THE GENERALIZED COGNITIVE DIAGNOSIS MODEL FRAMEWORK FOR POLYTOMOUS ATTRIBUTES

JIMMY DE LA TORRE

THE UNIVERSITY OF HONG KONG

XUELAN QIU

AUSTRALIAN CATHOLIC UNIVERSITY

KEVIN CARL P. SANTOS

UNIVERSITY OF THE PHILIPPINES - DILIMAN

Correspondence should be sent to Jimmy de la Torre, Faculty of Education, Pokfulam Road, Hong Kong.

E-Mail: j.delatorre@hku.hk

Phone: +852 3917 2539

Abstract

For classroom teaching and learning, classifying students' skills into more than two categories (e.g., no, basic, and advanced masteries) is more instructionally relevant. Such classifications require polytomous attributes, for which most existing cognitive diagnosis models (CDMs) are inapplicable. This paper proposes the *saturated polytomous cognitive diagnosis model* (sp-CDM), a general model that subsumes existing CDMs for polytomous attributes as special cases. The generalization is shown by mathematically illustrating the relationships between the proposed and existing CDMs. Moreover, algorithms to estimate the proposed model is proposed. A simulation study is conducted to evaluate the parameter recovery of the sp-CDM using the proposed estimation algorithms, as well as to illustrate the consequences of improperly fitting constrained or unnecessarily complex polytomous-attribute CDMs. A real-data example involving polytomous attributes is presented to demonstrate the practical utility of the proposed model.

Key words: cognitive diagnosis models, polytomous attributes, G-DINA, parameter estimation

1. Introduction

A wide variety of cognitive diagnosis models (CDMs) exist in the literature. Typically, these models were proposed for dichotomous attributes appropriate for determining, say, skill mastery or nonmastery. However, in many applications, classifying students into more than two categories is more instructionally relevant. Such classifications require polytomous attributes, where the attribute levels can be ordinal categories (e.g., no mastery, basic mastery, and advanced mastery). For example, the proportional reasoning (PR) assessment developed to measure the PR skills for middle school (equivalently, secondary) students (Tjoe & de la Torre, 2013, 2014) involves two three-level polytomous attributes, namely, (a) comparing and ordering of fractions, where level 0 represents nonmastery of the attribute, level 1 the ability to compare two fractions, and level 2 the ability to order three or more fractions; and (b) constructing ratios and proportions, where level 0 represents nonmastery, level 1 the ability to construct a single ratio, and level 2 the ability to construct a proportion, which is made up of two ratios. Attribute levels can also be nominal categories representing different content domains. For example, level 1 represents the prerequisite skills (e.g., add and subtract) in the attribute hierarchy, whereas level 2 the advanced skills (e.g., multiply and divide). In general, any M-level polytomous attribute can be equivalently represented by $M-1$ dichotomous attributes that follow a linear hierarchy (Leighton et al., 2004). In the example above, the polytomous attribute comparing/ordering fractions can be split into two dichotomous attributes, where the first deals with two fractions and the second with three or more fractions, and the mastery of the former is a prerequisite to the mastery the latter. The prerequisite relationship will constrain the number of possible mastery combinations to three, namely, 00, 10, and 11, which is equivalent to levels 0, 1, and 2 of the original polytomous attribute. Finally, it is important to underscore that this paper focuses on the attributes with more than two categories (i.e., polytomous *attributes*), rather than responses with more than two categories (i.e., polytomous *responses*). This distinction is necessary because both CDMs for polytomous attributes and those for polytomous responses have been both referred to as polytomous CDMs in the literature.

To accommodate polytomous attributes, several CDMs have been developed in the literature,

which are summarized and shown in Table 1 according to several key features such as general or constrained, link function, and core assumptions. For example, Templin (2004) extended the reparameterized unified model (RUM; Hartz, 2002) for polytomous attributes (RUM-PA) and proposed a constrained version (cRUM-PA), Karelitz (2004) proposed the ordered category attribute coding (OCAC) framework in conjunction with the *deterministic input, noisy* "and" *gate* (DINA) model (Junker & Sijtsma, 2001) to define the mastery levels as multiple ordered categories. By defining accuracy with fast speed as the highest level of an attribute, accuracy with slow speed as intermediate level, and nonmastery as the lowest level, Wang and Chen (2020) extended the DINA model to be the response accuracy model (RAM) model to measure students' fluency in answering the test items. Recently, Yakar et al. (2021) developed a fully additive model for polytomous attributes (fA-M) which accounts for the effects of each attribute levels. However, these models are deemed to be not general enough mainly because the models focused on a specific and constrained CDM.

The existing general CDMs for polytomous attributes include those proposed within the log-linear cognitive diagnosis model (LCDM; Henson et al., 2009), that within the general diagnostic model (GDM; von Davier, 2008), and that within the generalized *deterministic input, noisy* "and" *gate* (G-DINA) model (de la Torre, 2011).

The polytomous diagnostic classification model (PDCM) framework (Bao, 2019) extends the measurement and structural models of the LCDM to the polytomous attribute setting. Note that in the PDCM framework, only the attribute patterns ($\boldsymbol{\alpha}_l$) are polytomous, whereas the Q-matrix entries remain binary. In contrast, both the attribute pattern and Q-matrix entries are polytomous, for other CDMs in this paper, such as the OCAC framework, GDM, and the proposed framework. The probabilities between different levels in PDCM can be varied for greater flexibility or be equal for smaller number of parameters. The PDCM uses the dummy coding approach in which the M levels of an attribute are coded as $(M-1)$ dummy variables and the combinations of the dummy variables - representing different knowledge states - are treated as the polytomous attribute levels. For example, for three levels of an attribute, they are coded with two dummy variables and as (0,0), (1,0), and (1,1) to represent nonmastery, intermediate mastery, and mastery. This coding approach might be workable when the number of levels in attributes

and the number of attributes in a test are moderate. It becomes tedious and hard to interpret the representation of the knowledge states when the number of levels and attributes are large.

With a proper choice of the central component function, as in, the function $h(\cdot)$ that maps the attribute levels using the Q-matrix entries, the GDM can flexibly accommodate polytomous attributes. For example, a useful and reasonable choice of $h(\cdot)$ is defined as $h = min(q_k, \alpha_k)$. As a result, an attribute level that is higher than $h(\cdot)$ will not increase the probability of solving an item, whereas that is lower than $h(\cdot)$ results in a lower success probability. In other words, while there is no distinction between groups who possess the required attribute level and who have an even higher level, there is distinction between groups whose attribute levels are lower than the required level. Nonetheless, the GDM for polytomous attributes (GDM-PA) has neither been examined with simulation studies in enough details nor applied to the real data.

With respect to G-DINA model for polytomous attributes, namely, the pG-DINA model (J. Chen & de la Torre, 2013), the model relies on a core assumption, which is referred to the *specific attribute level mastery* (SALM), where each item is assumed to separates examinees into two reduced latent groups - those who are on or above a specific attribute level, and those who are below it. With the SALM assumption, some levels in the pG-DINA do not increase the success probability. Such a constraint may be too stringent because attribute vectors in the same reduced latent group are very likely to have varying levels with respect to the required attributes and thus their probabilities of success may not be identical.

The first and primary aim of this paper is to propose a general CDM framework for polytomous attributes analogous to the G-DINA model for dichotomous attributes. Specifically, the proposed model extends the pG-DINA model by relaxing the SALM assumption and allows for the different attribute levels to contribute differentially to the success probability. This work also aims to derive the special cases of the sp-CDM under different constraints and show the mathematical relationships between the sp-CDM and the existing CDMs for polytomous attributes and its special cases. Third, this work aims to address the estimation of the sp-CDM, to examine parameter recovery using the proposed estimation algorithms, and the consequences of fitting constrained and unnecessarily complex models across a range of conditions. Finally, the study aims to demonstrate the application of the sp-CDM with a real data of PR assessment.

This paper contributes to the literature by developing a unified framework for polytomous attributes. The proposed model has three unique features: (1) Compared to the existing CDMs for polytomous attributes, where some attribute levels share identical success probabilities, the sp-CDM allows for different attribute levels to have their unique contributions to the success probability; (2) the sp-CDM is formulated with alternative link functions, thus, making it more general; and (3) due to the different model formulations, the existing models can be mathematically shown to be special cases of the various forms of the sp-CDM with appropriate constraints. Despite the similar structure of this work to that for G-DINA (de la Torre, 2011) or the pG-DINA (J. Chen & de la Torre, 2013), the fundamental differences are substantial. Specifically, the formulations, the estimations, and the implications of three models are substantially different.

2. The Generalized Cognitive Diagnosis Model Framework for Polytomous Attributes

The generalized CDM framework for polytomous attributes can be expressed as three saturated models under different link functions. Let J be the number of items, K the number of attributes, and M_k the number of levels of attribute k. For notational convenience, but without loss of generality, it can be assumed that $M_k = M$, indicating the number of levels is identical for all attributes. Thus, there will be a total of $\prod_{k=1}^{K} M_k = M^K$ attribute patterns or latent classes. Let $K_j^* = \sum_{k=1}^{K} I(q_{jk} > 0)$ be the number of required attributes for the item j, $j=1, \ldots, J$. Again, for notational convenience, let the first K_j^* attributes be the required attributes for item j. We use $\mathbf{q}_j = (q_1, \cdots q_{K_j^*}, \mathbf{0})_{1 \times K}$ to denote the required levels in the K_j^* attributes to answer the item j correctly, and $\boldsymbol{\alpha}_l^* = (\alpha_1, \cdots, \alpha_k, \cdots, \alpha_{K_j^*}^*)_{1 \times K_j^*}$ the lth reduced attribute pattern or latent group. As can be seen from above, the entries in both the $\mathbf{q}$ and $\boldsymbol{\alpha}^*$ can have more than two categories.

To illustrate, consider $K = 3$, $M = 3$, and $\mathbf{q} = (1, 2, 0)$, which indicates that the first level of α_1 and the second level of α_2 are required for the item, hence, $K_j^* = 2$. For this example, there are $M^K = 27$ latent classes, which will be partitioned into $M^{K_j^*} = 9$ latent groups. Specifically, for this item, the latent classes $\boldsymbol{\alpha}_l$ and $\boldsymbol{\alpha}_{l'}$ are classified in the same latent group when $\alpha_{l1} = \alpha_{l'1}$ and $\alpha_{l2} = \alpha_{l'2}$. For example, the latent classes 000, 001, and 002 all belong to the latent group 00.

The item response function (IRF) of the proposed model using the identity link function is given by

$$
\begin{aligned}
P_j(\boldsymbol{\alpha}_l^*) = \delta_{j0} &+ \sum_{k=1}^{K_j^*} \sum_{m=1}^{M-1} \delta_{jkm} I\big[\alpha_{lk} = m\big] + \\
&\sum_{k'>k}^{K_j^*} \sum_{k=1}^{K_j^*-1} \sum_{m=1}^{M-1} \sum_{m'=1}^{M-1} \delta_{jkmk'm'} I\big[\alpha_{lk} = m\big] I\big[\alpha_{lk'} = m'\big] + \cdots + \\
&\sum_{m_1=1}^{M-1} \sum_{m_2=1}^{M-1} \cdots \sum_{m_{K_j^*}=1}^{M-1} \delta_{j1m_12m_2\ldots K_j^*m_{K_j^*}} \prod_{k=1}^{K_j^*} I\big[\alpha_{lk} = m_k\big],
\end{aligned}
\tag{1}
$$

where δ_{j0} is the intercept, δ_{jkm} is the main effect of the mth level of attribute k, $\delta_{jkmk'm'}$ is the two-way interaction effect of the mth level of attribute k and the m'th level of attribute k', and $\delta_{j1m_12m_2\ldots K_j^*m_{K_j^*}}$ is the K_j^*-way interaction effect of the m_1th level of the α_1, m_2th level of the α_2, up to the $m_{K_j^*}$th level of $\alpha_{K_j^*}$. It can be further noted that the subscript m of δ_{jkm} indicates that each attribute level in α_k contributes differentially to the success probability, as in, the steps between adjacent levels vary (e.g., the step between "no mastery" and "basic mastery" is different from that between "basic mastery" and "advanced mastery"). To reduce the number of parameters and, hence, model complexity, it can be assumed that the steps between levels within α_k are identical, which reduces δ_{jkm} to δ_{jk}.

In addition to the identity link function, the sp-CDM can also be formulated with the logit and log links. Despite the similar forms, the models using different link functions are essentially different in terms of the values and interpretations of the parameters. For this reason, different notations are used for parameters under formulations with logit and log link functions.

