As a statistical tool to assist formative assessments in educational settings, diagnostic classification models (DCMs) have been increasingly used to provide diagnostic information regarding examinees’ attributes. DCMs often adopt a dichotomous division such as the mastery and non-mastery of attributes to express the mastery states of attributes. However, many practical settings involve different levels of mastery states rather than a simple dichotomy in a single attribute. Although this practical demand can be addressed by polytomous-attribute DCMs, their computational cost in a Markov chain Monte Carlo estimation impedes their large-scale application due to the larger number of polytomous-attribute mastery patterns than that of binary-attribute ones. This study considers a scalable Bayesian estimation method for polytomous-attribute DCMs and developed a variational Bayesian (VB) algorithm for a polytomous-attribute saturated DCM—a generalization of polytomous-attribute DCMs—by building on the existing literature on polytomous-attribute DCMs and VB for binary-attribute DCMs. Furthermore, we proposed the configuration of parallel computing for the proposed VB algorithm to achieve better computational efficiency. Monte Carlo simulations revealed that our method exhibited the high performance in parameter recovery under a wide range of conditions. An empirical example is used to demonstrate the utility of our method.

Keywords: polytomous attribute, polytomous-attribute saturated DCM, variational Bayesian inference, parallel computing

The recent demand for formative assessments has motivated the development of psychometric models providing diagnostic information on students’ latent traits. A class of these models is termed diagnostic classification models (DCMs). It postulates that latent traits generally consist of multidimensional discrete skills called “attributes,” and students belong to one of the mastery profiles for these attributes, each of which presents a combination of the mastery and non-mastery of attributes (Rupp et al., 2010). Based on this assumption, DCMs aim to estimate the mastery probabilities of attributes for every individual and classify them into attribute mastery profiles. Thanks to this information, practitioners can design classroom activities that emphasize students’ educational needs.

Although most of the DCMs assign binary attributes in their parameterization, a few DCMs with polytomous attributes have been developed to perform more finer-grained diagnoses (for example, Chen & de la Torre, 2013; Karelitz, 2004; Templin & Bradshaw, 2013; Templin, 2004; von Davier, 2007; 2008). The seminal contribution to polytomous-attribute DCMs is Karelitz’s (2004) work, which introduced the ordered-category attribute coding framework to treat different attribute mastery levels in an attribute. In contrast to binary attributes that only consider a dichotomous division between their mastery and non-mastery, this framework broadens the dichotomous treatment and addresses the qualitative ordering of cognitive complexity within an attribute. For instance, Tjoe and de la Torre (2014) incorporated polytomous attributes as a part of the specified attributes for proportional reasoning, where “constructing ratios” and “constructing proportions” are assumed to present different mastery levels. Both attributes fall within the scope of a ratio. However, constructing proportions is more difficult to master than constructing ratios because calculating proportions requires the understanding of ratios. The inclusion of such attributes forms a new class of attribute mastery profiles untouched by binary attributes and helps practitioners measure students’ levels of mastery over attributes with more refined granularity.

This study considers the polytomous-attribute extension of an ordinary binary-attribute saturated DCM. A saturated DCM is a generalization of DCMs in the sense that all the main and interaction effects of the attributes are allowed, and each possible attribute mastery profile can have a unique correct-response probability. Examples of such binary-attribute models include the general diagnostic model (GDM: von Davier, 2008), log-linear cognitive diagnostic model (LCDM: Henson et al., 2009), and generalized deterministic input, noisy “and” gate
(G-DINA) model (de la Torre, 2011). For the generalized model of polytomous-attribute DCMs, Chen and de la Torre (2013) developed the polytomous G-DINA (pG-DINA) model by introducing a method to reduce polytomous-attribute mastery profiles to item-specific binary profiles pertaining to the unique correct-response probabilities for an item. This enables estimating the parameters of the pG-DINA model in the same manner as the G-DINA model.

Regardless of polytomy in an attribute, a class of saturated DCMs holds substantial utility. First, any sub-models in DCMs such as the deterministic input, noisy “and” gate (DINA; Junker & Sijtsma, 2001) model, deterministic input, noisy “or” gate (DINO; Templin & Henson, 2006) model, reduced reparameterized unified model (RRUM; Hartz & Roussos, 2008), and compensatory RUM (CRUM; Rupp et al., 2010) can be expressed within a saturated DCM by imposing appropriate restrictions on its parameters. For example, a saturated DCM with only the intercept and highest-order interaction terms of required attributes allowed is effectively the DINA model, which assumes that examinees must master all the necessary attributes to answer an item correctly. This flexibility in modeling different types of item-responding processes brings about the second advantage. Since a saturated DCM nests its sub-models, we can infer the nature of the item-responding process behind an item by inspecting its item parameter estimates. As is the case with the DINA model, if the intercept and highest-order interaction terms of a saturated DCM are estimated to be high, and other terms are estimated to be close to zero, then the item-responding process behind a corresponding item is likely to be the DINA model. As such, the item parameters in a saturated DCM richly portray the item-responding processes behind the items, informing practitioners quantitatively about how examinees respond to given items.

Despite the merits of a saturated DCM mentioned above, the computational speed for its Bayesian estimation can be painstakingly slow at large-scale settings where the numbers of examinees and attributes are sizable. This is because Markov chain Monte Carlo (MCMC)—a commonly used method for Bayesian estimation—is generally not scalable to such large-scale settings owing to its stochastic search for the parameter space of the targeted posterior (Gelman et al., 2013). To reduce this computational cost, a variational Bayesian (VB) inference method is often employed as an alternative to an MCMC estimation. VB inference is a deterministic and fast approach to approximate posterior distributions. This deterministic nature in an estimation procedure enables a scalable Bayesian estimation, and many applied researchers have utilized it for their Bayesian modeling to capitalize on its high scalability (Blei et al., 2017). In DCM literature, several VB inference algorithms have been developed in recent years (Yamaguchi, 2020; Yamaguchi & Okada, 2020; 2021). In particular, Yamaguchi and Okada (2021) developed a VB algorithm for a binary-attribute saturated DCM by introducing a G-matrix to reformulate it as a Bernoulli mixture model so that the priors for its model parameters become conditionally conjugate; this conditional conjugacy simplifies the derivation of a coordinate ascent mean-field VB inference (Wang & Blei, 2013). Additionally, Yamaguchi and Okada (2021) confirmed the sound accuracy and fast computation of their VB algorithm and showed the superior computational efficiency of VB estimation over maximum likelihood estimation based on an expectation-maximization (EM) algorithm under the condition that the response data are obtained consecutively, such as in computerized adaptive testing.

By contrast, none of these studies have worked on the problem of the scalability for polytomous-attribute DCMs despite the fact that these models would present a more severe computational problem in application settings. This particular severity in their MCMC computation results from the nature of polytomous attributes, in which the increase of mastery levels generally leads to expanding the parameter space of attribute mastery profiles from $2^K$ to $M^K$. Here, $K$ and $M$ denote the number of attributes and mastery levels, respectively. This expansion causes more intensive MCMC computation compared to binary-attribute DCMs. Therefore, it is desirable to develop a scalable Bayesian estimation algorithm for a polytomous-attribute saturated
DCM. Accordingly, in this study, we develop a VB algorithm for a polytomous-attribute saturated DCM, which builds on the prior work by Chen and de la Torre (2013) and Yamaguchi and Okada (2021), and propose its parallelized algorithm based on the configuration of the parallel-E parallel-M algorithm for generalized latent variable models (von Davier, 2016). Furthermore, we conduct simulation and empirical studies to assess the utility of our algorithm.

The remainder of this paper is structured as follows. In Section 2, we first introduce the two foundational models of Chen and de la Torre (2013) and Yamaguchi and Okada (2021) and then explain the polytomous-attribute G-matrix to derive a VB algorithm for a polytomous-attribute saturated DCM and the configuration of its parallel computing. Subsequently, our simulation and empirical studies are presented in Sections 3 and 4. Lastly, we discuss the limitations and future directions of this study.

1 METHODS

1.1 PG-DINA MODEL

Let $i$ $(1, \ldots, N)$, $j$ $(1, \ldots, J)$, and $k$ $(1, \ldots, K)$ express respondents, items, and attributes, respectively. For ease of notation, we assume that all the attributes have the same mastery level $M_k = M$. The number of attribute mastery profiles for polytomous-attribute DCMs is $M^K$, each of which is denoted as $l$ $(1, \ldots, L = M^K)$. An implementation of DCMs requires a $J \times K$ Q-matrix that specifies the relationships between the items and attributes. An entry $q_{jk}$ of a Q-matrix can take values from 0 to $M - 1$. If the value of $q_{jk}$ is $m$ $(0, \ldots, M - 1)$, more than the $m$-th level of mastery in attribute $k$ is necessary for a correct response to item $j$. A set of relevant attributes for item $j$ is indicated by $q_{j}^* = (q_{j1}, \ldots, q_{jK})^T$, which is the $j$-th vector of a Q-matrix $Q = (q_{1}, \ldots, q_{J})^T$. In addition, let $\alpha_l = (\alpha_{1l}, \ldots, \alpha_{Jl})^T$ be the $l$-th vector of an $L \times K$ attribute mastery profile matrix $A_{\text{profile}} = (\alpha_{1}, \ldots, \alpha_{L})^T$ that comprises all the possible attribute mastery profiles. An entry of $\alpha_{il}$ can also take values from 0 to $M - 1$. The superscript $T$ represents the transpose.

Moreover, following the notations adopted in Chen and de la Torre (2013) and de la Torre (2011), we use $K^*_j = \sum_{k=1}^{K} I(q_{jk} > 0)$ to denote the number of relevant attributes for item $j$. $I(\cdot)$ is an indicator function that takes the value of 1 when a given condition is satisfied. Using the notation of $K^*_j$, we can reduce each vector of an attribute mastery profile matrix $A_{\text{profile}}$ to the reduced vector $\alpha_{jl} = (\alpha_{j1*}, \ldots, \alpha_{jK^*_j*})^T$ for item $j$, where $l = 1, \ldots, M^{K^*_j*}$ and each entry of $\alpha_{j1*}$ corresponds to the attributes satisfying $q_{jk} > 0$ on a vector $q_{j}$. Besides, Chen and de la Torre (2013) modified the reduced vector to the collapsed attribute vector $\alpha_{jl}^* = (\alpha_{j1**}, \ldots, \alpha_{jK^*_j**})^T$ for further simplification, where each element of the reduced attribute vector $\alpha_{jl}^*$ is collapsed into a binary element:

$$
\alpha_{jlk}^* = \begin{cases} 
0 & \text{if } \alpha_{jlk} < q_{jk} \\
1 & \text{otherwise}
\end{cases}.
$$

The example of this transformation based on the one shown in Chen and de la Torre (2013) is illustrated in Table 1. Consider the case of $K = 3$, $M = 3$, and the item with $q_{j} = (2, 1, 0)^T$. Since $K^*_j$ for this $q$ vector is 2, the number of unique reduced attribute vectors $\alpha_{jl}^*$ is $M^{K^*_j} = 3^2 = 9$. Those vectors can be further collapsed into $L_j^* = 2^{K^*_j} = 2^2 = 4$ unique collapsed attribute vectors $\alpha_{jl}^*$. These collapsed vectors are essentially the patterns that can be distinguished by item $j$. 

The table depicts the reduced attribute vectors for a set of item masteries and the corresponding collapsed attribute vectors.
Accordingly, the number of unique correct-response probabilities for this item becomes 4, and each of them is expressed by the item response function of the pG-DINA model:

\[
P(x_{ij} = 1|\alpha^{***}_{jl}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha^{**}_{jk} + \sum_{k' > k}^{K_j^*} \delta_{jkk'}\alpha^{***}_{jkk'} + \cdots + \delta_{j1\ldots K_j^*} \prod_{k=1}^{K_j} \alpha^{**}_{jkk},
\]

where \( l = 1, \ldots, L_j^* (= 2^{K_j^*}) \). Here, \( \delta_{j0} \) and \( \delta_{jk} \) denote the intercept and main-effect terms of the relevant attributes, respectively. The terms after the second one denote all the combinations of the interaction effects of the relevant attributes. With the simplified formulation of attribute mastery profiles elaborated in Table 1, the item response function of the pG-DINA model is formulated to be the same as the G-DINA model, except for the fact that the attribute mastery profiles embedded in the item response function of the pG-DINA model are the collapsed attribute vectors, whereas those of the G-DINA model are the binary reduced attribute vectors (Chen & de la Torre, 2013). Because of such equivalence in their item response functions, the estimation procedure of the G-DINA model can be applied directly to that of the pG-DINA model.

