JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS OnlineFirst, published on August 10, 2015 as doi:10.3102/1076998615595403

Journal of Educational and Behavioral Statistics
Vol. XX, No. X, pp. 1–23
DOI: 10.3102/1076998615595403
© 2015 AERA. http://jebs.aera.net

Bayesian Estimation of the DINA Model With Gibbs Sampling

Steven Andrew Culpepper
University of Illinois at Urbana-Champaign

A Bayesian model formulation of the deterministic inputs, noisy "and" gate (DINA) model is presented. Gibbs sampling is employed to simulate from the joint posterior distribution of item guessing and slipping parameters, subject attribute parameters, and latent class probabilities. The procedure extends concepts in Béguin and Glas, Culpepper, and Sahu for estimating the guessing and slipping parameters in the three- and four-parameter normal-ogive models. The ability of the model to recover parameters is demonstrated in a simulation study. The technique is applied to a mental rotation test. The algorithm and vignettes are freely available to researchers as the "dina" R package.

Keywords: *cognitive diagnosis; bayesian statistics; markov chain monte carlo; spatial cognition*

Introduction

Cognitive diagnosis models (CDMs) remain an important methodology for diagnosing student skill acquisition. A variety of CDMs have been developed. In particular, existing CDMs include the deterministic inputs, noisy "and" gate (DINA) model (de la Torre, 2009a; de la Torre & Douglas, 2004, 2008; Doignon & Falmagne, 1999; Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977), the noisy inputs, deterministic, "and" gate (NIDA) model (Maris, 1999), the deterministic input, noisy "or" gate (DINO) model (Templin & Henson, 2006b), the reduced reparameterized unified model (RUM; Roussos, DiBello, et al., 2007), the generalized DINA (de la Torre, 2011), the log-linear CDM (Henson, Templin, & Willse, 2009), and the general diagnostic model (von Davier, 2008). The aforementioned models differ in complexity, and prior research demonstrated how models such as the DINA, NIDA, and DINO can be estimated in more generalized frameworks (e.g., see de la Torre, 2011; von Davier, 2014). The interpretability of the DINA model makes it a popular CDM (de la Torre, 2009a; von Davier, 2014), and the purpose of this article is to introduce a Bayesian formulation for estimating the DINA model parameters with Gibbs sampling.

This article includes five sections. The first section introduces the DINA model and discusses benefits of the proposed Bayesian formulation in the context

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

of prior research. The second section presents the Bayesian model formulation and full conditional distributions that are used to estimate person (i.e., latent skill/attribute profiles) and item parameters (i.e., slipping and guessing) using Markov Chain Monte Carlo (MCMC) with Gibbs sampling. The third section reports results from a simulation study on the accuracy of recovering the model parameters. The fourth section presents an application of the model using responses to a mental rotation test (Guay, 1976; Yoon, 2011). The final section includes discussion, future research directions, and concluding remarks.

The DINA Model

Overview

Throughout this article, let $i = 1, \ldots, N$ index individuals, $j = 1, \ldots, J$ index items, $k = 1, \ldots, K$ index skills, and $c = 1, \ldots, C$ index latent classes. CDMs assume that a set of K known skills underlie each item. The skills needed to succeed on each item are collected into a $J \times K$ matrix $\mathbf{Q}$, which is a structure matrix with elements that are akin to factor loadings relating skills to items. However, one exception is that the elements of $\mathbf{Q}$ are either 0 or 1 (i.e., $q_{jk} \in \{0, 1\}$) depending upon whether skill k is needed for success on item j. That is, q_{jk} equals 1 if skill k is needed to correctly answer item j and 0 otherwise. Let $\mathbf{q}'_j = (q_{j1}, \ldots, q_{jK})$ be the j th row of $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_J)'$, which indicates the requisite skills for item j.

The DINA model posits that individual test takers are classified into latent classes based upon mastery and nonmastery of the K skills. Let α_{ik} equal 1 if individual i has mastered skill k and zero otherwise. Furthermore, let the K dimensional latent attribute profile of binary indicators for individual i be $\boldsymbol{\alpha}_i = (\alpha_{i1}, ..., \alpha_{iK})'$. Given the categorical nature of the latent classes, $\boldsymbol{\alpha}_i$ can equal one of $C = 2^K$ attribute profiles. It will be useful subsequently to define $\boldsymbol{\alpha}_c = (\alpha_{c1}, \ldots, \alpha_{cK})'$ as the attribute profile for members of class c such that α_{ck} is 1 if members of class c possess skill k and zero otherwise.

The DINA model is a conjunctive model, which implies that students need all of the skills in $\mathbf{q}_j$ to correctly answer item j without guessing. We let the vector of ideal responses for student i be $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})'$ with $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} = \mathcal{I}(\boldsymbol{\alpha}'_i \mathbf{q}_j = \mathbf{q}'_j \mathbf{q}_j)$, where $\mathcal{I}$ denotes the indicator function. Notice that η_{ij} is 1 if student i has the requisite skills to answer item j and 0 if the student is missing at least one skill. The probability of individual i's success on item j for the DINA model is a function of individual latent class membership (i.e., $\boldsymbol{\alpha}_i$) and item parameters,

$$P(y_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{\Omega}_j) = \begin{cases} g_j, & \eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} = 0 \\ 1 - s_j, & \eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} = 1 \end{cases}, \tag{1}$$

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

where $\mathbf{\Omega}_j = (g_j, s_j)$ is a row vector of item parameters. By definition, $g_j = P(y_{ij} = 1|\eta_{ij} = 0)$ is the probability of correctly guessing item j for students who do not have the requisite skills needed for item j. In contrast, $s_j = P(y_{ij} = 0|\eta_{ij} = 1)$ is the probability of "slipping," which is defined as the chance that students who have the required skills to answer item j record an incorrect response.

The DINA item response function can be used to construct a likelihood of observing a sample of item responses. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})'$ denote a J dimensional vector of item responses for individual i, where y_{ij} is coded as 1 for a correct response and 0 for an incorrect response. Let $\mathbf{\Omega} = (\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_J)$ be a row vector of slipping and guessing parameters for J items. Furthermore, let $P_{jc} = P(y = 1|\boldsymbol{\alpha}_c, \mathbf{\Omega}_j)$ be the probability that members of class c correctly answer item j and note that $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c$ implies that individual i is in attribute class c. Assuming that item responses are conditionally independent given $\boldsymbol{\alpha}_i$, the probability of observing $\mathbf{y}_i$ given membership in class c (i.e., $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c$) is,

$$p(\mathbf{y}_i|\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \mathbf{\Omega}) = \prod_{j=1}^{J} P_{jc}^{y_{ij}}(1 - P_{jc})^{1-y_{ij}}. \tag{2}$$

Let $\pi_c = p(\boldsymbol{\alpha}_c)$ be the marginal probability of belonging to class c in the population and define $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_C)'$ as a C dimensional vector of class membership probabilities. The Law of Total Probability implies that the probability of observing $\mathbf{y}_i$ given item parameters $\mathbf{\Omega}$ and latent class probabilities $\boldsymbol{\pi}$ is given by,

$$p(\mathbf{y}_{ai}|\mathbf{\Omega}, \boldsymbol{\pi}) = \sum_{c=1}^{C} \pi_c p(\mathbf{y}_i|\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \mathbf{\Omega}). \tag{3}$$

The likelihood of observing a sample of N responses to J items is given by,

$$p(\mathbf{y}|\mathbf{\Omega}, \boldsymbol{\pi}) = \prod_{i=1}^{N} \sum_{c=1}^{C} \pi_c p(\mathbf{y}_i|\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{ac}, \mathbf{\Omega}), \tag{4}$$

where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)'$ is an $N \times J$ matrix of observed responses.

