PSYCHOMETRIKA—VOL. 83, NO. 1, 89–108
MARCH 2018
https://doi.org/10.1007/s11336-017-9579-4

BAYESIAN ESTIMATION OF THE DINA Q MATRIX

YINGHAN CHEN

UNIVERSITY OF NEVADA, RENO

STEVEN ANDREW CULPEPPER, YUGUO CHEN AND
JEFFREY DOUGLAS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Cognitive diagnosis models are partially ordered latent class models and are used to classify students into skill mastery profiles. The deterministic inputs, noisy "and" gate model (DINA) is a popular psychometric model for cognitive diagnosis. Application of the DINA model requires content expert knowledge of a Q matrix, which maps the attributes or skills needed to master a collection of items. Misspecification of Q has been shown to yield biased diagnostic classifications. We propose a Bayesian framework for estimating the DINA Q matrix. The developed algorithm builds upon prior research (Chen, Liu, Xu, & Ying, in J Am Stat Assoc 110(510):850–866, 2015) and ensures the estimated Q matrix is identified. Monte Carlo evidence is presented to support the accuracy of parameter recovery. The developed methodology is applied to Tatsuoka's fraction-subtraction dataset.

Key words: cognitive diagnosis models, deterministic inputs, noisy "and" gate (DINA) model, Q matrix, Bayesian statistics, fraction-subtraction data.

1. Introduction

Cognitive diagnosis models (CDMs) were developed to provide educators and researchers with instructionally relevant assessments (Huff & Goodman, 2007; Leighton & Gierl, 2007). CDMs use cognitive theory within a psychometric framework to, "…explain achievement test performance by providing insight into whether it is students' understanding (or lack of it) or something else that is the primary cause of their performance" (Norris, Macnab, & Phillips, 2007, p. 61). That is, CDMs are partially ordered latent class models that assist researchers in classifying students as either masters or non-masters on a collection of skills deemed important for success on educational tasks. A benefit is that the CDM framework provides educators more detailed diagnostic information regarding student skills/attributes than is available with more broadly defined continuous traits in item response models.

The application of CDMs is dependent upon the availability of cognitive theory that specifies the skills and/or attributes necessary for success on a collection of tasks. In particular, let $j = 1, \ldots, J$ and $k = 1, \ldots, K$ index items and skills, respectively. Content expert knowledge is codified in a $J \times K$ $\boldsymbol{Q}$ matrix with individual elements denoted by $q_{jk} = 1$ if skill k is needed to respond correctly to item j as a master and zero otherwise. Given a specified $\boldsymbol{Q}$ and item responses, the CDM item and latent class parameters can be accurately estimated. In fact, CDMs have received broader application among educational and psychological researchers in large part due to the availability of computer code and software (de la Torre, 2009; Chiu & Köhn, 2016;

Electronic supplementary material The online version of this article (doi:10.1007/s11336-017-9579-4) contains supplementary material, which is available to authorized users.

Correspondence should be made to Steven Andrew Culpepper, Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL 61820, USA. Email: sculpepp@illinois.edu

Chiu, Köhn, & Wu, 2016; Rupp, 2009; Rupp, Templin, & Henson, 2010; Templin & Hoffman, 2013).

However, the application of CDMs is not without limitations and model parameter estimates may be inaccurate in cases where researchers employ a misspecified $\boldsymbol{Q}$. For instance, Henson and Templin (2007) and Rupp and Templin (2008) provided Monte Carlo evidence that employing a misspecified $\boldsymbol{Q}$ results in inaccurate skill classification and biased item parameter estimators.

Clearly, a fundamental problem for CDM research is the estimation of $\boldsymbol{Q}$ and prior research considered several general strategies for estimating $\boldsymbol{Q}$. First, several studies assumed partial knowledge of $\boldsymbol{Q}$ and developed procedures and algorithms for refining (Chiu, 2013; de la Torre, 2008; de la Torre & Chiu, 2016) and estimating (DeCarlo, 2012; Templin & Henson, 2006) elements of $\boldsymbol{Q}$. For example, de la Torre and Chiu (2016) offer a validation procedure that first estimates item parameters from a provisional $\boldsymbol{Q}$ matrix and then updates elements of $\boldsymbol{Q}$ using a search algorithm. One limitation of the aforementioned strategy is that the partial knowledge of $\boldsymbol{Q}$ may be inaccurate, which could compromise the consistency of $\boldsymbol{Q}$ matrix refinement (Liu, 2017). Second, recent efforts considered Bayesian (Chung, 2014) and frequentist (Chen, Liu, Xu, & Ying, 2015; Liu, Xu, & Ying, 2012, 2013; Xiang, 2013) estimators of $\boldsymbol{Q}$. The first consistent estimator of $\boldsymbol{Q}$ is proposed in Liu et al. (2013), in the case of the deterministic inputs, noisy "and" gate (DINA) response model. This major breakthrough in understanding was limited in practice because the estimator required evaluating an objective function at all possible values of $\boldsymbol{Q}$. In real problems this would rarely be possible. For instance, an exam with $J = 40$ items and $K = 5$ skills would require evaluating 2^{200} candidates. To address this Liu et al. (2012) considered the case when an initial expert generated estimate of $\boldsymbol{Q}$ is available and is in a neighborhood of the true $\boldsymbol{Q}$. The objective function of Liu et al. (2013) is then evaluated row-by-row, considering all possible values each row could take. Using the example of 40 items and 5 skills, this reduces the complexity from 2^{200} objective function evaluations to only 40×2^5, which can be done quickly. However, the drawback is that it may not be reasonable to assume that the true $\boldsymbol{Q}$ can be found among so few candidates.

A fundamentally different approach to estimation of $\boldsymbol{Q}$ is introduced in Chen et al. (2015) in which the necessary and sufficient identifiability conditions for the DINA and deterministic input, noisy "or" gate (DINO) models are first derived before applying regularization techniques to estimate $\boldsymbol{Q}$ within the framework of statistical model selection. This technique afforded the chance to reach beyond the DINA and DINO models and use known methods for fitting generalized linear models, recognizing the DINA and DINO and several other models may be expressed in the log-linear cognitive diagnosis (LCDM) model formulation of Henson, Templin, and Willse (2009). By representing these models as linear through a link function, $\boldsymbol{Q}$ estimation was addressed as a problem of identifying the nonzero regression coefficients. This expanded the scope of possibilities in a mostly rigorous fashion, except that it did not directly allow for enforcing the identifiability conditions that were found for the DINA and DINO and not yet determined for more general models in the LCDM framework. Also, it is not clear how well the technique scales to large numbers of attributes, due to the number of interactions that must be modeled in order to have the generalized linear representation that is required.

This paper develops a Bayesian approach for estimating $\boldsymbol{Q}$ for the DINA model. The proposed Bayesian formulation in this study improves upon existing strategies for estimating $\boldsymbol{Q}$. First, the Bayesian estimator does not require partial knowledge of $\boldsymbol{Q}$ and is accordingly an exploratory technique similar to Chen et al. (2015). Second, the Bayesian approach discussed in this paper offers an efficient Markov chain Monte Carlo (MCMC) algorithm for estimating an identified $\boldsymbol{Q}$ for the DINA model. Specifically, the method of Chen et al. (2015) does not explicitly enforce the identifiability conditions. Chen et al.'s (2015) method performed well in simulation studies; however, the regularization techniques did not always produce identified $\boldsymbol{Q}$ matrices in empirical applications. Though other Bayesian approaches have been attempted, such as Chung (2014), this study proposes an MCMC strategy for sampling elements of $\boldsymbol{Q}$ restricted to the space of identified $\boldsymbol{Q}$ matrices.

The remainder of this paper includes three sections. The first section describes a Bayesian formulation for the DINA model and the MH sampler for $\boldsymbol{Q}$. Additionally, a theorem is included to demonstrate the algorithm for constraining samples to the identified space is irreducible and satisfies convergence requirements of MCMC methods. This paper discusses Bayesian estimation of $\boldsymbol{Q}$ for the DINA model, but it is important to note that the proposed model can be directly extended to the DINO model as well, given the duality of the DINA and DINO models (Xu & Zhang, 2016). The second section reports results from a Monte Carlo simulation study that examines the recoverability of $\boldsymbol{Q}$ for different sample sizes and population attribute distributions and compares the constrained MCMC sampler to Chung's (2014) unconstrained Gibbs sampler. The third section includes an application of the developed methodology to estimate $\boldsymbol{Q}$ for the classic Tatsuoka's fraction-subtraction dataset (K. K. Tatsuoka, 1984; C. Tatsuoka, 2002). The last section provides concluding remarks and implications for future research.

2. Bayesian Model Formulation

2.1. Model Formulation

As introduced in the previous section, there are various CDMs to explain achievement test performance. This paper focuses on the DINA model. Suppose there are N subjects, each of whom provides a binary response to J items. The model intuitively assumes that a subject's responses are stochastically related to the unobserved latent attributes. The attribute profile for individual i is denoted as $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iK})'$ where $\alpha_{ik} = 1$ means subject i has attribute k. Let $\boldsymbol{a}_c \in \{0, 1\}^K$ denote one of the 2^K attribute profile configurations. Let Y_{ij} equal one for a correct response and zero otherwise. Furthermore, define $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)'$ as an $N \times J$ matrix of dichotomous random variables where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})'$ is the J-dimensional vector of binary responses for individual i. Throughout the paper Y refers to a random variable, whereas y denotes a realized value.

