regularization, researchers may have an a priori belief concerning the highest order for the GDM interaction-effects. For instance, for the FS expert-specified $\boldsymbol{Q}$ all but one of the items required three or fewer skills. Consequently,

TABLE 3.
Expert-specified fraction-subtraction data $\boldsymbol{Q}$ matrix with 8 attributes.

	Item	I	II	III	IV	V	VI	VII	VIII
1.	$\frac{5}{3}-\frac{3}{4}$	0	0	0	1	0	1	1	0
2.	$\frac{3}{4}-\frac{3}{8}$	0	0	0	1	0	0	1	0
3.	$\frac{5}{6}-\frac{1}{9}$	0	0	0	1	0	0	1	0
4.	$3\frac{1}{2}-2\frac{3}{2}$	0	1	1	0	1	0	1	0
5.	$4\frac{3}{5}-3\frac{4}{10}$	0	1	0	1	0	0	1	1
6.	$\frac{6}{7}-\frac{4}{7}$	0	0	0	0	0	0	1	0
7.	$3-2\frac{1}{5}$	1	1	0	0	0	0	1	0
8.	$\frac{2}{3}-\frac{2}{3}$	0	0	0	0	0	0	1	0
9.	$3\frac{7}{8}-2$	0	1	0	0	0	0	0	0
10.	$4\frac{4}{12}-2\frac{7}{12}$	0	1	0	0	1	0	1	1
11.	$4\frac{1}{3}-2\frac{4}{3}$	0	1	0	0	1	0	1	0
12.	$\frac{11}{8}-\frac{1}{8}$	0	0	0	0	0	0	1	1
13.	$3\frac{3}{8}-2\frac{5}{6}$	0	1	0	1	1	0	1	0
14.	$3\frac{4}{5}-3\frac{2}{5}$	0	1	0	0	0	0	1	0
15.	$2-\frac{1}{3}$	1	0	0	0	0	0	1	0
16.	$4\frac{5}{7}-1\frac{4}{7}$	0	1	0	0	0	0	1	0
17.	$7\frac{3}{5}-\frac{4}{5}$	0	1	0	0	1	0	1	0
18.	$4\frac{1}{10}-2\frac{8}{10}$	0	1	0	0	1	1	1	0
19.	$4-1\frac{4}{3}$	1	1	1	0	1	0	1	0
20.	$4\frac{1}{3}-1\frac{5}{3}$	0	1	1	0	1	0	1	0

we might also expect that item response probabilities relate to at most a three-way interaction of the attributes. Accordingly, we estimated exploratory GDMs with $K = 5$ to 8 with models that included all main-effects in addition to two- and three-way interactions. The number of estimated item parameters for non-saturated models is in general $J \times \sum_{k=0}^{m} \binom{K}{k}$ where m denotes the highest order of the interaction-effects. For our FS analyses with $K = 5$ to 8 and $m = 3$, the number of estimated item parameters equaled 520, 840, 1280, and 1860 and we see that regularization is necessary given that the total number of parameters (i.e., item and structural parameters) exceeds the FS sample size for all models.

We implemented the Gibbs samplers using a chain of length 80,000 and a burnin of 20,000. We found that running a chain length of 80,000 with a sample size of 536 and $K = 7$ required roughly 891 seconds for a laptop with a 2.21 GHz processor. Additionally, we estimated the models by fixing the "spike" precision parameters to $\omega_0 = 500$ for all coefficients. In Monte Carlo simulation studies, we found evidence of improved recovery when using different values for the "slab" precision parameter, ω_1, for main- and interaction-effects. We set $\omega_1 = 1$ for the main-effects and $\omega_1 = 10$ for all interaction-effects.

TABLE 4.
Posterior average of model selection indicator matrix $\mathbf{\Delta}$ for GDM application to the fraction-subtraction data with $K = 7$.

v	Expert		Residual				
	1	2	3	4	5	6	7
I	0.59	0.01	0.02	0.01	0.01	0.01	0.01
II	0.03	0.02	0.09	0.03	0.02	0.02	0.02
III	0.04	0.02	0.06	0.02	0.02	0.05	0.02
IV	0.00	0.83	0.00	0.00	0.00	0.00	0.00
V	0.02	0.01	0.35	0.01	0.01	0.02	0.01
VI	0.06	0.05	0.03	0.03	0.04	0.03	0.03
VIII	0.02	0.05	0.03	0.05	0.04	0.02	0.04

Expert = validated attributes from pre-specified $\boldsymbol{Q}$; Residual = attributes that were not pre-specified.

5.2. *Results*

5.2.1. Model Fit We assessed relative model fit with the DIC and found evidence the GDM with $K = 7$ provided the best relative fit. That is, the DIC for GDMs with $K = 5, 6, 7$, and 8 were 8762, 8678, 8676, and 8740, respectively, which provides evidence the model with $K = 7$ provides the best relative fit. Note for comparison we estimated $\boldsymbol{Q}$ with the expert-predictors using the DINA measurement model. The DIC for the DINA with $K = 5$ equaled 10804, which suggests the more flexible GDM is needed to describe the latent class probabilities.

5.2.2. GDM Results for $K = 7$ The GDM with $K = 7$ provided the best relative fit to the fraction-subtraction data, and this subsection accordingly examines the parameter estimates. Table 4 reports the estimated $\overline{\mathbf{\Delta}}$ to provide information about which expert-predictors were active versus inactive. First, there was evidence that two expert-predictors were active as indicated by posterior averages for δ_{vk} exceeding the 0.5 threshold. More specifically, the first two columns of $\boldsymbol{Q}$ related to expert-predictors (I) and (IV) given that $\bar{\delta}_{1,1}$ and $\bar{\delta}_{4,2}$ both exceeded 0.5. Xu and Shang (2018) used an exploratory GDM with a reduced set of FS items (i.e., they omitted items 6, 8, and 9 from their analyses) and uncovered expert-predictors (I), (IV), and (V). Our results are similar to theirs given that we found evidence for (I) and (IV) and we see that expert-predictor (V) is trending toward activeness with $\bar{\delta}_{5,3} = 0.35$.