### 1.2 Binary-Attribute Saturated DCM Using G-Matrices

Since each item possesses its own correct-response probabilities associated with item-specific attribute mastery profiles, it can be construed that a set of binary responses \( x_j \) for item \( j \) is generated from heterogeneous populations with different Bernoulli distribution functions, where each item-specific attribute mastery profile has a unique Bernoulli probability. Based on this insight, Yamaguchi and Okada (2021) leveraged the idea of mixture modeling for heterogeneous populations and reformulated a binary-attribute saturated DCM as a Bernoulli mixture model by introducing a latent indicator vector \( z_i \) and a G-matrix \( G_j \).

In this section, for improved clarity of notations between binary-attribute and polytomous-attribute DCMs, we use \( H = 2^K \) to denote the number of attribute mastery profiles for binary-attribute DCMs, although we use \( L = M^K \) to denote that for polytomous-attribute DCMs in Section 1.1. In addition, let \( h^* (1, \ldots, H_j^*) \) and \( l^* (1, \ldots, L_j^*) \) represent the indices for the item-specific number of reduced attribute vectors in binary-attribute DCMs and that of collapsed attribute vectors in polytomous-attribute DCMs, respectively. A latent indicator vector \( z_i = (z_{i1}, \ldots, z_{iH})^T \) specifies the attribute mastery pattern to which examinee \( i \) belongs and is a

| Original \( \alpha_l \) | Reduced \( \alpha_{jl}^* \) |Collapsed \( \alpha_{jl}^{**} \) |
|------------------------|-------------------------|------------------------|
| \((0,0,0),(0,0,1),(0,0,2)\) | \((0,0)\) | \((0,0)\) |
| \((1,0,0),(1,0,1),(1,0,2)\) | \((1,0)\) | \((0,0)\) |
| \((2,0,0),(2,0,1),(2,0,2)\) | \((2,0)\) | \((1,0)\) |
| \((0,1,0),(0,1,1),(0,1,2)\) | \((0,1)\) | \((0,1)\) |
| \((1,1,0),(1,1,1),(1,1,2)\) | \((1,1)\) | \((1,1)\) |
| \((0,2,0),(0,2,1),(0,2,2)\) | \((0,2)\) | \((0,1)\) |
| \((1,2,0),(1,2,1),(1,2,2)\) | \((1,2)\) | \((1,1)\) |
| \((2,1,0),(2,1,1),(2,1,2)\) | \((2,1)\) | \((1,1)\) |
| \((2,2,0),(2,2,1),(2,2,2)\) | \((2,2)\) | \((1,1)\) |

Table 1: The example of collapsed attribute vectors when \( q_j = (2,1,0) \)
vector with $H = 2^K$ elements, where the $h$-th element takes the value of 1 when examinee $i$ belongs to class $h$ so that the values of $z_{ih}$ satisfy $z_{ih} = 1$ and $\sum_{h=1}^{H} z_{ih} = 1$. A G-matrix $G_j$, which is the $H_j^* (= 2^{K_j^*}) \times H$ matrix, is introduced in Yamaguchi and Okada (2021) to reduce a latent indicator vector $z_i$ to an item-specific latent indicator vector $z_{ji} = (z_{j1}, \ldots, z_{jiH_j^*})^T$ with $H_j^*$ elements, indicating to which item-specific attribute mastery pattern examinee $i$ belongs. The example of a G-matrix $G_j$ based on Yamaguchi and Okada (2021) is illustrated in Table 2.

| Attribute mastery patterns | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----------------------------|---|---|---|---|---|---|---|---|
| $\alpha_1$                | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| $\alpha_2$                | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 |
| $\alpha_3$                | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |

| Item-specific attribute mastery patterns | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | G-matrix |
|------------------------------------------|-------------|-------------|-------------|-----------|
| 1 | 0 | 0 | * | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | * | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 1 | * | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 4 | 1 | 1 | * | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |

Consider the case of $K = 3$ and the item with $q_j = (1, 1, 0)^T$. Since the number of attributes measured by this item is $K_j^* = 2$, the number of reduced attribute mastery patterns becomes $H_j^* = 2^{K_j^*} = 4$. Thus, $G_j$ becomes a $4 \times 8$ matrix. With this matrix, an item-specific latent indicator vector $z_{ji}$ is computed by multiplying $G_j$ by $z_i$ such that $z_{ji} = G_j z_i$. Elements of $z_{ji}$ are $z_{jih} = \sum_{h=1}^{H} g_{jh} z_{jh}$, where $h^* = 1, \ldots, H_j^* (= 2^{K_j^*})$ and $h = 1, \ldots, H (= 2^K)$. These elements take the value of 1 when examinee $i$ belongs to the $h^*$-th item-specific attribute mastery pattern and satisfy $\sum_{h=1}^{H_j^*} z_{jih} = 1$. For instance, under the specification in Table 2 and $z_i = (0, 0, 0, 0, 0, 0, 0, 0, 1)^T$, $z_{ji} = G_j z_i$ is computed as $z_{ji} = (0, 0, 0, 1)^T$, where $G_j$ correctly converts $z_i$ to an item-specific latent indicator vector $z_{ji}$ that has the value of 1 on its fourth element. Based on $G_j$ and $z_i$, Yamaguchi and Okada (2021) formulated the item response function of the binary-attribute saturated DCM with G-matrices as follows:

$$P(x_{ij} = 1|z_i, \theta_j, G_j, q_j) = \prod_{h^*=1}^{H_j^*} \theta_{jih^*}^{z_{jih^*}},$$

(3)

where $\theta_{jih^*}$ denotes the correct-response probability of item $j$ for an examinee with the $h^*$-th item-specific attribute mastery pattern. In addition, under the assumption of local independence given a latent indicator vector $z_i$, Yamaguchi and Okada (2021) provided the likelihood function of the binary-attribute saturated DCM using G-matrices:

$$P(X|Z, \Theta, G, Q) = \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{h^*=1}^{H_j^*} P(x_{ij}|z_i, \theta_j, G_j, q_j)$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{h^*=1}^{H_j^*} \left\{ \theta_{jih^*}^{x_{ij}} (1 - \theta_{jih^*})^{1-x_{ij}} \right\}^{z_{jih^*}}.$$  

(4)
Since the above likelihood function is in a Bernoulli mixture formulation, we can apply the well-known procedure of VB inference for Bernoulli mixture models.

### 1.3 Formulate a G-Matrix for a Polytomous-Attribute Saturated DCM

Similar to a G-matrix for the binary-attribute saturated DCM in Section 1.2, a poltomous-attribute G-matrix is obtained by introducing collapsed attribute vectors into the construction of G-matrices. An example of a poltomous-attribute G-matrix based on collapsed attribute vectors is shown in Table 3. In the same manner as a binary-attribute G-matrix, a poltomous-attribute G-matrix $G_{j}^{\text{poly}}$ converts a latent indicator vector for poltomous attributes $z_{i} = (z_{i1}, \ldots, z_{iL})^{T}$ to an item-specific indicator vector based on collapsed attribute vectors $z_{ji} = (z_{ji1}, \ldots, z_{jiL_{j}^{*}})^{T}$.

Table 3: The example of a poltomous-attribute G-matrix based on collapsed attribute vectors when $K = 2$, $M = 3$, and $q_{j} = (2, 0)$

| Attribute mastery patterns | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|----------------------------|---|---|---|---|---|---|---|---|---|
| $\alpha_{1}$              | 0 | 1 | 0 | 1 | 2 | 0 | 1 | 2 | 2 |
| $\alpha_{2}$              | 0 | 0 | 1 | 1 | 0 | 2 | 2 | 1 | 2 |

| Collapsed attribute mastery patterns | $\alpha_{1}$ | $\alpha_{2}$ | Polytomous-attribute G-matrix |
|-------------------------------------|--------------|--------------|-------------------------------|
| 1 0 *                               | 1 1 1 1 0 1 1 0 0 |
| 2 1 *                               | 0 0 0 0 1 0 0 1 1 |

Based on $G_{j}^{\text{poly}}$ and $z_{i}$, the item response function of the poltomous-attribute saturated DCM is defined as

$$P(x_{ij} = 1|z_{i}, \theta_{j}, G_{j}^{\text{poly}}, q_{j}) = \prod_{l^{*}=1}^{L_{j}^{*}} \theta_{j}^{x_{ij}z_{ijl^{*}}} (1 - \theta_{j}^{z_{ijl^{*}}})^{1-x_{ij}}. \quad (5)$$

Its likelihood function is also defined as

$$P(X|Z, \Theta, G^{\text{poly}}, Q) = \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l^{*}=1}^{L_{j}^{*}} P(x_{ij}|z_{i}, \theta_{j}, G_{j}^{\text{poly}}, q_{j})$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l^{*}=1}^{L_{j}^{*}} \left\{ \theta_{j}^{x_{ij}z_{ijl^{*}}} (1 - \theta_{j}^{z_{ijl^{*}}})^{1-x_{ij}} \right\}^{z_{ijl^{*}}}.$$

(6)

As Yamaguchi and Okada (2021) noted for the binary-attribute saturated DCM using G-matrices, the intercept, main, and interaction effects of the attributes in the pG-DINA model can also be obtained by transforming the estimated $\theta_{j}$ to $\delta_{j}$ using the least-square estimation elaborated in de la Torre (2011).

### 1.3.1 Extending a Polytomous-Attribute G-Matrix with Collapsed Attribute Vectors to that with Reduced Attribute Vectors

The utility of a poltomous-attribute G-matrix is not limited only to modeling correct-response probabilities based on collapsed attribute vectors. A G-matrix based on reduced attribute vectors
can also be formulated, allowing us to flexibly capture the subtle differences in correct-response probabilities at different levels of mastery states. An example of a polytomous-attribute G-matrix using reduced attribute vectors with \( K = 2, M = 3 \), and \( q_j = (2, 0) \) is presented in Table 4.

Table 4: The example of a polytomous-attribute G-matrix based on reduced attribute vectors when \( K = 2, M = 3 \), and \( q_j = (2, 0) \)

| Attribute mastery patterns | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-----------------------------|---|---|---|---|---|---|---|---|---|
| \( \alpha_1 \)              | 0 | 1 | 0 | 1 | 2 | 0 | 1 | 2 | 2 |
| \( \alpha_2 \)              | 0 | 0 | 1 | 1 | 0 | 2 | 2 | 1 | 2 |

| Reduced attribute mastery patterns | \( \alpha_1 \) | \( \alpha_2 \) | Polytomous-attribute G-matrix |
|-----------------------------------|-------------|-------------|-------------------------------|
| 1                                 | 0           | 1           | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 2                                 | 1           | *           | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 3                                 | 2           | *           | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |

Although the G-matrix based on collapsed attribute vectors in Table 3 assigns the same correct-response probability to respondents with \( \alpha_1 = 0 \) and those with \( \alpha_1 = 1 \), the G-matrix using reduced attribute vectors in Table 4 allows such respondents to have different correct-response probabilities. This G-matrix with the more relaxed assumption of item-responding processes enables a finer-grained inspection of how correct-response probabilities change with different levels of attribute mastery states.

1.4 BAYESIAN FORMULATION

In this section, we provide the Bayesian formulation for the polytomous-attribute saturated DCM with its G-matrix using collapsed attribute vectors. Its formulation for the saturated DCM with a polytomous-attribute G-matrix using reduced attribute vectors is the same as the one based on collapsed attribute vectors, except for the number of item-specific attribute mastery patterns with different correct-response probabilities for the items.

Because the difference in the formulation between the binary- and polytomous-attribute saturated DCMs lies only in the specification of a G-matrix, we follow the Bayesian formulation of the binary-attribute saturated DCM in Yamaguchi and Okada (2021).