As noted above, the $\boldsymbol{Q}$ matrix catalogs the attributes needed to master each item. Note that $\boldsymbol{Q} = \left(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_J\right)' = \left(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K\right)$ where $\boldsymbol{q}_j' = \left(q_{j1}, \ldots, q_{jK}\right)$ is the j-th row of $\boldsymbol{Q}$ and the italicized $\boldsymbol{Q}_k$ denotes the k-th column. The DINA model specifies a conjunctive relationship between latent attributes $\boldsymbol{\alpha}_i$ and observed responses Y_{ij} in the sense that success on an item occurs with highest probability for those students that possess all of the necessary attributes. The conjunctive nature of the DINA model is apparent with the ideal response for individual i on item j,

$$\eta_{ij} = \mathcal{I}\left(\alpha_{ik} \geq q_{jk} \text{ for all } k\right) = \mathcal{I}\left(\boldsymbol{\alpha}_i'\boldsymbol{q}_j = \boldsymbol{q}_j'\boldsymbol{q}_j\right), \tag{1}$$

where $\mathcal{I}(\cdot)$ denotes the indicator function. η_{ij} equals one if individual i possesses all of the necessary attributes and zero if at least one is missing. The item response function (IRF) for the DINA model is based upon the relationship between η_{ij} and Y_{ij} through the introduction of guessing and slipping parameters. That is, guessing is the probability of correctly answering item j when at least one attribute is lacking, $g_j = P\left(Y_{ij} = 1|\eta_{ij} = 0\right)$, and slipping is the probability of an incorrect response for individuals with all of the required attributes, $s_j = P\left(Y_{ij} = 0|\eta_{ij} = 1\right)$. A necessary constraint to ensure monotonicity of the IRF is $0 \leq g_j < 1 - s_j \leq 1$ (Junker & Sijtsma, 2001). Therefore, the probability that subject i correctly answers item j is,

$$P\left(Y_{ij} = 1|\boldsymbol{\alpha}_i, s_j, g_j, \boldsymbol{q}_j\right) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}. \tag{2}$$

The likelihood of observing $\boldsymbol{Y}$ under the assumption of conditional independence on the $N \times K$ matrix of attributes $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)'$ is,

$$p\left(\boldsymbol{Y}|\boldsymbol{s},\boldsymbol{g},\boldsymbol{\pi},\boldsymbol{Q}\right)=\prod_{i=1}^{N}\sum_{\boldsymbol{\alpha}_c\in\{0,1\}^K}\pi_c\prod_{j=1}^{J}\left[(1-s_j)^{\eta_{cj}}g_j^{1-\eta_{cj}}\right]^{y_{ij}}\left[s_j^{\eta_{cj}}(1-g_j)^{1-\eta_{cj}}\right]^{1-y_{ij}},\quad(3)$$

where $\eta_{cj}=\mathcal{I}\left(\boldsymbol{\alpha}_c'\boldsymbol{q}_j=\boldsymbol{q}_j'\boldsymbol{q}_j\right)$, $\boldsymbol{s}=(s_1,\ldots,s_J)'$, and $\boldsymbol{g}=(g_1,\ldots,g_J)'$. Furthermore, $\pi_c=P\left(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_c|\boldsymbol{\pi}\right)$ is the probability of the latent attribute configuration $\boldsymbol{\alpha}_c$, and the 2^K dimensional vector of latent class probabilities can be denoted by $\boldsymbol{\pi}=\left(\pi_1,\ldots,\pi_{2^K}\right)'$.

The definitions for the DINA CDM imply the following Bayesian formulation,

$$\begin{aligned}
Y_{ij}\left|\boldsymbol{\alpha}_i,s_j,g_j,\boldsymbol{q}_j\right.&\sim\text{Bernoulli}\left((1-s_j)^{\eta_{ij}}g_j^{1-\eta_{ij}}\right) &(4)\\
\eta_{ij}&=\mathcal{I}\left(\boldsymbol{\alpha}_i'\boldsymbol{q}_j=\boldsymbol{q}'_j\boldsymbol{q}_j\right),\\
p\left(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_c|\boldsymbol{\pi}\right)&\propto\prod_{c=1}^{2^K}\pi_c^{\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_c)} &(5)\\
0<\pi_c<1,&\ \sum_{c=1}^{C}\pi_c=1,\\
\boldsymbol{\pi}&\sim\text{Dirichlet}\left(\boldsymbol{\delta}_0\right) &(6)\\
\boldsymbol{\delta}_0&=\left(\delta_{01},\ldots,\delta_{0C}\right),\\
p\left(s_j,g_j\right)&\propto s_j^{\alpha_s-1}\left(1-s_j\right)^{\beta_s-1}g_j^{\alpha_g-1}\left(1-g_j\right)^{\beta_g-1}\mathcal{I}\left(0\le g_j<1-s_j\le 1\right), &(7)\\
p\left(\boldsymbol{Q}\right)&\propto\mathcal{I}\left(\boldsymbol{Q}\in\mathcal{Q}\right). &(8)
\end{aligned}$$

Equations 4 through 7 were discussed in prior research (Culpepper, 2015). In particular, Eq. 4 is the DINA model IRF, Eq. 5 is a categorical prior for $\boldsymbol{\alpha}_i$ conditioned upon $\boldsymbol{\pi}$, Eq. 6 is a conjugate Dirichlet prior for latent class probabilities $\boldsymbol{\pi}$, and Eq. 7 is a truncated Beta prior for item parameters that enforces the monotonicity restriction.

The innovation of this study is the specification of a prior for $\boldsymbol{Q}$, so that MCMC can be used to sample $\boldsymbol{Q}$ from the posterior distribution, which is defined as,

$$P(\boldsymbol{Q}|\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})\propto P(\boldsymbol{Y}|\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g},\boldsymbol{Q})\mathcal{I}(\boldsymbol{Q}\in\mathcal{Q}),\quad(9)$$

where

$$P(\boldsymbol{Y}|\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g},\boldsymbol{Q})=\prod_{i=1}^{N}\prod_{j=1}^{J}\left((1-s_j)^{\eta_{ij}}g_j^{1-\eta_{ij}}\right)^{y_{ij}}\left(s_j^{\eta_{ij}}\left(1-g_j\right)^{1-\eta_{ij}}\right)^{1-y_{ij}}.\quad(10)$$

Equation 8 specifies a uniform prior for $\boldsymbol{Q}$ in the space $\mathcal{Q}$ of identified models.

Chen et al. (2015) proved that $\boldsymbol{Q}\in\mathcal{Q}$ for the DINA model and is identified if the following additional conditions are met:

1. For a $J\times J$ permutation matrix $\mathbf{P}$, $\boldsymbol{Q}$ can be expressed as,

$$\mathbf{P}\boldsymbol{Q}=\begin{bmatrix}\mathbf{I}_K\\ \mathbf{I}_K\\ \tilde{\boldsymbol{Q}}\end{bmatrix},\quad(11)$$

where $\mathbf{I}_K$ is a $K \times K$ identity matrix and $\widetilde{\boldsymbol{Q}}$ is a $(J-2K) \times K$ sub-matrix of $\boldsymbol{Q}$ with column k denoted by $\widetilde{\boldsymbol{Q}}_k$.

2. Each skill loads onto at least three items, which implies $\boldsymbol{Q}_k'\mathbf{1}_J \geq 3$ where $\mathbf{1}_J$ is a J dimensional vector of ones. Similarly, if c_k is the kth column margin for $\widetilde{\boldsymbol{Q}}$ (i.e., $c_k = \widetilde{\boldsymbol{Q}}_k'\mathbf{1}_{J-2K}$), this condition is equivalent to $c_k > 0$.
3. Each item loads onto at least one skill such that $\boldsymbol{q}_j'\mathbf{1}_K > 0$.

2.2. *Identifiability and Consistency*

Our exploratory procedures for estimating the DINA $\boldsymbol{Q}$ enforce Chen et al.'s (2015) identifiability conditions. Accordingly, as noted by an anonymous reviewer, our procedures assume that the true $\boldsymbol{Q}$ satisfies the identifiability conditions. It is possible that the true $\boldsymbol{Q}$ for a collection of items does not satisfy Chen et al.'s (2015) identifiability conditions and it might be desirable to use an exploratory method, such as Chung's (2014) unconstrained approach, that freely estimates $\boldsymbol{Q}$ to determine whether $\boldsymbol{Q}$ in fact satisfies the identifiability conditions. However, Chen et al. (2015) showed that a consistent estimator exists for $\boldsymbol{Q}$ only if the aforementioned identifiability conditions are satisfied. Consequently, our Bayesian estimator is consistent if the true $\boldsymbol{Q}$ satisfies the identifiability conditions. In contrast, there is no known consistent estimator if the true $\boldsymbol{Q}$ is outside the space of identified $\boldsymbol{Q}$ matrices.

Restricting $\boldsymbol{Q}$ to the identified space is also important for estimating other model parameters, such as guessing and slipping probabilities. If the model is identified, the item parameters can be estimated by computing averages in the posterior because there is a single mode (up to a column permutation of $\boldsymbol{Q}$) with a one-to-one mapping between values of the item parameters and the corresponding value of the likelihood function. In contrast, if the model parameters are not identified, it is possible that two different sets of values for the item parameters correspond to the same likelihood (i.e., different modes for different values of the item parameters). In this case there is no longer a clear strategy for estimating the item parameters because a posterior average would be computed across multiple modes.