In Bayesian formulations, the primary focus is on the parameter posterior distribution, which for the DINA model is,

$$p(\mathbf{\Omega}, \boldsymbol{\pi}|\mathbf{y}) \propto p(\mathbf{y}|\Omega, \boldsymbol{\pi}) p(\mathbf{\Omega}) p(\boldsymbol{\pi}), \tag{5}$$

where $p(\mathbf{\Omega})$ is a prior distribution for item parameters and $p(\boldsymbol{\pi})$ is a prior distribution for latent class probabilities. Maris (1999) noted the intractable nature of the CDM parameter posterior distributions and alluded to how Monte Carlo procedures could be used to estimate posterior distributions. Accordingly, subsequent research considered MCMC procedures for estimating posterior distributions of CDM parameters (de la Torre & Douglas, 2004; Hartz, 2002; Henson et al., 2009; Templin, 2004). Roussos, Templin, and Henson (2007) provide an

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

overview of estimation methods (both Bayesian and non-Bayesian) prior researchers have employed to estimate CDM parameters. Furthermore, prior research discussed how applied researchers can estimate CDMs with standard statistical software (de la Torre, 2009b; Henson et al., 2009; Rupp, 2009; Rupp, Templin, & Henson, 2010).

Benefits of the Proposed Bayesian Formulation

The Bayesian model formulation for the DINA CDM developed in this article offers several benefits to the existing procedures. First, the model formulation yields analytically tractable full conditional distributions for the person and item parameters that enable the use of Gibbs sampling for efficient computations. Prior applications of the DINA model and other CDMs employed Metropolis-Hastings (MH) sampling to estimate model parameters, which require tuning of proposal distribution parameters for $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$. Consequently, one limitation with using MH sampling is that researchers need to tune proposal parameters each time a new data set is analyzed. More specifically, researchers who use MH sampling to estimate DINA model parameters need to choose proposal parameters to ensure no more than 50% of candidates are accepted. Specifying tuning parameters may require only a few attempts for a given data set, but the process could become tedious if the goal is to disseminate software for widespread use by applied researchers who have less experience with MH sampling. In contrast, the procedures discussed subsequently employ Gibbs sampling from parameter full conditional distributions and can be easily implemented without manual tuning.

As another example, consider the additional requirements needed to employ MH sampling in a simulation study similar to the one reported subsequently. After developing estimation code and deciding upon simulation conditions, it would be necessary to select proposal parameters used to sample candidates for $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$. Proposal parameters must be tuned using trial and error, where a collection of chains are run with different proposal parameter values to ensure 25–50% of candidates are accepted. It would be necessary to apply the tuning process for all conditions of the simulation study, which, depending upon the number of conditions, could require considerable effort to verify and catalogue proposal distribution parameter values. In contrast, employing the Gibbs sampler developed in this article would only require decisions about prior parameters. The procedure introduced subsequently employs uninformative uniform priors for $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$, so the developed algorithm can be immediately applied.

Second, the proposed formulation is advantageous to prior Gibbs samplers. For instance, prior CDM research employed Gibbs sampling to estimate the reparameterized DINA model using OpenBUGS (DeCarlo, 2012) and the higher order DINA (HO-DINA) model (de la Torre & Douglas, 2004) using WinBUGS (Li, 2008). However, one disadvantage is that OpenBUGS and WinBUGS do not explicitly impose the monotonicity restriction for the DINA model that

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

$0 \leq g_j + s_j < 1$ for all items (Junker & Sijtsma, 2001). One consequence is that prior Gibbs samplers did not strictly enforce the identifiability condition, which could impact convergence in some applications. The model discussed subsequently extends the Bayesian formulation of the three-parameter normal-ogive (3PNO) model of Béguin and Glass (2001) and Sahu (2002) and the four-parameter model of Culpepper (in press) to explicitly enforce the monotonicity requirement. Accordingly, the developed strategy could be extended to other CDMs that include guessing and slipping parameters such as the NIDA, DINO, and RUM.

Third, one documented advantage of using Gibbs sampling is improved mixing of the chain in comparison to procedures based upon acceptance/rejection sampling such as MH sampling. That is, prior research suggests that MH sampling mixes best (i.e., the posterior distribution is efficiently explored) if candidates are accepted between 25% and 50% of the iterations (e.g., see chapter 6 of Robert & Casella, 2009). Accordingly, MH sampling may contribute to slower mixing among posterior samples, given that posterior samples change in less than half the total iterations. In contrast, the Gibbs sampler has the potential to improve mixing by sampling values from full conditional distributions. However, as noted by an anonymous reviewer, Gibbs sampling does not guarantee faster mixing for all models and parameters (e.g., see Cowles, 1996, for discussion regarding threshold parameters in polytomous normal-ogive models). The Monte Carlo evidence given subsequently suggests that the proposed MCMC algorithm converges to the parameter posterior distribution within 350 to 750 iterations. In contrast, prior research that employed Bayesian estimation (DeCarlo, 2012; Henson et al., 2009; Li, 2008) reported requiring a burn-in between 1,000 and 5,000 iterations to achieve convergence.

Fourth, as discussed subsequently, students' latent attribute profiles are modeled using a general categorical prior. One advantage is that specifying a categorical prior provides a natural approach for computing the probability of skill acquisition and classifying students into latent classes (Huebner & Wang, 2011). That is, the posterior distribution for each student's skill distribution can be summarized to show the probability of membership in the various latent classes.

Fifth, the developed algorithm is made freely available to researchers as the "dina" R package. The Gibbs sampler was written in C++ and provides an efficient approach for estimating model parameters. In fact, the developed R package includes documentation and vignettes to instruct applied researchers.

Bayesian Estimation of the DINA Model

This section presents a Bayesian formulation for the DINA model. The first section presents the model and describes the priors and the second section presents the full conditional distributions for implementation in an MCMC Gibbs sampler.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

Bayesian Model Formulation

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

$$y_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{\Omega}_j \sim \text{Bernoulli}\left((1-s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}\right), \eta_{ij} = \mathcal{I}(\boldsymbol{\alpha}'_i \mathbf{q}_j = \mathbf{q}'_j \mathbf{q}_j), \tag{6}$$

$$p(\boldsymbol{\alpha}_\mathbf{i}|\boldsymbol{\pi}) \propto \prod_{c=1}^{C} \pi_c^{\mathcal{I}(\alpha_i = \alpha_c)}, 0 \leq \pi_c \leq 1, \sum_{c=1}^{C} \pi_c = 1. \tag{7}$$

$$p(\boldsymbol{\pi}) = p(\pi_1, \ldots, \pi_C) \propto \text{Dirichlet}\ (\boldsymbol{\delta}_0), \boldsymbol{\delta}_0 = (\delta_{01}, \ldots, \delta_{0C}). \tag{8}$$

$$p(\boldsymbol{\Omega}_j) = p(s_j, g_j) \propto s_j^{\alpha_s - 1}(1-s_j)^{\beta_s - 1} g_j^{\alpha_g - 1}(1-g_j)^{\beta_g - 1} \mathcal{I}\left((s_j, g_j) \in \mathcal{P}\right). \tag{9}$$

$$\mathcal{P} = \{(s, g) : 0 \leq s + g < 1, 0 \leq s < 1, 0 \leq g < 1\}.$$

Equation 6 assumes a Bernoulli item response function for y_{ij} when conditioned upon the latent attribute profile $\boldsymbol{\alpha}_{ai}$ and item parameters $\boldsymbol{\Omega}_j$. In fact, Equation 6 simply restates the item response function in Equation 1. Equation 7 assumes a categorical prior for attribute vector $\boldsymbol{\alpha}_i$, with the probability of membership in class c as $\pi_c = p(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c|\boldsymbol{\pi})$.

Equation 8 shows that the conjugate prior for $\boldsymbol{\pi}$ is a Dirichlet with parameter vector $\boldsymbol{\delta}'_0 = (\delta_{01}, \ldots, \delta_{0C})$. Recall the Dirichlet distribution is a multivariate version of the beta distribution for random variables with a support on the unit interval. The values of the elements within $\boldsymbol{\delta}_0$ specify prior information about the prevalence of class membership in the population. An uninformative prior can be implemented by setting the elements of $\boldsymbol{\delta}_0$ to 1 if there is no a priori information about $\boldsymbol{\pi}$. Alternatively, researchers can modify the values within $\boldsymbol{\delta}_0$ to reflect the relative prevalence of various attribute profiles. For instance, the marginal distribution for π_c under the prior is a Beta$(\delta_{0c}, \sum_{c' \neq c} \delta_{0c'})$, and the values of δ_0 can be chosen to reflect prior knowledge about the probability of observing attribute profile π_c. Note that the results from the Monte Carlo simulation discussed in the next section provides evidence that the model is robust to employing an uninformative prior with $\delta_{0c} = 1$ for all c.