We first write the distribution of \( Z \) as a categorical distribution with the mixing proportions \( \pi = (\pi_1, \ldots, \pi_L)^T \), which is given as

\[
P(Z|\pi) = \prod_{i=1}^{N} \prod_{l=1}^{L} \pi_i^{z_{il}}. \tag{7}
\]

The prior over \( \pi \) is selected to be a Dirichlet distribution with the parameter \( \delta^0 = (\delta^0_1, \ldots, \delta^0_L)^T \):

\[
P(\pi|\delta^0) = \prod_{l=1}^{L} \pi_l^{\delta^0_l - 1}. \tag{8}
\]

We choose a Beta distribution for the prior over the correct-response probability parameter \( \theta_{jl^*} \):

\[
P(\theta_{jl^*}|a_{jl^*}, b_{jl^*}) \propto \theta_{jl^*}^{a_{jl^*} - 1} (1 - \theta_{jl^*})^{b_{jl^*} - 1}, \tag{9}
\]
where \((a^0_{jl^*}, b^0_{jl^*})\) are the hyperparameters for \(\theta_{jl^*}\). In addition, the assumption of conditional independence on the correct-response probability parameters enables the joint probability of \(\theta_{jl^*}\) to be formulated as follows:

$$
P(\Theta|A^0, B^0) \propto \prod_{j=1}^{J} \prod_{l^*=1}^{L^*_j} \theta^0_{jl^*}^{-1} (1 - \theta_{jl^*})^{b^0_{jl^*}-1}.
$$

(10)

Based on the likelihood and priors, the joint posterior can be obtained as

$$
P(Z, \Theta, \pi|X, G^{\text{poly}}, Q, \delta^0, A^0, B^0)
\propto P(X|Z, \Theta, G^{\text{poly}}, Q) P(Z|\pi) P(\pi|\delta^0) P(\Theta|A^0, B^0).
$$

(11)

### 1.5 Variational Bayesian Inference

As Blitzstein and Hwang (2019) said “conditioning is the soul of statistics” (p. 46), the quintessence of Bayesian statistics is to update beliefs in current reasoning toward phenomena of interest based on observations. The mathematical formulation of this update is expressed as the Bayes rule in Equation (12), where the posterior distributions capture the uncertainty of parameters:

$$
P(\Psi|X) = \int \frac{P(X|\Psi) P(\Psi)}{P(X|\Psi) P(\Psi) d\Psi}.
$$

(12)

Although the Bayes rule offers the base of principled statistical inference, computing posterior distributions requires the marginalization term in the denominator. The closed-form solution for this term is usually unavailable in practice, and its numerical integration incurs a prohibitive computational cost (Bishop, 2006).

VB inference transforms intractable posterior computations into an optimization problem in which the objective is to search for the parametric form of distributions, often referred to as “variational distributions,” that produces the best approximation to posteriors (Galdo et al., 2020). In many cases, Kullback–Leibler (KL) divergence is selected as a metric to measure the discrepancy between the variational distribution \(q(\Psi)\) and posterior distribution \(P(\Psi|X)\). KL divergence emerges from the well-known decomposition of the log marginal likelihood:

$$
\log P(X) = \int q(\Psi) \log \frac{P(X, \Psi)}{q(\Psi)} d\Psi - \int q(\Psi) \log \frac{P(\Psi|X)}{q(\Psi)} d\Psi
= L(q) + \text{KL}[q(\Psi) \parallel P(\Psi|X)]
\geq L(q).
$$

(13)

Here, \(L(q)\) is the lower bound of the log marginal likelihood (variational lower bound; VLB). The minimization of KL divergence corresponds to the maximization of the VLB, and \(q(\Psi)\) equals \(P(\Psi|X)\) when KL divergence becomes 0 (Zhang et al., 2019). To alleviate the complexity in optimizing KL divergence, we assume that the variational distributions can be factorized into \(S\) components, which is known as the mean-field assumption on \(q(\Psi)\):

$$
q(\Psi) = \prod_{s=1}^{S} q(\psi_s).
$$

(14)
Iteratively, we update a set of parameters for each \( q(\psi_s) \) in a manner that satisfies \( q(\psi_s) \propto \exp(E_{s \neq s}[\log P(X, \Psi)]) \) until the changes of the VLB during iteration achieve the predefined stopping criterion.

In the case of the polytomous-attribute saturated DCM using polytomous-attribute G-matrices, we apply the mean-field assumption to the variational distributions of the model parameters \( \Psi = \{Z, \Theta, \pi\} \) such that

\[
q(\Psi) = q(Z)q(\Theta, \pi) = \left( \prod_{i=1}^{N} q(z_i) \right) \left( \prod_{j=1}^{J} \prod_{l^*=1}^{L_{j^*}} q(\theta_{jl^*}) \right) q(\pi).
\] (15)

As mentioned in Section 2.4, the difference between binary- and polytomous-attribute saturated DCMs is in the formulation of a G-matrix and not in the formulation of their joint posteriors. Accordingly, the derivation of the VB algorithm for the binary-attribute saturated DCM elaborated in Yamaguchi and Okada (2021) can be applied directly to the proposed method. The details of the derivation are shown in Supplementary Material A.

1.6 PARALLELIZATION OF THE PROPOSED ALGORITHM

VB inference generally consists of two steps: the variational E-step (VE-step) and variational M-step (VM-step). This is because the expectation step (E-step) in an EM algorithm corresponds to updating the variational posteriors of latent variables, and the maximization step (M-step) in an EM algorithm corresponds to updating the variational posteriors of parameters. Hence, the configuration of the parallelized EM algorithm for generalized latent variable models originally developed in von Davier (2016) is applicable to the proposed method.

After the initialization of the parameters, the parallel-E parallel-M algorithm in von Davier’s study (2016) starts with parallelizing the E-step using \( C \) cores, subdivides examinees into \( C \) groups, computes the posteriors of the latent variables given the responses of examinees, and calculates the expected counts of examinees in latent classes in parallel using \( C \) worker processes. These computed posteriors are aggregated in a master process. Subsequently, the parallel-M-step proceeds by subdividing items into \( C \) groups, computes the gradients of the parameters, and updates the parameters with these computed gradients in parallel using \( C \) worker processes. These updated parameters are aggregated in a master process. Finally, a convergence criterion is evaluated using the latent variables and parameters updated in the previous steps. These steps continue until the predefined stopping criterion is achieved.

In the case of the proposed method, we first allocate \( C \) cores for parallelization. Then, we subdivide examinees into \( C \) groups, update the variational posteriors of the latent variables \( q(z_i) \) in parallel, and aggregate these results in a master process. Subsequently, the variational posteriors of the mixing proportions \( q(\pi) \) are computed based on the values from the previous step. In the VM-step, we subdivide the items into \( C \) groups and update the variational posteriors of the correct-response probability parameters \( q(\theta_{jl^*}) \) in parallel. These results are aggregated in a master process, and the value of the VLB is computed using the results from the VE- and VM-steps. These steps continue until the predefined stopping criterion is achieved.
2 Simulation Study

2.1 Simulation Study 1: Performance of the Polytomous-Attribute Saturated DCM with Its G-Matrices Using Collapsed Attribute Vectors

2.1.1 Simulation Design

To confirm whether the proposed algorithm can recover the true parameter values, we conducted a simulation study under large-scale conditions. Specifically, we considered a sample size of 10000 or 30000, a number of items of 60 or 120, correlation coefficients among the attributes of .1 or .5, a number of attributes of $K = 4$ or 7, and a number of mastery levels of $M = 3$. Additionally, we set the maximum number of attributes measured by one item to $K_j^* = 3$ for $K = 4$ and to $K_j^* = 4$ for $K = 7$. With respect to the true Q-matrices, we provide their detailed specifications in Supplementary Material B. The true lowest and highest values of the correct-response probability parameters for each item were randomly generated from the following distributions: $p_j^{\text{lowest}} \sim \text{Uniform}(0.05, 0.25)$ and $p_j^{\text{highest}} \sim \text{Uniform}(0.75, 0.95)$. $p_j^{\text{lowest}}$ corresponds to the guessing probability when an examinee does not master any of the required attributes and $1 - p_j^{\text{highest}}$ corresponds to the slip probability when an examinee masters all the required attributes. Given these lowest and highest probability parameters, we monotonically increased the lowest value of the parameters as an examinee acquired the necessary attributes on collapsed attribute vectors. For each artificial dataset, we computed the expected a posteriori (EAP) estimates for the correct-response probability parameters and the maximum a posteriori (MAP) estimates for the attribute mastery profiles from the corresponding variational posterior distributions. The results are provided in Supplementary Material C.

To generate the attribute mastery profiles, we applied the following criteria using a multivariate standard normal distribution $\lambda_i \sim \text{MVM} (\mathbf{0}, \Sigma)$:

$$
\alpha_{ik} = \begin{cases} 
M - 1 & \text{if } \lambda_{ik} \geq \phi^{-1}\left(\frac{M-1}{M}\right) \\
\vdots \\
1 & \text{if } \phi^{-1}\left(\frac{2}{M}\right) > \lambda_{ik} \geq \phi^{-1}\left(\frac{1}{M}\right) \\
0 & \text{otherwise}
\end{cases},
$$

where $\lambda_i$ is a $K$-dimensional vector. Correlation coefficients among the attributes of .1 or .5 were assigned in the off-diagonal entries of $\Sigma$. This generation method was modified from the original criteria used in Chiu et al. (2009) and Liu et al. (2012) to produce polytomous attributes. For each artificial dataset, we computed the expected a posteriori (EAP) estimates for the correct-response probability and mixing proportion parameters and the maximum a posteriori (MAP) estimates for the attribute mastery profiles from the corresponding variational posterior distributions. The 100 datasets were simulated for all the conditions.
We assessed the performance of the parameter recovery using the bias and root-mean-square error (RMSE). The bias and RMSE for the correct-response probability parameters $\theta_{jl}$ were computed as

$$\text{Bias}_{\theta_{jl}} = \frac{1}{100} \sum_{t=1}^{T=100} \left( \hat{\theta}_{\text{est}}^{(t)} - \theta_{\text{true}} \right),$$

$$\text{RMSE}_{\theta_{jl}} = \sqrt{\frac{1}{100} \sum_{t=1}^{T=100} \left( \hat{\theta}_{\text{est}}^{(t)} - \theta_{\text{true}} \right)^2},$$

where $\theta_{\text{true}}$ denotes the true value of the correct-response probability parameter. We averaged these measures across the related items. In addition, we computed the bias and RMSE for the mixing proportion parameters $\pi_l$ in the same manner as the correct-response probability parameters. Since we could not directly specify the true values of the mixing proportion parameters because of the attribute-profile generation method based on a multivariate standard normal distribution, we first sampled 100 million attribute mastery profiles based on the above multivariate standard normal distribution. Then, the observed proportion of each attribute mastery pattern from this large random sample served as the true value of the corresponding mixing proportion parameter.

Furthermore, we also calculate the element-wise attribute classification rate (EACR) and pattern-wise attribute classification rate (PACR). These rates are defined as

$$EACR_{\alpha_k} = \frac{1}{100} \sum_{t=1}^{T=100} \frac{1}{N} \sum_{i=1}^{N} I(\hat{\alpha}_{tik} = \alpha_{tik}^{\text{true}}),$$

$$PACR_{\alpha_k} = \frac{1}{100} \sum_{t=1}^{T=100} \frac{1}{N} \sum_{i=1}^{N} I(\hat{\alpha}_{ti} = \alpha_{ti}^{\text{true}}).$$

2.1.2 Estimation Settings

We adopted the same settings for the hyperparameters $\delta^0$, $A^0$, and $B^0$ as in Yamaguchi and Okada (2021). These hyperparameters control the parametric forms of the prior distributions for the mixing proportion and correct-response probability parameters, and the superscript 0 indicates that these hyperparameters are relevant to the settings of the prior distributions. Specifically, a vector of ones was assigned to $\delta^0$. This specification yields a non-informative prior, meaning that the prior expectation of the mixing proportion for each attribute mastery pattern is $1/L$. For $A^0$ and $B^0$, we set weakly informative priors such that the prior expectation of the correct-response probability in the pattern with no relevant attributes mastered is below 0.5, and that of the pattern with all relevant attributes mastered is above 0.5. This setting for $A^0$ and $B^0$ aims to include the prior information of the monotonicity constraints on the correct-response probabilities into their posterior inference, which helps satisfy the monotonicity constraints in the estimation procedure. We investigated the effects of such a prior in Simulation Study 2.

Regarding the initial values of the proposed method, the latent indicator variables $z_{il}$ were set to be $1/L$, meaning that the probability of examinee $i$ belonging to pattern $l$ is equivalent across all the patterns. The stopping criteria for the proposed algorithm were specified such that the iteration stopped when the maximum change in the VLB became less than $10^{-4}$ or the number of iterations reached 2000. Finally, the number of cores for parallel computing was set to $C = 8$. 
We wrote all the programs for this study in the Julia programming language (1.66; Bezanson et al., 2017). You can access the estimation code at the open science framework: https://osf.io/fgn3t/?view_only=7d9df56df89446ef9ed88ed016be7fb4.

2.1.3 Results

Table 5 shows the biases and RMSEs for the correct-response probability parameters in the simulation conditions with \( K = 4 \). Owing to a large number of correct-response probability parameters, we summarized their biases and RMSEs in the same manner as Yamaguchi and Okada (2021), where the biases and RMSEs were averaged according to the number of attributes measured by one item. We observed that the values of the biases were almost zero in these conditions. Concerning the RMSEs, the increase in sample size lowered their values. Additionally, as the number of attributes measured by one item increased, the values of RMSEs worsened. These values also deteriorated as the value of the correlation coefficients among the attributes increased under the condition where the number of attributes measured by one item was two or more. However, the degree of such deterioration was small, and the parameter recovery in the correct-response probability parameters was satisfactory across all the conditions.