For the logit link,

$$
\begin{aligned}
\mathrm{logit}[P_j(\boldsymbol{\alpha}_l^*)] = \lambda_{j0} &+ \sum_{k=1}^{K_j^*} \sum_{m=1}^{M-1} \lambda_{jkm} I\big[\alpha_{lk} = m\big] + \\
&\sum_{k'>k}^{K_j^*} \sum_{k=1}^{K_j^*-1} \sum_{m=1}^{M-1} \sum_{m'=1}^{M-1} \lambda_{jkmk'm'} I\big[\alpha_{lk} = m\big] I\big[\alpha_{lk'} = m'\big] + \cdots + \\
&\sum_{m_1=1}^{M-1} \sum_{m_2=1}^{M-1} \cdots \sum_{m_{K_j^*}=1}^{M-1} \lambda_{j1m_12m_2\ldots K_j^*m_{K_j^*}} \prod_{k=1}^{K_j^*} I\big[\alpha_{lk} = m_k\big].
\end{aligned}
\tag{2}
$$

For the log link,

$$\log[P_j(\boldsymbol{\alpha}_l^*)] = \nu_{j0} + \sum_{k=1}^{K_j^*} \sum_{m=1}^{M-1} \nu_{jkm} I\big[\alpha_{lk} = m\big] +$$

$$\sum_{k'>k}^{K_j^*} \sum_{k=1}^{K_j^*-1} \sum_{m=1}^{M-1} \sum_{m'=1}^{M-1} \nu_{jkmk'm'} I\big[\alpha_{lk} = m\big] I\big[\alpha_{lk'} = m'\big] + \cdots + \quad (3)$$

$$\sum_{m_1=1}^{M-1} \sum_{m_2=1}^{M-1} \cdots \sum_{m_{K_j^*}=1}^{M-1} \nu_{j1m_12m_2\ldots K_j^*m_{K_j^*}} \prod_{k=1}^{K_j^*} I\big[\alpha_{lk} = m_k\big].$$

Equations 1, 2, and 3 are referred to as the *saturated polytomous cognitive diagnosis model* (sp-CDM) under the identity, logit, and log link functions, respectively. The number of parameters for item j for the three models is equal to the number of latent groups (i.e., $M^{K_j^*}$). Thus, the models offer much greater generality compared to the existing CDMs for polytomous attributes. Although flexible, the large number of parameters in these models can make their estimation challenging. Therefore, simpler and more interpretable models with fewer parameters are sometimes warranted. Note that the number of parameters for the saturated models does not take into account the required attribute levels - it is computed as the product of the maximum levels of the required attributes. Thus, in the example above, in addition to $(1, 2, 0)$, the q-vectors $(1, 1, 0)$, $(2, 1, 0)$, and $(2, 2, 0)$ will result in the same saturated models.

3. Special Cases

This section introduces several simplified CDMs for polytomous attributes with different assumptions, namely, the conjunctive, disjunctive, and additive assumptions, and how they can be derived from the sp-CDM by imposing appropriate constraints.

The Conjunctive version of the sp-CDM

In the conjunctive version of the sp-CDM (conj-sp-CDM), it is assumed that examinees should possess levels that are equal to or higher than the required levels in *all* the required attributes are expected to answer the item correctly. Alternatively, persons who lack at least one of the required attributes, or possess levels lower than those required in at least one of the

required attributes are expected to answer the item incorrectly. Hence, the IRF of the conj-sp-CDM can be expressed as

$$P_j(\boldsymbol{\alpha}_l^*) = \begin{cases} g_j & \text{if } I[\{\boldsymbol{\alpha}_l^* \geq \mathbf{q}_j\}] \prec \mathbf{1}_{K_j^*}, \\ 1 - s_j & \text{otherwise}, \end{cases} \tag{4}$$

where the symbol $\{\}$ insides $I[\cdot]$ indicates that the operation be carried out attribute by attribute and $I[\cdot] = 1$ if $\boldsymbol{\alpha}_l^* \geq \mathbf{q}_j$ and 0 otherwise. $\mathbf{1}_{K_j^*}$ is a vector of ones and of length K_j^*. The $\prec$ symbol indicates with respect to K_j^* required attributes, which is a partially order set of K attributes, at least one of the elements in the results of $I[\cdot]$ is less than 1. As shown in Equation 4, the conj-sp-CDM has two parameters for item j.

On the surface, the formulation of the conj-sp-CDM is similar to that of the DINA model. However, the parameters in Equation 4 require more complicated interpretations. For example, the g_j in the equation is the probability of correctly answering item j for individuals who lack at least one of the prescribed attributes, *or* who possess the required attributes, but with levels lower than required levels in the prescribed attributes; the $1 - s_j$ represents the probability of individuals who have attribute levels that are all at least equal to the required levels answering the item incorrectly. Thus, in the conj-sp-CDM model, the $M^{K_j^*}$ attribute vectors are classified into two latent groups– attribute vectors that jointly satisfy the required levels prescribed for item are classified in one group, and the rest of the attribute vectors in the other group.

The conj-sp-CDM model can be derived from the identity sp-CDM (i.e., Equation 1) by imposing the following three constraints:

(1) All the $(M - 1)K_j^*$ main effects are equal to zero, as in, $\delta_{jkm} = 0$ for $k = 1, \ldots, K_j^*$ and $m = 1, \ldots, M - 1$;

(2) All the $\prod_{k=1}^{K_j^*}(M - q_k)$ interaction terms involving attribute levels at least equal to the required attribute levels are identical, as in, $\delta_{j1m_1^{(+)}2m_2^{(+)}\ldots K_j^*m_{K_j^*}^{(+)}} = \delta_{j1q_12q_2\ldots K_j^*q_{K_j^*}}$, where $m_k^{(+)} = q_k, \cdots, M - 1$; and

(3) The remaining $M^{K_j^*} - (M - 1)K_j^* - \prod_{k=1}^{K_j^*}(M - q_k) - 1$ interaction effects involving at least one attribute level below the required level are all equal to zero.

With the conjunctive assumptions imposed on the identity sp-CDM, the IRF of the

conj-sp-CDM can also be expressed as:

$$P_j(\boldsymbol{\alpha}_l^*) = \delta_{j0} + \delta_{j1q_12q_2\ldots K_j^*q_{K_j^*}} \prod_{k=1}^{K_j^*} I\big[\alpha_{lk} \geq q_{jk}\big]. \tag{5}$$

The Disjunctive version of the sp-CDM

In the disjunctive version of the sp-CDM (disj-sp-CDM), the IRF is given by

$$P_j(\boldsymbol{\alpha}_l^*) = \begin{cases} 1 - s_j' & \text{if } I[\{\boldsymbol{\alpha}_l^* \geq \mathbf{q}_j\}] \succ \mathbf{0}_{K_j^*}, \\ g_j' & \text{otherwise,} \end{cases} \tag{6}$$

where $1 - s_j'$ is the probability of not slipping for persons who posses levels that are equal to or higher than the required levels in at least one required attribute, and g_j' is the success probability that persons who possess none of required attributes, or possess at least one of the required attributes, but all of which have the levels that lower than the required levels. As such, the disj-sp-CDM has two parameters for each item.

The disj-sp-CDM can be derived from the identity sp-CDM with the following constraints:

(1) The main effects $\delta_{jkm^{(-)}} = 0$ and $\delta_{jkm^{(+)}} = \delta_{jkq_k}$, where $m_k^{(-)} = 1, \cdots, q_k - 1$ represents levels of attribute k that are lower than the required level q_k, and $m_k^{(+)}$ is as defined as above.

(3) The remaining interaction effects (i.e., those involving at least one $m_k^{(-)}$) are of equal to zero.

With these assumptions, the identity sp-CDM can be reduced to be the disj-sp-CDM as:

$$P_j(\boldsymbol{\alpha}_l^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jkq_k} I\big[\alpha_{lk} \geq q_k\big] + \sum_{k'>k}^{K_j^*} \sum_{k=1}^{K_j^*} \delta_{jkq_kk'q_{k'}} I\big[\alpha_{lk} \geq q_k\big] I\big[\alpha_{lk'} \geq q_{k'}\big] + \\ \cdots + \delta_{j1q_12q_2\cdots K_j^*q_{K_j^*}} I\big[\alpha_{l1} \geq q_1\big] I\big[\alpha_{l2} \geq q_2\big] \cdots I\big[\alpha_{lK_j^*} \geq q_{K_j^*}\big]. \tag{7}$$

The Fully Additive Model for Polytomous Attributes

The fully additive model for polytomous attributes (fA-M; Yakar et al., 2021) assumes that mastering level m of required attribute k increases the success probability on item j by δ_{jkm}, and its contribution is independent of the contributions of the levels of the other other attributes. By

retaining only the main effects in Equation 1, the fA-M can be obtained, which has the following IRF:

$$P_j(\boldsymbol{\alpha}_l^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \sum_{m=1}^{M-1} \delta_{jkm} I[\alpha_{lk} = m]. \tag{8}$$

The fA-M has $K_j^*(M-1)+1$ parameters for item j. As with the saturated models, the number of parameters the fA-M does not depend on the required levels, only the maximum levels of the required attributes. Using the required level q_k as a cutoff value, two simplified and interpretable models can be obtained from the fA-M. Specifically, in the first simplified fA-M, mastering an attribute level that is higher than q_k will contribute to higher probability of answering an item correctly, whereas mastering those that are lower than q_k will have equal success probability. Hence, q_k serves as a minimum bar and will be denoted as min-fA-M. The IRF for the min-fA-M can be expressed as:

$$P_j(\boldsymbol{\alpha}_l^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} (\delta_{jk1_k} I[\alpha_{lk} < q_{jk}] + \delta_{jkm} I[\alpha_{lk} \geq q_{jk}]), \tag{9}$$

where δ_{jk1_k} represents the main effect of $m=1$ in attribute k. The number of parameters for item j in the min-fA-M reduces to $\sum_k^{K_j^*}(M - q_k + 1) + 1$.

In contrast to the min-fA-M, the second simplified fA-M assumes q_k is a maximum requirement and will be denoted as max-fA-M. The model is similar to the GDM in that the success probabilities for attribute levels that are higher than q_k are equal to each other in both models, which equal to the probability of level q_k. However, those of levels that are lower than q_k are different. The lower the level, the lower the success probability. The IRF for the max-fA-M can be expressed as:

$$P_j(\boldsymbol{\alpha}_l^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} (\delta_{jkm} I[\alpha_{lk} < q_k] + \delta_{jkq_k} I[\alpha_{lk} \geq q_k]). \tag{10}$$

where δ_{jkq_k} is the main effect of level $m = q_k$ in attribute k. The max-fA-M has $\sum_k^{K_j^*} q_k + 1$ parameters for item j.

4. The Connections between the sp-CDM and the Existing Models

This section shows, both mathematically and graphically, the connections between the sp-CDM and the existing models. Specifically, the existing CDMs for polytomous attributes can be formulated as special cases of the sp-CDM using different functions. A diagram (i.e., Figure 1) and a table (i.e., Table 2) help illustrate the connections between the sp-CDM and the existing models, and those among the existing models.

4.1 The identity model: pG-DINA and fA-M

The pG-DINA model can be obtained from the sp-CDM by replacing $I\big[\alpha_{l\cdot\cdot} = m\big]$ with $I\big[\alpha_{l\cdot} \geq q_{jk}\big]$. With respect to the fA-M, as shown in Equation 8, it is the identity sp-CDM that retains the main effects only.

4.2 The logit model: PDCM and GDM-PA

In the PDCM framework (Bao, 2019), the logit of the success probably in answering item j correctly is given by

$$
\log\frac{P(X_j = 1|\tilde{\boldsymbol{\alpha}})}{P(X_j = 0|\tilde{\boldsymbol{\alpha}})} = \lambda_{j0} + \sum_{k=1}^{K}\sum_{m=1}^{M-1} \lambda_{jk}^{m}\tilde{\alpha}_k^{m}q_{jk} +
$$
$$
\sum_{k=1}^{K-1}\sum_{m=1}^{M-1}\sum_{k'=k+1}^{K}\sum_{m'=1}^{M-1} \lambda_{jkk'}^{mm'}\tilde{\alpha}_k^{m}\tilde{\alpha}_{k'}^{m'}q_{jk}q_{jk'} + \cdots, \tag{11}
$$

where $\tilde{\alpha}_k^{m}$ is the dummy variable for level m of attribute k, q_{jk} is equal to 1 if kth attribute is measured by item j. λ_{jk}^{m} is the main effect of level m for attribute k and $\lambda_{jkk'}^{mm'}$ is the two-way interaction effect for level m of attribute k and level m' of attribute k'. As mentioned earlier, q_{jk} in the PDCM are still binary values. Hence, the PDCM is a special case of the logit sp-CDM.