Equation 9 includes the joint prior for $\boldsymbol{\Omega}_j$. Junker and Sijtsma (2001) established that the DINA model requires $0 \leq s_j + g_j < 1$ for all j so that each item provides information to discriminate among the latent attribute classes. The prior for $\boldsymbol{\Omega}_j$ in Equation 9 explicitly imposes the monotonicity restriction on the item parameters. In particular, the prior for $p(\boldsymbol{\Omega}_j)$ is constructed by taking the product of Beta densities for s_j and g_j with parameters α_s and β_s and α_g and β_g and imposing the identifiability restriction with linear truncation on the space defined by $\mathcal{P}$ for s_j and g_j. Prior information regarding the slipping and guessing parameters can be incorporated into the estimation by choosing values for α_s, β_s, α_g, and β_g. Note that choosing an uninformative prior with $\alpha_s = \beta_s = \alpha_g = \beta_g = 1$ is identical to employing a linearly truncated bivariate uniform prior for s_j and g_j. The simulation results

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

discussed in the next section support the use of uninformative priors for the slipping and guessing parameters.

Full Conditional Distributions

The full conditional distributions for $\boldsymbol{\alpha}_i$, $\boldsymbol{\pi}$, and $\boldsymbol{\Omega}_j$ are,

$$p(\boldsymbol{\alpha}_i|\mathbf{y}_i, \boldsymbol{\pi}, \boldsymbol{\Omega}) = \prod_{c=1}^{C} \widetilde{\pi}_{ic}^{\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_c)}, \widetilde{\pi}_{ic} = p(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c|\mathbf{y}_i, \boldsymbol{\pi}, \boldsymbol{\Omega}). \tag{10}$$

$$\boldsymbol{\pi}|\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N \propto \text{Dirichlet}(\widetilde{\mathbf{N}} + \boldsymbol{\delta}_0). \tag{11}$$

$$\widetilde{\mathbf{N}}' = (\widetilde{N}_1, \ldots, \widetilde{N}_C) = \left(\sum_{i=1}^{N} \mathcal{I}(\alpha_i = \alpha_1), \ldots, \sum_{i=1}^{N} \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_C)\right)$$

$$p(\boldsymbol{\Omega}_j|\mathbf{y}_j, \eta_{1j}, \ldots, \eta_{Nj}) \propto s_j^{\widetilde{\alpha}_s - 1}(1 - s_j)^{\widetilde{\beta}_s - 1} g_j^{\widetilde{\alpha}_g - 1}(1 - g_j)^{\widetilde{\beta}_s - 1} \mathcal{I}\Big((s_j, g_j) \in P\Big)$$

$$\begin{aligned}
\widetilde{\alpha}_s &= \mathcal{S}_j + \alpha_s, \widetilde{\beta}_s = \mathcal{T}_j - \mathcal{S}_j + \beta_s \\
\widetilde{\alpha}_g &= \mathcal{G}_j + \alpha_g, \widetilde{\beta}_g = N - \mathcal{T}_j - \mathcal{G}_j + \beta_g \\
\mathcal{T}_j &= \sum_{i=1}^{N} \eta_{ij}, \mathcal{S}_j = \sum_{i|y_{ij}=0}^{N} \eta_{ij}, \mathcal{G}_j = \sum_{i|y_{ij}=1}^{N} (1 - \eta_{ij}).
\end{aligned} \tag{12}$$

Equation 10 shows that the full conditional for $\boldsymbol{\alpha}_i$ is a categorical distribution with the probability of membership in category c as $\widetilde{\pi}_{ic}$. Specifically, Bayes's theorem implies that the probability that student i belongs to class c given $\mathbf{y}_i$, $\boldsymbol{\Omega}$, and $\boldsymbol{\pi}$ is,

$$\widetilde{\pi}_{ic} = p(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c|\mathbf{y}_i, \boldsymbol{\pi}, \boldsymbol{\Omega}) = \frac{p(\mathbf{y}_i|\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \boldsymbol{\Omega})\pi_c}{\sum_{c=1}^{C} \pi_c p(\mathbf{y}_i|\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \boldsymbol{\Omega})}. \tag{13}$$

The full conditional for $\boldsymbol{\alpha}_i$ is a categorical distribution with probabilities $\widetilde{\pi}_{i1}, \ldots, \widetilde{\pi}_{iC}$.

Equation 11 reports the full conditional distribution for $\boldsymbol{\pi}$. Specifically, the full conditional distribution for $\boldsymbol{\pi}$ can be derived as,

$$\begin{aligned}
p(\boldsymbol{\pi}|\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) \propto p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N|\boldsymbol{\pi})p(\boldsymbol{\pi}) &= \left[\prod_{i=1}^{N} p(\boldsymbol{\alpha}_i|\boldsymbol{\pi})\right] p(\boldsymbol{\pi}) \\
&= \left[\prod_{i=1}^{N}\prod_{c=1}^{C} \pi_c^{\mathcal{I}(\alpha_i=\alpha_c)}\right] \prod_{c=1}^{C} \pi_c^{\delta_{0c}} = \prod_{c=1}^{C} \pi_c^{\widetilde{N}_c + \delta_{0c}}.
\end{aligned} \tag{14}$$

Equation 14 shows that the full conditional for $\boldsymbol{\pi}$ is the product of the categorical prior for $\boldsymbol{\alpha}_i$, and the Dirichlet prior for $\boldsymbol{\pi}$. Let the number of students within the C latent classes for a given MCMC iteration be denoted by the vector $\widetilde{\mathbf{N}}' = (\widetilde{N}_1, \ldots, \widetilde{N}_C)$ such that $\widetilde{N}_c = \sum_{i=1}^{N} \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c)$ is the number of subjects within attribute class c. Accordingly, the full conditional for the latent class probabilities in Equation 14 is Dirichlet$(\widetilde{\mathbf{N}} + \boldsymbol{\delta}_0)$.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

Equation 12 indicates that the full conditional distribution for the item parameters s_j and g_j is a truncated bivariate Beta distribution. More precisely, the full conditional distribution for s_j and g_j is derived as,

$$\begin{aligned} p(s_j, g_j | y_{1j}, \ldots, y_{Nj}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) &\propto \left[\prod_{i=1}^{N} p(y_{ij}|\boldsymbol{\alpha}_i, s_j, g_j)\right] p(s_j, g_j) \\ &\propto \left\{\prod_{i=1}^{N} [(1-s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}]^{y_{ij}} [s_j^{\eta_{ij}} (1-g_j)^{1-\eta_{ij}}]^{1-y_{ij}}\right\}. \\ & s_j^{\alpha_s - 1}(1-s_j)^{\beta_s - 1} g_j^{\alpha_g - 1}(1-g_j)^{\beta_g - 1} \mathcal{I}\Big((s_j, g_j) \in \mathcal{P}\Big) \\ &= s_j^{\tilde{\alpha}_s - 1}(1-s_j)^{\tilde{\beta}_s - 1} g_j^{\tilde{\alpha}_g - 1}(1-g_j)^{\tilde{\beta}_g - 1} \mathcal{I}\Big((s_j, g_j) \in \mathcal{P}\Big) \end{aligned} \tag{15}$$

The final result in Equation 15 was established by simplifying the product over individuals and collecting terms in the exponents for s_j, $1 - s_j$, g_j, and $1 - g_j$. Recall that $\tilde{\alpha}_s$, $\tilde{\beta}_s$, $\tilde{\alpha}_g$, and $\tilde{\beta}_g$ are defined after Equation 12. For a given MCMC iteration, let $\mathcal{T}_j = \sum_{i=1}^{N} \eta_{ij}$ be the number of students with the attributes necessary to correctly answer item j, $\mathcal{S}_j = \sum_{i|y_{ij}=0}^{N} \eta_{ij}$ is the number of students who unsuccessfully applied their skills, and $\mathcal{G}_j = \sum_{i|y_{ij}=1}^{N} (1 - \eta_{ij})$ is the number of students who correctly guess the answer to item j. The full conditional for s_j and g_j is a truncated Beta distribution with parameters $\tilde{\alpha}_s$, $\tilde{\beta}_s$, $\tilde{\alpha}_g$, and $\tilde{\beta}_g$. The Appendix describes an approach for jointly sampling s_j and g_j subject to the monotonicity restriction of $0 \leq s_j + g_j < 1$.