Table 5: Biases and RMSEs for the correct-response probability parameters in the simulation conditions with \( K = 4 \)

| Number of Items | Sample Size | Correlation | Bias | RMSE | Bias | RMSE | Bias | RMSE |
|-----------------|-------------|-------------|------|------|------|------|------|------|
| 60              | 10000       | .1          | 0.0000 | 0.0061 | -.0002 | .0110 | .0000 | .0178 |
|                 | .5          | 0.0001 | 0.0060 | -.0003 | .0134 | .0000 | .0229 |
|                 | 30000       | .1          | 0.0000 | 0.0035 | -.0001 | .0063 | .0002 | .0103 |
|                 | .5          | 0.0001 | 0.0034 | -.0002 | .0077 | .0001 | .0132 |
| 120             | 10000       | .1          | -.0001 | .0055 | -.0001 | .0102 | -.0002 | .0160 |
|                 | .5          | -.0001 | .0054 | -.0001 | .0123 | -.0001 | .0204 |
|                 | 30000       | .1          | 0.0001 | 0.0031 | 0.0000 | .0059 | 0.0000 | .0092 |
|                 | .5          | 0.0000 | 0.0031 | 0.0000 | .0071 | 0.0000 | .0118 |

Note. RMSE = root-mean-square error.

The biases and RMSEs for the mixing proportion parameters are given in Table 6. Because of the large number of these parameters, which amounts to \( M^K \), we only show the maximum and minimum values of their biases and RMSEs among the \( M^K \) parameters. In all the conditions, the absolute values of the largest positive and negative biases were not greater than .0008 and the highest values of the RMSEs were no more than .0033. These results corroborate the sound accuracy of estimating the mixing proportion parameters.

Table 7 presents the EACRs and PACRs for the attribute profile estimation in the simulation conditions with \( K = 4 \). The EACRs were more than 92% in all the relevant conditions; however, the PACRs in the conditions with \( J = 60 \) were below 80%. This relatively unsatisfactory performance in such conditions is attributed to a large number of possible attribute mastery patterns with polytomous attributes and limited number of items. Nonetheless, when the number of items increased to \( J = 120 \), the PACRs climbed above 93%.

Subsequently, Table 8 shows the biases and RMSEs for the correct-response probability parameters in the simulation conditions with \( K = 7 \). Similar to the conditions with \( K = 4 \), the values of the biases were quite small across all the conditions, where their absolute largest value was only .0022 in the condition with \( J = 60, N = 10000 \), and the correlation coefficient
Table 6: Maximum and minimum values of the biases and RMSEs for the mixing proportion parameters in the simulation conditions with $K = 4$

| Number of Items | Sample Size | Correlation | Bias Maximum | Bias Minimum | RMSE Maximum | RMSE Minimum |
|-----------------|-------------|-------------|--------------|--------------|--------------|--------------|
| 60              | 10000       | .1          | .0003        | -.0005       | .0020        | .0011        |
|                 |             | .5          | .0004        | -.0007       | .0032        | .0006        |
|                 | 30000       | .1          | .0002        | -.0002       | .0010        | .0006        |
|                 |             | .5          | .0002        | -.0003       | .0010        | .0003        |
| 120             | 10000       | .1          | .0003        | -.0004       | .0015        | .0009        |
|                 |             | .5          | .0003        | -.0008       | .0033        | .0005        |
|                 | 30000       | .1          | .0002        | -.0002       | .0009        | .0004        |
|                 |             | .5          | .0002        | -.0003       | .0017        | .0003        |

Note. Maximum and minimum values of the biases correspond to the largest positive and negative biases among the $M^R$ mixing proportion parameters, respectively. RMSE = root-mean-square error.

Table 7: EACRs and PACRs for the attribute profile estimation in the simulation conditions with $K = 4$

| Number of Items | Sample Size | Correlation | EACR$_{\alpha_1}$ | EACR$_{\alpha_2}$ | EACR$_{\alpha_3}$ | EACR$_{\alpha_4}$ | PACR$_{\alpha}$ |
|-----------------|-------------|-------------|--------------------|--------------------|--------------------|--------------------|-----------------|
| 60              | 10000       | .1          | .922              | .922               | .922               | .932               | .745            |
|                 |             | .5          | .932              | .932               | .933               | .941               | .772            |
|                 | 30000       | .1          | .922              | .922               | .922               | .933               | .746            |
|                 |             | .5          | .933              | .932               | .933               | .942               | .773            |
| 120             | 10000       | .1          | .983              | .983               | .981               | .980               | .933            |
|                 |             | .5          | .985              | .985               | .983               | .983               | .940            |
|                 | 30000       | .1          | .983              | .983               | .981               | .980               | .933            |
|                 |             | .5          | .985              | .985               | .984               | .983               | .941            |

Note. EACR = element-wise attribute classification rate, PACR = pattern-wise attribute classification rate.

In contrast to the biases, the relatively large values of the RMSEs were observed when the number of attributes measured by one item was not less than three, where these values ranged from .0103 to .0606. Nonetheless, the values of the RMSEs in other conditions were not more than .0186. The patterns in the effects of the simulation factors were also the same as those with $K = 4$, where the increase in the number of attributes measured by one item and value of the correlation coefficients worsened the values of the RMSEs. We also provide the maximum and minimum values of the biases and RMSEs for the mixing proportion parameters in Table 9. Similar to those with $K = 4$, the small values of the absolute largest positive and negative biases were observed, where these absolute values were less than or equal to .0085. The highest values of RMSEs were also not greater than .0090 in all the conditions, supporting the sound accuracy of parameter estimation in correct-response probability parameters.

Table 10 presents the EACRs and PACRs for the attribute profile estimation in the simulation conditions with $K = 7$. The EACRs ranged from 79.9% to 88.4% under the conditions with $J = 60$, whereas those under the conditions with $J = 120$ ranged from 91.6% to 95.4%. As the number of items positively contributes to the accuracy of attribute profile estimation, the conditions with a large number of items presented more accurate attribute profile estimation than those with a relatively small number of items. Regarding the PACRs, their values were less
than or equal to .372 under the conditions with $J = 60$ and .695 under those with $J = 120$. Although the increase in the number of items significantly ameliorated the parameter recovery of attributes, the PACRs were significantly worse than those under the conditions with $K = 4$ because of the large number of attributes.

Table 8: Biases and RMSEs for the correct-response probability parameters in the simulation conditions with $K = 7$

| Number of Items | Sample Size | Correlation | Number of Attributes Measured by One Item | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
|-----------------|-------------|-------------|------------------------------------------|------|------|------|------|------|------|------|------|
|                 |             |             | 1                                        |      |      |      |      |      |      |      |      |
|                 | 60          | .1          | .0002                                    | .0104| -.0002| .1414| -.0008| .2344| -.0003| .0411|
|                 |             | .5          | .0012                                    | .0091| -.0022| .1868| -.0006| .3254| -.0014| .0606|
|                 | 120         | .1          | -.0002                                  | .0061| -.0002| .0122| -.0005| .1784| -.0001| .0290|
|                 |             | .5          | -.0004                                  | .0060| -.0007| .0146| -.0004| .2414| -.0004| .0431|

Note. RMSE = root-mean-square error.

Table 9: Maximum and minimum values of the biases and RMSEs for the mixing proportion parameters in the simulation conditions with $K = 7$

| Number of Items | Sample Size | Correlation | Mixing Proportion Parameter | Bias | RMSE |
|-----------------|-------------|-------------|-----------------------------|------|------|
|                 |             |             | Maximum | Minimum | Maximum | Minimum |
|                 | 60          | .1          | .0001  | -.0005 | .0010   | .0001   |
|                 |             | .5          | .0001  | -.0085 | .0090   | .0001   |
|                 | 120         | .1          | .0001  | -.0003 | .0006   | .0001   |
|                 |             | .5          | .0001  | -.0333 | .0036   | .0001   |

Note. Maximum and minimum values of the biases correspond to the largest positive and negative biases among the $M^K$ mixing proportion parameters, respectively. RMSE = root-mean-square error.

Finally, we present the CPU time and convergence rate for all the conditions in Table 11. The proposed method completed the parameter estimation within approximately 50 seconds, and the convergence rates of the estimation were 100% under the conditions with $K = 4$. When it comes to $K = 7$, the computational speed differs greatly between $J = 60$ and $J = 120$. Counterintuitively, the CPU time under the conditions with $J = 120$ was much shorter than under those with $J = 60$. Specifically, the CPU times under the conditions with $J = 120$ and $N = 30000$ were less than around 1400 to 1500 seconds, whereas these times under those with $J = 60$ and $N = 30000$ were less than around 3000 to 3100 seconds. The longer CPU time in the conditions with $J = 60$ than those with $J = 120$ would be attributed to an unsuitable setting for the specified number of cores in the estimation. Lastly, the convergence rates were
Table 10: EACRs and PACRs for the attribute profile estimation in the simulation conditions with $K = 7$

| Number of Items | Sample Size | Correlation | EACR$_{\alpha_1}$ | EACR$_{\alpha_2}$ | EACR$_{\alpha_3}$ | EACR$_{\alpha_4}$ | EACR$_{\alpha_5}$ | EACR$_{\alpha_6}$ | PACR$_{\alpha}$ |
|----------------|-------------|-------------|-------------------|------------------|------------------|------------------|------------------|------------------|----------------|
| 60             | 10000       | .1          | 0.799             | 0.823            | 0.862            | 0.858            | 0.821            | 0.830            | 0.813           |
|                |             | .5          | 0.832             | 0.849            | 0.882            | 0.879            | 0.848            | 0.855            | 0.842           |
|                | 30000       | .1          | 0.802             | 0.826            | 0.864            | 0.861            | 0.822            | 0.832            | 0.816           |
|                |             | .5          | 0.835             | 0.852            | 0.884            | 0.881            | 0.851            | 0.857            | 0.845           |
| 120            | 10000       | .1          | 0.941             | 0.935            | 0.944            | 0.934            | 0.941            | 0.932            | 0.916           |
|                |             | .5          | 0.950             | 0.945            | 0.953            | 0.943            | 0.950            | 0.942            | 0.929           |
|                | 30000       | .1          | 0.942             | 0.938            | 0.946            | 0.936            | 0.943            | 0.934            | 0.918           |
|                |             | .5          | 0.951             | 0.946            | 0.954            | 0.945            | 0.951            | 0.944            | 0.931           |

Note. EACR = element-wise attribute classification rate, PACR = pattern-wise attribute classification rate.

also 100% under the conditions with $K = 7$. These results generally support the sound stability and computational speed of our method for parameter estimation.

Table 11: CPU time and convergence rate of the proposed method in Simulation Study 1

| $K = 4$ | Number of Items | Sample Size | Correlation | CPU Time (seconds) | Convergence Rate |
|---------|----------------|-------------|-------------|--------------------|------------------|
| 60      | 10000          | .1          | 13.51       | 100%               |
|         |                | .5          | 12.80       | 100%               |
|         | 30000          | .1          | 45.79       | 100%               |
|         |                | .5          | 41.51       | 100%               |
| 120     | 10000          | .1          | 13.65       | 100%               |
|         |                | .5          | 13.92       | 100%               |
|         | 30000          | .1          | 43.43       | 100%               |
|         |                | .5          | 41.48       | 100%               |

| $K = 7$ | Number of Items | Sample Size | Correlation | CPU Time (seconds) | Convergence Rate |
|---------|----------------|-------------|-------------|--------------------|------------------|
| 60      | 10000          | .1          | 695.95      | 100%               |
|         |                | .5          | 682.68      | 100%               |
|         | 30000          | .1          | 3058.70     | 100%               |
|         |                | .5          | 3011.15     | 100%               |
| 120     | 10000          | .1          | 591.55      | 100%               |
|         |                | .5          | 559.19      | 100%               |
|         | 30000          | .1          | 1506.23     | 100%               |
|         |                | .5          | 1396.65     | 100%               |
2.2 SIMULATION STUDY 2: EFFECTS OF PRIOR INFORMATION ON MONOTONICITY CONSTRAINTS

One concern about the proposed method is that the monotonicity constraints on correct-response probabilities cannot be directly incorporated into the estimation procedure. Since these constraints are essential for the interpretable clustering of respondents based on attribute mastery profiles and meaningful item parameter estimates (Yamaguchi & Templin, 2021), the issue of the constraints must be addressed in the proposed method. To cope with this issue, we employed the weakly informative priors of the monotonicity constraints in Simulation Study 1. The same approach was also adopted in Yamaguchi and Okada (2020)'s study of the VB estimation for the DINA model, where they specified the priors for the slip and guessing parameters in such a manner that the prior expectations of both parameters were less than .50. In this simulation study, we assessed the effects of such prior information on monotonicity in the correct-response probability parameters. Specifically, we compared the estimation performance between the proposed methods equipped with and without these prior specifications to investigate how such priors can stabilize estimating the correct-response probability parameters with monotonicity.