In the GDM (von Davier, 2008), the IRF can be expressed as

$$
\text{logit}\left[P(X_j = 1|\beta, q_k, \gamma, \alpha_k)\right] = \beta + \gamma^{T}h(q_k, \alpha_k). \tag{12}
$$

As mentioned earlier, for polytomous attributes with $q_k \in \{0, 1, 2, \cdots, m\}$ and $\alpha_k \in \{0, 1, 2, \cdots, m\}$, a useful and reasonable choice of $h(\cdot)$ in GDM is defined as

$$
h(q_k, \alpha_k) = \begin{cases} q_k & \text{if } \alpha_k > q_k \\ \alpha_k & \text{if } \alpha_k \leq q_k \end{cases} \tag{13}
$$
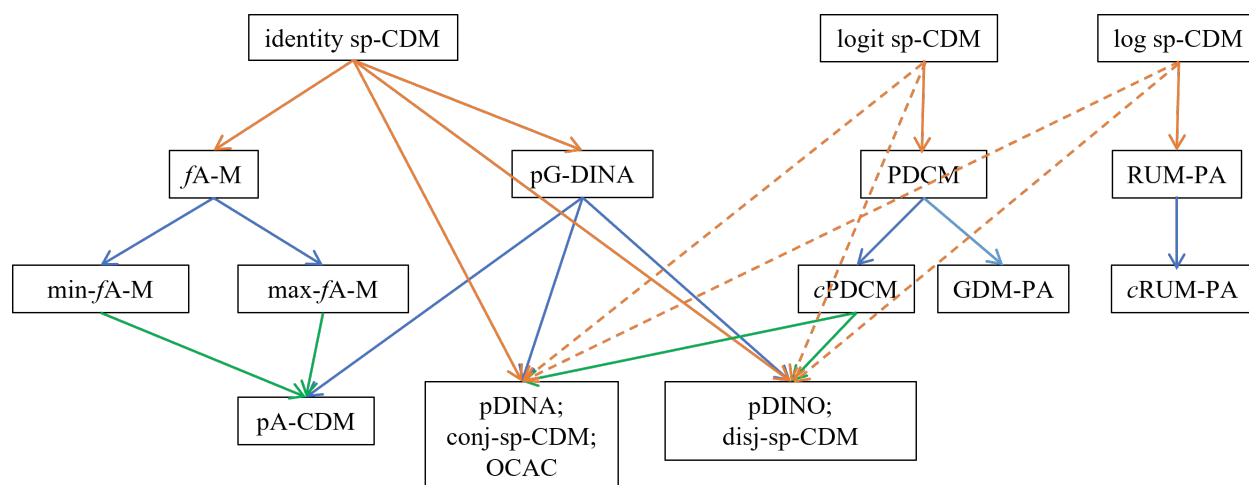
FIGURE 1.

The generalized cognitive diagnosis model framework for polytomous attributes

Note: sp-CDM: saturated polytomous cognitive diagnosis model; fA-M: fully additive model for polytomous attributes; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption; PDCM: saturated polytomous-attribute diagnostic classification model; RUM-PA: reparameterized unified model for polytomous attributes; min-fA-M: fA-M using q_k as a minimum requirement; max-fA-M: fA-M using q_k as a maximum requirement; cPDCM: constrained PDCM; GDM-PA: general diagnostic model for polytomous attributes; cRUM-PA: constrained RUM-PA; pA-CDM: additive model for polytomous attributes; pDINA: deterministic input, noisy "and" gate model for polytomous attributes; pDINO: deterministic input, noisy "or" gate model for polytomous attributes; conj-sp-CDM: conjunctive version of sp-CDM; disj-sp-CDM: disjunctive version of sp-CDM; OCAC: ordered category attribute coding framework. The colors orange, blue and green can be interpreted as the number of steps (i.e., 1, 2 and 3 steps) for the reduced models to be derived from the saturated model of a particular link function. For example, the pA-CDM can be derived from the identity sp-CDM through pG-DINA (two steps) or through fA-M then either min-fA-M or max-fA-M (three steps). The dashed lines indicate that the reduced models can also be shown to be special cases of sp-CDM with logit or log link functions.

To this end, Equation 12 is equivalent to

$$\text{logit}\left[P(X_j = 1|\beta, q_k, \gamma, \alpha_k)\right] = \beta + \sum_{k=1}^{K_j^*} \sum_{m=1}^{M} \gamma_{jk} \cdot h(q_k, \alpha_k), \tag{14}$$

where γ_{jk} is the increase in the logit of the probability of success for every level of α_k mastered up to the required level (i.e., q_k).

4.3 The log model: RUM-PA

In the polytomous attribute RUM (Templin, 2004) with $q_k \in \{0,1\}$ and $\alpha_k \in \{0, 1, 2, \cdots, m\}$, the success probability is given by:

$$P(X_j = 1|q_k, \alpha_k, \pi^*, r_k^*) = \pi^* \prod_{k=1}^{K} (r_k^*)^{f_k(q_k, \alpha_k)}, \tag{14}$$

where

$$f(q_k, \alpha_k) = \begin{cases} 1 & \text{if } q_k = 1, \alpha_k = 0 \\ 0 & \text{if } q_k = 1 \leq \alpha_k = m \end{cases} \tag{15}$$

Like in the PDCM, the Q-matrices in the RUM-PA are assumed to be dichotomous, whereas the person attributes are assumed to be polytomous. Equation 14 can be written with respect to different attribute level m as:

$$P(X_j = 1|q_k, \alpha_k, \pi^*, r_k^*) = \pi^* \prod_{k=1}^{K} \prod_{m=1}^{M} (r_k^*)^{f_{km}(q_{km}, \alpha_{km})}. \tag{16}$$

Equation 16 can be rewritten using α_k^* with $q_k = 1$ as:

$$P(X_j = 1|q_k, \alpha_k^*, \pi^*, r_k^*) = \pi^* \prod_{k=1}^{K} \prod_{m=1}^{M} r_{km}^* \times \prod_{k=1}^{K} \prod_{m=1}^{M} \left(\frac{1}{r_{km}^*}\right)^{\alpha_{km}}. \tag{17}$$

Thus,

$$\log\left[P(X_j = 1|q_k, \alpha_k^*, \pi^*, r_k^*)\right] = \log(\pi^*) + \sum_{k=1}^{K} \sum_{m=1}^{M} \log(r_{km}^*) - \sum_{k=1}^{K} \sum_{m=1}^{M} \log(r_{km}^*)\alpha_{km}. \tag{18}$$

By setting $\log(\pi^*) + \sum_{k=1}^{K} \sum_{m=1}^{M} \log(r_{km}^*) = \nu_0$ and $-\sum_{k=1}^{K} \sum_{m=1}^{M} \log(r_{km}^*) = -\sum_{k=1}^{K} \sum_{m=1}^{M} \nu_{km}$, Equation 18 is a special case of log sp-CDM without the interaction terms.

Table 2.

The relationships between sp-CDMs and the existing CDMs for polytomous attributes

Existing CDMs	Constraints on the sp-CDM	Special Cases
pG-DINA	sp-CDM with $I\left[\alpha_{l.} \geq q_{jk}\right]$	OCAC
OCAC	pG-DINA without interaction terms	—
fA-M	identity sp-CDM with main effects of levels	—
PDCM	logit sp-CDM	cPDCM; GDM-PA
GDM-PA	logit sp-CDM without interaction terms	—
RUM-PA	log sp-CDM with $\log(\pi^*) + \sum_{k=1}^{K} \sum_{m=1}^{M} \log(r_{km}^*) = \nu_0$ and $-\sum_{k=1}^{K} \sum_{m=1}^{M} \log(r_{km}^*) = -\sum_{k=1}^{K} \sum_{m=1}^{M} \nu_{km}$	cRUM-PA

Note. sp-CDM: saturated polytomous cognitive diagnosis models; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption; OCAC: ordered category attribute coding framework; fA-M: fully additive model for polytomous attributes; PDCM: saturated polytomous-attribute diagnostic classification models; cPDCM: constrained PDCM; GDM-PA: general diagnostic model for polytomous attributes; RUM-PA: reparameterized unified model for polytomous attributes; cRUM-PA: constrained RUM-PA.

5. Model Estimation

The estimation algorithms for the parameters, the corresponding standard errors (SEs), and the person attribute patterns in the sp-CDM model are primarily similar to those in estimating the parameters of the G-DINA model described in de la Torre (2011) and those of the DINA model in de la Torre (2009). Specifically, parameters for the saturated forms of the sp-CDM can be estimated via the marginal maximum likelihood estimation method with an expectation-maximization (MMLE/EM), and those of reduced models can be estimated by incorporating appropriate design matrix in the MMLE/EM procedure. Details can be found in Appendix A.

6. Simulation Study

Design

Two research questions were investigated in the simulation study. First, how well can the parameters of the sp-CDM be recovered with the proposed estimation algorithms? And second, how does the fit of the sp-CDM compare with those of constrained and simplified models across different data generation assumptions. Due to time and space constraints, we focus on the identity sp-CDM and its two special cases, namely, the pG-DINA and fA-M, in this study. The details of the design are summarized in Table 3. The levels for the manipulated factors followed previous studies (e.g., de la Torre et al., 2021). In particular, the levels of item quality, defined as a function of the guessing and slip parameters, were computed as p_0 and $(1 - p_1)$. Specifically, items with $(p_0, 1 - p_1) \in U(.05, .15)$, $U(.10, .20)$, and $U(.15, .30)$ were classified as high, moderate, and low quality items, respectively. To this end, the generating values for the intercept parameters were set to be .05, .10, and .15 under the three types of quality items. For the main effect parameters, the mean of the generating values are .16, .15, and .13, respectively, and for the interaction effect parameters, they are .1, .08, and .07, respectively. Attribute patterns were generated from a uniform distribution where all possible attribute patterns were equally likely. The Q-matrix for $K = 3$ and $K = 5$ with 26 items are shown in Tables 4 and B.1, respectively, and the Q-matrix with 52 items are duplicate of respective Q-matrices. The Q-matrix in the

simulation study was specified to satisfy the sufficient conditions similar to Theorem 4 in Fang et al. (2019). The GDINA package (Ma & de la Torre, 2019) and a customized program were used to generate the data and estimate the models. The monotonic constraints were imposed when estimating the models. A total of 168 conditions were examined and each condition was replicated 100 times.

To answer the research questions, the simulation study was carried out in three steps. Step 1 was designed to answer the first research question. In this step, data under different conditions were generated following the sp-CDM (i.e., Equation 1) and fitted with the true model (i.e., the sp-CDM). Steps 2 and 3 were designed to answer the second research question. In the second step, the pG-DINA model and fA-M were also fitted to the generated data in step 1 to investigate the consequences of fitting reduced models when neither the SALM nor the additive assumption holds. In step 3, two sets of data were generated following the pG-DINA model and fA-M, and fitted with sp-CDM, as well as their respective true model to investigate the consequences of fitting an unnecessarily complicated model (i.e., the sp-CDM) when the SALM or the additive assumption holds. For the sake of discussion, step 1 is referred to as the parameter recovery study, whereas steps 2 and 3 as the comparison study.

Evaluation Criteria

The dependent variables in the parameter recovery study were the bias and root mean square error (RMSE) of the estimated success probabilities of the reduced attribute patterns in item j (denoted as $\hat{P}_j(\boldsymbol{\alpha}_l^*)$), and were defined as

$$\text{Bias}\left(\hat{P}_j(\boldsymbol{\alpha}_l^*)\right) = \sum_{j=1}^{J}\sum_{l=1}^{L_j^*}\left[\bar{\hat{P}}_j(\boldsymbol{\alpha}_l^*) - P_j(\boldsymbol{\alpha}_l^*)\right]/\sum_{j=1}^{J}L_j^* \tag{19}$$

and

$$\text{RMSE}\left(\hat{P}_j(\boldsymbol{\alpha}_l^*)\right) = \sqrt{\sum_{j=1}^{J}\sum_{l=1}^{L_j^*}\sum_{r=1}^{R}\left[\hat{P}_j^{(r)}(\boldsymbol{\alpha}_l^*) - P_j(\boldsymbol{\alpha}_l^*)\right]^2/\left[R\times\sum_{j=1}^{J}L_j^*\right]}, \tag{20}$$

respectively, where $P_j(\boldsymbol{\alpha}_l^*)$ is the generating probability of $\boldsymbol{\alpha}_l^*$ in item j, $\hat{P}_j^{(r)}(\boldsymbol{\alpha}_l^*)$ is the estimate of $\hat{P}_j(\boldsymbol{\alpha}_l^*)$ in the rth replication, $\bar{\hat{P}}_j(\boldsymbol{\alpha}_l^*)$ is the mean of $\hat{P}_j(\boldsymbol{\alpha}_l^*)$ across R replications, and L_j^* is the

TABLE 3.