2.3. *Sampling Candidates of $\boldsymbol{Q} \in \mathcal{Q}$*

$\boldsymbol{Q}$ can be sampled in a manner that satisfies the aforementioned identifiability restrictions. That is, the viable candidates for $\boldsymbol{Q}$, $\boldsymbol{Q}_*$, are in the set:

$$\boldsymbol{Q}_* \in \mathcal{Q} = \left\{\boldsymbol{Q} : \boldsymbol{Q}_k'\mathbf{1}_J \geq 3 \;\forall\, k,\; \boldsymbol{q}_j'\mathbf{1}_K > 0 \;\forall\, j,\; (\mathbf{P}\boldsymbol{Q})' = \left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'\right]\right\}. \tag{12}$$

This section describes three approaches for drawing candidates $\boldsymbol{Q}_*$ for a MH sampler.

2.3.1. Independence Sampler The first strategy is to construct independent samples of $\boldsymbol{Q}$ by generating a random $\mathbf{PA}$ matrix and then randomly permuting it with $\mathbf{P}'$ such that $\boldsymbol{Q}_* = \mathbf{P}'(\mathbf{PA})$. Specifically, consider the following steps to draw independent candidates $\boldsymbol{Q}_*$:

1. Fix the first $2K$ rows of $\mathbf{PA}$ to be $[\mathbf{I}_K, \mathbf{I}_K]'$.
2. Draw a sample of $\widetilde{\boldsymbol{Q}}$ that satisfies $\widetilde{\boldsymbol{Q}}_k'\mathbf{1}_{J-2K} > 0$ and $\boldsymbol{q}_j'\mathbf{1}_K > 0$:
   (a) For column k where $k = 1, \ldots, K-1$, sample c_k uniformly in $[1, J-2K]$.
   (b) Then sample $\widetilde{\boldsymbol{Q}}_k$ uniformly from one of the $\binom{J-2K}{c_k}$ configurations.
   (c) For the last column K, count the number of rows that have all zero elements excluding the last column, denoted as l, fix the elements of the l rows in the last column to be 1.
   (d) Sample c_K uniformly in $[\max(1, l), J-2K]$.

(e) Sample $\widetilde{\boldsymbol{Q}}_K$ uniformly from one of the $\binom{J-2K-l}{c_K-l}$ configurations with the l elements fixed in 2(c).

3. Draw a $J \times J$ permutation matrix $\mathbf{P} = \left(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_J}\right)$ where $(i_1, \ldots, i_J)$ is a random permutation of $(1, \ldots, J)$ and $\mathbf{e}_i$ is the vector of

$$\mathbf{e}_i(l) = \begin{cases} 0 & l \neq i \\ 1 & l = i \end{cases}. \tag{13}$$

4. Set the candidate as $\boldsymbol{Q}_* = \mathbf{P}'(\mathbf{PA})$.

2.3.2. Dependence Sampler #1 (DS1) The independence sampler described above ensures that each $\boldsymbol{Q}_*$ will be independent of the previous draw. A disadvantage of using an independence sampler is that it has a relatively low acceptance rate and likely requires more iterations to explore the posterior distribution in the MH steps. To introduce some dependence, we propose a random walk movement from the current state:

1. Initialize with any identified $\boldsymbol{Q}^{(0)} \in \mathcal{Q}$.
2. Update candidate $\boldsymbol{Q}^{(t)}$ at iteration t under identifiability constraints:

(a) Randomly select a column, say the k-th column.
(b) Let k_1 be the number of rows that have a single 1 at column k. Fix the 1's of those k_1 elements in the k-th column.
(c) For every $i \neq k, i = 1, \ldots, K$, let l_i denote the number of rows that equal to 1 at the i-th column and equal to 0 at every column other than column i and column k. Because of the identifiability constraints, $l_i \geq 2$. Randomly select 2 elements from those l_i rows in the k-th column and fix them as 0.
(d) There are $J - k_1 - 2(K-1)$ remaining elements in $\boldsymbol{Q}_k$ that are not fixed in 2(b) or 2(c) and the column sum c_k should be at least 3. Sample $\boldsymbol{Q}_{*k}$ uniformly from one of the $\sum_{c=\max\{3,k_1\}}^{J-2(K-1)} \binom{J-k_1-2(K-1)}{c-k_1}$ configurations for the remaining elements in $\boldsymbol{Q}_{k,B}$.

2.3.3. Dependence Sampler #2 (DS2) Preliminary Monte Carlo studies suggested that the aforementioned dependence sampler accepts candidates with a low probability. In practice, we can adapt the range of each movement to adjust the acceptance rate, that is to say, we can update a subset of size B in a selected column:

1. Initialize with any identified $\boldsymbol{Q}^{(0)} \in \mathcal{Q}$.
2. Update candidate $\boldsymbol{Q}^{(t)}$ under identifiability constraints:

(a) Randomly select one column and a subset of size B within it, say $\boldsymbol{Q}_{k,B}$ in the k-th column.
(b) Let k_1 be the number of rows within $\boldsymbol{Q}_{k,B}$ that have a 1 at column k and zeros in the other columns. Fix the k_1 elements of those rows as 1 in $\boldsymbol{Q}_{k,B}$.
(c) For every $i \neq k, i = 1, \ldots, K$, let l_i be the number of rows outside $\boldsymbol{Q}_{k,B}$ that have a single 1 at the i-th column; let b_i be the number of rows in $\boldsymbol{Q}_{k,B}$ that equal to 1 at the i-th column and equal to 0 for all columns other than columns i and k. If $l_i < 2$, randomly select $2 - l_i$ elements of the b_i rows in $\boldsymbol{Q}_{k,B}$ and fix as 0.
(d) Let $0 \leq k_0 \leq \min\{B - k_1, 2K - 2\}$ be the number of elements in $\boldsymbol{Q}_{k,B}$ that are fixed as 0 in 2(c).

(e) Let m be the number of elements equal to 1 in the k-th column outside $\boldsymbol{Q}_{k,B}$.

(f) There are $B - k_0 - k_1$ remaining elements in $\boldsymbol{Q}_{k,B}$ that are not fixed in 2(b) or 2(c) and the column sum c should be at least 3. Sample $\boldsymbol{Q}_{*k,B}$ uniformly from one of the $\sum_{c=\max\{3,m+k_1\}}^{B-k_0+m} \binom{B-k_0-k_1}{c-m-k_1}$ configurations for the remaining elements in $\boldsymbol{Q}_{k,B}$.

Example. We provide a step-by-step illustration of DS2 for steps 2(a) to 2(f) with a simple example using the following matrix $\boldsymbol{Q} \in \mathcal{Q}$:

$$\boldsymbol{Q} = \begin{pmatrix} 1 & \mathbf{0}(\mathrm{I}) & 0 \\ 0 & \mathbf{1}(\mathrm{II}) & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & \mathbf{0}(\mathrm{III}) & 1 \\ 0 & \mathbf{1}(\mathrm{IV}) & 0 \\ 0 & \mathbf{0}(\mathrm{V}) & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{Q}_* = \begin{pmatrix} 1 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & \mathbf{1} & 1 \\ 0 & \mathbf{1} & 0 \\ 0 & \mathbf{0} & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Consider the following steps to perform the DS2:

(a) Suppose $\boldsymbol{Q}_{2,5}$ in the 2nd column is randomly selected to be updated. Note that entries of $\boldsymbol{Q}_{2,5}$ are indexed by Roman numerals.

(b) Entry (IV) in $\boldsymbol{Q}_{2,5}$ must be fixed to 1 according to 2(b), which implies that $k_1 = 1$.

(c) Examining rows of $\boldsymbol{Q}$ outside of $\boldsymbol{Q}_{2,5}$ shows that the third and fifth rows have a single one in the first and third columns, respectively, so $l_1 = 1$ and $l_3 = 1$. The first row of $\boldsymbol{Q}$ is the only row within $\boldsymbol{Q}_{2,5}$ where there is a 1 in the first column and a zero in the third, which implies $b_1 = 1$. Similarly, $b_3 = 2$ given that the second and eighth rows of $\boldsymbol{Q}$ in $\boldsymbol{Q}_{2,5}$ have a 1 in column three and zero in column one. $l_1 = 1 < 2$ and $b_1 = 1$ imply that entry (I) is fixed to 0. $l_3 = 1$ and $b_3 = 2$, so one of entry (II) or (V) is randomly set to 0. For purposes of illustration suppose entry (V) is randomly set to 0.

(d) In the previous step, entry (I) and (V) were fixed to 0, so $k_0 = 2$.

(e) $m = 2$ for the ones in the fourth and last position of column two.

(f) The column sum c needs to be at least 3. Based on the fact that $m + k_1 = 3$, the possible configurations for entries (II) and (III) are (0, 0), (0, 1), (1, 0) and (1, 1). In this case we sample the configuration with entries (II) and (III) equal to 0 and 1, respectively.

2.4. *Irreducibility and Symmetry of Proposed Transition Function*

A necessary condition for MH sampling is that the Markov chain is irreducible, in that any $\boldsymbol{Q} \in \mathcal{Q}$ can be reached from any other state. It is trivial to show that the independence sampler can propose every possible $\boldsymbol{Q} \in \mathcal{Q}$. For dependence samplers, we prove the irreducibility of the proposal that updates only one element at a time. This would imply irreducibility of both DS1 and DS2, as shown in the remarks after the proof.

Theorem 1. *When $J - 2K \geq 2$, the dependence sampler #2 with $B = 1$ can reach every possible candidate $\boldsymbol{Q}_* \in \mathcal{Q}$ in finite steps, regardless of the current state $\boldsymbol{Q}^{(t)}$ at time t.*

Proof. It is obvious the dependence sampler #2 ensures the candidate $\boldsymbol{Q}$ satisfies the identifiability constraints. When $B = 1$, DS2 is equivalent to updating a randomly chosen element in $\boldsymbol{Q}^{(t)}$. If the selected element is 1 from the row vector $\boldsymbol{e}_k$, it remains unchanged in step 2(b); if it is 0 from the row vector $\boldsymbol{e}_k$, it remains unchanged in step 2(c) if there are only 2 $\boldsymbol{e}_k$ in $\boldsymbol{Q}$. Alternatively, the selected element can be updated to 1 later in 2(f) if there are more than 2 $\boldsymbol{e}_k$ in $\boldsymbol{Q}$. If the selected element is 1 in a column with column sum equal to 3, it remains unchanged due to step 2(f). For all other positions, the selected element can be updated to either 1 or 0.