Regarding the simulation condition, we used the same datasets under the setting $K = 4$, $J = 60$, and $N = 10000$ or $30000$; the correlation coefficient among the attributes is .1 as in Simulation Study 1. Additionally, for the proposed method with non-informative priors, we assigned $a_{jl*}^0 = 1$ and $b_{jl*}^0 = 1$ to the priors on all the correct-response probability parameters, which gives no prior information to posterior inference. All the other specifications were set to be the same as those in Simulation Study 1.

2.2.1 RESULTS

The results of Simulation Study 2 are provided in Table 12. As shown in the table, the proposed method with non-informative priors performed the parameter estimation as accurately as that with the weakly informative priors of the monotonicity constraints. Additionally, we found that the methods with both types of priors did not violate monotonicity in the correct-response probability parameters under any of the relevant conditions.
Table 12: The results in Simulation Study 2 ($K = 4$, correlation coefficient .1)

### Biases and RMSEs for the correct-response probability parameters

| Prior Settings          | Number of Items | Sample Size | 1   | 2   | 3   |       |       |       |
|-------------------------|-----------------|-------------|-----|-----|-----|-------|-------|-------|
|                         |                 |             | Bias | RMSE| Bias| RMSE  | Bias | RMSE  |
| Weakly Informative Prior| 60              | 10000       | .0000 | .0061| -.0002| .0110  | .0000 | .0178  |
|                         |                 | 30000       | .0000 | .0035| -.0001| .0063  | .0002 | .0103  |
| Non-Informative Prior   | 60              | 10000       | .0002 | .0061| .0000  | .0110  | .0000 | .0179  |
|                         |                 | 30000       | .0001 | .0035| .0000  | .0063  | .0002 | .0103  |

### Maximum and minimum values of the biases and RMSEs for the mixing proportion parameters

| Prior Settings          | Number of Items | Sample Size | Bias             | RMSE             |
|-------------------------|-----------------|-------------|------------------|------------------|
|                         |                 |             | Maximum | Minimum | Maximum | Minimum |
| Weakly Informative Prior| 60              | 10000       | .0003  | -.0005  | .0020   | .0011   |
|                         |                 | 30000       | .0002  | -.0002  | .0010   | .0006   |
| Non-Informative Prior   | 60              | 10000       | .0003  | -.0005  | .0020   | .0011   |
|                         |                 | 30000       | .0002  | -.0002  | .0010   | .0006   |

### EACRs and PACRs for the attribute profile estimation

| Prior Settings          | Number of Items | Sample Size | EACR$_{a_1}$ | EACR$_{a_2}$ | EACR$_{a_3}$ | EACR$_{a_4}$ | PACR$_{a}$ |
|-------------------------|-----------------|-------------|--------------|--------------|--------------|--------------|------------|
| Weakly Informative Prior| 60              | 10000       | .922         | .922         | .922         | .932         | .745       |
|                         |                 | 30000       | .922         | .922         | .922         | .932         | .746       |
| Non-Informative Prior   | 60              | 10000       | .922         | .922         | .922         | .932         | .745       |
|                         |                 | 30000       | .922         | .922         | .922         | .932         | .746       |

Note. Maximum and minimum values of the biases correspond to the largest positive and negative biases among the $M^K$ mixing proportion parameters, respectively. RMSE = root-mean-square error, EACR = element-wise attribute classification rate, PACR = pattern-wise attribute classification rate.
3 Empirical Study

We conducted an empirical study to show the utility of the proposed method by comparing it with the performance of an MCMC estimation. The dataset from the standardized achievement test for Grade 9 Mathematics garnered in 2019 by TOKYO SHOSEKI CO., LTD. was used for the empirical evaluation. The contents of this examination target four aspects of mathematics: numbers and algebraic expressions, geometrical figures, functions, and making use of data. Examples of items modified from the original ones include “simplify: \((-12)xy^2 \div 6xy \times (-4xy)\)” and “solve for \(a\) in the following equation: \(4a - 7 = 8a - 9\).” The sample size is 21,888, and the number of items is 34.

Regarding the attribute specification, the second author identified three binary and polytomous attributes for this test based on perspectives from cognitive and educational psychology. This Q-matrix was validated by two educational psychologists, and its inconsistencies in the validation process were discussed between them. The first attribute is binary and this represents a computational skill. The second attribute is polytomous with the levels of \(M = 3\) and this represents the (1) comprehension of the procedure and (2) conceptual understanding of the procedure. The third attribute is binary and this represents the comprehension of the terminology. The detailed specification of the Q-matrix for this examination is shown in Table 13.

Concerning the MCMC estimation, we developed the Gibbs sampler for the polytomous-attribute saturated DCM using its G-matrices based on collapsed attribute vectors. The specifications of the hyperparameters were set to be the same as those in Simulation Study 1. When implementing this Gibbs sampler, we run three chains of 5000 iterations with 2000 as the burn-in. We assessed the convergence of MCMC chains via the convergence diagnostics called the rank-normalized split-\(\hat{R}\) proposed in Vehtari et al. (2021) and confirmed that all the parameters satisfied \(\hat{R} < 1.05\). Additionally, we parallelized generating MCMC chains using three cores. For the proposed method, we assigned the same specifications as in Simulation Study 1 and set the same number of cores as the Gibbs sampler, which corresponds to \(C = 3\).

First, the computational times of the Gibbs sampler and the proposed method with parallel computing were 777.21 and 12.45 seconds, respectively. As expected, the parallelized VB algorithm presented a superior computational speed over the parallelized Gibbs sampler, and the computational time of our method was about 60 times faster than that of the Gibbs sampler. With respect to the effect of parallel computing, the computational time of the proposed method without parallel computing was 30.30 seconds. As the number of attributes is three, the parallelized proposed method reduces the computational time without parallel computing by only about 60%. Nonetheless, the parallelization would yield much better computational efficiency as the attribute dimensionality rises.

Second, we show the results of the parameter estimation for the correct-response probability parameters. Owing to space limitations, we provide their estimates of EAP and posterior standard deviation (SD) for each item in Supplementary Material D. Here, we present the scatter plot of the EAP and posterior SD for the parameter estimates. Figure 1 shows that the EAP estimates of the correct-response probability parameters from our method are closely aligned with those from the Gibbs sampler. Subsequently, Figure 2 shows their posterior SD estimates. In general, the estimates from our method were close to those from the Gibbs sampler. However, our method moderately underestimated the posterior SDs compared with the Gibbs sampler for the items for which the corresponding MCMC-estimated posterior SDs were relatively large. This underestimation of posterior variance is a well-known tendency of VB inference (Bishop, 2006). Nonetheless, the largest absolute difference in the EAP estimates for the correct-response probability parameters between these two estimation methods was .0180, which coincides with \(\theta_{jl}\) with \(j = 30\) and \(l^* = 4\), while that for the posterior SD estimates was .0730, which corresponds to \(\theta_{jl}\) with \(j = 28\) and \(l^* = 4\). These results indicate that the estimated values of the parameters from the proposed method were highly consistent with those from the Gibbs sampler.
Table 13: The specification of the Q-matrix for the empirical study

| Item ID | Attribute 1 (binary) | Attribute 2 (polytomous) | Attribute 3 (binary) |
|---------|----------------------|--------------------------|----------------------|
|         | Computational Skill (1) | Comprehension of Procedure (1) and Conceptual Understanding of Procedure (2) | Comprehension of Terminology (1) |
| 1       | 1                    | 0                        | 0                    |
| 2       | 1                    | 0                        | 0                    |
| 3       | 1                    | 0                        | 0                    |
| 4       | 1                    | 0                        | 0                    |
| 5       | 1                    | 0                        | 0                    |
| 6       | 1                    | 0                        | 0                    |
| 7       | 1                    | 0                        | 0                    |
| 8       | 0                    | 1                        | 1                    |
| 9       | 0                    | 1                        | 0                    |
| 10      | 0                    | 1                        | 1                    |
| 11      | 1                    | 1                        | 0                    |
| 12      | 0                    | 2                        | 0                    |
| 13      | 0                    | 1                        | 0                    |
| 14      | 1                    | 1                        | 1                    |
| 15      | 1                    | 1                        | 1                    |
| 16      | 1                    | 1                        | 1                    |
| 17      | 1                    | 1                        | 0                    |
| 18      | 1                    | 1                        | 0                    |
| 19      | 1                    | 1                        | 0                    |
| 20      | 1                    | 1                        | 0                    |
| 21      | 1                    | 1                        | 0                    |
| 22      | 1                    | 1                        | 0                    |
| 23      | 1                    | 1                        | 0                    |
| 24      | 0                    | 0                        | 1                    |
| 25      | 0                    | 2                        | 0                    |
| 26      | 0                    | 0                        | 1                    |
| 27      | 1                    | 1                        | 0                    |
| 28      | 1                    | 1                        | 1                    |
| 29      | 1                    | 1                        | 1                    |
| 30      | 1                    | 1                        | 1                    |
| 31      | 0                    | 1                        | 1                    |
| 32      | 0                    | 1                        | 0                    |
| 33      | 0                    | 2                        | 0                    |
| 34      | 0                    | 2                        | 0                    |

Third, Table 14 presents the EAP and posterior SD estimates for the mixing proportion parameters. Similar to the correct-response probability parameters, the EAP and posterior SD estimates from our method were closely aligned with those from the Gibbs sampler. The largest absolute difference in the EAP estimates for the mixing proportion parameters was .00024 and that in the posterior SD estimate was .00216. Both these values correspond to the mixing proportion parameter of the attribute mastery class with $\alpha = (1, 1, 1)$. These results indicate that our method can estimate model parameters as accurately as the Gibbs sampler.

Lastly, we computed the percentage of the matched entries in the estimated attribute mastery patterns between the proposed method and Gibbs sampler. The element-wise matched ratios for the attributes were .9993 (A1), .9989 (A2), and .9990 (A3) and the pattern-wise matched ratio was .9981. This result also corroborates the high similarity of the parameter estimates between these two methods.
Figure 1: The scatter plot of the EAP estimates for the correct-response probability parameters from the proposed method and Gibbs sampler.

Table 14: The estimates of EAP and posterior standard deviation for the mixing proportion parameters

| Attribute Mastery Patterns | 000 | 001 | 010 | 011 | 020 | 021 | 100 | 101 | 110 | 111 | 120 | 121 |
|----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| **Gibbs**                  |     |     |     |     |     |     |     |     |     |     |     |     |
| Mean                       | .259| .007| .018| .001| .004| .001| .064| .232| .020| .022| .002| .368|
| SD                         | .003| .001| .002| .000| .001| .001| .003| .004| .002| .003| .001| .005|
| **VB**                     |     |     |     |     |     |     |     |     |     |     |     |     |
| Mean                       | .259| .007| .018| .001| .004| .001| .064| .232| .020| .022| .002| .368|
| SD                         | .003| .001| .001| .000| .000| .000| .002| .003| .001| .001| .000| .003|

4 Conclusion and Discussion

To evaluate the performance of the proposed method, we conducted a Monte Carlo simulation under the simulation design with large-scale conditions. Three main findings were observed in Simulation Study 1. First, the biases of the correct-response probability parameters were small across all the conditions. Second, the following general tendencies were found: (1) sample size reduced the values of the RMSEs, (2) as the number of attributes measured by one item increased, the values of the RMSEs worsened, and (3) as the value of the correlation coefficients among the attributes increased, the RMSEs deteriorated when the number of attributes measured by one item was greater than or equal to two. Finally, the relatively high values of the RMSEs were observed when the number of attributes measured by one item increased. However, the effect of the factors deteriorating the parameter estimation was small, and estimation accuracy was generally satisfactory under a wide range of conditions. In addition, we conducted a simulation
to investigate the effect of the weakly informative priors of monotonicity constraints in Simulation Study 2. The findings from this investigation show that the estimation methods with the weakly informative and non-informative priors produced identical parameter estimates. They also did not yield estimates of the correct-response probability parameters without monotonicity. These results suggest that, regardless of the prior specifications, the proposed method can stably estimate monotonic correct-response probability parameters when the true values of these probability parameters satisfy monotonicity.