Table 5 reports the entry-wise posterior means for $\overline{\boldsymbol{Q}}$. A comparison between Tables 3 and 5 shows that expert-predictor (I) was recovered exactly by $\widehat{\boldsymbol{Q}}$. Similarly, expert-predictor (IV) is similar to $\widehat{\boldsymbol{Q}}_2$ in all entries but for items 6 and 10. Furthermore, $\widehat{\boldsymbol{Q}}_3$ shares common elements with $\boldsymbol{X}_5$, but, as noted above, $\bar{\delta}_{5,3} = 0.35$, which suggests that $\boldsymbol{X}_5$ is not an active predictor of $\boldsymbol{Q}_3$. Instead, we consider $\widehat{\boldsymbol{Q}}_3$ as a residual attribute in the sense that it was not represented in the pre-specified expert $\boldsymbol{Q}$. We note that $\widehat{\boldsymbol{Q}}_3$ shares a similar pattern of ones and zeros as a column from Chen et al.'s (2018) estimated $\boldsymbol{Q}$ matrix using the DINA model. Specifically, Chen et al. (2018) found evidence for an attribute that loaded on items 10 through 20 in a manner similar to $\widehat{\boldsymbol{Q}}_3$ in Table 5.

The remaining four columns of $\overline{\boldsymbol{Q}}$ are residual attributes that do not match any of the pre-specified expert-predictors. Three of the residual attributes load onto just two items. For instance, Table 5 indicates that the fourth attribute is present in only items 5 and 9, the fifth attribute loads on items 6 and 12, and the sixth attribute relates to items 4 and 8. Attribute five may capture the fact that items 6 and 12 are the only items that can be solved by simply subtracting the numerators. Furthermore, the sixth attribute loads onto two items (i.e., 4 and 8) that involve the subtraction of equivalent fractions, which suggests the attribute could be interpreted as "identify equivalent

TABLE 5.
Posterior average of GDM $\boldsymbol{Q}$ for fraction-subtraction data with $K = 7$.

j	Item	$\overline{\boldsymbol{Q}}$						
		Expert		Residual				
		1	2	3	4	5	6	7
1.	$\frac{5}{3}-\frac{3}{4}$	0.02	1.00	0.01	0.01	0.03	0.01	0.00
2.	$\frac{3}{4}-\frac{3}{8}$	0.03	1.00	0.03	0.00	0.01	0.01	0.01
3.	$\frac{5}{6}-\frac{1}{9}$	0.04	1.00	0.00	0.00	0.01	0.00	0.03
4.	$3\frac{1}{2}-2\frac{3}{2}$	0.01	0.00	1.00	0.00	0.00	1.00	0.00
5.	$4\frac{3}{5}-3\frac{4}{10}$	0.02	0.99	0.01	0.94	0.02	0.00	0.01
6.	$\frac{6}{7}-\frac{4}{7}$	0.00	1.00	0.26	0.01	0.69	0.00	1.00
7.	$3-2\frac{1}{5}$	1.00	0.02	0.42	0.01	0.01	0.07	0.00
8.	$\frac{2}{3}-\frac{2}{3}$	0.17	0.01	0.52	0.39	0.20	1.00	0.42
9.	$3\frac{7}{8}-2$	0.06	0.13	0.01	1.00	0.13	0.00	0.06
10.	$4\frac{4}{12}-2\frac{7}{12}$	0.05	0.50	1.00	0.03	0.00	0.01	0.00
11.	$4\frac{1}{3}-2\frac{4}{3}$	0.01	0.01	1.00	0.01	0.01	0.03	0.01
12.	$\frac{11}{8}-\frac{1}{8}$	0.03	0.01	1.00	0.00	0.52	0.01	1.00
13.	$3\frac{3}{8}-2\frac{5}{6}$	0.01	1.00	1.00	0.08	0.00	0.01	0.00
14.	$3\frac{4}{5}-3\frac{2}{5}$	0.43	0.15	1.00	0.01	0.00	0.00	1.00
15.	$2-\frac{1}{3}$	1.00	0.05	0.99	0.03	0.01	0.01	0.02
16.	$4\frac{5}{7}-1\frac{4}{7}$	0.06	0.08	1.00	0.02	0.01	0.01	1.00
17.	$7\frac{3}{5}-\frac{4}{5}$	0.18	0.00	1.00	0.02	0.02	0.00	0.00
18.	$4\frac{1}{10}-2\frac{8}{10}$	0.67	0.21	1.00	0.01	0.14	0.01	0.30
19.	$4-1\frac{4}{3}$	1.00	0.01	1.00	0.00	0.00	0.00	0.00
20.	$4\frac{1}{3}-1\frac{5}{3}$	0.04	0.00	1.00	0.01	0.01	0.00	0.00

Expert = validated attributes from pre-specified $\boldsymbol{Q}$; Residual = attributes that were not pre-specified.

fractions." The element-wise posterior means for the seventh attribute suggest that $\widehat{\boldsymbol{Q}}_7$ relates to four items (6, 12, 14, 16), which are the items that could be solved with either only numerator subtraction or both numerator and whole number subtraction.