Because any $\boldsymbol{Q} \in \mathcal{Q}$ can be expressed as $\boldsymbol{Q} = \mathbf{P}'\left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'\right]'$, if we can show that any $\boldsymbol{Q}_* = \left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'_*\right]'$ can be reached from $\boldsymbol{Q}^{(t)} = \left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'\right]'$ and that any row permutation is allowed, then the irreducibility is proved. We first claim that if current state $\boldsymbol{Q}^{(t)} = \left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'\right]'$, any candidate $\boldsymbol{Q}_* = \left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'_*\right]'$ can be reached by the proposed transition strategy. With the additional condition that $J - 2K \geq 2$, we can choose to update elements in one row of $\widetilde{\boldsymbol{Q}}_*$, say row l, sequentially to all 1's in finite steps. This is always feasible because if row $l = \boldsymbol{e}_k$, there exist two other $\boldsymbol{e}_k$ rows in $\boldsymbol{Q}^{(t)}$, then 0's in row l can be updated to 1 in step 2(f); if row l is in any other form, 0's in it can be updated to 1. Since the new row l ensures every column sum is at least 3, the elements in the remaining rows of $\widetilde{\boldsymbol{Q}}$ can be updated sequentially to any possible configuration. We can choose to update them to be the same as the corresponding rows in $\widetilde{\boldsymbol{Q}}_*$. After that, row l can also be updated sequentially to be the same as row l in $\widetilde{\boldsymbol{Q}}_*$, so that $\widetilde{\boldsymbol{Q}}_*$ satisfying identifiability constraints 2 and 3 can be reached. This indicates that the Markov chain can reach all possible candidates where the position of the two identity matrices remains the same as the current state.

It is left to show that any random row permutation can be achieved by this transition, and it suffices to show that any pair of rows i and j can exchange. Without loss of generality, let the current state $\boldsymbol{Q}^{(t)} = \left[\mathbf{I}_K, \mathbf{I}_K, \widetilde{\boldsymbol{Q}}'\right]'$, and we need to prove that for any i and j, $i \neq j$, candidate $\boldsymbol{Q}_*$ of rows i and j swapped can be achieved through the transition procedure. If both $i, j > 2K$, i.e., they are rows in $\widetilde{\mathbf{Q}}$, based on the argument above, $\boldsymbol{Q}_*$ can be reached.

Now consider the case where $i \leq 2K$, $j > 2K$, i.e., one row is in $[\mathbf{I}_K, \mathbf{I}_K]'$ and one is in $\widetilde{\boldsymbol{Q}}$. Suppose row $i = \boldsymbol{e}_k$ that has a single 1 in the k-th column, and row j could be any other nonzero row vector. There exists another row $m = \boldsymbol{e}_k$ from the identity matrix, $m \neq i, j$. We can start the transition by updating another row l from $\widetilde{\boldsymbol{Q}}$, $l \neq j$, sequentially to be a vector of all 1's. The new row l ensures that all column sum is at least 3, and then, row j can be updated to $\boldsymbol{e}_k$ sequentially. If the original row j equals to 0 in column k, we can choose to update the elements in other columns of row i to be the same as the original row j, and update 1 in column k of row i to be 0 at last. Any 0 in row i can be updated to 1 because there exist two other rows m and j of form $\boldsymbol{e}_k$. If the original row j equals to 1 in column k, we can update the remaining elements in row i sequentially to elements of the original row j; thus, we have switched the rows i and j, and then, we can update row l back to its original configuration.

If both $i, j \leq 2K$, we can find another row l from $\widetilde{\boldsymbol{Q}}$. As noted above, we can switch row i with row l, switch row j with the current row i, and then, switch the current row j with the current row l; thus, we have switched rows i and j. □

DS2 with $B > 1$ can reach every possible state with similar procedures. Every transition of updating 1 element at a time is a special transition of DS2 with size B, where we choose B elements but just update one of them. Therefore, any state can be sequentially achieved as shown in Theorem 1 through DS2 with size $B > 1$. Similarly, updating a single element is also a special transition of DS1. Therefore, DS1 and DS2 guarantee that the Markov chain is irreducible.

When $J - 2K = 1$, the only possible form of $\boldsymbol{Q} \in \mathcal{Q}$ is that $\boldsymbol{Q} = \mathbf{P}[\mathbf{I}_K, \mathbf{I}_K, \mathbf{1}_K]'$, where $\mathbf{P}$ is a $J \times J$ permutation matrix. To sample random candidates from $\mathcal{Q}$, we can simply switch two rows in every transition so as to draw dependent samples. Furthermore, we will show that the transition proposals DS1 and DS2 are both symmetric.

Theorem 2. *DS2 proposes a symmetric transition from current $\boldsymbol{Q}^{(t)}$ to any proposed candidate $\boldsymbol{Q}_*$, i.e., $T(\boldsymbol{Q}_*, \boldsymbol{Q}^{(t)}) = T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*)$ where $T(x, y)$ is the transition probability from state x to y.*

Proof. It is obvious that for $\boldsymbol{Q}_* = \boldsymbol{Q}^{(t)}$, $T(\boldsymbol{Q}_*, \boldsymbol{Q}^{(t)}) = T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*)$ always holds. We just need to prove the equation still holds when $\boldsymbol{Q}_* \neq \boldsymbol{Q}^{(t)}$.

When $B = 1$, if $\boldsymbol{Q}_* \neq \boldsymbol{Q}^{(t)}$, then the element selected cannot be any of the following three cases: 1) a 1 from the row vector $\boldsymbol{e}_k$, 2) a 1 from the column with column sum exactly 3, and 3) a 0 from the row vector $\boldsymbol{e}_k$ when there are only two $\boldsymbol{e}_k$'s in $\boldsymbol{Q}$, because for all these cases the selected element would remain unchanged based on DS2. Other than the three cases, the selected block can have two configurations: 0 or 1, and then, $T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*) = \frac{1}{2} = T(\boldsymbol{Q}_*, \boldsymbol{Q}^{(t)})$ holds.

For $B > 1$, let p_i denote a specific path of updating steps in DS2 that changes $\boldsymbol{Q}^{(t)}$ to $\boldsymbol{Q}_*$. Then, the transition probability of a given path p_i based on steps 2(a), 2(c), and 2(f) should be,

$$T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*|p_i) = \overbrace{\frac{1}{K\binom{J}{B}}}^{2(a)} \times \overbrace{\frac{1}{\prod_{i \neq k} \binom{b_i}{2-l_i}^{\mathcal{I}(l_i<2)}}}^{2(c)} \times \overbrace{\frac{1}{\sum_{c=\max\{3, m+k_1\}}^{B-k_0+m} \binom{B-k_0-k_1}{c-m-k_1}}}^{2(f)}. \tag{14}$$

When the block is chosen in step 2(a), k_1 in step 2(b) is determined, and b_i and l_i in step 2(c) are also determined because the number positions need to be fixed as 0 or 1 remain the same for the given block. So k_1, b_i, l_i and k_0 are all determined by the block, and m is also determined because m is the number of 1's outside the block. Therefore, the transition probability (14) remains the same for all the paths from $\boldsymbol{Q}^{(t)}$ to $\boldsymbol{Q}_*$.

For every path p_i changing $\boldsymbol{Q}^{(t)}$ to $\boldsymbol{Q}_*$ with the chosen block B_i and chosen positions in steps 2(b) and 2(c), there is a corresponding path q_i that chooses the same block and positions and changes $\boldsymbol{Q}_*$ to $\boldsymbol{Q}^{(t)}$. Based on the argument above, the transition probability of path q_i is the same as (14), i.e., $T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*|p_i) = T(\boldsymbol{Q}_*, \boldsymbol{Q}^{(t)}|q_i)$. So the transition probability of DS2 satisfies:

$$T(\boldsymbol{Q}_*, \boldsymbol{Q}^{(t)}) = \sum_{\substack{q_i \in \{\text{all paths from} \\ \boldsymbol{Q}_* \text{ to } \boldsymbol{Q}^{(t)}\}}} T(\boldsymbol{Q}_*, \boldsymbol{Q}^{(t)}|q_i) = \sum_{\substack{p_i \in \{\text{all paths from} \\ \boldsymbol{Q}^{(t)} \text{ to } \boldsymbol{Q}_*\}}} T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*|p_i) = T(\boldsymbol{Q}^{(t)}, \boldsymbol{Q}_*). \tag{15}$$

□

2.5. *Metropolis Sampling for $\boldsymbol{Q}$*

The proposed transition functions to generate matrix $\boldsymbol{Q}$ are irreducible and symmetric, which implies Metropolis sampling can be implemented to draw $\boldsymbol{Q}$ from the posterior $p(\boldsymbol{Q}|\boldsymbol{Y}, \boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\alpha})$. Given the current iteration at time t:

1. Draw a candidate $\boldsymbol{Q}_*$ through DS2.
2. Draw $U \sim Uniform(0, 1)$ and update

$$\boldsymbol{Q}^{(t)} = \begin{cases} \boldsymbol{Q}_* & \text{if } U \leq r(\boldsymbol{Q}^{(t-1)}, \boldsymbol{Q}_*) \\ \boldsymbol{Q}^{(t-1)} & \text{otherwise} \end{cases}$$

where $r(\boldsymbol{Q}^{(t-1)}, \boldsymbol{Q}_*) = \min\left\{1, \frac{p(\boldsymbol{Q}_*|\boldsymbol{Y},\boldsymbol{s},\boldsymbol{g},\boldsymbol{\alpha})}{p(\boldsymbol{Q}^{(t-1)}|\boldsymbol{Y},\boldsymbol{s},\boldsymbol{g},\boldsymbol{\alpha})}\right\}$.