Monte Carlo Simulation Study

This section presents Monte Carlo simulation results to demonstrate the accuracy of the MCMC algorithm and to provide additional information about convergence. The simulation study is similar to the designs of de la Torre and Douglas (2004), de la Torre (2009), and Huebner and Wang (2011). Specifically, the simulation considers $K = 5, J = 30$, and $N = 1,000$ with a $\mathbf{Q}$ matrix from de la Torre and Douglas (2004) and de la Torre (2009; e.g., $\mathbf{Q}'$ is included in the x-axis of Figures 2 and 3). The simulation study considers three issues. First, following Huebner and Wang (2011), two conditions are studied for item guessing and slipping parameters to understand how the relationship between latent skills and items impacts parameter recovery. A high diagnosticity case (i.e., a stronger relationship between latent skills and items) is considered with $g_j = s_j = 0.1$, whereas a low diagnosticity case is $s_j = g_j = 0.2$ for all j.

Second, two cases are considered for latent class probabilities. Recall that $\boldsymbol{\pi}$ is the population skill distribution and π_c denotes the proportion of the population with attribute profile $\boldsymbol{\alpha}_c$. In the first case, the probability of class membership is constant with $\pi_c = 2^{-K}$ for all c (i.e., the examinee skill distribution is flat). In the second case, the probability of belonging to a class with three or more skills is

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

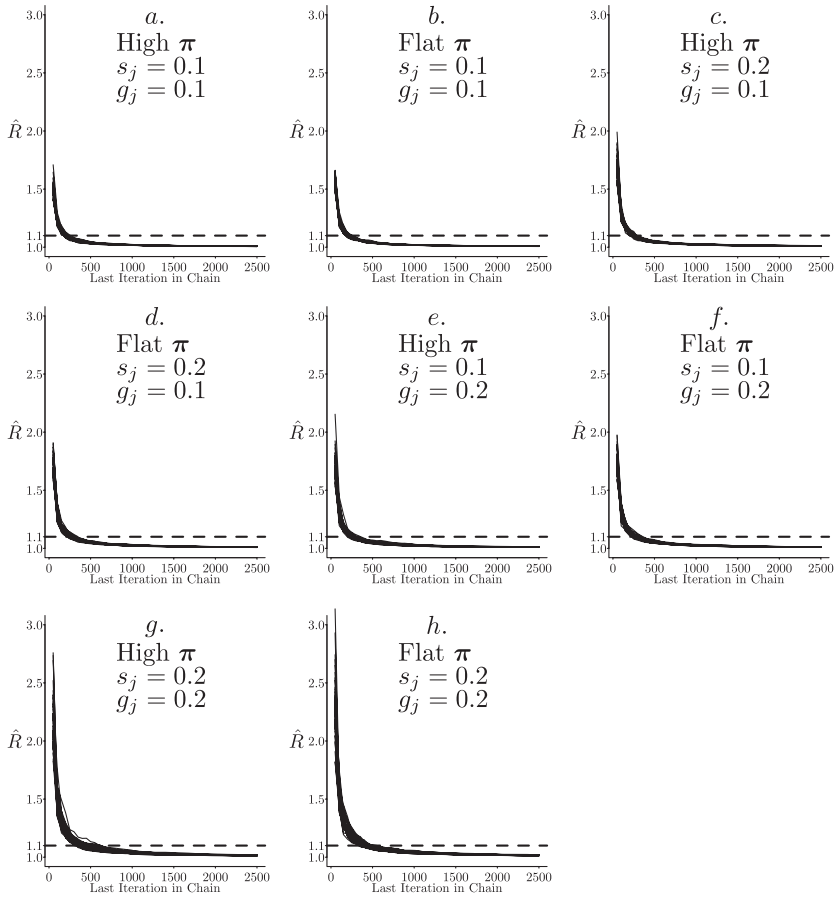

FIGURE 1. *Plots of Brooks-Gelman multivariate potential scale reduction factor $\hat{R}$ for Monte Carlo simulation conditions.* Note. *Each panel includes plots of $\hat{R}$ from 50 replications. $\hat{R}$ is calculated from five independent chains with different starting values and assesses convergence of item parameters, $\mathbf{\Omega}$, and latent class probabilities, $\boldsymbol{\pi}$. The plots are based upon chain lengths from 50 to 2,500 in increments of 50.*

twice as likely as belonging to a class with zero, one, or two skills (i.e., a high skill distribution). That is, $\pi_c = .021$ for all $\boldsymbol{\alpha}_c$ with two or fewer skills and $\pi_c = .042$ for $\boldsymbol{\alpha}_c$ with three or more skills.

Third, the simulation study assessed the number of iterations needed to achieve convergence in the multivariate posterior distribution. For each replication, five chains were run and the multivariate potential scale reduction factor $\hat{R}$ was computed (Brooks & Gelman, 1998) using the R "coda" package (Plummer,

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

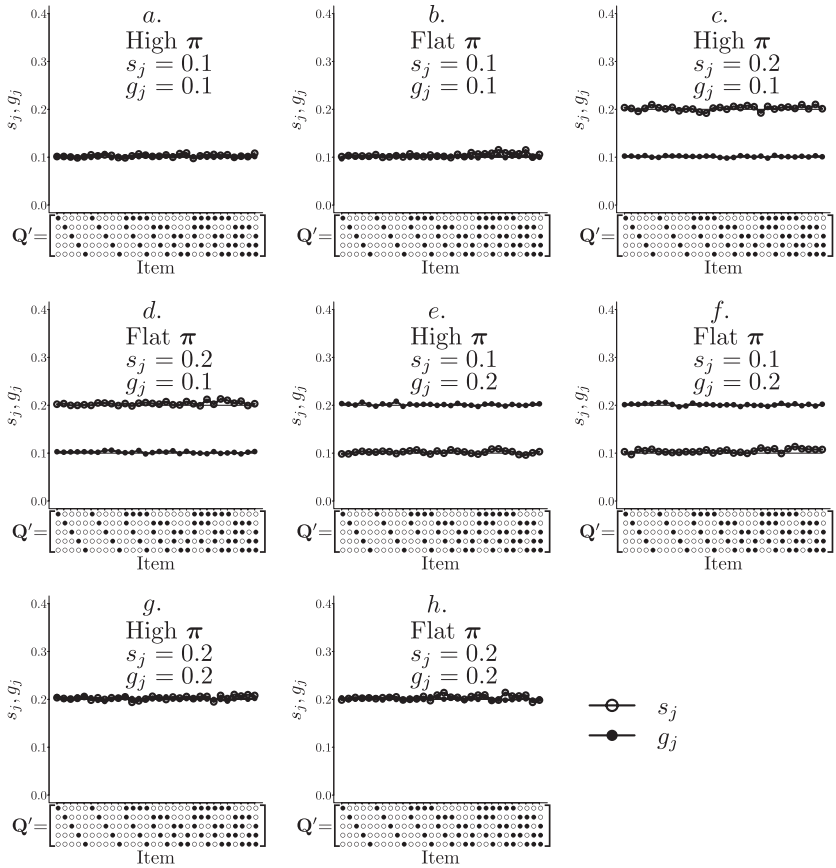

FIGURE 2. *Monte Carlo expected slipping and guessing parameters for different population values for s_j, g_j, and $\boldsymbol{\pi}$.* Note. *The skills required for each item are listed in the $K \times J$ $\mathbf{Q}'$ matrix along the x-axis where $\circ =$ "0" and $\bullet =$ "1". Expected values were estimated from 50 replications.*

Best, Cowles, &, Vines 2006) to assess convergence of $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$. Values of $\hat{R}$ less than 1.1 or 1.2 are considered evidence that the chain has converged (Brooks & Gelman, 1998). Note that $\hat{R}$ was computed for chain lengths from 50 to 2,500 to assess the burn-in needed to achieve convergence.