Through the empirical study, we observed that our method presented computational superiority over the Gibbs sampler; in fact, the computational time of our method was about 60 times faster than that of the Gibbs sampler. Moreover, our method estimated the model parameters using only a small amount of memory in the computer during the estimation, whereas the Gibbs sampler presented substantial memory usage for its estimation. Accordingly, our method was significantly more computationally efficient than the Gibbs sampler in terms of both computational time and memory usage. Additionally, parallelization moderately reduced the computational time of our method compared with that without parallel computing. Although the effect of parallelization was relatively modest, its degree would increase when the numbers of samples and attributes rise. Regarding the estimates of EAP and posterior SD for the correct-response probability and mixing proportion parameters, both estimation methods produced highly similar estimates. The high similarity of estimates was also found in the attribute profile estimation. Since the observed convergence diagnostics corroborated that the MCMC samples from the Gibbs sampler accurately constructed targeted posteriors, this result provides strong evidence of the sound accuracy of our method in variational approximation to targeted posteriors. However, the attribute dimensionality of the Q-matrix in the empirical study is three, and only one attribute is polytomous. Accordingly, this simple specification of the Q-matrix may contribute to the high
similarity of the parameter estimates between the VB algorithm and the Gibbs sampler. Thus, the proposed method may provide parameter estimates differing from those of a Gibbs sampler when the dimensionality and polytomy of attributes become larger.

The limitations of the proposed method are twofold. First, although our method incorporates the weakly informative priors of the monotonicity constraints into the posterior inference for the correct-response probability parameters, it does not ensure that these probability parameter estimates always hold monotonicity. Rather, monotonicity can still be violated even when we use such informative priors. Hence, a VB algorithm that can strictly enforce the monotonicity constraints in its estimation procedure should be developed in future research.

Second, the susceptibility of the parameter estimates to the initial values can be observed in the estimation. Yamaguchi and Okada (2021) reported in their simulation study that the choices of initial values hardly influenced the biases of the parameter estimates but did influence their RMSEs. This varying effect of the initial values on the parameter estimation can also be expected in the proposed method because the estimation procedure of our method is comparable with the binary-attribute saturated DCM except for the implementation of parallel computing. This issue can be easily addressed by introducing simulated annealing in the optimization process of the VLB, such as deterministic annealing variational Bayes (Katahira et al., 2008). However, in practice, we would recommend implementing the proposed method with different initial values multiple times and choosing the result with the maximum value of the VLB.

Concerning future directions of this study, it is practically important to develop a scalable Q-matrix estimation method for polytomous-attribute DCMs. Although a Q-matrix is assumed to be known in the proposed method, it is commonly provisional or unknown in real-world settings. To leverage the utility of polytomous-attribute DCMs in such settings, estimation methods to learn a polytomous-attribute Q-matrix from data are required to conduct accurate diagnosis because misspecifications of a Q-matrix are known to be detrimental to parameter estimation (Kunina-Habenicht et al., 2012; Rupp & Templin, 2008). Although accurate Q-matrix estimation is computationally demanding or difficult even in binary-attribute DCMs (e.g., Chung, 2019; Liu et al., 2020; Xu & Shang, 2018), it is a promising direction to extend existing Q-matrix estimation methods in binary-attribute DCMs to polytomous-attribute DCMs or to develop a VB-based Q-matrix estimation method for polytomous-attribute DCMs.

In conclusion, the present study developed a parallelized VB algorithm for a polytomous-attribute saturated DCM and demonstrated its utility through simulation and empirical studies. The proposed method serves as a scalable psychometric method that can be applicable to big educational data.

**STATEMENTS AND DECLARATIONS**

**ACKNOWLEDGEMENTS**

The authors would like to express their sincere gratitude to TOKYO SHOSEKI CO., LTD. for generously sharing the data used in our empirical study, to Dr. Haruhiko Mitsunaga for his support under JSPS KAKENHI (B) Grant Number JP20H01720, and to Dr. Kazuhiro Yamaguchi for helpful discussions and comments on the proposed method. A preliminary report of this study was presented at the World Meeting of the International Society for Bayesian Analysis 2021.

**FUNDING**

This work was supported by JSPS KAKENHI (Grant Numbers 19H00616, 21H00936, and 20H01720).
CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98. https://doi.org/10.1137/141000671

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Taylor & Francis.

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement*, 37(6), 419–437. https://doi.org/10.1177/0146621613479818

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74(4), 633–665. https://doi.org/10.1007/s11336-009-9125-0

Chung, M. (2019). A Gibbs sampling algorithm that estimates the Q-matrix for the DINA model. *Journal of Mathematical Psychology*, 93, 102275. https://doi.org/10.1016/j.jmp.2019.07.002

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

Galdo, M., Bahg, G., & Turner, B. M. (2020). Variational bayesian methods for cognitive science. *Psychological Methods*, 25(5), 535–559. https://doi.org/10.1037/met0000242

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press. https://doi.org/10.1201/b16018

Hartz, S., & Roussos, L. (2008). The fusion model for skills diagnosis: Blending theory with practicality. *ETS Research Report Series*, 2008(2), 1–57. https://doi.org/10.1002/j.2333-8504.2008.tb02157.x

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

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258–272. https://doi.org/10.1177/01466210122032064

Karelitz, T. (2004). *Ordered category attribute coding framework for cognitive assessments* [Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign]. http://hdl.handle.net/2142/82064

Katahira, K., Watanabe, K., & Okada, M. (2008). Deterministic annealing variant of variational Bayes method. *Journal of Physics: Conference Series*, 95, 012015. https://doi.org/10.1088/1742-6596/95/1/012015

Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models: Detection of model misspecification in DCMs. *Journal of Educational Measurement*, 49(1), 59–81. https://doi.org/10.1111/j.1745-3984.2011.00160.x

Liu, C.-W., Andersson, B., & Skrondal, A. (2020). A constrained Metropolis–Hastings Robbins–Monro algorithm for Q matrix estimation in DINA models. *Psychometrika*, 85(2), 322–357. https://doi.org/10.1007/s11336-020-09707-4
Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement, 36*(7), 548–564. https://doi.org/10.1177/0146621612456591

Rupp, A. A., & Templin, J. (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. https://doi.org/10.1177/0013164407301545

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

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

Templin, J. L. (2004). *Generalized linear mixed proficiency models for cognitive diagnosis* [Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign]. http://hdl.handle.net/2142/82077

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Tjoe, H., & de la Torre, J. (2014). The identification and validation process of proportional reasoning attributes: An application of a cognitive diagnosis modeling framework. *Mathematics Education Research Journal, 26*(2), 237–255. https://doi.org/10.1007/s13394-013-0090-7

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, Advance Publication, 1–38. https://doi.org/10.1214/20-BA1221

von Davier, M. (2007). Mixture distribution diagnostic models. *ETS Research Report Series, 2007*(2), i–21. https://doi.org/10.1002/j.2333-8504.2007.tb02074.x

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology, 61*(2), 287–307. https://doi.org/10.1348/000711007X193957

von Davier, M. (2016). High-performance psychometrics: The parallel-E parallel-M algorithm for generalized latent variable models. *ETS Research Report Series, 2016*(2), 1–11. https://doi.org/10.1002/ets2.12120

Wang, C., & Blei, D. M. (2013). Variational inference in nonconjugate models. *Journal of Machine Learning Research, 14*, 1005–1031. https://www.jmlr.org/papers/volume14/wang13b/wang13b.pdf

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

Yamaguchi, K. (2020). Variational bayesian inference for the multiple-choice DINA model. *Behaviomtrika, 47*(1), 159–187. https://doi.org/10.1007/s41237-020-00104-w

Yamaguchi, K., & Okada, K. (2020). Variational Bayes inference for the DINA model. *Journal of Educational and Behavioral Statistics, 45*(5), 569–597. https://doi.org/10.3102/1076998620911934

Yamaguchi, K., & Okada, K. (2021). Variational Bayes inference algorithm for the saturated diagnostic classification model. *Psychometrika, 85*(4), 973–995. https://doi.org/10.1007/s11336-020-09739-w

Yamaguchi, K., & Templin, J. (2021). A Gibbs sampling algorithm with monotonicity constraints for diagnostic classification models. *Journal of Classification*. https://doi.org/10.1007/s00357-021-09392-7
Zhang, C., Butepage, J., Kjellstrom, H., & Mandt, S. (2019). Advances in variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(8), 2008–2026. https://doi.org/10.1109/TPAMI.2018.2889774
In Supplementary Material A, we provide the derivation of the conditional variational posterior for each of the parameters and the lower bound of the log marginal likelihood (variational lower bound; VLB) in the polytomous-attribute DCM.

First, the log joint posterior of the polytomous-attribute DCM is defined as

\[
\log P(Z, \Theta, \pi|X, G^{\text{poly}}, Q, \delta^0, A^0, B^0) 
\propto \log P(X|Z, \Theta, G^{\text{poly}}, Q) P(Z|\pi) P(\pi|\delta^0) P(\Theta|A^0, B^0) 
\]

\[
= \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l'=1}^{L'} z_{ijl'} \left\{ x_{ij} \log \theta_{jl'} + (1 - x_{ij}) \log(1 - \theta_{jl'}) \right\} 
+ \sum_{i=1}^{N} \sum_{l=1}^{L} z_{il} \log \pi_l + \sum_{l=1}^{L} (\delta^0_l - 1) \log \pi_l 
+ \sum_{j=1}^{J} \sum_{l'=1}^{L'} \left\{ (a^0_{jl'} - 1) \log \theta_{jl'} + (b^0_{jl'} - 1) \log(1 - \theta_{jl'}) \right\} 
\]

\[
= \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{l'=1}^{L'} \left( \sum_{l=1}^{L} g_{jl'l} \right) \left\{ x_{ij} \log \theta_{jl'} + (1 - x_{ij}) \log(1 - \theta_{jl'}) \right\} 
\cdot z_{ijl'} = \sum_{l=1}^{L} g_{jl'l} \sum_{l'=1}^{L'} \left( a^0_{jl'} - 1 \right) \log \theta_{jl'} + (b^0_{jl'} - 1) \log(1 - \theta_{jl'}) . \quad (21)
\]

Next, under the mean-field assumption of the variational distribution such that \( q(\Psi) = q(Z) q(\Theta, \pi) = \prod_{i=1}^{N} q(z_{i}) \prod_{j=1}^{J} \prod_{l'=1}^{L'} q(\theta_{jl'}) q(\pi) \), where \( \Psi = \{ Z, \Theta, \pi \} \), we derive the conditional variational posterior for each of the parameters in the polytomous-attribute saturated DCM.

1The derivation in Supplementary Material A is primarily an adaptation of that of the binary-attribute DCMs shown in Yamaguchi and Okada (2021).
A1. Inference of $z_i$

By applying the formula for updating the parameters of the variational posteriors in Section 2.5, the log variational posterior $\log q(z_i)$ is obtained as

$$
\log q(z_i) = E_q(\Theta, \pi) \left[ \sum_{j=1}^{J} \sum_{l=1}^{L_j^*} \sum_{l^*=1}^{L} g_{jl^* l} z_{il^* l} \left\{ x_{ij} \log \theta_{jl^* l} + (1 - x_{ij}) \log (1 - \theta_{jl^* l}) \right\} + \sum_{l=1}^{L} z_{il} \log \pi_l \right] + \text{const}
$$

$$
= E_q(\Theta, \pi) \left[ \sum_{l=1}^{L} z_{il} \left( \sum_{j=1}^{J} \sum_{l^*=1}^{L_j^*} g_{jl^* l} \left\{ x_{ij} \log \theta_{jl^* l} + (1 - x_{ij}) \log (1 - \theta_{jl^* l}) \right\} + \log \pi_l \right) \right] + \text{const}
$$

$$
= \sum_{l=1}^{L} z_{il} \log \rho_{il} + \text{const}, \quad (22)
$$

where

$$
\log \rho_{il} = \sum_{j=1}^{J} \sum_{l^*=1}^{L_j^*} g_{jl^* l} \left\{ x_{ij} E_q(\theta_{jl^* l}) [\log \theta_{jl^* l}] + (1 - x_{ij}) E_q(\theta_{jl^* l}) [\log (1 - \theta_{jl^* l})] \right\} + E_q(\pi_l)[\log \pi_l]. \quad (23)
$$

Thus, the conditional variational posterior of $z_i$ is expressed as the following categorical distribution:

$$
q(z_i) \propto \prod_{l=1}^{L} r_{il}^{z_{il}}, \quad (24)
$$

where $r_{il} = \frac{\rho_{il}}{\sum_{l=1}^{L} \rho_{il}}$.

A2. Inference of $\pi$

Similarly, the log variational posterior $\log q(\pi)$ is obtained as

$$
\log q(\pi) = E_q(Z, \Theta) \left[ \sum_{i=1}^{N} \sum_{l=1}^{L} z_{il} \log \pi_l + \sum_{l=1}^{L} (\delta_l^0 - 1) \log \pi_l \right] + \text{const}
$$

$$
= \sum_{l=1}^{L} \log \pi_l \left( \left\{ \sum_{i=1}^{N} E_q(z_i) [z_{il}] \right\} + \delta_l^0 - 1 \right) + \text{const}. \quad (25)
$$
This is equivalent to the following Dirichlet distribution:

\[ q(\pi) \propto \prod_{l=1}^{L} \pi_l^{\delta_l^* - 1}, \]  

(26)

where \( \delta_l^* = \left\{ \sum_{i=1}^{N} E_{q(z_i)} z_{il} \right\} + \delta_0^l \).