6. Discussion

The goal of this manuscript was to develop procedures that use expert knowledge to estimate CDM $\boldsymbol{Q}$ matrices. We proposed a fully Bayesian approach that borrows strategies from the statistics model selection literature and assessed the performance of these methods in Monte Carlo experiments and a data application. Overall, the results support incorporating expert knowledge in $\boldsymbol{Q}$ estimation and in this section we summarize key findings and discuss future research directions.

First, we developed a framework for validating expert knowledge using the GDM framework, which generalizes other models such as the DINA model and the rRUM. A benefit of using the

GDM is that we are able to validate expert knowledge and uncover possibly different cognitive processes (e.g., conjunctive, compensatory, disjunctive, etc.) for the items.

Second, the Monte Carlo simulation studies offer evidence of accurate expert-predictor selection and $\boldsymbol{Q}$ matrix estimation. In particular, we found evidence that using the MVN prior to predict elements of $\boldsymbol{Q}$ significantly improves recovery. The prediction of $\boldsymbol{Q}$ with the MVN prior parameters is ideal for cases where the expert-predictors describe the underlying structure. In cases where the expert-predictors are weakly related to $\boldsymbol{Q}$ prediction with Method 2 may be less preferred given that the expert-predictors do not explain the latent structure. In these cases, the multivariate intercepts may be the only nonzero coefficients and the predicted probabilities would reflect the overall prevalence of ones in columns of $\boldsymbol{Q}$, which may be different from the posterior probability given the observed data. In cases where expert-predictors weakly relate to the underlying structure, researchers should instead estimate Q with alternative summaries of the posterior, such as the element-wise mode (i.e., Method 1 in Table 2).

The DINA simulation study results in "Appendix B" provide additional insights. Namely, both the active and inactive predictors were selected with high probability whenever the true and estimated K were closely matched. The Monte Carlo results provide evidence that inactive predictors are nearly always identified. Active predictors are also selected with high probability, but the recovery rate declines when the estimated K is over-specified and the attributes are more highly correlated.

Third, the application to the FS dataset demonstrated how expert knowledge can be used to improve interpretation of the uncovered attributes. The posterior means for the $\delta'_{vk}s$ provide evidence that two expert-predictors [i.e., (I) and (IV)] related to $\boldsymbol{Q}$ and one expert-predictor was trending toward activeness [i.e., (V)]. Additionally, we uncovered an additional four residual attributes that were not previously specified by experts. Larger item pools with additional fraction-subtraction items that are representative of the content domain are needed to assess the extent to which the uncovered residual attributes generalize more broadly. Furthermore, future studies should, whenever possible, use a set of items that have an orthogonal expert-predictor matrix to maximize power. For instance, expert-predictor (VII) for the FS dataset had minimal variability and future studies should include additional items that do not require subtraction of numerators. Additionally, the interpretation of attribute 5 as "identify equivalent fractions" may be improved with a few more problems that require the skill.

There are several directions for future research. First, the Monte Carlo results suggested expert-predictor selection was less accurate when attributes are more correlated. Future research should consider using a higher-order model for attributes to improve performance when the attribute distribution is structured with greater dependence. Second, research may study model identifiability issues. That is, the prior in this study did not explicitly enforce identifiability conditions and there may be value in understanding whether there are conditions that $\boldsymbol{\Gamma}$ should satisfy. Third, the prior proposed in this paper is directly applicable to other latent variable models, such as exploratory factor analysis, and future research may consider benefits of using the developed methods in other areas of psychometrics to validate expert knowledge about the underlying structure.

Fourth, additional research is needed regarding the practical implications of exploratory CDMs. For instance, the methods we discuss in this paper are designed to infer which portions of an expert-specified $\boldsymbol{Q}$ matrix describe the underlying structure. A next step is to converse with experts regarding the similarities and differences between the pre-specified and expert $\boldsymbol{Q}$ matrices, as well as residual attributes. An important goal of follow-up discussions with experts is to determine whether there is evidence to refine the cognitive theory used to create the original pre-specified $\boldsymbol{Q}$ matrix. Similarly, an opportunity for methodologists and psychometricians is to ensure that the employed CDMs capture the underlying phenomenon of interest so that subsequent dialogue with experts facilitates the refinement of statistical models. The ultimate value of research in this area will be determined by the instructional value of such conversations.

Fifth, we considered a fully exploratory framework, but there may be cases where researchers want to explicitly fix values of $\boldsymbol{Q}$ to reflect partial expert knowledge. Additional research is needed to evaluate the accuracy of various strategies for incorporating partial knowledge about $\boldsymbol{Q}$. Finally, in our application we used the DIC to infer the number of attributes, K. Another strategy to explore in future research is to consider K as a random variable using Bayesian nonparametric methods (e.g., see Gershman & Blei, 2012; Griffiths & Ghahramani, 2011).

In conclusion, CDMs provide a useful framework for integrating cognitive theory into psychometric models. The general unavailability of the $\boldsymbol{Q}$ matrix encouraged the development and application of exploratory CDMs. We contribute to the literature by creating a framework that explicitly incorporates expert knowledge to improve estimation and enhance interpretation of exploratory CDM analyses.

Acknowledgments

This research was partially supported by National Science Foundation Methodology, Measurement, and Statistics Program Grants #1632023 and # 1758631 and Spencer Foundation Grant #201700062.

Appendix A: Monte Carlo Estimates of Bias and Mean Absolute Deviation for the GDM $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$

See Tables A1, A2, A3, A4, A5, A6, A7, A8, A9 and A10 .

Table A1.
Summary of bias for GDM item parameters, $\boldsymbol{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 500$), and $\rho = 0$.