The MCMC algorithm was implemented with 5,000 iterations, and the first 1,000 were discarded. Estimation accuracy was assessed by computing the expected values and root mean squared error (RMSE) for s_j, g_j, and $\boldsymbol{\pi}$ over 50 replications. Uninformative priors were used for the Beta and Dirichlet distributions with $\alpha_g = \alpha_s = \beta_g = \beta_s = 1$ and $\boldsymbol{\delta}_0 = \mathbf{1}_{32}$, where $\mathbf{1}_{32}$ is a 32-dimensional vector with elements equal to one.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

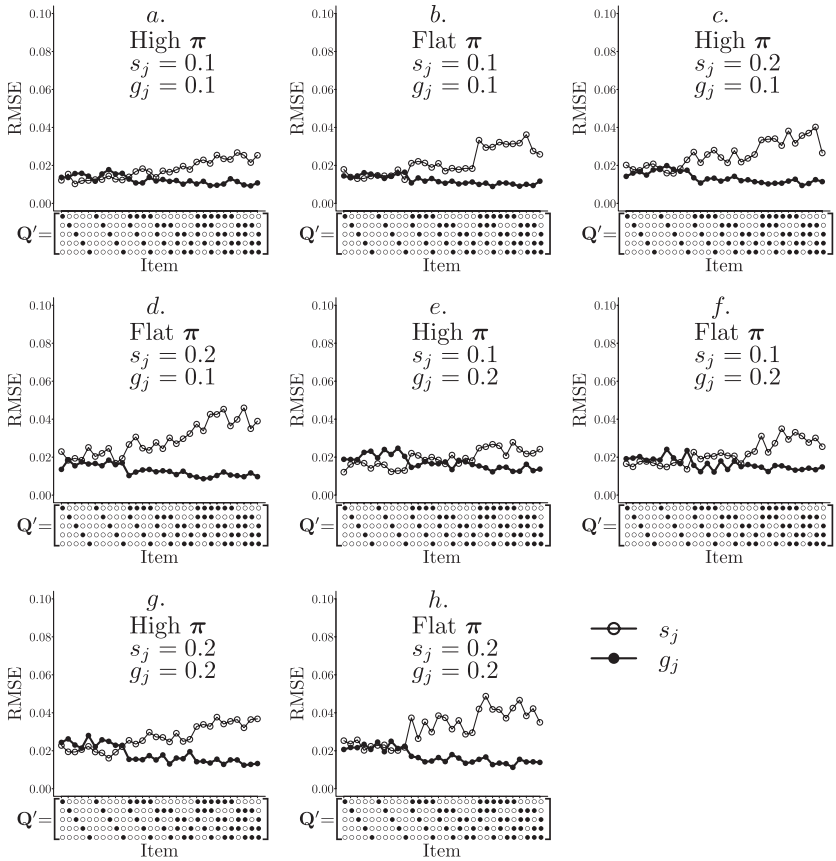

FIGURE 3. *Monte Carlo root mean squared error (RMSE) of slipping and guessing parameters for different population values for s_j, g_j, and $\boldsymbol{\pi}$.* Note. *The skills required for each item are listed in the $K \times J$ $\mathbf{Q}'$ matrix along the x-axis where $\circ$ = "0" and $\bullet$ = "1". RMSE values were estimated from 50 replications.*

Figure 1 provides evidence of convergence for the simulation posterior distributions. Specifically, Figure 1 includes eight panels plotting the Brooks–Gelman $\hat{R}$ for 50 replications. Figure 1 provides evidence that the diagnosticity condition impacts convergence, whereas the nature of the population skill distribution (i.e., high vs. flat) does not impact $\hat{R}$. For instance, in the high diagnosticity case (i.e., $s_j = g_j = 0.1$), Figure 1 shows that 300 iterations yielded $\hat{R} < 1.1$. In contrast, Figure 1 provides evidence that in the low diagnosticity condition (i.e., $s_j = g_j = 0.2$) that $\hat{R} < 1.1$ after roughly 750 iterations. Consequently, the Monte Carlo evidence suggests that employing a burn-in of 1,000 should be adequate for the simulation study to achieve convergence to the posterior distribution.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

Figures 2 and 3 plot the Monte Carlo estimated expected values and RMSE of the slipping and guessing parameters across populations values for s_j, g_j, and $\boldsymbol{\pi}$. Figure 2 provides evidence that the proposed algorithm yielded minimal bias for item parameters across the eight simulation conditions. Figure 3 plots the slipping and guessing RMSE for the 30 items across the eight simulation conditions. Figure 2 demonstrated the algorithm yielded unbiased estimates of item parameters, which implies that the RMSE values in Figure 3 quantify sampling variability. Similar to prior research (de la Torre, 2009a), Figure 3 provides evidence that sampling variability for slipping parameters is associated with the number of requisite skills (i.e., $\mathbf{q}'_j\mathbf{q}_j$). The sampling variability of s_j was nearly twice as large for some items that required three skills as opposed to items that required one skill. Furthermore, the sampling variability of s_j was larger in cases where $\boldsymbol{\pi}$ had a flat versus high skill distribution. Consequently, the results in Figure 3 provide evidence of an interaction effect involving the number of ones in $\mathbf{q}_j$ and the skill distribution on the sampling variability of slipping parameters. In contrast, Figure 3 suggests that the sampling variability for g_j was relatively smaller for items that required more skills.

The RMSE for slipping parameters exceeded 0.04 for items that require three skills, which, given that bias is near 0, translates to a margin of error of approximately $\pm .08$ (i.e., $2 \cdot 0.04$ if the sampling distribution for slipping parameters with $N = 1{,}000$ is approximately normal). There are at least two interpretations. One possibility is that the proposed algorithm, while unbiased, may not be efficient. Alternatively, the observed sampling variability in Figure 3 could be characteristic of the DINA model. Note that de la Torre (2009) found a similar relationship between the requisite skills and the RMSE for item parameters with the exception that the maximum standard error for slipping and guessing parameters was 0.031 and 0.017, respectively, for $N = 2{,}000$ and a flat π distribution. That is, de la Torre (2009) found a negative relationship between $\mathbf{q}'_j\mathbf{q}_j$ and RMSE for g_j and a positive relationship between $\mathbf{q}'_j\mathbf{q}_j$ and the RMSE for s_j. An additional simulation study replicated de la Torre (2009) with 100 samples to gather evidence as to the plausibility of either explanation. In the follow-up simulation study, the maximum RMSE for slipping and guessing was 0.032 and 0.017, which provides evidence for the latter explanation that slipping (guessing) parameters are more variable for items that require more (fewer) skills.

Figures 4 and 5 present Monte Carlo evidence concerning the bias and RMSE for the latent class skill distribution $\boldsymbol{\pi}$. Similar to Figure 2, the proposed algorithm yielded unbiased estimates of $\boldsymbol{\pi}$ across the simulation conditions. Figure 5 plots RMSE for $\boldsymbol{\pi}$ and provides information about sampling variability. Unlike Figure 3, Figure 5 shows that RMSE for $\boldsymbol{\pi}$ was unrelated to the diagnosticity condition and shape of the latent class skill distribution. However, in panel *h* of Figure 5 (i.e., the Flat $\boldsymbol{\pi}$ and low diagnosticity case), the sampling variability of π_c was elevated for less skilled latent classes.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