Summary of the simulation design

Factor	Level
M (attribute levels)	3
N (sample size)	1,000, 2,000
K (number of attributes)	3, 5
J (number of items)	26, 52
Item Quality	High, Moderate, Low
True/Fitted model	sp-CDM/sp-CDM
	sp-CDM/pG-DINA, sp-CDM/fA-M
	pG-DINA/sp-CDM, pG-DINA/pG-DINA,
	fA-M/sp-CDM, fA-M/fA-M

Note. sp-CDM: saturated polytomous cognitive diagnosis models; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption; fA-M: fully additive model for polytomous attributes.

TABLE 4.

Q-matrix for conditions of three attributes in simulation study

Item	α_1	α_2	α_3	Item	α_1	α_2	α_3	Item	α_1	α_2	α_3
1	1	0	0	11	1	0	1	21	1	2	1
2	0	1	0	12	1	0	2	22	1	2	2
3	0	0	1	13	2	0	1	23	2	1	1
4	2	0	0	14	2	0	2	24	2	1	2
5	0	2	0	15	0	1	1	25	2	2	1
6	0	0	2	16	0	1	2	26	2	2	2
7	1	1	0	17	0	2	1				
8	1	2	0	18	0	2	2				
9	2	1	0	19	1	1	1				
10	2	2	0	20	1	1	2				

number of attribute patterns in item j.

In the comparison study, the dependent variables were the proportion of correctly classified attributes (PCA) and vectors (PCV), which were computed as:

$$\text{PCA}_r = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{m=1}^{M} I[\hat{\alpha}_{nkm} = \alpha_{nkm}]}{N \times K}, \tag{21}$$

and

$$\text{PCV}_r = \frac{\sum_{n=1}^{N} I[\hat{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_n]}{N}, \tag{22}$$

respectively, where $I[\hat{\alpha}_{nkm} = \alpha_{nkm}]$ was used to evaluate the match between the estimated and generated attribute in the rth replication, and $I[\hat{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_n]$ to attribute vectors. For both studies, the results are summarized using the average values of the variables across replications.

Results

For $K = 3$, the bias and RMSE under different conditions are shown in Figures 2 and 3, respectively, and PCA and PCV under the high, moderate, and low quality item conditions in Tables 5, 6, and 7, respectively. In particular, the upper panels of Figures 2 and 3 give the biases and RMSEs from fitting the sp-CDM to the sp-CDM data under different numbers of attributes, item qualities, test lengths, and sample sizes. The figures show that the parameters of the sp-CDM can be well recovered with the proposed estimation algorithms, particularly when high quality items were involved. For example, the mean biases were between -0.007 and 0.000 and the mean RMSEs were between 0.000 and 0.009 across all conditions. The upper panels of Tables 5 to 7 reveal that the classification of attributes and vectors are satisfactory under the sp-CDM. For example, for the high quality item conditions (i.e., Table 5), the PCAs were between 88.2% and 96.9%, and the PCVs were between 71.9% and 89.3%.

In comparison, a closer inspection of the panels of Figures 2 and 3 and Tables 5 to 7 reveals that fitting the reduced models, either the fA-M or the pG-DINA, to the sp-CDM data resulted in larger biases and RMSEs, or lower PCAs and PCVs, particularly when high quality items were used. Dramatically different results were obtained when the sp-CDM was fitted to the data generated using the reduced models, the biases, RMSEs, PCAs and PCVs were similar to those

obtained when the corresponding true models were fitted. These results can have important practical implications - when it is unclear which is the true model for an item, it is safer to fit the more general sp-CDM, rather than a particular reduced model.

Due to space constraints, the results for $K = 5$ are give in Appendix B, where Figures B.1 and B.2 contain the bias and RMSE, respectively, and Tables B.2, B.3, and B.4 the PCA and PCV under the three item quality conditions, respectively. in general, the findings for $K = 5$ were similar to those for $K = 3$. However, it should be noted that for two conditions in Figure B.2 (i.e., $K = 5$, $J = 26$, $N = 1000$, and items are either of low or moderate quality), the sp-CDM produced larger RMSEs than f-AM even though the data were generated from the sp-CDM. This could be attributed the instability of estimating the sp-CDM when a large number of parameters are involved and the data are not sufficiently informative.
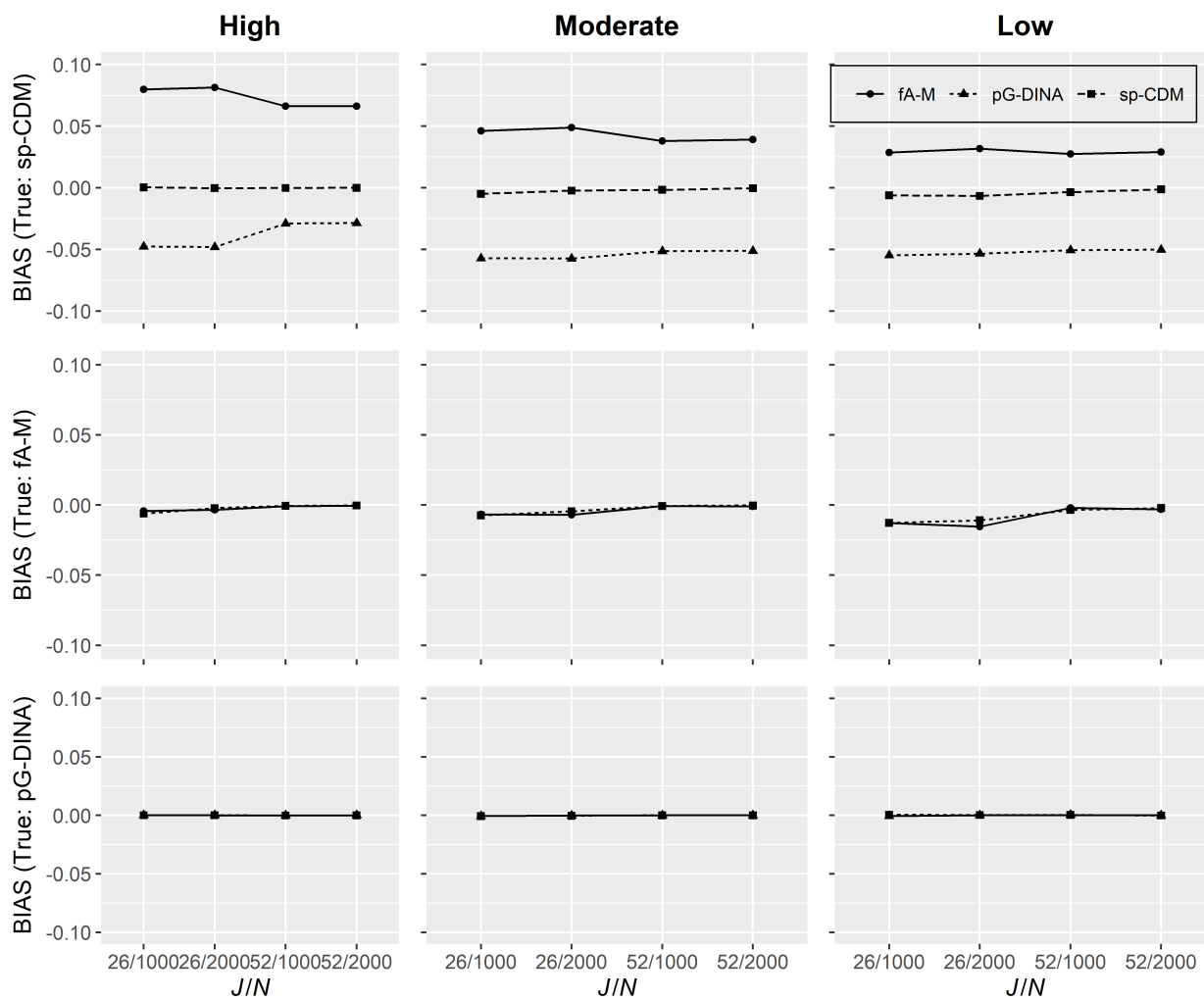
FIGURE 2.

Bias in parameter recovery with three attributes

Note: sp-CDM: saturated polytomous cognitive diagnosis models; fA-M: fully additive model for polytomous attributes; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption. J: test length; N: sample size
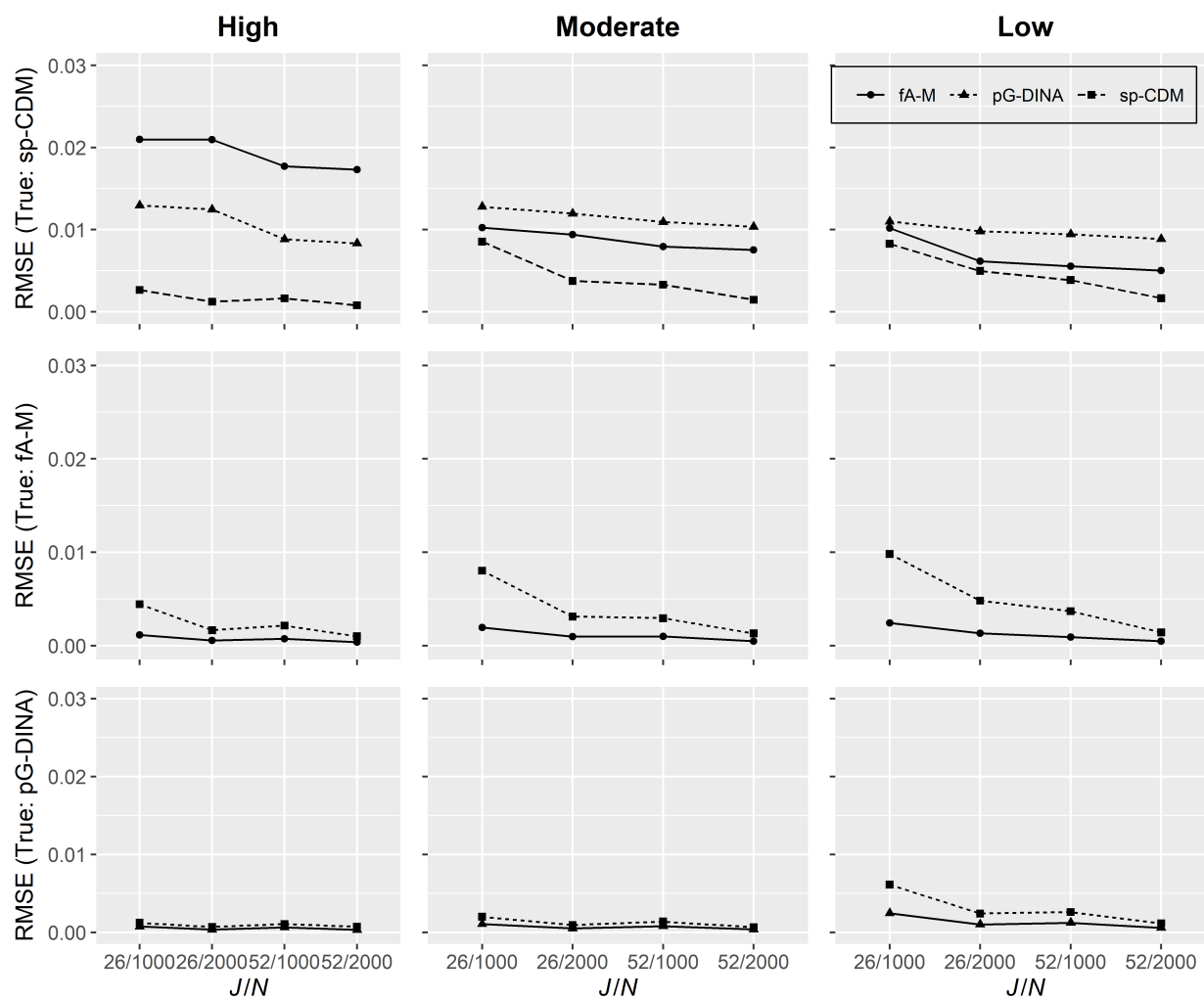
FIGURE 3.