Because of the ratio, we can use the posteriors with unknown normalizing constant to update $\boldsymbol{Q}$.

Furthermore, a Metropolis within Gibbs algorithm can be implemented to sample the parameters $\boldsymbol{Q}$, $\boldsymbol{s}$, $\boldsymbol{g}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\pi}$ from the posterior distribution. The method is described as follows:

1. Randomly initialize $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\pi}^{(0)}, \boldsymbol{Q}^{(0)} \in \mathcal{Q}, \boldsymbol{g}^{(0)}$ and $\boldsymbol{s}^{(0)}$ such that $0 \leq g_j^{(0)} \leq 1-s_j^{(0)} \leq 1$ for $j = 1, \cdots, J$.
2. At iteration t, update $\boldsymbol{s}^{(t)}$, $\boldsymbol{g}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$, $\boldsymbol{\pi}^{(t)}$, and $\boldsymbol{Q}^{(t)}$:
   (a) Update $\boldsymbol{g}^{(t)} \sim p\left(\boldsymbol{g}|\boldsymbol{Y}, \boldsymbol{s}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{Q}^{(t-1)}\right)$.
   (b) Update $\boldsymbol{s}^{(t)} \sim p\left(\boldsymbol{s}|\boldsymbol{Y}, \boldsymbol{g}^{(t)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{Q}^{(t-1)}\right)$.
   (c) Update $\boldsymbol{\alpha}^{(t)} \sim p\left(\boldsymbol{\alpha}|\boldsymbol{Y}, \boldsymbol{s}^{(t)}, \boldsymbol{g}^{(t)}, \boldsymbol{\pi}^{(t-1)}, \boldsymbol{Q}^{(t-1)}\right)$.
   (d) Update $\boldsymbol{\pi}^{(t)} \sim p\left(\boldsymbol{\pi}|\boldsymbol{\alpha}^{(t)}\right)$.
   (e) Update $\boldsymbol{Q}^{(t)}$ using Metropolis steps as aforementioned.

Note that steps 2(a) through 2(d) are Gibbs updates with full conditional distributions discussed in Culpepper (2015).

2.6. *Constrained Gibbs Sampling for* $\boldsymbol{Q}$

Theorem 1 has proved that updating a single element in $\boldsymbol{Q}$ through DS2 ensures the irreducibility of the Markov chain. Based on the sampling strategy of DS2 with special case of $B = 1$, we propose an alternative Gibbs sampling algorithm for $\boldsymbol{Q}$ subject to identifiability constraints. Given the current iteration at time t the constrained Gibbs sampling algorithm proceeds as follows:

1. Update $\boldsymbol{s}^{(t)}$, $\boldsymbol{g}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$, $\boldsymbol{\pi}^{(t)}$ through Gibbs sampling same as aforementioned.
2. Update $q_{jk}^{(t)}$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ through Gibbs sampling:
   (a) If $q_{jk}^{(t-1)}$ is in one of the following three positions: 1) a 1 from the row vector $\boldsymbol{e}_k$, 2) a 1 from the column with column sum exactly 3, and 3) a 0 from the row vector $\boldsymbol{e}_k$ where there are only two $\boldsymbol{e}_k$'s in the current $\boldsymbol{Q}$, $q_{jk}^{(t)}$ remains the same as $q_{jk}^{(t-1)}$.
   (b) Otherwise, $q_{jk}^{(t-1)}$ can be updated to 0 or 1. Sample $q_{jk}^{(t)} = i$ with probability proportional to $p(\boldsymbol{Y}|\boldsymbol{s}^{(t)}, \boldsymbol{g}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{Q}_{new}^{(t)}, q_{jk}^{(t)} = i, \boldsymbol{Q}_{old}^{(t-1)})$, $i = 0, 1$, where $\boldsymbol{Q}_{new}^{(t)}$ refers to the elements of the current $\boldsymbol{Q}$ that have already been updated, and $\boldsymbol{Q}_{old}^{(t-1)}$ refers to the elements of the current $\boldsymbol{Q}$ that have not been updated.

As shown in Theorem 1, the constrained Gibbs sampling ensures that the Markov chain is irreducible and the sampled $\boldsymbol{Q}$ always satisfies identifiability constraints.

2.7. *Computing* $\hat{\boldsymbol{Q}}$*, Column Permuting*

In the Monte Carlo study, the estimated $\hat{\boldsymbol{Q}}$ is taken as the mode of the sampled $\boldsymbol{Q}$ matrices for posterior inference. In order to find the mode, we use a bijective mapping from binary sequence to integer number (e.g., see Chung, 2014; von Davier, 2014), and transform the column vectors of sampled $\boldsymbol{Q}$ into integer numbers. For instance, $\boldsymbol{v}'\boldsymbol{Q}_k$ is a bijection for $\boldsymbol{v}' = \left(2^{J-1}, 2^{J-2}, \ldots, 2, 1\right)$. Every $\boldsymbol{Q}$ can be recorded as a K dimensional integer vector. We permute the vector in a decreasing order so that column permutation of $\boldsymbol{Q}$ is allowed, and count the occurrence of each vector. The estimated $\hat{\boldsymbol{Q}}$ corresponds to a vector that appears most, and it satisfies the identifiability constraints. Note that finding the posterior mode in this manner eliminates the problem of columns permuting when summarizing $\hat{\boldsymbol{Q}}$.

3. Monte Carlo Simulation Studies

This section reports results from a Monte Carlo simulation study on the relative performance of the proposed constrained MCMC procedures for estimating $\boldsymbol{Q}$ with Chung's (2014) unconstrained Gibbs sampler.

3.1. Overview

The Monte Carlo simulation study examined recovery of $\boldsymbol{Q}$ for different sample sizes (i.e., N = 500, 1000, 2000, and 4000), numbers of attributes (i.e., K = 3, 4), and correlations among the latent attributes (i.e., ρ = 0, 0.05, 0.15, and 0.25). For the $\rho = 0$ case attribute profiles were generated uniformly from the 2^K configurations. For cases where $\rho > 0$ attribute dependence is introduced using the method of Chiu, Douglas, and Li (2009). Suppose $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ follows multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ where the covariance matrix $\boldsymbol{\Sigma}$ has unit variance and a common correlation ρ:

$$\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_K + \rho\mathbf{1}\mathbf{1}'.$$

We consider the situation where $\rho = 0.05, 0.15$ and 0.25 and the attribute α for each subject is given by

$$\alpha_k = \begin{cases} 1 & \text{if } \theta_k \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad k = 1, \ldots, K.$$

The simulation study specified the true unknown $\boldsymbol{Q}$ matrices, $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, as:

$$\boldsymbol{Q}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{Q}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

which both satisfy the aforementioned identifiability constraints.

We generated 100 independent simulated datasets for each condition to assess the performance of three MCMC methods. The independence sampler and the DS1 sampler mixed more slowly than the DS2 sampler with acceptance rates for $\boldsymbol{Q}$ less than 0.5% and 5%, respectively. In contrast, the acceptance rate for DS2 proposal was between 18% and 25%. In preliminary investigations we considered $B = K$ and $B = 2K$ and the simulation study and application employ $B = 2K$ rather than $B = K$ given the acceptance rate was higher and approximately 20%.

The simulation study also examined the performance of the constrained Gibbs sampler with block size of $B = 1$ and Chung's (2014) unconstrained Gibbs sampler. In all tables of results, "MH" refers to our method of Metropolis sampling DS2 with a block size of $B = 2K$, "CGibbs" refers to our developed constrained Gibbs sampler that sequentially updates each element of $\boldsymbol{Q}$, and "Gibbs" refers to the unconstrained sampler. For each replication we ran a Markov chain of length 30,000 with a burn-in length of 15,000 for all MCMC methods. The starting value for $\boldsymbol{s}$, $\boldsymbol{g}$, $\boldsymbol{\pi}$ is sampled from the prior distributions, and the starting value of $\boldsymbol{Q}$ is randomly drawn from the sample space $\mathcal{Q}$. Each replication recorded the $\boldsymbol{Q}$ that maximized the posterior mode to compute the overall accuracy as well as entry-wise accuracy of the MCMC methods.

For the diagnostic of convergence, we used Geweke's diagnostic (Geweke, 1992) on -2 times log of the posterior density to determine the burn-in period (Cowles & Carlin, 1996). For iterations from 15,000 to 30,000 in chains from MH and CGibbs sampler, the reported Z-scores are -0.218 and -0.628, respectively, for $K = 3$, $N = 4000$, and -1.36 and -0.705, respectively, for $K = 4$, $N = 4000$, indicating that there is no evidence against the convergence of the joint posterior.

3.2. *Results*

Table 1 reports the accuracy of the three MCMC methods (i.e., MH, CGibbs, and Gibbs) across N, ρ, and K. Specifically, Table 1 presents the number of replications where the Bayesian methods correctly recovered all elements of $\boldsymbol{Q}$ (i.e., $\hat{\boldsymbol{Q}} = \boldsymbol{Q}$) in addition to the average entry-wise accuracy, which was computed as the percentage of correctly estimated entries. In general, our methods (i.e., MH and CGibbs) successfully estimated $\boldsymbol{Q}$ and demonstrated greater accuracy than the unconstrained Gibbs sampler.