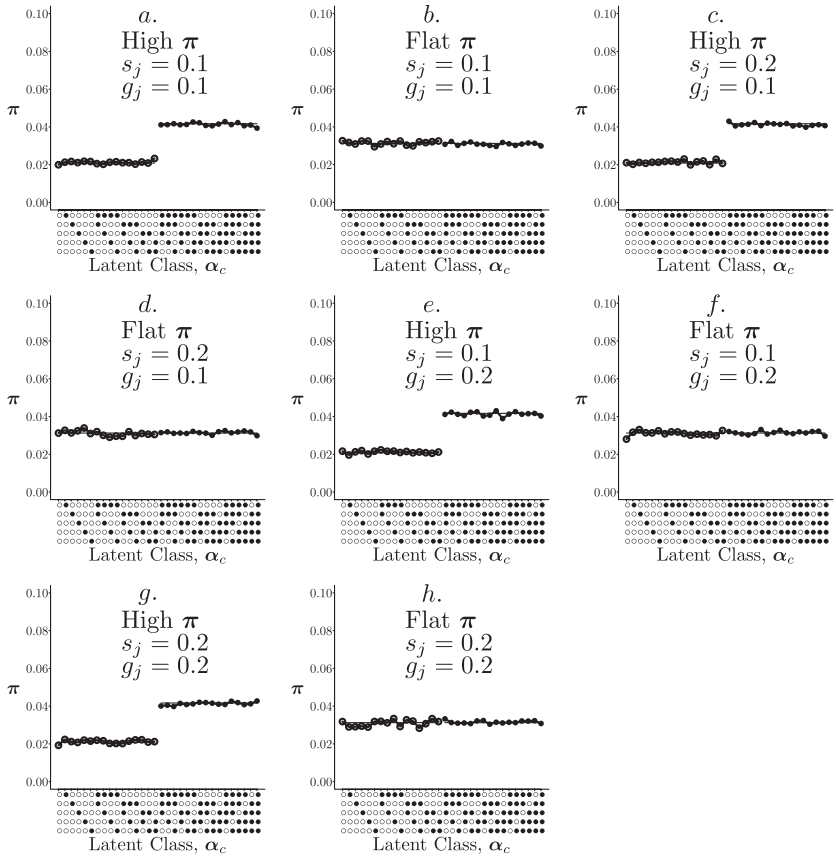

FIGURE 4. *Monte Carlo expected latent class skill distribution for different population values for s_j, g_j, and* $\boldsymbol{\pi}$. Note. α_c *is included in the x-axis label where* $\circ$ = "0" *for the absence of a skill and* $\bullet$ = "1" *for the presence of a skill. Expected values were estimated from 50 replications.*

In short, the Monte Carlo simulation study supports several conclusions. First, the new Gibbs sampler accurately estimates model parameters. Second, the algorithm is accurate when employing uninformative priors. Third, the results suggest that the model converges in a reasonable number of iterations.

Application: Revised Purdue Spatial Visualization Tests–Visualization of Rotations

This subsection reports an application of the DINA model to 388 responses to the Revised Purdue Spatial Visualization Tests–Visualization of Rotations

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

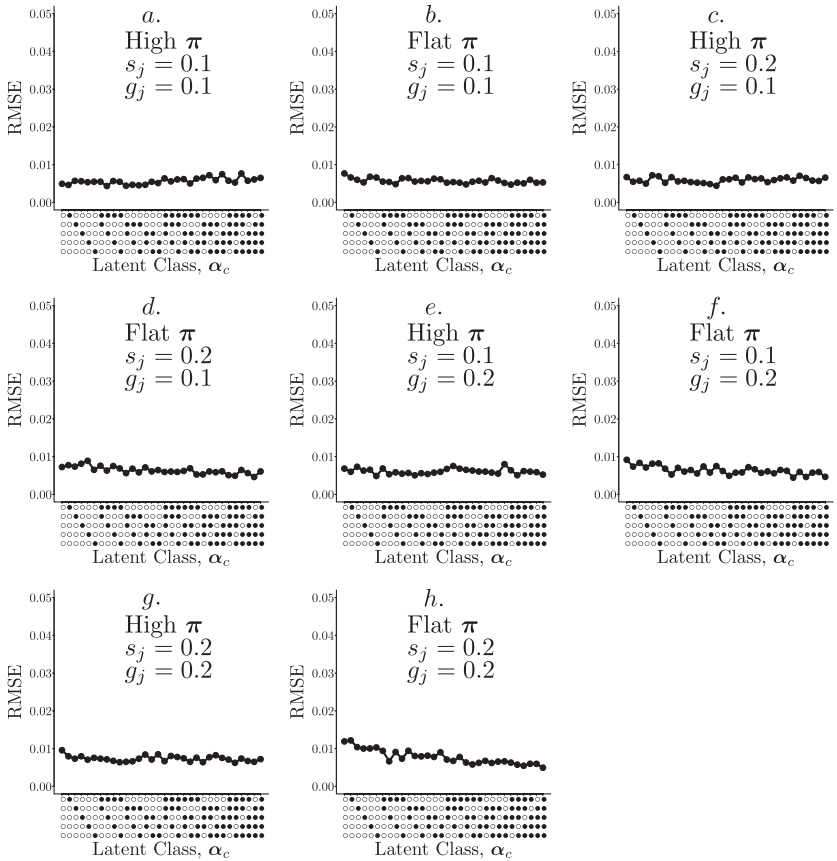

FIGURE 5. *Monte Carlo root mean squared error (RMSE) of latent class skill distribution for different population values for s_j, g_j, and* **π**. Note. α_c *is included in the x-axis label where ○ = "0" for the absence of a skill and ● = "1" for the presence of a skill. RMSE values were estimated from 50 replications.*

(PSVT-R; see Maeda, Yoon, Kim-Kang, & Imbrie, 2013; Maeda & Yoon, 2013; Yoon, 2011). The PSVT-R is a revision of Guay (1976) and assesses test takers' ability to mentally rotate three-dimensional objects along the x- and y-axes. Figure 6 includes a practice PSVT-R item. For the PSVT-R, items include a reference object that is rotated and then subjects are presented a new object and must determine which of the five options corresponds to the rotated version. For the practice item in Figure 6, subjects must be able to perform 90° rotations on both the x- and the y-axes (the correct answer is "D"). All items include x- and y-axes rotations with objects of varying complexity. Accordingly, the DINA model was applied to the PSVT-R by defining four mental rotation skills: (1) 90° x-axis (i.e.,

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

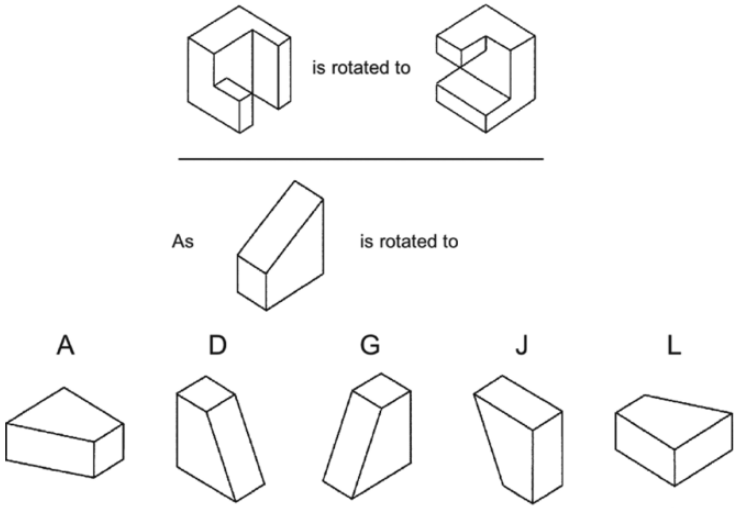

FIGURE 6. *Practice item from the Revised Purdue Spatial Visualization Tests–Visualization of Rotations.*

$k = 1$), (2) 90° y-axis (i.e., $k = 2$), (3) 180° x-axis (i.e., $k = 3$), and (4) 180° y-axis (i.e., $k = 4$). The defined skills imply a hierarchical structure. Namely, students who have mastery of 180° rotations should also be skilled at 90° rotations. Consequently, the hierarchical structure of the spatial rotation tasks requires 9 of the 16 possible latent classes. More specifically, the following classes are omitted: "0010," "0001," 1001," "0110," "0011," "1011," and "0111," because the 180° rotation skills appear without the corresponding 90° rotation skills.

Figure 7 plots the Brooks-Gelman multivariate $\hat{R}$ to provide information about posterior convergence for $\mathbf{\Omega}$ and $\boldsymbol{\pi}$. The computed $\hat{R}$ was calculated from five independent chains that employed random starting values. Specifically, starting values for s_j and g_j were drawn from uniform distributions, $\boldsymbol{\pi}$ was sampled from a uniform Dirichlet, and $\boldsymbol{\alpha}_i$ was sampled from a uniform categorical distribution. The plot is based upon chain lengths from 50 to 2,500 in increments of 50. A standard rule of thumb is that convergence to a posterior distribution has been reached once $\hat{R}$ is below 1.1 or 1.2. Figure 7 provides evidence $\hat{R} < 1.1$ after 350 iterations. The following discussion interprets parameter estimates using a burn-in of 1,000 iterations.