Root mean square error (RMSE) in parameter recovery with three attributes

Note: sp-CDM: saturated polytomous cognitive diagnosis models; fA-M: fully additive model for polytomous attributes; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption. J: test length; N: sample size

TABLE 5.

Correctly classified attributes (PCA) and vectors (PCV) (in %) with three attributes and high quality items

| | | Test Length/Sample Size | | | | | | | | | | | |
| | | 26/1000 | | | 26/2000 | | | 52/1000 | | | 52/2000 | | |
Gen	EC	1	2	3	1	2	3	1	2	3	1	2	3
sp-CDM	PCA1	90.6	86.3	87.2	91.0	86.2	87.3	96.6	91.7	95.4	96.9	91.6	95.4
	PCA2	88.9	84.3	82.9	89.4	84.5	82.7	95.7	90.8	92.4	96.1	91.1	92.8
	PCA3	88.2	82.5	82.2	88.8	82.6	82.6	95.4	89.7	92.2	95.7	90.2	92.4
	PCV	71.9	58.3	58.9	73.2	58.3	59.2	88.4	73.2	81.4	89.3	73.8	81.9
fA-M	PCA1	89.7	90.6	81.3	90.6	90.9	81.6	96.5	96.7	88.2	96.7	96.8	88.3
	PCA2	86.0	87.4	77.9	87.2	87.7	77.7	94.1	94.6	79.4	94.5	94.8	79.4
	PCA3	84.2	86.3	77.3	85.9	86.5	77.2	93.1	93.8	77.7	93.8	94.1	77.9
	PCV	65.4	69.1	47.4	68.5	69.7	47.3	84.8	86.2	51.8	86.0	86.6	52.0
pG-DINA	PCA1	96.8	95.1	97.0	97.0	95.2	97.1	99.5	99.0	99.5	99.5	99.1	99.6
	PCA2	96.1	94.2	96.3	96.3	94.3	96.4	99.3	98.9	99.4	99.3	98.9	99.4
	PCA3	95.5	93.4	95.6	95.6	93.5	95.7	99.1	98.5	99.2	99.1	98.6	99.2
	PCV	89.5	83.9	90.0	89.9	84.3	90.2	97.9	96.5	98.2	98.1	96.6	98.2

Note. Gen: Generating model; EC: Evaluation criteria; 1: sp-CDM; 2: fA-M; 3: pG-DINA; PCAk: PCA of attribute k

<p style="text-align:center">TABLE 6.</p>

<p style="text-align:center">Correctly classified attributes (PCA) and vectors (PCV) (in %) with three attributes and moderate quality items</p>

| | | Test Length/Sample Size | | | | | | | | | | | |
| | | 26/1000 | | | 26/2000 | | | 52/1000 | | | 52/2000 | | |
Gen	EC	1	2	3	1	2	3	1	2	3	1	2	3
sp-CDM	PCA1	79.1	78.2	75.9	79.9	78.4	76.2	89.3	87.7	83.2	90.2	87.8	83.2
	PCA2	76.5	75.5	73.4	77.4	76.2	74.1	87.1	85.5	79.6	88.3	86.0	80.2
	PCA3	75.0	74.7	74.0	76.7	75.0	74.4	86.6	84.9	81.3	87.9	85.0	81.2
	PCV	44.9	43.0	40.9	48.5	43.7	41.6	68.0	62.9	52.2	70.9	63.3	52.6
fA-M	PCA1	83.7	85.7	79.5	85.2	86.0	79.7	93.6	94.2	87.0	94.1	94.4	87.0
	PCA2	79.4	82.1	76.0	81.2	82.6	76.5	90.5	91.6	81.7	91.4	91.9	82.2
	PCA3	75.2	78.8	75.3	77.6	79.3	75.4	87.2	88.8	81.0	88.5	89.2	81.2
	PCV	50.4	56.4	45.1	54.5	57.2	45.7	74.4	77.3	56.3	76.7	78.0	56.7
pG-DINA	PCA1	92.5	90.0	92.9	92.6	89.9	92.8	97.9	96.7	98.1	98.1	96.8	98.1
	PCA2	91.5	88.4	92.0	91.8	88.7	92.1	97.6	96.3	97.8	97.8	96.3	97.9
	PCA3	90.6	87.1	91.1	91.1	87.7	91.4	97.1	95.6	97.3	97.4	95.8	97.5
	PCV	78.4	69.5	79.5	79.1	70.3	79.7	93.2	89.3	93.7	93.8	89.7	94.0

Note. Gen: Generating model; EC: Evaluation criteria; 1: sp-CDM; 2: fA-M; 3: pG-DINA; PCAk: PCA of attribute k

TABLE 7.

Correctly classified attributes (PCA) and vectors (PCV) (in %) with three attributes and low quality items

| | | Test Length/Sample Size | | | | | | | | | | | |
| | | 26/1000 | | | 26/2000 | | | 52/1000 | | | 52/2000 | | |
Gen	EC	1	2	3	1	2	3	1	2	3	1	2	3
sp-CDM	PCA1	73.2	71.8	72.1	74.0	73.9	72.3	84.2	83.6	79.8	85.4	83.9	80.1
	PCA2	70.1	69.2	68.8	71.2	70.8	69.9	81.2	80.8	76.1	82.6	81.5	76.7
	PCA3	67.1	66.0	66.2	67.9	67.0	67.8	77.3	77.2	74.1	79.2	77.3	74.2
	PCV	33.4	33.0	33.1	36.3	34.5	34.2	53.2	51.8	43.7	56.9	52.4	44.3
fA-M	PCA1	76.7	79.4	74.6	78.3	79.9	74.8	88.4	89.6	81.0	89.3	89.8	81.2
	PCA2	73.3	74.9	71.7	73.4	75.6	72.3	83.7	85.9	78.8	85.2	86.1	78.9
	PCA3	69.7	70.5	69.3	68.8	71.6	69.6	77.8	81.3	75.5	80.4	81.8	75.9
	PCV	43.0	43.1	37.5	40.2	44.2	37.9	58.2	63.6	47.6	62.0	64.3	48.0
pG-DINA	PCA1	80.6	73.8	82.4	82.0	74.3	82.9	90.9	83.7	91.9	91.6	83.9	92.1
	PCA2	80.5	75.3	82.2	81.9	76.4	82.8	91.0	86.4	91.8	91.7	86.9	92.1
	PCA3	80.6	74.5	82.4	82.1	75.1	83.0	91.1	84.6	91.9	91.6	84.1	92.0
	PCV	56.5	39.9	60.0	59.2	41.1	61.0	77.4	60.3	79.4	79.0	60.5	79.9

Note. Gen: Generating model; EC: Evaluation criteria; 1: sp-CDM; 2: fA-M; 3: pG-DINA; PCAk: PCA of attribute k

One reviewer noted that the RMSE results from both the $K = 3$ and $K = 5$ conditions for the sp-CDM and fAM tend to yield decreased RMSEs when tests contain more items and use smaller samples. For example, in the top panel of Figure 3, the RMSE results under the condition $J52/N1000$ are smaller than those for the condition $J26/N2000$. To investigate this phenomena, an additional simulation was conducted where the dataset was generated from the sp-CDM with $K = 3$ and high item quality for the two conditions (i.e., $J52/N1000$ and $J26/N2000$). Three analyses were carried out as follows: In the first analysis, for the condition of $J52/N1000$, the true attribute patterns for each person were assigned a posterior probability of 0.950, and the remaining 26 patterns for the person a probability of $(1 - 0.950)/26 \approx 0.002$. In contrast, for the condition of $J26/N2000$, the true attribute patterns for each person were assigned a posterior probability of 0.700, and other patterns a probability of $(1 - 0.700)/26 \approx 0.011$. This analysis mimics a scenario where the attribute patterns are better estimated in a smaller sample size condition than in a larger one. In the second analysis, the setting for the patterns' posterior probabilities were reversed for the two conditions to mimic a scenario where the attribute patterns are better estimated in a larger sample size than in a smaller one. Finally, the patterns' posterior probabilities were set using the results of the original simulation study. Specifically, the mean of the posterior probabilities for the true patterns across persons was computed for the two conditions–which are 0.601 and 0.456 for the conditions of $J52/N1000$ and $J26/N2000$, respectively–and are used in the third analysis.

The biases and RMSEs for the additional simulation study are shown in Table 8. It was found that, for each analysis, the test that was assigned more accurate attribute pattern estimates resulted in better item parameter estimates (i.e., smaller bias and RMSE), even when the sample size was supposedly smaller. For example, in the first and third analyses, the biases and RMSEs under $J52/N1000$ are smaller than those under $J26/N2000$. These results demonstrate how more informative (i.e., longer) tests calibrated with a smaller sample size can produce better item parameter estimates.

TABLE 8.

Additional simulation study: Bias and root mean square error (RMSE)

| Analysis | $J52/N1000$ | | $J26/N2000$ | |
	Bias	RMSE	Bias	RMSE
First	0.003	0.038	0.020	0.101
Second	0.020	0.103	0.003	0.029
Third	0.026	0.134	0.036	0.180

Note. In the first analysis, the posterior probabilities for the true attribute patterns are set to be 0.95 and 0.70 for the conditions of $J52/N1000$ and $J26/N2000$, respectively. In the second analysis, they are 0.70 and 0.95 for the two conditions, respectively. In the third analysis, the probabilities are 0.601 and 0.456, respectively.

7. Real Data Example

Data and Analysis

The responses in this example consisted of 1,408 middle school students in Hong Kong to a PR assessment described earlier. The assessment uses 31 multiple-choice items measuring six PR attributes, namely, (1) prerequisite skills and concepts required in proportional reasoning, (2) comparing and ordering fractions, (3) constructing ratios and proportions, (4) identifying a multiplicative relationship between sets of values, (5) differentiating a proportional relationship from a non-proportional relationship, and (6) applying algorithms in solving proportional reasoning problems, among which, the second and third attributes are polytomous with $M = 3$ and other attributes are dichotomous. The Q-matrix for the empirical example is provided in Table 9, where each item requires one to four attributes (i.e., $1 \leq K_j^* \leq 4$). This Q-matrix satisfies the identifiability conditions by Fang et al. (2019). The sp-CDM, fA-M, and pG-DINA model were fitted to the data. No monotonicity constraint was imposed in this analysis. The deviance, Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare the three fitted models.

TABLE 9.

Q-matrix for the proportional reasoning data and the number of parameters under the sp-CDM, fAM, and pG-DINA model

Item	α_1	α_2	α_3	α_4	α_5	α_6	#Par			Item	α_1	α_2	α_3	α_4	α_5	α_6	#Par		
							M_1	M_2	M_3								M_1	M_2	M_3
1	1	0	0	0	0	1	4	3	4	17	1	2	0	0	0	0	6	4	4
2	1	1	1	0	0	0	18	6	8	18	1	0	2	0	1	1	24	6	16
3	1	0	2	1	1	0	24	6	16	19	1	0	2	1	1	0	24	6	16
4	1	1	1	0	0	0	18	6	8	20	0	0	2	1	1	0	12	5	8
5	1	1	1	0	0	0	18	6	8	21	1	1	1	0	0	0	18	6	8
6	1	0	2	1	1	0	24	6	16	22	1	2	0	0	0	0	6	4	4
7	1	0	0	0	0	1	4	3	4	23	1	1	1	0	0	0	18	6	8
8	1	2	1	0	0	0	18	6	8	24	1	0	2	0	1	1	24	6	16
9	1	0	0	1	0	0	4	3	4	25	1	1	1	0	0	0	18	6	8
10	0	2	0	0	0	0	3	3	2	26	0	1	1	0	0	0	9	5	4
11	1	0	0	0	0	0	2	2	2	27	1	0	0	1	1	0	8	4	8
12	0	0	0	0	1	0	2	2	2	28	1	0	2	0	1	1	24	6	16
13	1	0	2	1	1	0	24	6	16	29	1	0	2	1	1	0	24	6	16
14	1	1	1	0	0	0	18	6	8	30	1	0	0	1	0	1	8	4	8
15	0	0	2	0	0	0	3	3	2	31	1	2	0	0	0	0	6	4	4
16	0	0	2	0	0	0	3	3	2	sum							416	148	254

Note. #Par = number of parameters; M_1 = sp-CDM, saturated generalized deterministic, input, noisy, "and" gate model for polytomous attributes; $M_2 = f$A-M, fully additive model for polytomous attributes; M_3 = pG-DINA, polytomous generalized DINA model.