Item	$\boldsymbol{Q}$			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	−0.036	−0.032	−0.015	−0.009	−0.010	−0.004	0.011	0.016
2	0	0	1	−0.034	−0.031	−0.011	−0.007	−0.009	−0.003	0.014	0.018
3	0	0	1	−0.036	−0.048	−0.010	−0.018	−0.014	−0.022	0.013	0.005
4	0	1	0	−0.034	−0.007	−0.035	−0.009	−0.008	0.020	−0.008	0.014
5	0	1	0	−0.034	−0.007	−0.031	−0.005	−0.010	0.018	−0.006	0.017
6	0	1	0	−0.034	−0.008	−0.035	−0.012	−0.009	0.018	−0.009	0.012
7	1	0	0	−0.038	−0.011	−0.011	0.017	−0.025	−0.001	−0.001	0.020
8	1	0	0	−0.038	−0.010	−0.015	0.014	−0.021	0.002	−0.000	0.020
9	1	0	0	−0.038	−0.006	−0.016	0.018	−0.025	0.002	−0.005	0.018
10	0	1	1	−0.039	−0.038	−0.025	−0.010	−0.015	−0.002	0.010	0.015
11	0	1	1	−0.034	−0.033	−0.019	−0.015	−0.009	0.002	0.015	0.010
12	1	0	1	−0.036	−0.042	−0.012	−0.009	−0.016	−0.010	0.013	0.009
13	1	0	1	−0.036	−0.042	−0.013	−0.009	−0.014	−0.013	0.015	0.007
14	1	0	1	−0.038	−0.038	−0.013	−0.004	−0.016	−0.003	0.014	0.016
15	1	1	0	−0.036	−0.008	−0.024	0.011	−0.026	0.006	−0.005	0.012
16	1	1	0	−0.038	−0.013	−0.034	−0.000	−0.017	0.014	−0.002	0.015
17	1	1	0	−0.037	−0.008	−0.022	0.018	−0.012	0.023	−0.004	0.017
18	1	1	1	−0.029	−0.072	−0.017	−0.032	0.002	−0.018	0.017	−0.022
19	1	1	1	−0.034	−0.072	−0.030	−0.039	−0.002	−0.020	0.016	−0.025
20	1	1	1	−0.031	−0.057	−0.030	−0.028	−0.001	−0.005	0.003	−0.024

The estimates are based upon 100 replications.

TABLE A2.
Summary of bias for GDM item parameters, $\mathbf{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 1500$), and $\rho = 0$.

Item	$\boldsymbol{Q}$			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	– 0.030	– 0.038	– 0.010	– 0.015	– 0.003	– 0.014	0.017	0.006
2	0	0	1	– 0.030	– 0.035	– 0.009	– 0.013	– 0.002	– 0.010	0.018	0.008
3	0	0	1	– 0.031	– 0.039	– 0.010	– 0.017	– 0.003	– 0.013	0.018	0.006
4	0	1	0	– 0.030	– 0.007	– 0.025	– 0.003	– 0.008	0.015	– 0.002	0.017
5	0	1	0	– 0.030	– 0.007	– 0.029	– 0.009	– 0.005	0.018	– 0.003	0.013
6	0	1	0	– 0.031	– 0.009	– 0.028	– 0.005	– 0.008	0.017	– 0.003	0.017
7	1	0	0	– 0.033	– 0.010	– 0.012	0.013	– 0.017	0.004	– 0.000	0.017
8	1	0	0	– 0.033	– 0.009	– 0.011	0.014	– 0.019	0.003	– 0.001	0.018
9	1	0	0	– 0.035	– 0.011	– 0.014	0.010	– 0.019	– 0.000	– 0.001	0.016
10	0	1	1	– 0.027	– 0.028	– 0.020	– 0.014	– 0.004	0.002	0.010	0.005
11	0	1	1	– 0.030	– 0.027	– 0.020	– 0.006	– 0.007	0.001	0.011	0.011
12	1	0	1	– 0.032	– 0.024	– 0.013	0.002	– 0.014	– 0.004	0.009	0.009
13	1	0	1	– 0.031	– 0.027	– 0.012	0.001	– 0.014	– 0.006	0.010	0.010
14	1	0	1	– 0.031	– 0.031	– 0.012	– 0.003	– 0.010	– 0.006	0.012	0.008
15	1	1	0	– 0.032	– 0.011	– 0.017	0.012	– 0.016	0.011	– 0.006	0.009
16	1	1	0	– 0.034	– 0.013	– 0.027	– 0.000	– 0.010	0.016	– 0.002	0.011
17	1	1	0	– 0.033	– 0.011	– 0.016	0.012	– 0.015	0.015	– 0.002	0.013
18	1	1	1	– 0.031	– 0.042	– 0.022	0.007	– 0.007	0.007	0.002	0.002
19	1	1	1	– 0.029	– 0.046	– 0.019	0.006	– 0.009	0.003	0.006	0.002
20	1	1	1	– 0.029	– 0.042	– 0.014	0.025	– 0.012	0.002	– 0.006	0.004

The estimates are based upon 100 replications.

TABLE A3.
Summary of bias for GDM item parameters, $\mathbf{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 500$), and $\rho = 0.5$.