The CGibbs approach tended to provide more accurate estimates of $\boldsymbol{Q}$ for smaller sample sizes (i.e., $N = 500$ and 1000), whereas the MH and Gibbs samplers were less accurate for smaller sample sizes and the difference in recovery rates was more pronounced for $K = 4$. It is important to note that the three Bayesian estimators outperformed Chen et al.'s (2015) L_1 penalty approach for sample sizes of 500 and 1000. For example, for the $\rho = 0$ and $N = 500$ case Chen et al.'s (2015) L_1 approach correctly recovered 38 out of 100 $\boldsymbol{Q}$ matrices for $K = 3$ and 20 $\boldsymbol{Q}$ matrices for $K = 4$. In contrast, all three Bayesian methods recovered more than 90% of $\boldsymbol{Q}$ matrices for $K = 3$ and the CGibbs approach recovered 97 out of 100 $\boldsymbol{Q}$ matrices when $K = 4$.

Sample size impacted the performance of the Bayesian methods. In fact, Chen et al.'s (2015) L_1 outperformed the Bayesian methods for sample sizes of $N = 2000$ and 4000. The CGibbs approach tended to be less accurate when $N = 4000$; however, the proposed MH sampler outperformed the unconstrained Gibbs sampler in these cases. Table 1 provides evidence that $\boldsymbol{Q}$ was most difficult to recover when $N = 4000$ and $K = 4$. Both the MH and Gibbs samplers became less accurate as ρ increased. For $K = 4$ and $\rho > 0$ cases, the constrained Gibbs sampler consistently reports the best recovery rates for all correlation settings.

Figures 1 and 2 report the mean squared error (MSE) of the estimated guessing and slipping item parameters. The constrained Gibbs sampler tends to have the smallest MSEs among the three Bayesian algorithms, and the unconstrained Gibbs sampler has the largest MSEs. For instance, Fig. 1 reports MSE for the $\rho = 0$ case and shows that the MH and CGibbs algorithms produced more accurate item parameter estimates than the unconstrained Gibbs method. For $N = 2000$ and $K = 3$, the MSEs of guessing parameters for items requiring single attributes are larger for the unconstrained Gibbs than items requiring multiple attributes. Additionally, the unconstrained Gibbs has larger MSE for slipping parameters in the $K = 4$ case. Furthermore, Fig. 2 reports the item parameter MSEs for $\rho = 0.05$ and 0.25. Similar to the independent case, the MSEs

TABLE 1.
Summary of Recovery Rate of $\boldsymbol{Q}$ by Sample Size (N), Number of Attributes (K), and Attribute Dependence (ρ) for the Metropolis–Hastings (MH), constrained Gibbs (CGibbs), and unconstrained Gibbs (Gibbs) Samplers.

K	N	ρ	$\hat{\boldsymbol{Q}} = \boldsymbol{Q}$			Entry-wise Average		
			MH	CGibbs	Gibbs	MH	CGibbs	Gibbs
3	500	0	91	95	94	95.89	98.22	97.12
3	1000	0	94	99	95	97.59	99.61	97.93
3	2000	0	96	92	90	98.52	97.33	95.11
3	4000	0	98	88	91	99.56	96.33	95.42
3	500	0.05	87	99	82	93.25	99.63	92.96
3	1000	0.05	92	99	90	94.67	99.64	94.41
3	2000	0.05	93	95	90	97.89	98.52	95.18
3	4000	0.05	94	88	88	98.78	96.33	95.20
3	500	0.15	93	99	91	97.42	99.63	95.44
3	1000	0.15	95	99	92	98.51	99.67	95.89
3	2000	0.15	96	95	94	99.29	98.10	97.65
3	4000	0.15	95	91	89	99.35	97.39	94.04
3	500	0.25	92	99	90	98.70	99.98	94.55
3	1000	0.25	96	98	94	99.42	99.65	96.20
3	2000	0.25	96	94	93	99.92	98.10	95.83
3	4000	0.25	97	89	91	99.88	95.57	95.76
4	500	0	60	97	59	91.42	99.92	89.81
4	1000	0	67	91	67	93.11	97.29	92.24
4	2000	0	76	79	73	94.52	94.54	93.96
4	4000	0	87	51	82	95.50	87.02	95.25
4	500	0.05	37	98	40	82.00	99.13	83.39
4	1000	0.05	52	94	58	88.57	98.36	89.87
4	2000	0.05	48	90	53	88.02	97.44	89.28
4	4000	0.05	53	53	51	89.50	89.62	88.94
4	500	0.15	34	96	40	81.61	99.43	83.09
4	1000	0.15	44	88	60	84.87	96.83	92.00
4	2000	0.15	55	90	53	89.13	97.17	88.97
4	4000	0.15	56	74	52	89.92	91.64	89.19
4	500	0.25	35	97	37	81.78	99.58	82.94
4	1000	0.25	43	96	58	84.67	98.57	90.24
4	2000	0.25	55	85	54	89.87	95.64	89.80
4	4000	0.25	55	79	51	90.09	94.07	89.30

$\hat{\boldsymbol{Q}} = \boldsymbol{Q}$ refers to the number of correctly recovered $\boldsymbol{Q}$ out of 100 replications. Entry-wise average refers to average percentage of correctly recovered entries of $\boldsymbol{Q}$.

for the unconstrained sampler are larger than for the constrained MH and CGibbs methods. One explanation for the difference in recovery of item parameters can be attributed to the fact that our methods restricted $\boldsymbol{Q}$ to be identifiable, whereas the unconstrained sampler can visit values for $\boldsymbol{Q}$ where the item parameters are not identified.

4. Application: Fraction-Subtraction Data

In this section we compare our methods and the unconstrained Gibbs sampler (Chung, 2014) in an application to estimate $\boldsymbol{Q}$ for Tatsuoka's fraction-subtraction data set (C. Tatsuoka, 2002; K. K. Tatsuoka, 1984). The data set contains responses to $J = 20$ items from $N = 536$ middle

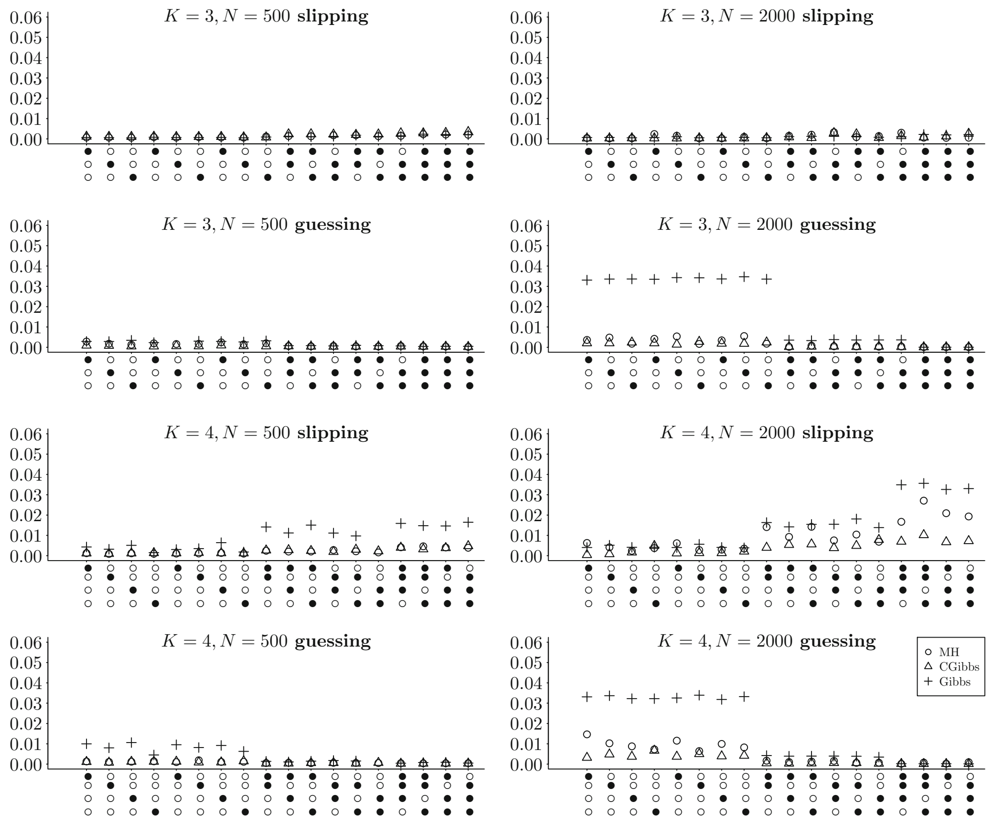

FIGURE 1.
Mean squared errors for slipping and guessing parameters for independent attributes with $K = 3, 4$, and $N = 500, 2000$. The x-axis labels refer to the corresponding entries in $\boldsymbol{Q}$, *open circle* refers to 0 and *filled circle* refers to 1. The legend at the last plot is the same for all plots.

school students and has been widely analyzed (e.g., Chen et al., 2015; DeCarlo, 2010, 2012; de la Torre, 2008; de la Torre & Douglas, 2004, 2008; Mislevy & Wilson, 1996). In fact, Table 2 presents an expert derived $\boldsymbol{Q}$ matrix with the following eight skills:

1. Convert a whole number to fraction,
2. Separate a whole number from fraction,
3. Simplify before subtraction,
4. Find a common denominator,
5. Borrow from the whole number part,
6. Column borrow to subtract the second numerator from the first,
7. Subtract numerators,
8. Reduce answers to simplest form.