Results

Table 10 shows the number of parameters and the fit statistics (i.e., deviance, AIC, and BIC) for the sp-CDM, fA-M, and pG-DINA models in the empirical example. All the fit statistics indicate that the sp-CDM fitted the data the best, and the fA-M the worst.

<div align="center">TABLE 10.</div>

<div align="center">Fit statistics of the sp-CDM, fA-M, and pG-DINA model for the empirical example</div>

| Fitted Model | Number of Parameters | | Deviance | AIC | BIC |
	Items	Population			
sp-CDM	416	143	43360.17	44192.17	46376.13
fA-M	148	143	47983.06	48565.06	50092.79
pG-DINA	254	143	45524.30	46318.30	48402.52

Note. sp-CDM = saturated generalized deterministic, input, noisy, "and" gate model for polytomous attributes; fA-M = fully additive model for polytomous attributes; pG-DINA = polytomous generalized DINA model.

To quantify the discrepancies between the parameter estimates of item j, the root mean squared difference (RMSD) between any model pair is computed as follows:

$$\text{RMSD}_j(M_1, M_2) = \sqrt{\sum_{l=1}^{L_j} w_{jl}(P_{jlM_1} - P_{jlM_2})^2}, \tag{23}$$

where M_1 and M_2 are pair of models, w_{jl} and L_j are the posterior probability of latent group l and the number of latent groups for item j based on the sp-CDM, respectively, and P_{jlm} is the success probability of latent group l on item j based on a model m. Given in Table 11 are the results of RMSD for the 31 items, as well as the average RMSD for the entire test. On the average, fA-M and pGDINA model had the most similar the item parameter estimates (average RMSD = 0.17), whereas the sp-CDM and pG-DINA model had the most disparate estimates (average RMSD = 0.23). These results suggest that, at least for this empirical example, the two reduced model behaved more similarly to each other than they did to the saturated model. However, this pattern did not necessarily hold for all items. For example, the discrepancies

between fA-M and pGDINA model turned out to be the largest for items 13 and 19.

TABLE 11.

Root mean squared differences between the sp-CDM, fA-M, and pG-DINA model for the empirical example

Item	(M_1, M_2)	(M_1, M_3)	(M_2, M_3)	Item	(M_1, M_2)	(M_1, M_3)	(M_2, M_3)
1	0.24	0.14	0.12	17	0.20	0.25	0.10
2	0.12	0.23	0.20	18	0.09	0.20	0.15
3	0.09	0.28	0.27	19	0.33	0.41	0.36
4	0.22	0.23	0.15	20	0.28	0.36	0.14
5	0.15	0.23	0.14	21	0.31	0.28	0.26
6	0.22	0.35	0.28	22	0.22	0.46	0.33
7	0.28	0.32	0.08	23	0.20	0.15	0.17
8	0.20	0.19	0.21	24	0.29	0.29	0.29
9	0.37	0.25	0.15	25	0.18	0.30	0.19
10	0.12	0.05	0.12	26	0.12	0.13	0.06
11	0.04	0.06	0.01	27	0.21	0.26	0.16
12	0.06	0.05	0.01	28	0.17	0.28	0.25
13	0.31	0.30	0.36	29	0.11	0.15	0.09
14	0.19	0.18	0.25	30	0.39	0.24	0.16
15	0.18	0.16	0.10	31	0.18	0.20	0.07
16	0.19	0.17	0.07	Average	0.20	0.23	0.17

Note. M_1 = sp-CDM, saturated generalized deterministic, input, noisy, "and" gate model for polytomous attributes; M_2 = fA-M, fully additive model for polytomous attributes; M_3 = pG-DINA, polytomous generalized DINA model.

To better understand how similar and disparate item parameter estimates look like, the side-by-side success probability bar graphs of two items requiring one dichotomous and one polytomous attributes are given in Figure 4. Item 17 had estimates that can be considered more similar, whereas item 22 more disparate. It should be note that because RMSD is computed using latent group weights, large differences in the success probability estimates can have a limited

impact on the RMSD, and vice versa. As can be seen from the upper panel, despite having more similar item parameter estimates, the success probabilities of item 17 for some latent groups (i.e., 10 and 11) can be quite different. In contrast, the lower panel for shows that item parameter estimates for item 22 were quite disparate, and huge discrepancies the success probabilities for latent groups for 01, 02, and 11 can be found, particularly, between sp-CDM and pG-DINA model.

The above comparisons were not meant to establish in general the similarities or differences between the three polytomous CDMs. Rather, it sought to better understand how the three models behave for a very particular empirical data set, which may provide insights into how future studies can be designed for the different models to be compared in a more systematic and comprehensive way.

The results can have important practical implications for diagnosing the mastery status of students. As shown in Figure 4, based on the sp-CDM, latent groups 10 and 11 have slightly higher success probabilities on item 17 than latent groups 00, 01, and 02. This suggests that mastering the prerequisite skills and concepts increases the probability of getting item 17 correctly; however, the success probability for this item is the highest when, in addition to mastering the prerequisite skills and concepts, an examinee also masters ordering fractions (level 2 of α_2). In comparison, again based on the sp-CDM, the estimated success probability on item 22 is the highest for latent group 12, whereas the success probabilities for other latent groups (i.e., 00, 01, 02, 10, and 11) are similar to each other, indicating a conjunctive process for the item. Overall, an examinee should master levels 1 and 2 of the two required attributes, respectively, to optimize their success probabilities on these two items.

Of the possible 144 latent classes, 25 did not have any student. However, it should be noted that, although some latent classes had no observations, all the latent groups, from which estimates were derived, were non-empty, albeit some were small. For example, one of the 24 latent groups of items 3 had an expected size of 7.12. Of the remaining latent classes, 122111 was the largest - about 35.5% of the students had this attribute pattern. Finally, the individual attribute prevalences were as follows: 88.2% mastered α_1; 16.2% and 68.5% were in levels 1 and 2 of α_2, respectively; 11.7% and 74.7% were in levels 1 and 2 of α_3, respectively; and 75.4% , 70.5%, and 63.5% mastered α_4, α_5, and α_6, respectively.
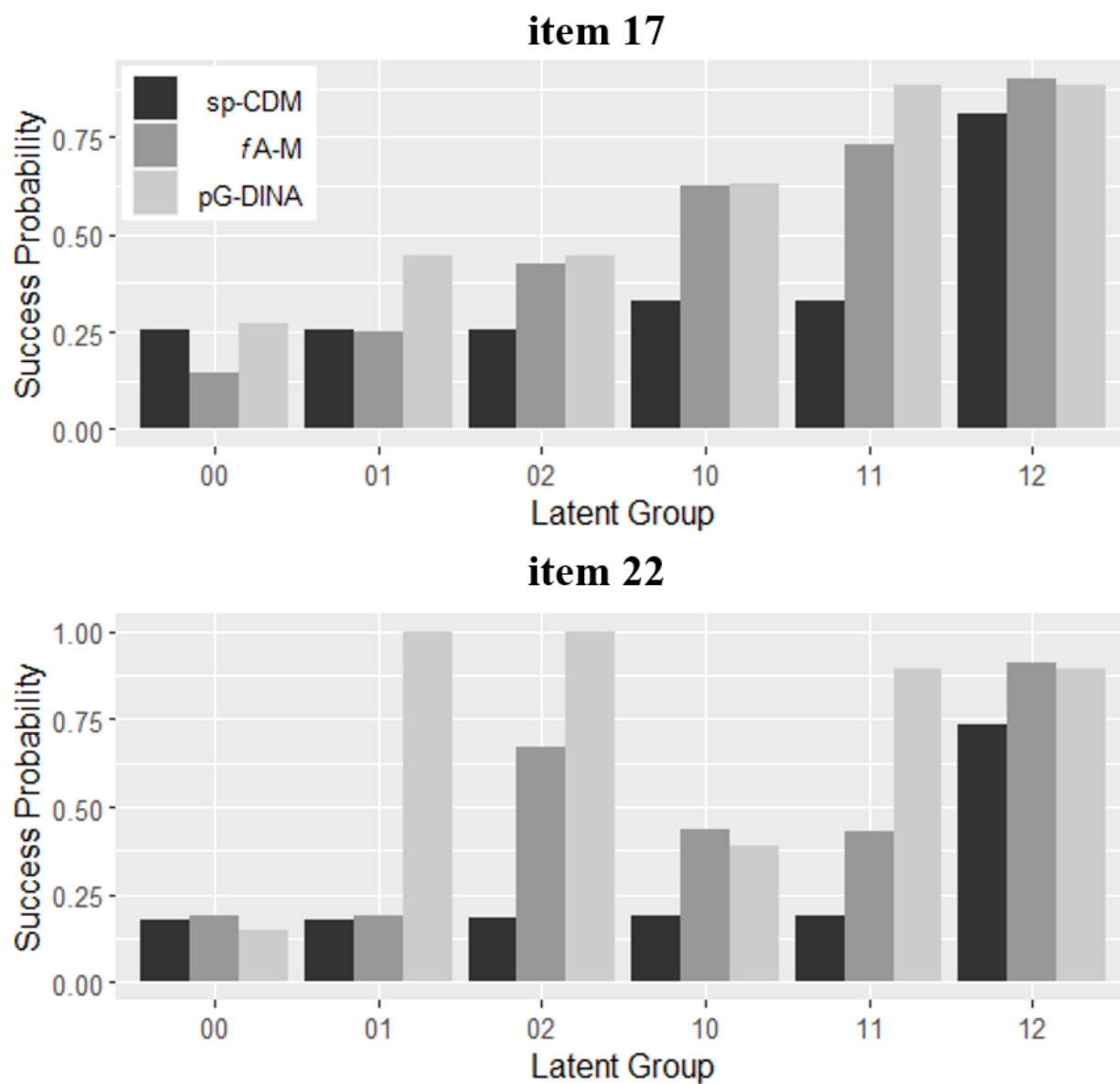
item 17

item 22

FIGURE 4.

Success probabilities of latent groups in two items in the empirical example

Note: sp-CDM: saturated polytomous cognitive diagnosis models; fA-M: fully additive model for polytomous attributes; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption.

8. Discussion

Finer grained feedback in the form of polytomous attributes can better inform classroom instruction and learning. However, the existing CDMs for polytomous attributes are deemed to be not general enough because most of them focused on a specific and constrained CDM or were proposed with very stringent assumptions. To this end, a more general framework, referred to as the sp-CDM, was proposed. The proposed model is a straightforward extension of the pG-DINA model (J. Chen & de la Torre, 2013), which itself is generalized from the G-DINA model for polytomous attributes, by relaxing its SALM assumption and can be formulated using the identity, logit, and log link functions. As such, the sp-CDM includes all the existing CDMs for polytomous attributes as its special cases. This paper has also illustrated the relationships between the sp-CDM and the existing CDMs mathematically and graphically.

In addition to the theoretical illustration, the estimation of the proposed model and the consequences of using constrained polytomous-attribute CDMs were examined via simulation study. The results showed that the parameter of the proposed model can be well recovered using the proposed estimation algorithms. On the other hand, improperly fitting a constrained polytomous-attribute CDM can lead to poor item parameter estimates and misdiagnosis of students' true mastery levels while unnecessarily fitting the complex sp-CDM does little harm to the item and person estimation, particularly when high quality items were used. Moreover, the PR assessment example demonstrated the applicability of the proposed model to real data and its advantages over the constrained models.

Despite the promising results, this study is not without its limitations. First, although the simulation study manipulated several important factors, other relevant factors such as the number of attribute levels, the distribution of attribute patterns, and the link functions were fixed. Additional simulation studies are needed in the future to investigate the performance of the proposed model across a wider range of conditions. For example, the current simulation studies used the same number of attribute levels (i.e., three) across the items, which might not always the case in practice. Nevertheless, both the model and estimation algorithms proposed in this work are sufficiently general to apply to varying and a larger number of attribute levels, as evidenced

by the empirical example. It would be interesting for future studies to extend the simulation design to incorporate attributes with more, as well as varying levels. This extension would be provide useful information on how the increased attribute levels, which will lead to a greater number of latent classes, affects the sample size and test length required for the parameters of the proposed model to be estimated accurately. Moreover, this work focused on the uniform distributed attribute patterns. Future studies can extend the proposed model to other attribute distributions (e.g., higher-order distribution; de la Torre & Douglas, 2004), to understand the model performance across a wider variety of conditions.