Item	$\boldsymbol{Q}$			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	– 0.034	– 0.060	– 0.010	– 0.031	– 0.010	– 0.028	0.015	– 0.001
2	0	0	1	– 0.033	– 0.061	– 0.009	– 0.031	– 0.007	– 0.028	0.018	0.000
3	0	0	1	– 0.034	– 0.052	– 0.011	– 0.025	– 0.010	– 0.020	0.015	0.005
4	0	1	0	– 0.035	– 0.007	– 0.049	– 0.022	– 0.006	0.024	– 0.013	0.011
5	0	1	0	– 0.032	– 0.003	– 0.044	– 0.017	– 0.004	0.026	– 0.013	0.009
6	0	1	0	– 0.034	– 0.006	– 0.043	– 0.015	– 0.007	0.023	– 0.010	0.015
7	1	0	0	– 0.035	0.001	0.001	0.038	– 0.028	– 0.001	– 0.002	0.021
8	1	0	0	– 0.035	0.001	– 0.000	0.038	– 0.027	– 0.001	– 0.000	0.022
9	1	0	0	– 0.035	0.002	0.003	0.040	– 0.030	– 0.004	– 0.003	0.018
10	0	1	1	– 0.030	– 0.042	– 0.035	– 0.041	– 0.002	– 0.004	0.006	– 0.008
11	0	1	1	– 0.032	– 0.037	– 0.037	– 0.030	– 0.006	– 0.001	– 0.001	– 0.000
12	1	0	1	– 0.033	– 0.049	– 0.002	– 0.009	– 0.019	– 0.020	0.015	0.003
13	1	0	1	– 0.034	– 0.048	– 0.004	– 0.009	– 0.020	– 0.022	0.014	0.001
14	1	0	1	– 0.034	– 0.058	– 0.004	– 0.020	– 0.015	– 0.021	0.017	0.001
15	1	1	0	– 0.032	0.000	– 0.023	0.017	– 0.013	0.025	– 0.007	0.014
16	1	1	0	– 0.035	0.002	– 0.031	0.010	– 0.018	0.025	– 0.006	0.015

TABLE A3.
continued

Item	Q			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
17	1	1	0	– 0.034	– 0.001	– 0.020	0.021	– 0.028	0.011	– 0.008	0.015
18	1	1	1	– 0.030	– 0.059	– 0.041	– 0.047	– 0.007	– 0.011	0.010	– 0.010
19	1	1	1	– 0.032	– 0.052	– 0.046	– 0.048	– 0.000	0.005	0.009	– 0.011
20	1	1	1	– 0.031	– 0.047	– 0.047	– 0.039	– 0.001	0.001	0.005	– 0.010

The estimates are based upon 100 replications.

TABLE A4.
Summary of bias for GDM item parameters, $\boldsymbol{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 1500$), and $\rho = 0.5$.

Item	Q			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	– 0.025	– 0.038	– 0.005	– 0.016	– 0.006	– 0.013	0.014	0.007
2	0	0	1	– 0.025	– 0.032	– 0.006	– 0.010	– 0.005	– 0.007	0.013	0.012
3	0	0	1	– 0.027	– 0.039	– 0.007	– 0.016	– 0.005	– 0.011	0.014	0.009
4	0	1	0	– 0.027	– 0.005	– 0.031	– 0.010	– 0.004	0.018	– 0.005	0.012
5	0	1	0	– 0.026	– 0.004	– 0.036	– 0.015	– 0.004	0.019	– 0.010	0.008
6	0	1	0	– 0.027	– 0.004	– 0.034	– 0.013	– 0.004	0.019	– 0.007	0.011
7	1	0	0	– 0.030	0.032	– 0.004	0.058	– 0.022	0.011	– 0.002	0.025
8	1	0	0	– 0.030	0.034	– 0.002	0.061	– 0.020	0.012	0.000	0.026
9	1	0	0	– 0.030	0.030	– 0.003	0.058	– 0.019	0.010	0.002	0.024
10	0	1	1	– 0.025	– 0.034	– 0.027	– 0.013	– 0.004	– 0.004	0.003	0.010
11	0	1	1	– 0.025	– 0.038	– 0.026	– 0.014	– 0.003	– 0.006	0.004	0.008
12	1	0	1	– 0.028	– 0.010	– 0.004	0.021	– 0.012	0.003	0.012	0.018
13	1	0	1	– 0.026	– 0.012	– 0.003	0.018	– 0.017	0.002	0.008	0.017
14	1	0	1	– 0.023	– 0.017	– 0.000	0.015	– 0.014	– 0.001	0.012	0.018
15	1	1	0	– 0.029	0.008	– 0.026	0.022	– 0.013	0.026	– 0.001	0.020
16	1	1	0	– 0.030	0.010	– 0.016	0.034	– 0.011	0.032	– 0.001	0.021
17	1	1	0	– 0.029	0.011	– 0.031	0.016	– 0.013	0.031	– 0.004	0.019
18	1	1	1	– 0.025	– 0.031	– 0.026	0.003	– 0.011	0.011	0.006	0.009
19	1	1	1	– 0.025	– 0.025	– 0.029	0.005	– 0.011	0.013	0.005	0.002
20	1	1	1	– 0.026	– 0.021	– 0.033	0.005	– 0.008	0.022	0.003	0.006

The estimates are based upon 100 replications.

TABLE A5.
Summary of mean absolute deviation for GDM item parameters, $\mathbf{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 500$), and $\rho = 0$.