In Table 2 almost all of the items require Attribute 7, and several attributes are required at the same time, indicating that 8 attributes might be too many for 20 items. Hence, this section reports an estimate of $\boldsymbol{Q}$ for $K = 3$ and 4.

Table 3 presents the estimated $\hat{\boldsymbol{Q}}$, slipping, and guessing parameters from both the MH and unconstrained Gibbs methods for $K = 3$. Note that $\hat{\boldsymbol{Q}}$ for the constrained Gibbs sampler was the same as reported in Table 3 for the MH sampler. Most entries of the estimated $\hat{\boldsymbol{Q}}$ are the

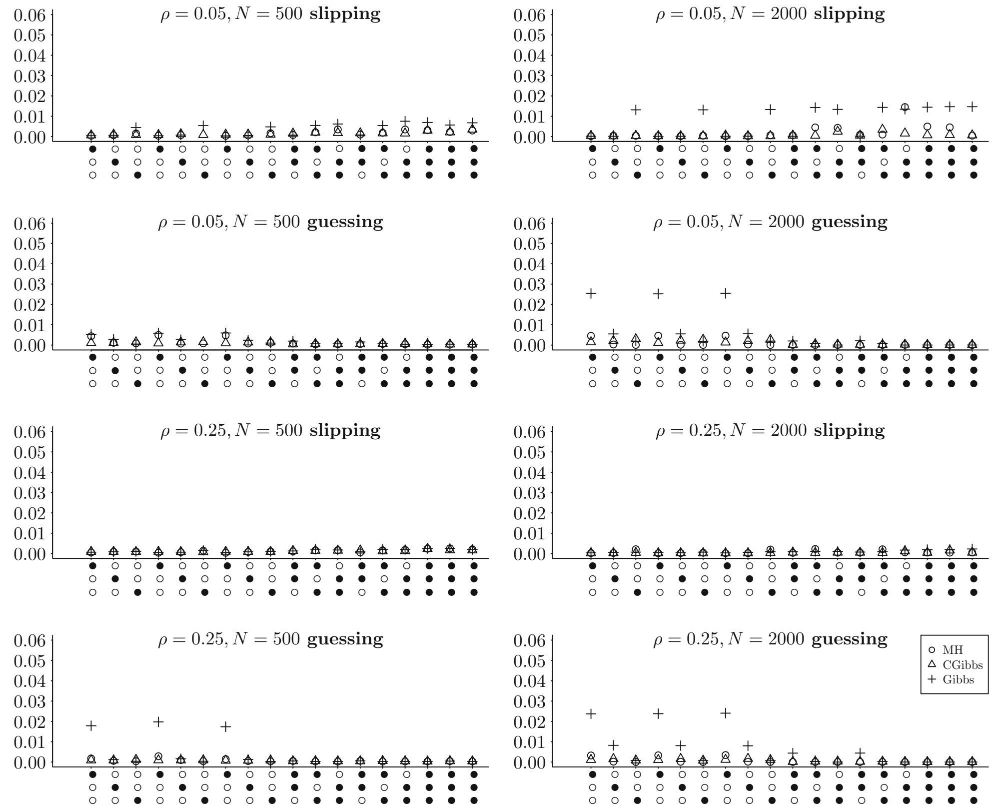

FIGURE 2.
Mean squared error for slipping and guessing parameters for dependent attributes with $K = 3$, $\rho = 0.05, 0.25$, and $N = 500, 2000$. The x-axis labels refer to the corresponding entries in $\boldsymbol{Q}$, *open circle* refers to 0 and *filled circle* refers to 1. The legend at the last plot is the same for all plots.

same for the constrained and unconstrained methods; however, an important difference is that the unconstrained Gibbs sampler did not yield a $\hat{\boldsymbol{Q}}$ that satisfies the identifiability constraints. By looking at items requiring only one attribute, we can interpret the three attributes as (1) finding common denominator, (2) borrowing from the integer part and (3) applying subtraction for integer and fraction parts separately. For items $3\frac{1}{2} - 2\frac{3}{2}$ and $4\frac{4}{12} - 2\frac{7}{12}$ our method suggests that these items only require the second attribute, whereas the unconstrained sampler suggests two attributes are needed for success. Similarly, our method suggests the item $3 - 2\frac{1}{5}$ requires the first and last attributes, whereas the unconstrained sampler estimated only a single attribute. The estimated slipping and guessing parameters intuitively correspond to the easiness of each item. For instance, item $\frac{2}{3} - \frac{2}{3}$ and item $3\frac{7}{8} - 2$ have large guessing parameters, and items $3\frac{3}{8} - 2\frac{5}{6}$ and $4 - 1\frac{4}{3}$ have large slipping parameters. Furthermore, most items requiring a single attribute tend to have smaller guessing parameters. Table 3 shows that the differences in estimates for $\boldsymbol{Q}$ translates into differences in item parameters between the two methods. For instance, the estimated slipping parameters of items $3 - 2\frac{1}{5}$, $3\frac{7}{8} - 2$, and $2 - \frac{1}{3}$ significantly differ between the two methods.

Table 4 presents the results for $K = 4$. Again, constrained Gibbs sampling estimated the same $\hat{\boldsymbol{Q}}$ as MH sampling and the results for MH and unconstrained Gibbs sampling are similar. Our method identifies items $3 - 2\frac{1}{5}$ and $2 - \frac{1}{3}$ as testing the same single attribute, whereas the unconstrained approach indicates the latter item involves more attributes. The estimated slipping parameters of items $3 - 2\frac{1}{5}$ and $3\frac{7}{8} - 2$ still differ here.

TABLE 2.
Expert Specified Fraction-Subtraction Data $\boldsymbol{Q}$ Matrix with 8 attributes.

	1	2	3	4	5	6	7	8
$\frac{5}{3}-\frac{3}{4}$	0	0	0	1	0	1	1	0
$\frac{3}{4}-\frac{3}{8}$	0	0	0	1	0	0	1	0
$\frac{5}{6}-\frac{1}{9}$	0	0	0	1	0	0	1	0
$3\frac{1}{2}-2\frac{3}{2}$	0	1	1	0	1	0	1	0
$4\frac{3}{5}-3\frac{4}{10}$	0	1	0	1	0	0	1	1
$\frac{6}{7}-\frac{4}{7}$	0	0	0	0	0	0	1	0
$3-2\frac{1}{5}$	1	1	0	0	0	0	1	0
$\frac{2}{3}-\frac{2}{3}$	0	0	0	0	0	0	1	0
$3\frac{7}{8}-2$	0	1	0	0	0	0	0	0
$4\frac{4}{12}-2\frac{7}{12}$	0	1	0	0	1	0	1	1
$4\frac{1}{3}-2\frac{4}{3}$	0	1	0	0	1	0	1	0
$\frac{11}{8}-\frac{1}{8}$	0	0	0	0	0	0	1	1
$3\frac{3}{8}-2\frac{5}{6}$	0	1	0	1	1	0	1	0
$3\frac{4}{5}-3\frac{2}{5}$	0	1	0	0	0	0	1	0
$2-\frac{1}{3}$	1	0	0	0	0	0	1	0
$4\frac{5}{7}-1\frac{4}{7}$	0	1	0	0	0	0	1	0
$7\frac{3}{5}-2\frac{4}{5}$	0	1	0	0	1	0	1	0
$4\frac{1}{10}-2\frac{8}{10}$	0	1	0	0	1	1	1	0
$4-1\frac{4}{3}$	1	1	1	0	1	0	1	0
$4\frac{1}{3}-1\frac{5}{3}$	0	1	1	0	1	0	1	0

5. Discussion

The fundamental building block of a cognitive diagnosis model is the $\boldsymbol{Q}$ matrix. However, estimation of this discrete parameter has been a statistical challenge. While many models have been developed for cognitive diagnosis, with challenging parameter estimation issues of their own, $\boldsymbol{Q}$ has traditionally been taken as known. This is limiting in practice, because it requires expert opinion and any two experts are likely to disagree on many entries, or even on the number of attributes. Statistical methods for exploring the structure of $\boldsymbol{Q}$ are desirable for much the same reason that exploratory factor analysis has proven useful. Specifying the nonzero loadings in confirmatory factor analysis is essentially the same as specifying $\boldsymbol{Q}$, but exploratory factor analysis has become more widely used because it allows for a wider scope of possibilities than confirmatory factor analysis. However, the discrete nature of $\boldsymbol{Q}$, and the reliance upon Boolean algebra rather than linear algebra, has made estimation of $\boldsymbol{Q}$ a difficult problem.

The introduction of a consistent estimator of $\boldsymbol{Q}$ in Liu et al. (2013) offered hope for solving this problem, and recent years have seen several advances in estimation of $\boldsymbol{Q}$. The Bayesian approach of this research has borrowed from the identifiability theory of Chen et al. (2015), in order to implement a fully Bayesian approach capable of exploring all possible $\boldsymbol{Q}$ matrices. By enforcing these constraints as well as constraints on permutations of columns, a trivial source of unidentifiability, the number of possible $\boldsymbol{Q}$ matrices is reduced from $2^{J\times K}$, but is still enormous.

TABLE 3.
Estimated $\boldsymbol{Q}$, Slipping ($\hat{s}$), and Guessing ($\hat{g}$) parameters for $K = 3$ from the Constrained Metropolis–Hastings (MH) and Unconstrained Gibbs Samplers.