Second, it has been recognized that failure to satisfy the Q-matrix identifiability conditions can result in poor parameter estimates. The existing necessary and sufficient conditions for the identifiability of CDMs in the literature focus on dichotomous attributes. For example, see Y. Chen et al. (2015), Y.-H. Chen et al. (2018), Chiu et al. (2009), DeCarlo (2011), Gu and Xu (2019), Liu et al. (2012), and Xu and Zhang (2016) for the conditions for the DINA model, and Fang et al. (2019), Gu and Xu (2021), Köhn and Chiu (2018), and Xu (2017) for general models. In contrast, at present only Fang et al. (2019) have discussed the identifiability conditions for the Q-matrix for polytomous attributes. However, the relevant results (i.e., Theorem 4) were limited to the sufficient conditions. To optimize the process of developing assessments that involve polytomous attributes, further research is needed to establish both the necessary and sufficient conditions specific to the identifiability of the sp-CDM and potentially its special cases.

Third, the current work focuses on polytomous attributes used in conjunction with dichotomous responses. Future research should extend the sp-CDM to also cover polytomous responses (e.g., Ma & de la Torre, 2020), as well as develop the associated estimation algorithms and computer program to implement such a model.

Fourth, although it has been noted that a polytomous attribute can be equivalently represented as a set of linearly structured dichotomous attributes, it is not clear to what extent the equivalence extends to methodologies that are specifically developed for each attribute type. For example, what modifications are needed for the empirical Q-matrix validation procedures developed for dichotomous attributes (e.g., de la Torre & Chiu, 2016) to be equivalent to Q-matrix validation procedures developed for polytomous attributes (e.g., de la Torre et al.,

2021). Incidentally, a more general Q-matrix validation procedure that can be used with proposed model needs to be considered in future research.

Finally, this work proposes an MMLE/EM algorithm for estimating the sp-CDM and its special cases. The results of the simulation study demonstrate that the algorithm provides accurate estimates, and is efficient in estimating the proposed models. However, challenges arise when the algorithm has to deal with the complexities associated with the sp-CDM in its most general form. For example, the estimation of standard errors in the saturated models becomes particularly challenging due to often encountered singular Hessian matrices.

Furthermore, the parameter estimation becomes notably challenging in situations when the sample size is small relative to the number of attributes and attribute levels. To investigate this, an additional simulation was conducted with a sample size of 500, maintaining the same settings as the primary simulation study. The results are given in Figure B.3 and Table B.5 in Appendix B. It was found that, under $K = 3$ conditions, item parameter recovery and the PCA and PCV are satisfactory. However, for $K = 5$, although item parameter recovery and the PCA are only marginally acceptable, the PCV exhibits a significant deterioration, particularly when the item quality was low. The deterioration in the PCV performance may be attributed to the sparse latent classes - the expected number of individuals are 18 $(500/3^3)$ and two $(500/3^5)$ when $K = 3$ and $K = 5$, respectively. These findings suggest that the sp-CDM may not be well-suited for stand-alone small-sample settings (e.g., classroom assessment). Nonetheless, small-sample applications are still possible provided items can be calibrated a priori using a sufficiently large pool of individuals. In future research, it would be beneficial to explore the use of alternative estimation procedures such as nonparametric methods (e.g., Chiu et al., 2018) or Bayesian modal estimation (Ma & Jiang, 2021) to obtain robust person classification when polytomous attributes and small sizes are involved. It can be noted that small sample sizes impact not only the quality of item parameter estimates and attribute classification accuracy, but more so the standard error estimates. Thus, exploring various estimators of the CDM standard errors (Philipp et al., 2018) in the context of the proposed model need to be considered, particularly when the sample size is small.

References

Bao, Y. (2019). *A diagnostic classification model for polytomous attributes* [Doctoral dissertation]. University of Georgia. http://getd.libs.uga.edu

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

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850–866. https://doi.org/10.1080/01621459.2014.934827

Chen, Y.-H., Culpepper, S. A., Chen, Y.-G., & Douglas, J. (2018). Bayesian estimation of the dina Q. *Psychometrika*, *83*(1), 89–108. https://doi.org/10.1007/s11336-017-9579-4

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

Chiu, C.-Y., Sun, Y., & Bian, Y. (2018). Cognitive diagnosis for small educational programs: The general nonparametric classification method. *Psychometrika*, *83*, 355–375.

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

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

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

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

de la Torre, J., Qiu, X., & Santos, C. K. (2021). An empirical Q-matrix validation method for the polytomous G-DINA model. *Psychometrika*, *87*, 693–724. https://doi.org/10.1007/s11336-021-09821-x

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The dina model, classification, class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*(1), 8–26. https://doi.org/10.1177/0146621610377081

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*, 19–40. https://doi.org/10.1007/s11336-018-09658-x

Gu, Y., & Xu, G. (2019). The sufficient and necessary conditions for the identifiability and estimability of the DINA model. *Psychometrika*, *84*, 468–483. https://doi.org/10.1007/s11336-018-9619-8

Gu, Y., & Xu, G. (2021). Sufficient and necessary conditions for the identifiability of the Q-matrix. *Statistica Sinica*, *31*, 449–472. https://doi.org/10.5705/ss.202018.0410

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

Henson, R. A., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 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. M. (2004). *Ordered category attribute coding framework for cognitive assessments* [Doctoral dissertation]. University of Illinois at Urbana-Champaign. https://hdl.handle.net/2142/82064

Köhn, H., & Chiu, C.-Y. (2018). How to build a complete Q-matrix for a cognitively diagnostic test. *Journal of Classification*, *35*, 273–299. https://doi.org/10.1007/s00357-018-9255-0

Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, *41*, 205–236.

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

Ma, W., & de la Torre, J. (2019). Gdina: The Generalized DINA model framework [R package version 2.3.2].

Ma, W., & de la Torre, J. (2020). An empirical Q-matrix validation method for the sequential generalized DINA model. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 142–163. https://doi.org/10.1111/bmsp.12156

Ma, W., & Jiang, Z. (2021). Estimating cognitive diagnosis models in small samples: Bayes modal estimation and monotonic constraints. *Applied Psychological Measurement*, *45*(2), 95–111. https://doi.org/10.1177/0146621620977681

Philipp, M., Strobl, C., de la Torre, J., & Zeileis, A. (2018). On the estimation of standard errors in cognitive diagnosis models. *Journal of Educational and Behavioral Statistics*, *43*(1), 88–115. https://doi.org/10.3102/107699861771972

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

Tjoe, H., & de la Torre, J. (2013). Designing cognitively-based proportional reasoning problems as an application of modern psychological measurement models. *Journal of Mathematics Education*, *6*(2), 17–22.

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*, 237–255. https://doi.org/10.1007/s13394-013-0090-7

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

Wang, S., & Chen, Y. (2020). Using response times and response accuracy to measure fluency within cognitive diagnosis models. *Psychometrika*, *85*(3), 600–629. https://doi.org/10.1007/s11336-020-09717-2

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, *45*(2), 675–707.

Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika, 81*, 625–649. https://doi.org/10.1007/s11336-015-9471-z

Yakar, L., Dogan, N., & de la Torre, J. (2021). Retrofitting of polytomous cognitive diagnosis and multidimensional item response theory models. *Journal of Measurement and Evaluation in Education and Psychology, 12*(2), 97–111. https://doi.org/10.21031/epod.778861

Appendix A

Item Parameter Estimation for the Saturated Polytomous Cognitive Diagnosis Model via Marginal Maximum Likelihood with an Expectation-Maximization Algorithm

This appendix provides details of item parameter estimation in the sp-CDM. Due to space constraints, it focuses on the identity sp-CDM in particular and the algorithms are ready to be implemented to the sp-CDM with the logit and log link functions.

Based on the IRF of the saturated identity sp-CDM (i.e., Equation 1) and assuming local independence, the marginalized likelihood of the data, denoted as $L(\boldsymbol{X})$, is given by

$$L(\boldsymbol{X}) = \prod_{n=1}^{N} L(\boldsymbol{X}_n) = \prod_{n=1}^{N} \sum_{l=1}^{L} L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) p(\boldsymbol{\alpha}_l), \tag{A1}$$

where $L(\boldsymbol{X}_n)$ is the marginalized likelihood of the response vector of examinee n, $L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta})$ is the joint probability of the examinee's response vector $\boldsymbol{X}_n$ conditional on attribute vectors $\boldsymbol{\alpha}_l$ and the item parameters $\boldsymbol{\delta}$, and $p(\boldsymbol{\alpha}_l)$ is the prior probability of the attribute vectors $\boldsymbol{\alpha}_l$.

The logarithm of $L(\boldsymbol{X})$ is

$$LL(\boldsymbol{X}) = logL(\boldsymbol{X}) = log \prod_{n=1}^{N} L(\boldsymbol{X}_n) = \prod_{n=1}^{N} log \sum_{l=1}^{L} L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) p(\boldsymbol{\alpha}_l). \tag{A2}$$

To find the marginal likelihood equation for $\boldsymbol{\delta}$ of item j, take

$$\frac{\partial}{\partial \boldsymbol{\delta}_j} LL(\boldsymbol{X}) = 0. \tag{A3}$$

Then,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\delta}_j} LL(\boldsymbol{X}) &= \sum_{n=1}^{N} \frac{\partial}{\partial \boldsymbol{\delta}_j} (logL(\boldsymbol{X}_n)) \\ &= \sum_{n=1}^{N} [L(\boldsymbol{X}_n)]^{-1} \frac{\partial}{\partial \boldsymbol{\delta}_j} L(\boldsymbol{X}_n) \\ &= \sum_{n=1}^{N} [L(\boldsymbol{X}_n)]^{-1} \sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) \frac{\partial}{\partial \boldsymbol{\delta}_j} L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}). \end{aligned} \tag{A4}$$

Assuming local independence, the term $L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta})$ in equation A4 is given by

$$L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) = \prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j}, \tag{A5}$$

where $P_j(\boldsymbol{\alpha}_l^*)$ is defined in Equation 1, $Q_j(\boldsymbol{\alpha}_l^*) = 1 - P_j(\boldsymbol{\alpha}_l^*)$, $X_j = 1$ for correct answer of examinee n in item j and 0 otherwise.

Hence, the derivative of $L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta})$ becomes

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\delta}_j} L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) &= \frac{\partial}{\partial \boldsymbol{\delta}_j} \left[P_1(\boldsymbol{\alpha}_l^*)^{X_1} Q_1(\boldsymbol{\alpha}_l^*)^{1-X_1} \cdots P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} \cdots P_J(\boldsymbol{\alpha}_l^*)^{X_J} Q_J(\boldsymbol{\alpha}_l^*)^{1-X_J} \right] \\
&= \left[\prod_{j' \neq j}^{J} P_{j'}(\boldsymbol{\alpha}_l^*)^{X_{j'}} Q_{j'}(\boldsymbol{\alpha}_l^*)^{1-X_{j'}} \right] \times \frac{\partial}{\partial \boldsymbol{\delta}_j} \left[P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} \right],
\end{aligned}
$$

(A6)

where the product is over $j' \neq j$ rather than j because the derivation is with respect to $\boldsymbol{\delta}_j$.