Item	$\boldsymbol{Q}$			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	0.036	0.053	0.017	0.043	0.017	0.041	0.019	0.041
2	0	0	1	0.034	0.053	0.015	0.042	0.016	0.043	0.020	0.041
3	0	0	1	0.036	0.062	0.017	0.045	0.017	0.045	0.020	0.040
4	0	1	0	0.034	0.017	0.042	0.029	0.017	0.025	0.026	0.029
5	0	1	0	0.034	0.015	0.037	0.029	0.016	0.022	0.024	0.029
6	0	1	0	0.034	0.017	0.042	0.031	0.017	0.025	0.027	0.031
7	1	0	0	0.038	0.018	0.020	0.022	0.029	0.022	0.020	0.027
8	1	0	0	0.038	0.018	0.018	0.021	0.025	0.018	0.018	0.027
9	1	0	0	0.038	0.022	0.019	0.024	0.028	0.020	0.019	0.024
10	0	1	1	0.039	0.063	0.036	0.046	0.019	0.058	0.030	0.045
11	0	1	1	0.034	0.061	0.035	0.040	0.017	0.056	0.035	0.036
12	1	0	1	0.036	0.077	0.017	0.070	0.028	0.037	0.031	0.037
13	1	0	1	0.036	0.070	0.016	0.063	0.032	0.038	0.030	0.039
14	1	0	1	0.038	0.068	0.018	0.064	0.031	0.038	0.030	0.039
15	1	1	0	0.036	0.020	0.053	0.052	0.040	0.034	0.028	0.028
16	1	1	0	0.038	0.016	0.060	0.055	0.035	0.035	0.025	0.028
17	1	1	0	0.037	0.018	0.054	0.053	0.035	0.042	0.026	0.029
18	1	1	1	0.029	0.094	0.056	0.088	0.040	0.068	0.039	0.051
19	1	1	1	0.034	0.099	0.061	0.090	0.038	0.073	0.044	0.054
20	1	1	1	0.031	0.081	0.069	0.088	0.039	0.067	0.034	0.055

The estimates are based upon 100 replications.

TABLE A6.
Summary of mean absolute deviation for GDM item parameters, $\mathbf{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 1500$), and $\rho = 0$.

Item	$\boldsymbol{Q}$			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	0.030	0.043	0.014	0.031	0.017	0.030	0.020	0.029
2	0	0	1	0.030	0.041	0.013	0.031	0.018	0.029	0.021	0.032
3	0	0	1	0.031	0.045	0.013	0.031	0.017	0.030	0.021	0.029
4	0	1	0	0.030	0.013	0.027	0.017	0.013	0.018	0.014	0.023
5	0	1	0	0.030	0.015	0.031	0.020	0.013	0.022	0.016	0.020
6	0	1	0	0.031	0.013	0.030	0.017	0.013	0.020	0.015	0.021
7	1	0	0	0.033	0.018	0.018	0.020	0.019	0.014	0.012	0.019
8	1	0	0	0.033	0.015	0.016	0.020	0.020	0.013	0.012	0.022
9	1	0	0	0.035	0.018	0.019	0.018	0.021	0.012	0.012	0.020
10	0	1	1	0.028	0.044	0.027	0.035	0.016	0.037	0.023	0.033
11	0	1	1	0.030	0.043	0.026	0.028	0.017	0.034	0.024	0.029
12	1	0	1	0.032	0.047	0.018	0.043	0.023	0.028	0.018	0.030
13	1	0	1	0.031	0.045	0.016	0.041	0.023	0.028	0.020	0.031
14	1	0	1	0.031	0.053	0.015	0.047	0.020	0.024	0.020	0.024

TABLE A6.
continued

Item	Q			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
15	1	1	0	0.032	0.015	0.036	0.034	0.025	0.022	0.016	0.019
16	1	1	0	0.034	0.018	0.039	0.032	0.019	0.023	0.016	0.020
17	1	1	0	0.033	0.015	0.034	0.034	0.022	0.025	0.016	0.022
18	1	1	1	0.031	0.055	0.036	0.054	0.022	0.034	0.023	0.033
19	1	1	1	0.029	0.060	0.031	0.059	0.024	0.041	0.026	0.035
20	1	1	1	0.029	0.057	0.037	0.060	0.023	0.033	0.021	0.037

The estimates are based upon 100 replications.

TABLE A7.
Summary of mean absolute deviation for GDM item parameters, $\boldsymbol{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 500$), and $\rho = 0.5$.

Item	Q			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	0.034	0.065	0.016	0.045	0.015	0.044	0.020	0.036
2	0	0	1	0.033	0.072	0.013	0.052	0.015	0.050	0.022	0.040
3	0	0	1	0.034	0.063	0.015	0.047	0.017	0.045	0.021	0.038
4	0	1	0	0.035	0.016	0.052	0.034	0.018	0.029	0.028	0.025
5	0	1	0	0.032	0.015	0.048	0.032	0.015	0.028	0.029	0.027
6	0	1	0	0.034	0.015	0.049	0.033	0.020	0.026	0.028	0.028
7	1	0	0	0.035	0.022	0.024	0.041	0.032	0.022	0.021	0.028
8	1	0	0	0.035	0.022	0.024	0.041	0.031	0.025	0.024	0.032
9	1	0	0	0.035	0.022	0.027	0.044	0.035	0.025	0.023	0.029
10	0	1	1	0.030	0.076	0.044	0.053	0.013	0.069	0.030	0.038
11	0	1	1	0.032	0.075	0.047	0.048	0.016	0.072	0.032	0.038
12	1	0	1	0.033	0.081	0.021	0.075	0.032	0.042	0.036	0.038
13	1	0	1	0.034	0.085	0.021	0.077	0.035	0.049	0.039	0.045
14	1	0	1	0.034	0.089	0.022	0.081	0.030	0.043	0.033	0.038
15	1	1	0	0.032	0.016	0.067	0.067	0.036	0.042	0.027	0.031
16	1	1	0	0.035	0.021	0.067	0.064	0.037	0.044	0.023	0.029
17	1	1	0	0.034	0.018	0.066	0.068	0.043	0.042	0.028	0.026
18	1	1	1	0.030	0.101	0.089	0.111	0.054	0.078	0.048	0.050
19	1	1	1	0.032	0.097	0.079	0.105	0.039	0.090	0.049	0.045
20	1	1	1	0.032	0.099	0.081	0.100	0.043	0.083	0.049	0.047

The estimates are based upon 100 replications.