A3. Inference of \( \theta_{jl^*} \)

Finally, the log variational posterior \( \log q(\theta_{jl}) \) is obtained as

\[
\log q(\theta_{jl}) = E_{q(z)q(\pi)} \left[ \sum_{i=1}^{N} \left( \sum_{l=1}^{L} g_{jl^* l} z_{il} \right) \left\{ x_{ij} \log \theta_{jl^*} + (1 - x_{ij}) \log (1 - \theta_{jl^*}) \right\} \right] + (a_{jl^*}^0 - 1) \log \theta_{jl^*} + (b_{jl^*}^0 - 1) \log (1 - \theta_{jl^*}) + \text{const}
\]

\[
= \left( \sum_{i=1}^{N} \sum_{l=1}^{L} g_{jl^* l} E_{q(z_i)} [z_{il}] x_{ij} + a_{jl^*}^0 - 1 \right) \log \theta_{jl^*} + \left( \sum_{i=1}^{N} \sum_{l=1}^{L} g_{jl^* l} E_{q(z_i)} [z_{il}] (1 - x_{ij}) + b_{jl^*}^0 - 1 \right) \log (1 - \theta_{jl^*}) + \text{const}. \tag{27}
\]

The right-hand side of the above equation is equivalent to the following Beta distribution:

\[ q(\theta_{jl^*}) \propto \theta_{jl^*}^{a_{jl^*}^* - 1} (1 - \theta_{jl^*})^{b_{jl^*}^* - 1}, \]  

(28)

where

\[
\begin{align*}
a_{jl^*}^* &= \sum_{i=1}^{N} \sum_{l=1}^{L} g_{jl^* l} E_{q(z_i)} [z_{il}] x_{ij} + a_{jl^*}^0, \\
b_{jl^*}^* &= \sum_{i=1}^{N} \sum_{l=1}^{L} g_{jl^* l} E_{q(z_i)} [z_{il}] (1 - x_{ij}) + b_{jl^*}^0. 
\end{align*} \tag{29}
\]
A4. Deriving the Variational Lower Bound

We provide the derivation of the VLB of the log marginal likelihood for the polytomous-attribute saturated DCM with G-matrices using collapsed attribute vectors. The VLB is defined as

\[ L(q) = \int q(\Psi) \log \frac{P(X, \Psi)}{q(\Psi)} d\Psi, \]  

(30)

where \( \Psi = \{Z, \Theta, \pi\} \). Thus, the VLB is written as

\[
L(q) = \sum Z \int \int q(Z) q(\Theta, \pi) \log \frac{P(X, Z, \Theta, \pi|G^{\text{poly}}, Q, \delta^0, A^0, B^0)}{q(Z, \Theta, \pi)} d\Theta d\pi \\
= \mathbb{E}_{q(Z)q(\Theta, \pi)} \left[ \log P(X, Z, \Theta, \pi|G^{\text{poly}}, Q, \delta^0, A^0, B^0) \right] \\
- \mathbb{E}_{q(Z)q(\Theta, \pi)} \left[ \log q(Z) q(\Theta, \pi) \right] \\
= \mathbb{E}_{q(Z)q(\Theta, \pi)} \left[ \log P(X|Z, \Theta, G^{\text{poly}}, Q) \right] + \mathbb{E}_{q(Z)q(\Theta, \pi)} \left[ \log P(Z|\pi) \right] \\
+ \mathbb{E}_{q(Z)q(\Theta, \pi)} \left[ \log P(\pi|\delta^0) \right] + \mathbb{E}_{q(Z)q(\Theta, \pi)} \left[ \log P(\Theta|A^0, B^0) \right] \\
- \mathbb{E}_{q(Z)} \left[ \log q(Z) \right] - \mathbb{E}_{q(\Theta)} \left[ \log q(\Theta) \right] - \mathbb{E}_{q(\pi)} \left[ \log q(\pi) \right]. \]  

(31)
Next, we show the derivations for seven terms in Equation (31). These terms are derived as

\[
E_q(Z)q(\Theta, \pi) \left[ \log P(X|Z, \Theta, G^{poly}, Q) \right]
= \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l=1}^{L} \left( \sum_{l^*}^{L} g_{jl^*} E_q(z_{il^*}) \right) \left( x_{ij} E_q(\theta_{jl^*}) \log \theta_{jl^*} \right)
+ (1 - x_{ij}) E_q(\theta_{jl^*}) \log(1 - \theta_{jl^*}) \right) \right)
\] (32)

\[
E_q(Z)q(\Theta, \pi) \left[ \log P(Z|\pi) \right] = \sum_{i=1}^{N} \sum_{l=1}^{L} E_q(z_{il}) E_q(\pi) \log \pi_l \] (33)

\[
E_q(Z)q(\Theta, \pi) \left[ \log P(\Theta|\pi^0) \right]
= - \log \left( \prod_{l=1}^{L} \Gamma(\delta_l^0) \right) + \sum_{l=1}^{L} (\delta_l^0 - 1) E_q(\pi) \log \pi_l \] (34)

\[
E_q(Z)q(\Theta, \pi) \left[ \log P(\Theta|A^0, B^0) \right]
= - \sum_{j=1}^{J} \sum_{l^*=1}^{L^*} \log B(a_{jl^*}^0, b_{jl^*}^0)
+ \sum_{j=1}^{J} \sum_{l^*=1}^{L^*} \left( (a_{jl^*}^0 - 1) E_q(\theta_{jl^*}) \log \theta_{jl^*} \right) + (b_{jl^*}^0 - 1) E_q(\theta_{jl^*}) \log(1 - \theta_{jl^*}) \right) \right) \] (35)

\[
E_q(Z) \left[ \log q(Z) \right] = \sum_{i=1}^{N} \sum_{l=1}^{L} \log E_q(z_{il}) z_{il} \] (36)

\[
E_q(\Theta) \left[ \log q(\Theta) \right]
= - \sum_{j=1}^{J} \sum_{l^*=1}^{L^*} \log B(a_{jl^*}^*, b_{jl^*}^*)
+ \sum_{j=1}^{J} \sum_{l^*=1}^{L^*} \left( (a_{jl^*}^* - 1) E_q(\theta_{jl^*}) \log \theta_{jl^*} \right) + (b_{jl^*}^* - 1) E_q(\theta_{jl^*}) \log(1 - \theta_{jl^*}) \right) \right) \] (37)

\[
E_q(\pi) \left[ \log q(\pi) \right] = - \log \left( \prod_{l=1}^{L} \Gamma(\delta_l^*) \right) + \sum_{l=1}^{L} (\delta_l^* - 1) E_q(\pi) \log \pi_l \] (38)
Then, by substituting Equations (32) to (38) with some algebra, the VLB is given as

\[
L(q) = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l^*=1}^{L^*_j} \left( \sum_{l=1}^{L} g_{jl^*} r_{it} \right) \left( x_{ij} \psi(a_{jl^*}^*) - \psi(a_{jl^*}^* + b_{jl^*}^*) \right) 
+ (1 - x_{ij}) \psi(b_{jl^*}^*) - \psi(a_{jl^*}^* + b_{jl^*}^*) \right) 
+ \sum_{i=1}^{N} \sum_{l=1}^{L} r_{it} \left( \psi(\delta_l^*) - \psi(\sum_{l'=1}^{L} \delta_l'^*) \log r_{it} \right) 
+ \log \left( \frac{\prod_{l=1}^{L} \Gamma(\delta_l^*)}{\prod_{l'=1}^{L} \Gamma(\sum_{l=1}^{L} \delta_l'^*)} \right) 
\log \left( \frac{\prod_{l=1}^{L} \Gamma(\delta_l^0)}{\prod_{l'=1}^{L} \Gamma(\sum_{l=1}^{L} \delta_l'^0)} \right) 
+ \sum_{l=1}^{L} (\delta_l^0 - \delta_l^*) E_q(\pi) [\log \pi_l] 
+ \sum_{j=1}^{J} \sum_{l^*=1}^{L^*_j} \left\{ \log B(a_{jl^*}^*, b_{jl^*}^*) - \log B(a_{jl^*}^0, b_{jl^*}^0) 
+ (a_{jl^*}^0 - a_{jl^*}^*) \left( \psi(a_{jl^*}^*) - \psi(a_{jl^*}^* + b_{jl^*}^*) \right) 
+ (b_{jl^*}^0 - b_{jl^*}^*) \left( \psi(b_{jl^*}^*) - \psi(a_{jl^*}^* + b_{jl^*}^*) \right) \right\}.
\]

(39)

where \( \Gamma(\cdot) \) and \( \psi(\cdot) \) denote the gamma and digamma functions, respectively.
The true Q-matrices used in all the simulation studies were specified in the following manner. The Q-matrix with $K = 4$ and $J = 60$ comprises the diagonal matrices with one and two, 28 $q$-vectors that require two attributes, and 24 $q$-vectors that require three attributes. The number of items measuring each attribute is 34 for all the attributes. The Q-matrix with $K = 4$ and $J = 120$ is the double stacking of the Q-matrix with $K = 4$ and $J = 60$. Similarly, the Q-matrix with $K = 7$ and $J = 60$ contains the diagonal matrices with one and two, 14 $q$-vectors that require two attributes, 21 $q$-vectors that require three attributes, and 11 $q$-vectors that require four attributes. The number of items that measure each attribute was set to be approximately equal across the attributes as follows: 20 (A1), 21 (A2), 22 (A3), 23 (A4), 22 (A5), 21 (A6), and 20 (A7). The Q-matrix with $K = 7$ and $J = 120$ is the double stacking of the Q-matrix with $K = 7$ and $J = 60$.

In the following figures for the specifications of the true Q-matrices, the white, gray, and black boxes denote the Q-matrix entries that take the values of 0, 1, and 2, respectively.
Q-matrix \((K=4, J=60)\)
Q- matrix (K=4, J=120)
Q-matrix (K=7, J=60)
Q- matrix (K=7, J=120)
Supplementary Material C

Additional Simulation Study: Performance of the Polytomous-Attribute Saturated DCM with its G-Matrices Using Reduced Attribute Vectors

We investigated the performance of the polytomous-attribute saturated DCM with its G-matrix based on reduced attribute vectors. In the additional simulation study, we considered a sample size of 10000 or 30000, a number of items of 60, a correlation coefficient among the attributes of .1, and a number of attributes of $K = 4$. Regarding the specification of the true values of the correct-response probability parameters, we used the same values of $p^{\text{lowest}}_j$ and $p^{\text{highest}}_j$ as the relevant conditions in Simulation Study 1. Then, for each item, we applied the same procedure to set the true values of these parameters given the item-specific reduced attribute vectors. The other specifications were set to be the same as those in Simulation Study 1.

Results

Table C1 shows the biases and RMSEs for the correct-response probability parameters in all the conditions of the additional simulation study. Similar to Simulation Study 1, the biases and RMSEs were averaged according to the number of attributes measured by one item. For the polytomous-attribute saturated DCM with its G-matrices using reduced attribute vectors, the number of correct-response probability parameters for each item becomes $M^{K_j}$, which is in contrast to that with the G-matrices using collapsed attribute vectors, where the number of these parameters for each item is $2^{K_j}$.

Regarding the correct-response probability parameters, the same pattern as for collapsed attribute vectors was found. Specifically, the values of the biases were close to zero, although they were larger than those with collapsed attribute vectors. In addition, sample size lowers the values of the RMSEs and these values increased as the number of attributes measured by one item rose. In particular, the values of the RMSEs were relatively large when the sample size was 10000 and the number of attributes measured by one item was three. However, the RMSEs were less than approximately .0700 across all the conditions. With respect to the mixing proportion parameters, the absolute values of the largest positive and negative biases were within .006 and the maximum value of the RMSEs was also less than about .008. These results suggest the satisfactory accuracy of the proposed method for estimating the mixing proportion parameters.