Table 1 reports the $\mathbf{Q}$ matrix and MCMC estimates of the posterior item parameters. First, note that the 30 PSVT-R items demonstrated variability in estimated guessing and slipping parameters. Specifically, g_j ranged from 0.084 to 0.733 and $1 - s_j$ ranged from 0.384 to 0.961. Consequently, heterogeneity in

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

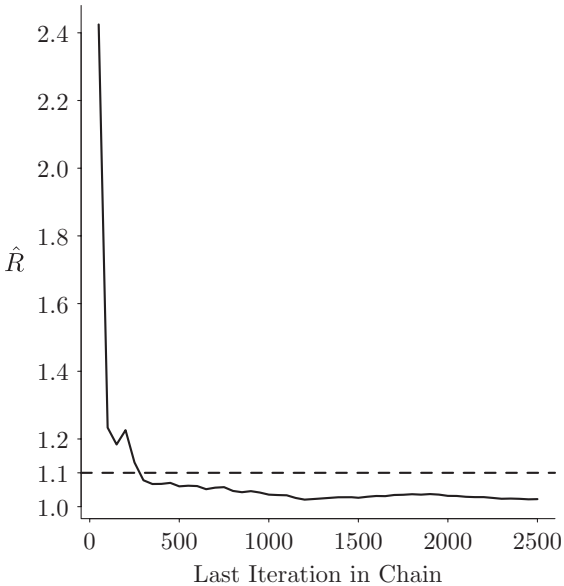

FIGURE 7. *Plot of Brooks-Gelman multivariate potential scale reduction factor, $\hat{R}$, for the Purdue Spatial Visualization Tests–Visualization of Rotations (PSVT-R) application.* Note. *$\hat{R}$ is based upon five independent chains with different starting values and assesses convergence of item parameters, $\mathbf{\Omega}$, and latent class probabilities, $\boldsymbol{\pi}$. The plot is based upon chain lengths from 50 to 2,500 in increments of 50.*

object complexity contributed to heterogeneity in slipping and guessing parameters. For instance, students who have the requisite skills are able to answer 21 of the 30 items with greater than an 80% probability. In contrast, students with the requisite skills have less than a 50% probability of correctly answering Items 29 and 30 (i.e., $1 - \hat{s}_{29} = 0.492$ and $1 - \hat{s}_{30} = 0.384$). The ability of subjects to correctly guess with high probability is expected because some items use simpler shapes, and the distractors may be easier to detect (e.g., see the practice problem in Figure 6). For instance, the probability of guessing the correct answer on Items 1 and 2 was $\hat{g}_1 = 0.617$ and $\hat{g}_2 = 0.733$, respectively. In contrast, the probability of correctly guessing Item 30, which had a more complex shape, was nearly 0 with $\hat{g}_{30} = 0.084$.

It is important to recall that items with larger g_j and s_j provide less diagnostic information. The results in Table 1 suggest that several PSVT-R items have larger g_j and s_j. One concern with applying CDMs is misspecification of $\mathbf{Q}$. The PSVT-R items require the skill to dynamically rotate objects, and the specified $\mathbf{Q}$ matrix is consistent with prevailing spatial cognition theory (Uttal et al., 2013). Consequently, additional research is needed to assess the PSVT-R item quality (e.g., revision of distractors or objects).

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

TABLE 1.
PSVT-R **Q** *Matrix and MCMC Estimates of* g_j *and* $1 - s_j$

	Q				$\hat{g}$		$1 - \hat{s}$	
Item	α_1	α_2	α_3	α_4	Mean	*SE*	Mean	*SE*
1	1	0	0	0	0.617	0.089	0.944	0.017
2	0	1	0	0	0.733	0.047	0.920	0.018
3	0	1	0	0	0.654	0.053	0.961	0.013
4	0	1	0	0	0.676	0.049	0.886	0.021
5	0	0	1	0	0.627	0.041	0.950	0.018
6	0	0	1	0	0.603	0.041	0.928	0.021
7	0	1	0	0	0.410	0.066	0.914	0.020
8	0	1	1	0	0.560	0.037	0.765	0.034
9	1	1	0	0	0.419	0.045	0.862	0.024
10	0	0	0	1	0.567	0.039	0.902	0.024
11	1	1	0	0	0.545	0.044	0.881	0.023
12	0	0	0	1	0.389	0.042	0.925	0.022
13	1	1	0	0	0.476	0.043	0.871	0.024
14	1	1	0	0	0.377	0.043	0.853	0.025
15	0	1	0	0	0.303	0.050	0.700	0.032
16	1	1	0	0	0.387	0.042	0.843	0.027
17	0	1	1	0	0.502	0.039	0.849	0.028
18	1	1	0	0	0.398	0.044	0.757	0.029
19	0	0	0	1	0.421	0.039	0.870	0.029
20	1	1	0	0	0.378	0.043	0.848	0.026
21	0	0	0	1	0.336	0.039	0.847	0.030
22	0	0	0	1	0.402	0.039	0.813	0.030
23	1	0	0	1	0.423	0.036	0.830	0.033
24	0	0	1	0	0.404	0.041	0.760	0.035
25	0	1	1	0	0.356	0.037	0.722	0.037
26	0	1	1	0	0.293	0.035	0.704	0.039
27	1	0	0	1	0.420	0.036	0.820	0.033
28	0	1	1	0	0.268	0.034	0.660	0.038
29	1	1	0	0	0.209	0.034	0.492	0.034
30	0	1	1	0	0.084	0.022	0.384	0.037

Note. The rotation skills are $\alpha_1 = 90^\circ$ x-axis, $\alpha_2 = 90^\circ$ y-axis, $\alpha_3 = 180^\circ$ x-axis, and $\alpha_4 = 180^\circ$ y-axis. PSVT-R = Purdue Spatial Visualization Tests–Visualization of Rotations; MCMC = Markov Chain Monte Carlo.

Table 2 reports information for $\hat{\pi}_g$ and shows that the majority of subjects were classified into two of the nine classes. Latent class "1111" was the most prevalent with $\hat{\pi}_9 = .456$, which provides evidence that nearly half of subjects were skilled at 180° rotations along the x- and y-axes. The second most prevalent latent class consisted of subjects who were not skilled at 90° rotations (i.e., latent

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

TABLE 2.
MCMC Posterior Class Probability Distribution for Spatial Rotation Test

Latent Classes				$\hat{\pi}$	
α_1	α_2	α_3	α_4	Mean	*SE*
0	0	0	0	0.188	0.053
1	0	0	0	0.083	0.051
0	1	0	0	0.073	0.029
1	1	0	0	0.089	0.028
1	0	1	0	0.023	0.018
0	1	0	1	0.038	0.016
1	1	1	0	0.029	0.019
1	1	0	1	0.018	0.015
1	1	1	1	0.456	0.033

Note. The rotation skills are $\alpha_1 = 90°$ x-axis, $\alpha_2 = 90°$ y-axis, $\alpha_3 = 180°$ x-axis, and $\alpha_4 = 180°$ y-axis. MCMC = Markov Chain Monte Carlo; *SE* = standard error.

class "0000") as evidenced by $\hat{\pi}_1 = .188$. The results in Table 2 also suggest that 24.6% of subjects were skilled at only 90° rotations (i.e., latent classes "1000," "0100," or "1100"). Finally, the results in Table 2 suggest that 2.3% of subjects were only skilled at x-axis rotations (i.e., "1010"), and 3.8% subjects were only skilled at y-axis rotations (i.e., "0101").

Discussion

This study presented a Bayesian formulation of the widely applied DINA model. An efficient Gibbs sampler was developed to estimate model parameters. The purpose of this section is to discuss the benefits of the current study in the context of existing research and to describe directions for future research.

First, the proposed formulation could be extended to estimate **Q**. CDMs assume that researchers are able to accurately articulate the skills needed to correctly respond to items in **Q**. Recent research considered statistical procedures for estimating **Q** (Chen, Liu, Xu, & Ying, 2014; Liu, Xu, & Ying, 2012, 2013) and for detecting a misspecified **Q** matrix (Chiu, 2013; de la Torre, 2008; Rupp & Templin, 2008). For instance, Chiu (2013) provides an overview of methods and procedures for refining **Q**. Furthermore, DeCarlo (2012) and Templin and Henson (2006a) described Bayesian approaches for modeling selected elements of **Q** as random rather than fixed. Note that Templin and Henson (2006a) employed MH sampling and Decarlo (2012) used Gibbs sampling with OpenBUGS. The proposed Bayesian formulation

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

can incorporate prior strategies for modeling elements of $\mathbf{Q}$ with the advantages of the proposed Bayesian formulation, which includes (1) improved mixing of $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$, (2) enforcement of the monotonicity requirement, and (3) ease of implementation without tuning.