TABLE A8.
Summary of mean absolute deviation for GDM item parameters, $\mathbf{\Theta}$, using the MVN prior by latent class and item for $K = 3$, $J = 20$, sample size ($N = 1500$), and $\rho = 0.5$.

Item	Q			Latent class, $\boldsymbol{\alpha}$							
	q_1	q_2	q_3	000	001	010	011	100	101	110	111
1	0	0	1	0.025	0.043	0.011	0.030	0.010	0.027	0.016	0.025
2	0	0	1	0.025	0.039	0.010	0.026	0.010	0.027	0.016	0.026
3	0	0	1	0.027	0.045	0.012	0.031	0.012	0.025	0.018	0.023
4	0	1	0	0.027	0.012	0.032	0.017	0.011	0.020	0.014	0.018
5	0	1	0	0.027	0.012	0.038	0.023	0.013	0.023	0.018	0.016
6	0	1	0	0.027	0.011	0.035	0.020	0.012	0.021	0.017	0.019
7	1	0	0	0.030	0.052	0.020	0.062	0.023	0.026	0.014	0.030
8	1	0	0	0.030	0.051	0.019	0.065	0.020	0.025	0.013	0.031
9	1	0	0	0.030	0.051	0.020	0.063	0.021	0.024	0.014	0.029
10	0	1	1	0.025	0.051	0.033	0.029	0.011	0.040	0.021	0.024
11	0	1	1	0.025	0.055	0.030	0.027	0.012	0.045	0.019	0.024
12	1	0	1	0.028	0.072	0.018	0.070	0.019	0.031	0.021	0.030
13	1	0	1	0.027	0.071	0.016	0.069	0.021	0.030	0.019	0.031
14	1	0	1	0.024	0.075	0.018	0.073	0.018	0.028	0.021	0.028
15	1	1	0	0.029	0.029	0.049	0.062	0.022	0.039	0.018	0.026
16	1	1	0	0.030	0.030	0.045	0.060	0.023	0.042	0.017	0.028
17	1	1	0	0.029	0.030	0.052	0.059	0.022	0.040	0.018	0.024
18	1	1	1	0.025	0.073	0.050	0.075	0.024	0.053	0.024	0.027
19	1	1	1	0.026	0.068	0.051	0.079	0.024	0.063	0.024	0.026
20	1	1	1	0.027	0.078	0.050	0.075	0.025	0.061	0.022	0.029

The estimates are based upon 100 replications.

TABLE A9.
Summary of bias for GDM structural parameters, $\boldsymbol{\pi}$, using the MVN prior by latent class, sample size (N), and attribute tetrachoric correlation ρ for $K = 3$.

N	ρ	Latent class, $\boldsymbol{\alpha}$							
		000	001	010	011	100	101	110	111
500	0.0	−0.020	0.005	0.004	0.005	−0.002	0.008	0.010	−0.010
1500	0.0	−0.020	0.003	0.004	0.004	−0.001	0.011	0.008	−0.009
500	0.5	−0.028	0.007	0.011	0.009	0.003	0.006	0.005	−0.013
1500	0.5	−0.024	0.007	0.010	0.015	0.009	0.001	0.005	−0.023

The estimates are based upon 100 replications.

TABLE A10.
Summary of mean absolute deviation for GDM structural parameters, $\boldsymbol{\pi}$, using the MVN prior by latent class, sample size (N), and attribute tetrachoric correlation (ρ) for $K = 3$.

N	ρ	Latent class, $\boldsymbol{\alpha}$							
		000	001	010	011	100	101	110	111
500	0.0	0.021	0.009	0.013	0.011	0.021	0.016	0.025	0.016
1500	0.0	0.020	0.006	0.009	0.006	0.013	0.013	0.016	0.014
500	0.5	0.029	0.008	0.017	0.010	0.020	0.013	0.025	0.025
1500	0.5	0.024	0.008	0.012	0.017	0.015	0.011	0.016	0.026

The estimates are based upon 100 replications.

Appendix B: DINA Monte Carlo Simulation Study

Overview

In this section, we review results from two simulation studies to assess expert-predictor selection accuracy when a DINA model is the data generating mechanism. The goal of the simulation studies is to assess how expert-predictor selection accuracy is affected by the specification of the number of attributes K, the size of attribute tetrachoric correlations, and degree of expert-predictor correlation. We focus our attention on model selection accuracy given that we found evidence in preliminary investigations that accurately selecting the expert-predictors for the more parsimonious DINA model typically coincided with accurately recovering $\boldsymbol{Q}$ and the other model parameters.

First, the degree of mismatch between the true K and the estimated K was manipulated to understand the extent to which over-specifying K impacts expert-predictor selection accuracy. The true number of attributes was fixed to $K = 3$ in both simulation studies and we varied the estimated K to be 3, 4, 5, or 6 (i.e., the amount by which K was over-specified was 0, 1, 2, or 3 attributes). Second, as for the GDM simulation study reported above, we manipulated attribute dependence by generating attributes using the multivariate normal probit model described by Chiu et al. (2009) with the population tetrachoric correlation taking values of $\rho = 0$ or 0.5. Third, the two simulation studies reported in this section differ in terms of the extent to which expert-predictor correlate. Study #1 sets $V = 7$ and generates orthogonal attributes by sampling $x_{jv} \sim \text{Bernoulli}(0.5)$ and $\boldsymbol{Q}$ is defined as the first three columns of $\boldsymbol{X}$. In contrast, Study #2 sets $V = 7$ and uses an expert-derived $\boldsymbol{X}$ based upon Tatsuoka (1990) specified $\boldsymbol{Q}$ for Tatsuoka's FS dataset (see Table 3). Note attribute (VII) was dropped given minimal variability and, in order to match the results of the FS application, the true $\boldsymbol{Q}$ in Study #2 was defined as attributes (I), (IV), and (V).