Lastly, the performance of our method for estimating attribute mastery patterns was less satisfactory than the model with G-matrices using collapsed attribute vectors owing to more notched item-specific correct-response probabilities. Since the number of item-specific attribute mastery patterns that can have different correct-response probabilities increases to $M^{K_j}$ from $2^{K_j}$ in the model with reduced attribute vectors, the values of the correct-response probabilities among the item-specific attribute mastery patterns differ little from each other, meaning that the item’s power to distinguish the likelihood at which a respondent belongs to a certain attribute mastery pattern also decreases compared with the model using collapsed attribute vectors. Hence, a large number of items would be required to ensure the accuracy of the attribute profile estimation in the model based on reduced attribute vectors.
Table C1: Results of the additional simulation study ($K = 4$, correlation coefficient .1)

| Number of Attributes Measured by One Item | 1 | 2 | 3 |
|------------------------------------------|---|---|---|
| Number of Items                         | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 60                                      | 10000 | -.0021 | .0228 | -.0021 | .0327 | -.0019 | .0701 |
|                                          | 30000 | -.0012 | .0093 | -.0010 | .0149 | -.0010 | .0325 |

Maximum and minimum values of the biases and RMSEs for the mixing proportion parameters

| Mixing Proportion Parameter | Bias | RMSE |
|-----------------------------|------|------|
| Number of Items             |      |      |
| 60                          |      |      |
| 10000                       | .0032 | -.0059 | .0081 | .0027 |
| 30000                       | .0008 | -.0010 | .0035 | .0012 |

EACRs and PACRs for the attribute profile estimation

| Attribute Classification Rate | Number of Items | Sample Size |
|-------------------------------|-----------------|-------------|
| EACR$_{\alpha_1}$           | 60              | 10000       |
| EACR$_{\alpha_2}$           |                 | .757        |
| EACR$_{\alpha_3}$           |                 | .761        |
| EACR$_{\alpha_4}$           |                 | .760        |
| PACR$_{\alpha}$             | .776            |
| EACR$_{\alpha_2}$           |                 | .783        |
| EACR$_{\alpha_3}$           |                 | .784        |
| EACR$_{\alpha_4}$           |                 | .795        |
| PACR$_{\alpha}$             | .419            |

Note. Maximum and minimum values of the biases correspond to the largest positive and negative biases among the $M^K$ mixing proportion parameters, respectively. RMSE = root-mean-square error, EACR = element-wise attribute classification rate, PACR = pattern-wise attribute classification rate.
SUPPLEMENTARY MATERIAL D

EMPIRICAL STUDY: ESTIMATES OF THE CORRECT-RESPONSE PROBABILITY PARAMETERS

Tables D1 and D2 present the estimates of the EAP and posterior standard deviation (SD) for the correct-response probability parameters from the Gibbs sampler and the proposed method, respectively. In the tables, \( P(\cdot) \) means the correct-response probability given collapsed attribute vectors for the items. For instance, when item \( j \) has \( q_j = (1, 0, 0) \), its collapsed attribute vectors are \( \alpha_{j1}^* = (0) \) and \( \alpha_{j2}^* = (1) \). Hence, \( P(0) \) for item \( j \) corresponds to \( P(x_{ij} = 1 | \alpha_{j1}^* = (0)) \).

Table D1: The estimates of EAP and posterior standard deviation for the correct-response probability parameters from the Gibbs sampler

| Items | \( P(0) \) (EAP) | \( P(1) \) (EAP) | \( P(00) \) (Post) | \( P(01) \) (Post) | \( P(10) \) (Post) | \( P(11) \) (Post) |
|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1     | .358 (.007)     | .931 (.002)     | .931 (.002)     | .931 (.002)     | .931 (.002)     | .931 (.002)     |
| 2     | .479 (.007)     | .935 (.002)     | .935 (.002)     | .935 (.002)     | .935 (.002)     | .935 (.002)     |
| 3     | .285 (.006)     | .852 (.003)     | .852 (.003)     | .852 (.003)     | .852 (.003)     | .852 (.003)     |
| 4     | .457 (.007)     | .946 (.002)     | .946 (.002)     | .946 (.002)     | .946 (.002)     | .946 (.002)     |
| 5     | .360 (.007)     | .931 (.002)     | .931 (.002)     | .931 (.002)     | .931 (.002)     | .931 (.002)     |
| 6     | .051 (.003)     | .758 (.004)     | .758 (.004)     | .758 (.004)     | .758 (.004)     | .758 (.004)     |
| 7     | .121 (.005)     | .818 (.003)     | .818 (.003)     | .818 (.003)     | .818 (.003)     | .818 (.003)     |
| 8     | .227 (.006)     | .603 (.008)     | .506 (.025)     | .879 (.004)     | .879 (.004)     | .879 (.004)     |
| 9     | .310 (.004)     | .825 (.004)     | .825 (.004)     | .825 (.004)     | .825 (.004)     | .825 (.004)     |
| 10    | .199 (.006)     | .820 (.007)     | .801 (.020)     | .963 (.002)     | .963 (.002)     | .963 (.002)     |
| 11    | .054 (.003)     | .161 (.022)     | .213 (.006)     | .750 (.005)     | .750 (.005)     | .750 (.005)     |
| 12    | .593 (.005)     | .978 (.002)     | .978 (.002)     | .978 (.002)     | .978 (.002)     | .978 (.002)     |
| 13    | .479 (.005)     | .909 (.003)     | .909 (.003)     | .909 (.003)     | .909 (.003)     | .909 (.003)     |
| 14    | .210 (.006)     | .159 (.060)     | .433 (.033)     | .715 (.129)     | .715 (.129)     | .715 (.129)     |
| 15    | .364 (.007)     | .448 (.070)     | .644 (.031)     | .768 (.118)     | .425 (.019)     | .656 (.008)     |
| 16    | .374 (.007)     | .523 (.069)     | .642 (.032)     | .530 (.141)     | .478 (.019)     | .656 (.008)     |
| 17    | .096 (.004)     | .661 (.030)     | .591 (.007)     | .909 (.003)     | .909 (.003)     | .909 (.003)     |
| 18    | .272 (.006)     | .467 (.029)     | .529 (.007)     | .859 (.004)     | .859 (.004)     | .859 (.004)     |
| 19    | .007 (.001)     | .108 (.018)     | .301 (.007)     | .798 (.005)     | .798 (.005)     | .798 (.005)     |
| 20    | .082 (.004)     | .180 (.022)     | .124 (.005)     | .499 (.006)     | .499 (.006)     | .499 (.006)     |
| 21    | .059 (.003)     | .483 (.030)     | .358 (.007)     | .921 (.003)     | .921 (.003)     | .921 (.003)     |
| 22    | .402 (.007)     | .863 (.020)     | .807 (.005)     | .946 (.003)     | .946 (.003)     | .946 (.003)     |
| 23    | .238 (.006)     | .634 (.029)     | .478 (.007)     | .854 (.004)     | .854 (.004)     | .854 (.004)     |
| 24    | .245 (.006)     | .783 (.004)     | .783 (.004)     | .783 (.004)     | .783 (.004)     | .783 (.004)     |
| 25    | .387 (.004)     | .657 (.006)     | .657 (.006)     | .657 (.006)     | .657 (.006)     | .657 (.006)     |
| 26    | .504 (.006)     | .806 (.004)     | .806 (.004)     | .806 (.004)     | .806 (.004)     | .806 (.004)     |
| 27    | .617 (.007)     | .941 (.014)     | .870 (.004)     | .968 (.002)     | .968 (.002)     | .968 (.002)     |
| 28    | .256 (.006)     | .816 (.066)     | .576 (.035)     | .539 (.147)     | .346 (.019)     | .635 (.008)     |
| 29    | .198 (.006)     | .540 (.072)     | .461 (.032)     | .503 (.142)     | .246 (.018)     | .552 (.009)     |
| 30    | .532 (.007)     | .965 (.024)     | .919 (.019)     | .885 (.088)     | .664 (.018)     | .865 (.006)     |
| 31    | .009 (.001)     | .115 (.005)     | .049 (.012)     | .558 (.006)     | .558 (.006)     | .558 (.006)     |
| 32    | .151 (.003)     | .329 (.005)     | .329 (.005)     | .329 (.005)     | .329 (.005)     | .329 (.005)     |
| 33    | .024 (.002)     | .466 (.007)     | .466 (.007)     | .466 (.007)     | .466 (.007)     | .466 (.007)     |
| 34    | .355 (.004)     | .779 (.006)     | .779 (.006)     | .779 (.006)     | .779 (.006)     | .779 (.006)     |

Note. We present the values of EAP and posterior standard deviation as “EAP (Posterior SD).”
Table D2: The estimates of EAP and posterior standard deviation for the correct-response probability parameters from the proposed method

| Items | P(00) | P(01) | P(10) | P(11) | P(000) | P(001) | P(010) | P(011) | P(100) | P(101) | P(110) | P(111) |
|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1     | .358 (.006) | .931 (.002) |       |       |        |        |        |        |        |        |        |        |
| 2     | .478 (.006) | .935 (.002) |       |       |        |        |        |        |        |        |        |        |
| 3     | .285 (.006) | .852 (.003) |       |       |        |        |        |        |        |        |        |        |
| 4     | .456 (.006) | .946 (.002) |       |       |        |        |        |        |        |        |        |        |
| 5     | .359 (.006) | .931 (.002) |       |       |        |        |        |        |        |        |        |        |
| 6     | .051 (.003) | .758 (.003) |       |       |        |        |        |        |        |        |        |        |
| 7     | .121 (.004) | .818 (.003) |       |       |        |        |        |        |        |        |        |        |
| 8     | .226 (.005) | .602 (.007) | .506 (.016) | .879 (.004) |        |        |        |        |        |        |        |        |
| 9     | .310 (.004) | .825 (.004) |       |       |        |        |        |        |        |        |        |        |
| 10    | .198 (.005) | .819 (.005) | .802 (.013) | .963 (.002) |        |        |        |        |        |        |        |        |
| 11    | .054 (.003) | .162 (.016) | .213 (.005) | .750 (.005) |        |        |        |        |        |        |        |        |
| 12    | .593 (.004) | .978 (.002) |       |       |        |        |        |        |        |        |        |        |
| 13    | .479 (.004) | .909 (.003) |       |       |        |        |        |        |        |        |        |        |
| 14    | .210 (.005) | .159 (.028) | .432 (.023) | .719 (.066) | .343 (.013) | .597 (.007) | .789 (.018) | .933 (.003) |        |        |        |        |
| 15    | .364 (.006) | .450 (.039) | .644 (.022) | .781 (.061) | .424 (.013) | .656 (.007) | .841 (.017) | .929 (.003) |        |        |        |        |
| 16    | .374 (.006) | .523 (.039) | .643 (.022) | .520 (.074) | .478 (.013) | .656 (.007) | .774 (.019) | .918 (.003) |        |        |        |        |
| 17    | .095 (.004) | .662 (.021) | .591 (.006) | .909 (.003) |        |        |        |        |        |        |        |        |
| 18    | .272 (.006) | .467 (.022) | .528 (.006) | .859 (.004) |        |        |        |        |        |        |        |        |
| 19    | .007 (.001) | .108 (.014) | .301 (.006) | .797 (.004) |        |        |        |        |        |        |        |        |
| 20    | .082 (.004) | .180 (.017) | .124 (.004) | .499 (.005) |        |        |        |        |        |        |        |        |
| 21    | .058 (.003) | .484 (.022) | .358 (.006) | .921 (.003) |        |        |        |        |        |        |        |        |
| 22    | .402 (.006) | .862 (.015) | .807 (.005) | .946 (.002) |        |        |        |        |        |        |        |        |
| 23    | .238 (.006) | .635 (.021) | .478 (.006) | .854 (.004) |        |        |        |        |        |        |        |        |
| 24    | .245 (.005) | .783 (.004) |       |       |        |        |        |        |        |        |        |        |
| 25    | .387 (.004) | .657 (.005) |       |       |        |        |        |        |        |        |        |        |
| 26    | .503 (.006) | .806 (.003) |       |       |        |        |        |        |        |        |        |        |
| 27    | .617 (.006) | .941 (.010) | .870 (.004) | .968 (.002) |        |        |        |        |        |        |        |        |
| 28    | .256 (.006) | .816 (.030) | .575 (.022) | .544 (.074) | .345 (.013) | .635 (.007) | .525 (.023) | .866 (.004) |        |        |        |        |
| 29    | .198 (.005) | .537 (.039) | .462 (.023) | .492 (.074) | .246 (.011) | .552 (.007) | .438 (.022) | .873 (.004) |        |        |        |        |
| 30    | .532 (.007) | .970 (.013) | .917 (.013) | .903 (.044) | .663 (.013) | .865 (.005) | .783 (.019) | .980 (.002) |        |        |        |        |
| 31    | .009 (.001) | .115 (.004) | .047 (.007) | .558 (.005) |        |        |        |        |        |        |        |        |
| 32    | .151 (.003) | .329 (.005) |       |       |        |        |        |        |        |        |        |        |
| 33    | .024 (.001) | .466 (.005) |       |       |        |        |        |        |        |        |        |        |
| 34    | .355 (.004) | .779 (.005) |       |       |        |        |        |        |        |        |        |        |

*Note.* We present the values of EAP and posterior standard deviation as "EAP (Posterior SD)."