Item	MH					Unconstrained Gibbs				
	$\hat{\boldsymbol{Q}}$			$\hat{s}$	$\hat{g}$	$\hat{\boldsymbol{Q}}$			$\hat{s}$	$\hat{g}$
$\frac{5}{3} - \frac{3}{4}$	1	0	0	0.14	0.04	1	0	0	0.14	0.04
$\frac{3}{4} - \frac{3}{8}$	1	0	0	0.07	0.05	1	0	0	0.07	0.05
$\frac{5}{6} - \frac{1}{9}$	1	0	0	0.14	0.01	1	0	0	0.14	0.01
$3\frac{1}{2} - 2\frac{3}{2}$	0	1	0	0.13	0.21	0	1	1	0.12	0.18
$4\frac{3}{5} - 3\frac{4}{10}$	1	0	1	0.21	0.30	1	0	1	0.22	0.31
$\frac{6}{7} - \frac{4}{7}$	0	0	1	0.04	0.31	0	0	1	0.04	0.30
$3 - 2\frac{1}{5}$	1	0	1	0.24	0.05	1	0	0	0.35	0.03
$\frac{2}{3} - \frac{2}{3}$	1	1	0	0.05	0.58	1	1	0	0.06	0.58
$3\frac{7}{8} - 2$	0	0	1	0.17	0.34	0	0	1	0.25	0.34
$4\frac{4}{12} - 2\frac{7}{12}$	0	1	0	0.23	0.03	0	1	1	0.23	0.03
$4\frac{1}{3} - 2\frac{4}{3}$	0	1	1	0.08	0.07	0	1	1	0.07	0.06
$1\frac{1}{8} - \frac{1}{8}$	0	0	1	0.09	0.21	0	0	1	0.09	0.20
$3\frac{3}{8} - 2\frac{5}{6}$	1	1	1	0.33	0.02	1	1	1	0.35	0.02
$3\frac{4}{5} - 3\frac{2}{5}$	0	0	1	0.07	0.10	0	0	1	0.07	0.09
$2 - \frac{1}{3}$	1	0	1	0.17	0.11	1	0	1	0.25	0.06
$4\frac{5}{7} - 1\frac{4}{7}$	0	0	1	0.11	0.13	0	0	1	0.11	0.12
$7\frac{3}{5} - 2\frac{4}{5}$	0	1	1	0.14	0.05	0	1	1	0.14	0.04
$4\frac{1}{10} - 2\frac{8}{10}$	0	1	1	0.16	0.13	0	1	1	0.16	0.13
$4 - 1\frac{4}{3}$	1	1	1	0.32	0.03	1	1	1	0.33	0.03
$4\frac{1}{3} - 1\frac{5}{3}$	0	1	1	0.19	0.02	0	1	1	0.18	0.02

$\hat{\boldsymbol{Q}}$ estimated from CGibbs is the same as the MH algorithm above.

Much of this research has concerned efficient implementation of MCMC to quickly search the space of identifiable models in order to provide a quality estimator in a practical amount of time. The sample sizes and test lengths we have studied proved quite manageable, and accurate estimates of all model parameters, including $\boldsymbol{Q}$, could be obtained. An ongoing question will be the scope of the method, and how the computational complexity changes with N, J and especially K.

An issue worthy of further investigation is how to best summarize the posterior distribution of the discrete $\boldsymbol{Q}$. The simulations considered estimation of $\boldsymbol{Q}$ by taking the most frequent value observed in the Markov chain. This proved effective, but may be problematic as K increases, considering the exponentially increasing number of candidates. The concern is that a tremendously long chain might be required when using the mode as the estimator in a parameter space of such great cardinality. Another promising method is to simply take the entry-wise modes to comprise the estimator and additional research is needed. Another fruitful topic for future research will be extensions to models beyond the DINA. The MCMC algorithms for other models should be similar, and the main difference will be to discern what restrictions are needed for Bayesian formulations of other models to be identifiable.

Table 4.
Estimated $\boldsymbol{Q}$, Slipping ($\hat{s}$), and Guessing ($\hat{g}$) parameters for $K = 4$ from the Constrained Metropolis–Hastings (MH) and Unconstrained Gibbs Samplers.

Item	MH						Gibbs					
	$\hat{\boldsymbol{Q}}$				$\hat{s}$	$\hat{g}$	$\hat{\boldsymbol{Q}}$				$\hat{s}$	$\hat{g}$
$\frac{5}{3}-\frac{3}{4}$	1	0	0	0	0.11	0.04	1	0	0	0	0.12	0.04
$\frac{3}{4}-\frac{3}{8}$	1	0	0	0	0.04	0.05	1	0	0	0	0.04	0.04
$\frac{5}{6}-\frac{1}{9}$	1	0	0	0	0.12	0.01	1	0	0	0	0.12	0.01
$3\frac{1}{2}-2\frac{3}{2}$	0	1	0	0	0.12	0.15	0	1	0	0	0.12	0.16
$4\frac{3}{5}-3\frac{4}{10}$	1	0	1	0	0.16	0.34	1	0	1	0	0.21	0.31
$\frac{6}{7}-\frac{4}{7}$	0	0	1	0	0.04	0.30	0	0	1	0	0.04	0.28
$3-2\frac{1}{5}$	0	0	0	1	0.24	0.02	0	0	0	1	0.20	0.01
$\frac{2}{3}-\frac{2}{3}$	0	0	1	0	0.06	0.54	0	1	0	0	0.06	0.56
$3\frac{7}{8}-2$	0	0	1	0	0.17	0.40	0	0	1	0	0.25	0.33
$4\frac{4}{12}-2\frac{7}{12}$	0	1	0	0	0.23	0.03	0	1	1	0	0.23	0.03
$4\frac{1}{3}-2\frac{4}{3}$	0	1	1	0	0.08	0.07	0	1	1	0	0.08	0.07
$1\frac{1}{8}-\frac{1}{8}$	0	0	1	0	0.09	0.21	0	0	1	0	0.09	0.17
$3\frac{3}{8}-2\frac{5}{6}$	1	1	1	1	0.33	0.02	1	1	1	0	0.34	0.02
$3\frac{4}{5}-3\frac{2}{5}$	0	0	1	0	0.07	0.10	0	0	1	0	0.07	0.06
$2-\frac{1}{3}$	0	0	0	1	0.11	0.04	0	0	1	1	0.11	0.04
$4\frac{5}{7}-1\frac{4}{7}$	0	0	1	0	0.11	0.13	0	0	1	0	0.12	0.12
$7\frac{3}{5}-2\frac{4}{5}$	0	1	1	0	0.14	0.06	0	1	1	0	0.14	0.05
$4\frac{1}{10}-2\frac{8}{10}$	0	1	1	1	0.20	0.13	0	1	1	0	0.16	0.13
$4-1\frac{4}{3}$	0	1	1	1	0.32	0.03	0	1	1	1	0.26	0.03
$4\frac{1}{3}-1\frac{5}{3}$	0	1	1	0	0.19	0.02	0	1	1	0	0.19	0.02

$\hat{\boldsymbol{Q}}$ estimated from CGibbs is the same as the MH algorithm above.

The value of exploratory methods for determining $\boldsymbol{Q}$ will ultimately be decided through applications. This research provides an alternative to relying on experts for specifying $\boldsymbol{Q}$. However, the ideal scenario would be to compare $\boldsymbol{Q}$ matrices developed by experts with those developed statistically, in order to reach compromises that allow both a good model fit and a useful interpretation. It is also quite conceivable that a portion of $\boldsymbol{Q}$ could be determined by experts, which would greatly enhance the interpretation and reduce the space to explore statistically. A special case of this would encompass the problem of introducing new items to an item bank. If $\boldsymbol{Q}$ is known for a subset of the items with an appropriate span, then calibration of the remainder of $\boldsymbol{Q}$ and other item parameters could readily be accomplished.

References

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.

Chiu, C.-Y., & Köhn, H.-F. (2016). The reduced RUM as a logit model: Parameterization and constraints. *Psychometrika*, *81*(2), 350–370.

Chiu, C.-Y., Köhn, H.-F., & Wu, H.-M. (2016). Fitting the reduced RUM with Mplus: A tutorial. *International Journal of Testing*, *16*, 1–21.

Chung, M. (2014). *Estimating the Q-matrix for Cognitive Diagnosis Models in a Bayesian Framework* (Unpublished doctoral dissertation). Columbia University.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*, 883–904.

Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *40*(5), 454–476.

DeCarlo, L. T. (2010). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8–26.

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

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. (2009). Estimation code for the G-DINA model. In *Presentation at the meeting of the American Educational Research Association*. San Diego, CA.

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.

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

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics, 4*, 169–188.

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

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210.

Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 19–60). Cambridge: Cambridge University Press.

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.

Leighton, J. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3–18). Cambridge: Cambridge University Press.

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.

Mislevy, R. J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, *61*(1), 41–71.

Norris, S. P., Macnab, J. S., & Phillips, L. M. (2007). Cognitive modeling of performance on diagnostic achievement tests. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 61–84). Cambridge: Cambridge University Press.

Rupp, A. A. (2009). Software for calibrating diagnostic classification models: An overview of the current state-of-the-art. In *Symposium conducted at the meeting of the American Educational Research Association*, San Diego, CA.

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.

Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.

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.

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., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, *32*(2), 37–50.

von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, *67*(1), 49–71.

Xiang, R. (2013). *Nonlinear Penalized Estimation of True Q-matrix in Cognitive Diagnostic Models (Unpublished doctoral dissertation)*. Columbia University.
Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, *81*(3), 625–649.

Manuscript Received: 2 MAY 2016
Final Version Received: 1 MAY 2017
Published Online Date: 31 AUG 2017