The simulation studies focus on conditions similar to the empirical application, so $N = 500$ and $J = 20$. Additionally, $\boldsymbol{Y}$ is generated from a DINA model with slipping and guessing parameters equal to 0.2 for all j. A Gibbs sampler was implemented to approximate the posterior distribution of $\boldsymbol{Q}$ and $\boldsymbol{\Gamma}$, as well as the other DINA model parameters. A Bayesian formulation for the DINA model (e.g., see Chen et al., 2018; Culpepper, 2015) was implemented to estimate attributes, using an unstructured Dirichlet prior for $\boldsymbol{\pi}$, and a uniform prior for slipping and guessing parameters. Note that 100 replications were executed for all conditions and a chain of length 40,000 was run with a burnin of 20,000. Model performance was measured by computing the matrix-wise accuracy of $\boldsymbol{\Delta}$ separately for the active and inactive predictors.

Simulation Study #1 Results

The third and fourth columns of Table B1 report expert-predictor selection accuracy for the randomly generated $\boldsymbol{X}$ and $\boldsymbol{Q}$. The results suggest that the proportion of correctly identified inactive expert-predictors exceeds 0.94 for all estimated K and ρ. The active predictors are accurately selected for $\rho = 0$, but the chance of selecting the active expert-predictors declined to 0.57 when $\rho = 0.5$ and three unnecessary attributes are estimated.

Table B1.
Summary of matrix-wise accuracy (i.e., $\widehat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$) for active and inactive expert-predictors across simulation study, estimated K, and ρ.

Est. K	ρ	Study #1		Study #2	
		Active	Inactive	Active	Inactive
3	0.0	0.99	1.00	1.00	1.00
4	0.0	1.00	0.96	1.00	0.99
5	0.0	1.00	0.94	0.93	0.96
6	0.0	0.92	0.98	0.77	0.95
3	0.5	0.91	1.00	0.89	1.00
4	0.5	0.84	0.96	0.87	0.99
5	0.5	0.79	0.98	0.57	0.93
6	0.5	0.57	1.00	0.37	0.96

Study #1 randomly generated $\boldsymbol{X}$ and $\boldsymbol{Q}$, whereas Study #2 defined $\boldsymbol{X}$ and $\boldsymbol{Q}$ based upon all attributes in Table 3 except (VII). Estimates are based upon 100 replications.

Simulation Study #2 Results

The accuracy of expert-predictor selection for an expert-derived $\boldsymbol{X}$ and $\boldsymbol{Q}$ is presented in the right two columns of Table B1. The results suggest that the proportion of correctly identified inactive expert-predictors exceeds 0.93 for all estimated K and ρ. The active predictors are accurately selected for $\rho = 0$, but the chance of selecting the accurate predictors declines to 0.37 when $\rho = 0.5$ and K is over-specified by three attributes.

References

Albert, J. (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.

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

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*, 651–673. https://doi.org/10.1214/06-BA122.

Chen, Y., & Culpepper, S. A. (2018). *A multivariate probit model for learning trajectories with application to classroom assessment*. New York: In Paper presentation at the International Meeting of the Psychometric Society.

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

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.

Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, *37*(8), 598–618.

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*(4), 633–665.

Chung, M. (2014). Estimating the Q-matrix for cognitive diagnosis models in a Bayesian framework (Unpublished doctoral dissertation). Columbia University.

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., & Chen, Y. (2018). Development and application of an exploratory reduced reparameterized unified model. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/1076998618791306.

Culpepper, S. A., & Hudson, A. (2018). An improved strategy for Bayesian estimation of the reduced reparameterized unified model. *Applied Psychological Measurement*, *42*, 99–115. https://doi.org/10.1177/0146621617707511.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*(4), 343–362.

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

de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, *81*(2), 253–273.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353.

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement*, *36*, 447–468.

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

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12.

Griffiths, T. L., & Ghahramani, Z. (2011). The Indian buffet process: An introduction and review. *Journal of Machine Learning Research*, *12*, 1185–1224.

Gu, Y., & Xu, G. (2018). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*. https://doi.org/10.1007/s11336-018-9619-8.

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

Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality (Unpublished doctoral dissertation)*. University of Illinois at Urbana-Champaign.

Henson, R. A., & Templin, J. (2007). *Importance of Q-matrix construction and its effects cognitive diagnosis model results*. Chicago, IL: In Annual Meeting of the National Council on Measurement. in Education.

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.

Ishwaran, H., & Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics*, *33*, 730–773.

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.

Klein, M. F., Birenbaum, M., Standiford, S. N., & Tatsuoka, K. K. (1981). Logical error analysis and construction of tests to diagnose student "bugs" in addition and subtraction of fractions. (Tech. Rep.). University of Illinois at Urbana-Champaign, Champaign, IL.

Liu, J. (2017). On the consistency of Q-matrix estimation: A commentary. *Psychometrika*, *82*(2), 523–527.

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, *36*(7), 548–564.

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

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187–212.

O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, *4*(1), 85–117.

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.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *51*(3), 337–350.

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

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis (Tech. Rep.). University of Illinois at Urbana-Champaign, Champaign, IL.

Templin, J. L. & Henson, R. A. (2006). A Bayesian method for incorporating uncertainty into Q-matrix estimation in skills assessment. In *Symposium conducted at the meeting of the American Educational Research Association*, San Diego, CA.

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.

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

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*. https://doi.org/10.1080/01621459.2017.

Manuscript Received: 16 NOV 2017
Published Online Date: 19 NOV 2018