Second, the DINA model is a special case of more general CDMs (e.g., see de la Torre, 2011; Henson et al., 2009; von Davier, 2008). Future research could extend the Bayesian formulation developed in this study to more general CDM models. Specifically, the model formulation described in this article demonstrated how to exploit the fact that CDMs include a finite number of latent classes. Future research should consider implementing a similar strategy for estimating other CDMs.

Third, evidence from the Monte Carlo simulation and application suggests that the proposed algorithm converged to the parameter posterior distribution within a reasonable number of MCMC iterations. For instance, the results in Figures 1 and 6 suggest the Gibbs sampler converged generally in under 750 iterations. Furthermore, the simulation study provided evidence that the algorithm is accurate when using uninformative priors for $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$.

Computing the full conditional distribution over the 2^K latent classes is not without some disadvantages and " ... some simplification might be desired when the dimension of $\boldsymbol{\alpha}$ is larger than $K = 3$ or perhaps $K = 4$" (de la Torre & Douglas, 2004, p. 337). The MCMC program was written in C++ using the Rcpp package (Eddelbuettel et al., 2011; Eddelbuettel, 2013) and completed 5,000 MCMC iterations with $J = 30$, $K = 5$, and $N = 1,000$ in approximately 128 seconds using a laptop with a 2.4 GHz processor and 6GB of RAM. However, the developed Gibbs sampler likely requires additional computational resources in comparison to procedures that approximate the latent class distribution. For example, de la Torre and Douglas (2004) employed concepts from nonlinear factor analysis to model higher order factors underlying latent skills. de la Torre and Douglas (2004) proposed the HO-DINA where skills are independent given a higher order $K^{'} \leq K$ dimensional continuous latent trait θ_i, such that $p(\boldsymbol{\alpha}_i|\boldsymbol{\theta}_i) = \prod_{k=1}^{K} p(\alpha_{ik}|\boldsymbol{\theta}_i)$. The HO-DINA approximates the 2^K latent classes with $K^{'}(K^{'}-1)/2$ parameters for the relationships among the higher order factors and $2K^{'}$ parameters for the latent intercepts and loadings if the factor model has simple structure. If $K = 20$, there are 1,048,576 latent classes that the developed algorithm would loop over for each student for every MCMC iteration. In contrast, employing an HO-DINA model with $K^{'} = 20$ would require the estimation of 230 parameters and could, if the approximation is adequate, be computationally advantageous. Additional research is needed for understanding best practice in applying CDMs in high-dimensional applications.

In conclusion, a new Bayesian model formulation was presented for the DINA model. Research regarding CDMs remains a priority, and the methods

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

developed in this article provide an alternative for easily estimating model parameters.

Appendix

Sampling s_j and g_j with the Monotonicity Restriction of $0 \leq s_j + g_j < 1$, the slipping and guessing parameters must be jointly sampled from the full conditional distributions to ensure the monotonicity of likelihood condition, $0 \leq s_j + g_j < 1$. Let f_{sg} denote the joint density function of s_j and g_j, let f_s and f_g be the respective marginal density functions, and let $f_{s|g}$ and $f_{g|s}$ be the conditional distributions. For the model formulation discussed in this article, the joint distribution f_{sg} is defined as,

$$f_{sg} \propto f_s f_g \mathcal{I}\Big((s,g) \in \mathcal{P}\Big), \tag{A1}$$

where $\mathcal{P} = \{(s,g) : 0 \leq s + g < 1, 0 \leq s < 1, 0 \leq g < 1\}$ and $\mathcal{I}$ denotes the indicator function. Accordingly, the marginal distribution of g_j is,

$$f_g \propto \int_0^{1-g} f_{sg} ds = f_g \int_0^{1-g} f_s ds = f_g F_s(1-g), \tag{A2}$$

where $F_s(1-g)$ is the cumulative distribution function of s_j evaluated at $1-g$. Note that f_g is a continuous mixture distribution and is analytically intractable (Pham-Gia & Turkkan, 1998). The conditional distribution of s_j given g_j is,

$$f_{s|g} = \frac{f_{sg}}{f_g} \propto \frac{f_s}{F_s(1-g)} \mathcal{I}(0 \leq s < 1-g), \tag{A3}$$

which is a right-truncated Beta distribution at $1-g$. A similar argument implies that $f_{g|s}$ is a right-truncated Beta distribution at $1-s$. Let $g_j^{(t)}$ and $s_j^{(t)}$ be the guessing and slipping parameters at iteration t. g_j and s_j can be sampled from the linearly truncated full conditional distribution in two steps,

1. Sample $g_j^{(t)} | s_j^{(t-1)} \sim \text{Beta}(\tilde{\alpha}_g, \tilde{\beta}_g) \mathcal{I}\Big(0 \leq g_j^{(t)} < 1 - s_j^{(t-1)}\Big)$.
2. Draw $s_j^{(t)} | g_j^{(t)} \sim \text{Beta}(\tilde{\alpha}_s, \tilde{\beta}_s) \mathcal{I}\Big(0 \leq s_j^{(t)} < 1 - g_j^{(t)}\Big)$.

Acknowledgments

This article benefited from the comments and suggestions of Jeffrey Douglas, Aaron Hudson, Li Cai, and three anonymous reviewers. Any remaining errors belong to the author.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

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

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

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*, 850–866.

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

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

Culpepper, S. A. (in press). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*.

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*, 343–362.

de la Torre, J. (2009a). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115–130.

de la Torre, J. (2009b). *Estimation code for the G-DINA model*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA.

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

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 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*, 595–624.

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin, Heidelberg: Springer.

Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. New York, NY: Springer.

Eddelbuettel, D., François, R., Allaire, J., Chambers, J., Bates, D., & Ushey, K. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*, 1–18.

Guay, R. (1976). *Purdue spatial visualization test*. West Lafayette, IN: Purdue University.

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

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*, 191–210.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, *71*, 407–419.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.

Li, F. (2008). *A modified higher-order DINA model for detecting differential item functioning and differential attribute functioning* (Unpublished doctoral dissertation). University of Georgia, Athens.

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

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

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational and Behavioral Statistics*, *2*, 99–120.

Maeda, Y., & Yoon, S. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue Spatial Visualization Tests: Visualization of rotations (PSVT:R). *Educational Psychology Review*, *25*, 69–94.

Maeda, Y., Yoon, S. Y., Kim-Kang, G., & Imbrie, P. (2013). Psychometric properties of the revised PSVT:R for measuring first year engineering students' spatial ability. *International Journal of Engineering Education*, *29*, 763–776.

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

Pham-Gia, T., & Turkkan, N. (1998). Distribution of the linear combination of two general beta variables and applications. *Communications in Statistics-Theory and Methods*, *27*, 1851–1869.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11. Retrieved from http://CRAN.R-project.org/doc/Rnews/

Robert, C., & Casella, G. (2009). *Introducing Monte Carlo methods with R*. New York, NY: Springer.

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnosis system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275–318). Cambridge, England: Cambridge University Press.

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, *44*, 293–311.

Rupp, A. A. (2009). *Software for calibrating diagnostic classification models: An overview of the current state-of-the-art*. 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*, 78–96.

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

Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, *72*, 217–232.

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015

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

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

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

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*, 352.

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

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*, 49–71.

Yoon, S. Y. (2011). *Psychometric properties of the revised Purdue Spatial Visualization Tests: Visualization of rotations (the revised PSVT-R)* (Unpublished doctoral dissertation). Purdue University, West Lafayette.

Author

STEVEN ANDREW CULPEPPER is an Assistant Professor at the University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL 61820; email: sculpepp@illinois.edu. His research interests include statistical modeling in the social sciences, Bayesian models and computation, and alternative standardized testing formats.

Manuscript received September 20, 2014
First revision received March 4, 2015
Second revision received May 22, 2015
Accepted May 27, 2015

Downloaded from http://jebs.aera.net at CMU Libraries - library.cmich.edu on December 8, 2015