The second term in the right hand side of Equation of A6 is

$$
\begin{aligned}
&\frac{\partial P_j(\boldsymbol{\alpha}_l^*)^{X_j}}{\partial \boldsymbol{\delta}_j} \times Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} + P_j(\boldsymbol{\alpha}_l^*)^{X_j} \times \frac{\partial Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j}}{\partial \boldsymbol{\delta}_j} \\
&= X_j P_j(\boldsymbol{\alpha}_l^*)^{X_j-1} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} + (1-X_j) P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j-1} \frac{\partial Q_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \\
&= P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{X_j - P_j(\boldsymbol{\alpha}_l^*)}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right].
\end{aligned}
$$

(A7)

By substituting Equation A7 to the right hand side of Equation A6 and rearranging yields

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\delta}_j} L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) &= \left[\prod_{j' \neq j}^{J} P_{j'}(\boldsymbol{\alpha}_l^*)^{X_{j'}} Q_{j'}(\boldsymbol{\alpha}_l^*)^{1-X_{j'}} \right] \times P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{X_j - P_j(\boldsymbol{\alpha}_l^*)}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] \\
&= \prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_l^*)^{X_j} Q_j(\boldsymbol{\alpha}_l^*)^{1-X_j} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{X_j - P_j(\boldsymbol{\alpha}_l^*)}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] \\
&= L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{X_j - P_j(\boldsymbol{\alpha}_l^*)}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right].
\end{aligned}
$$

(A8)

Substituting Equation A8 to Equation A4 yields

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\delta}_j} LL(\boldsymbol{X}) &= \sum_{n=1}^{N} [L(\boldsymbol{X}_n)]^{-1} \sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{X_j - P_j(\boldsymbol{\alpha}_l^*)}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] \\
&= \sum_{l=1}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] \sum_{n=1}^{N} \frac{L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) p(\boldsymbol{\alpha}_l)}{L(\boldsymbol{X}_n)} [X_j - P_j(\boldsymbol{\alpha}_l^*)] \\
&= \sum_{l=1}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] \sum_{n=1}^{N} p(\boldsymbol{\alpha}_l^*|\boldsymbol{X}_n) [X_j - P_j(\boldsymbol{\alpha}_l^*)] \qquad \text{(A9)} \\
&= \sum_{l=1}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] \left[\sum_{n=1}^{N} p(\boldsymbol{\alpha}_l^*|\boldsymbol{X}_n) X_j - P_j(\boldsymbol{\alpha}_l^*) \sum_{n=1}^{N} p(\boldsymbol{\alpha}_l^*|\boldsymbol{X}_n) \right] \\
&= \sum_{l=1}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] [R_{jl} - P_j(\boldsymbol{\alpha}_l^*) T_l],
\end{aligned}
$$

where $p(\boldsymbol{\alpha}_l^*|\boldsymbol{X}_n) = \frac{L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta}) p(\boldsymbol{\alpha}_l)}{L(\boldsymbol{X}_n)}$ is the posterior probability that examinee n is in the latent group $\boldsymbol{\alpha}_l^*$, $R_{jl} = \sum_{n=1}^{N} p(\boldsymbol{\alpha}_l^*|\boldsymbol{X}_n) X_j$ is the number of examinees in the latent group $\boldsymbol{\alpha}_l^*$ expected to answer the item j correctly, and $T_l = \sum_{n=1}^{N} p(\boldsymbol{\alpha}_l^*|\boldsymbol{X}_n)$ is the number of examinees expected to be in the latent group $\boldsymbol{\alpha}_l^*$.

Therefore, the marginal likelihood Equation A3 can be written as follows:

$$
\sum_{l=1}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l^*)}{\partial \boldsymbol{\delta}_j} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l^*) Q_j(\boldsymbol{\alpha}_l^*)} \right] [R_{jl} - P_j(\boldsymbol{\alpha}_l^*) T_l] = 0. \qquad \text{(A10)}
$$

Solving Equation A10 yields the MML estimation of $P(\boldsymbol{\alpha}_{lj}^*)$, which can be expressed as

$$
\hat{P}(\boldsymbol{\alpha}_{lj}^*) = \frac{R_{\boldsymbol{\alpha}_{lj}^*}}{T_{\boldsymbol{\alpha}_{lj}^*}}, \qquad \text{(A11)}
$$

To convert the estimates of $\hat{P}(\boldsymbol{\alpha}_{lj}^*)$ into the item parameters $\boldsymbol{\delta}$, the design matrix $\boldsymbol{M}_j$ is needed. With $\boldsymbol{M}$, the estimates of $\boldsymbol{\delta}_j = \{\delta_{j0}, \delta_{jkm}, \cdots, \delta_{jkmk'm'}, \cdots, \delta_{j1m_12m_2...K_j^*m_{K_j^*}}\}'$ can be computed as

$$
\hat{\boldsymbol{\delta}}_j = (\boldsymbol{M}_j' \boldsymbol{M}_j)^{-1} \boldsymbol{M}_j' \hat{\boldsymbol{P}}_j, \qquad \text{(A12)}
$$

where $\hat{\boldsymbol{P}}_j = \{\hat{P}(\boldsymbol{\alpha}_{lj}^*)\}$. This is identical to the design matrix used with G-DINA model (de la Torre, 2011). Due to the important role the design matrix $\boldsymbol{M}$ plays in the estimation of the sp-CDM, the following shows how $\boldsymbol{M}$ can be constructed.

Under the sp-CDM, the dimension size of design matrix is $M^{K_j^*} \times P_j$, where P_j is the number of the parameters of the model of interest, which is $M^{K_j^*}$ when converting $\hat{\boldsymbol{P}}_j$ to $\hat{\boldsymbol{\delta}}_j$. To illustrate, let $K_j^* = 2$ and $M = 3$ for item j. This item has one intercept parameter, four main effect parameters for each level of the two nonzero levels of the required attributes, and four two-way interaction parameters between the levels of the required attributes, leading to a total of nine parameters. Hence, $P_j = 9$ in this example. The corresponding saturated design matrix is

$$\boldsymbol{M}_{9\times9} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \tag{A13}$$

where rows 1 through 9 of $\boldsymbol{M}$ correspond to the latent groups 00, 10, 20, 01, 02, 11, 21, 12, and 22, respectively; columns 1 through 9 represent the intercept followed by the four main effects, and then by the four two-way interaction effect. For example, the success probability of the latent group 11 (i.e., row 6) $\delta_{j0} + \delta_{j1m_1} + \delta_{j2m_1} + \delta_{j1m_12m_1}$, whereas that of the latent group 22 (i.e., row 9) is $\delta_{j0} + \delta_{j1m_1} + \delta_{j1m_2} + \delta_{j2m_1} + \delta_{j2m_2} + \delta_{j1m_12m_1} + \delta_{j1m_22m_1} + \delta_{j1m_12m_2} + \delta_{j1m_22m_2}$.

The implementation of the MMLE/EM algorithm is as follows:

Step 1: The expectation (E) step

(1) Use the Equation A5 and provisional item parameter estimates to compute the likelihood of each examinee's response vector at each of the L attribute patterns.

(2) Use the Equation A1 to compute the likelihood of the whole data. For convenience, the uniform distribution is usually chosen to initialize the prior distribution, as in, $p(\boldsymbol{\alpha}_l) = 1/L$ in the first iteration. In subsequent iterations, the prior distribution is updated by replacing it with the posterior distribution, which itself is updated at the end of each iteration.

It is not uncommon that many examinees can have the same response vectors to the J items in a data set and the computations of $L(\boldsymbol{X}_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta})$ are often replicated. To increase the computational efficiency, an alternative way to compute the likelihood of data starts from grouping the N examinees' response vectors into 2^J possible response patterns. Let the item response patterns are denoted as U_s ($s = 1, \cdots, 2^J$) and the number of examinees possessing response pattern s is given by f_s. The likelihood of data can be calculated as

$L(\boldsymbol{X}) = \left[\sum_{l=1}^{L} L(U_s|\boldsymbol{\alpha}, \boldsymbol{\delta})p(\boldsymbol{\alpha_l}) \right]^{f_s}$.

(3) Count the number of examinees expected to be in the latent group $\boldsymbol{\alpha}_l^*$ among $2^{K_j^*}$ latent groups and the number of examinees in the latent group $\boldsymbol{\alpha}_l^*$ expected to answer the item j correctly and use them as the values of T_{jl} and R_{jl} in Equation A9, respectively.

Step 2: The maximization (M) step

Solve the likelihood equation A3 using the values of T_{jl} and R_{jl}. Because these values depend on $L(X_n|\boldsymbol{\alpha}_l, \boldsymbol{\delta})$, which in turn, depends on the unknown item parameters, the likelihood equations are implicit and must be solved with an iterative procedure(e.g., Newton-Raphson procedure).

The E step and M step will be repeated unless certain criteria are met (e.g., the change of likelihood between two successive cycles is less than 0.001, the maximum number of iterations, say, 100 is reached).

Appendix B

Q-matrix with Five Attributes and Additional Results in Simulation Study

TABLE B.1.

Q-matrix for conditions of five attributes in simulation study

Item	α_1	α_2	α_3	α_4	α_5	Item	α_1	α_2	α_3	α_4	α_5	Item	α_1	α_2	α_3	α_4	α_5
1	1	0	0	0	0	11	1	1	0	0	0	21	0	0	1	2	1
2	0	1	0	0	0	12	1	2	0	0	0	22	0	0	1	2	2
3	0	0	1	0	0	13	2	1	0	0	0	23	0	0	2	1	1
4	0	0	0	1	0	14	2	2	0	0	0	24	0	0	2	1	2
5	0	0	0	0	1	15	1	0	1	0	0	25	0	0	2	2	1
6	2	0	0	0	0	16	1	0	2	0	0	26	0	0	2	2	2
7	0	2	0	0	0	17	2	0	1	0	0						
8	0	0	2	0	0	18	2	0	2	0	0						
9	0	0	0	2	0	19	0	0	1	1	1						
10	0	0	0	0	2	20	0	0	1	1	2						

FIGURE B.1.

Bias in parameter recovery with five attributes

Note: sp-CDM: saturated polytomous cognitive diagnosis models; fA-M: fully additive model for polytomous attributes; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption. J: test length; N: sample size
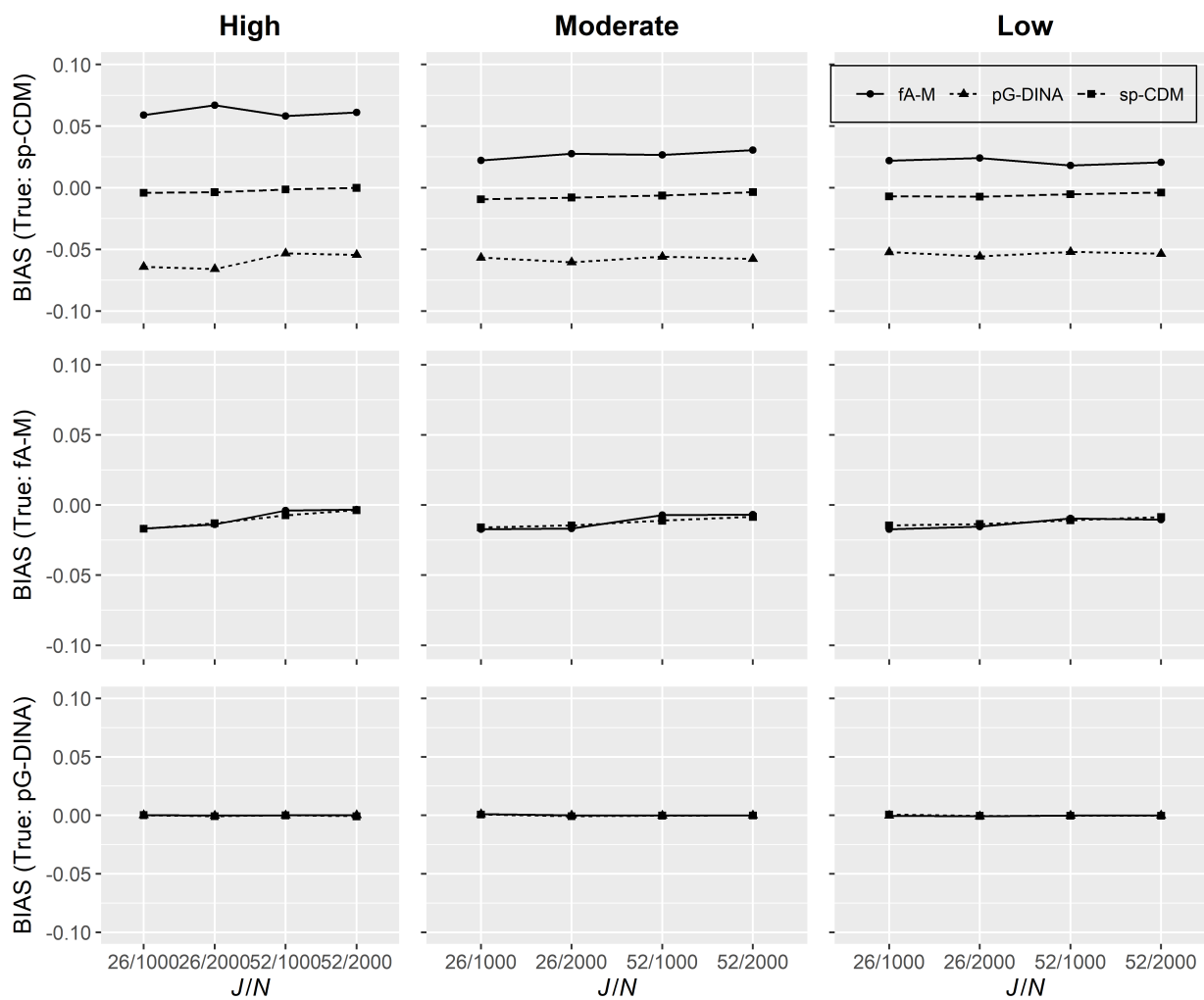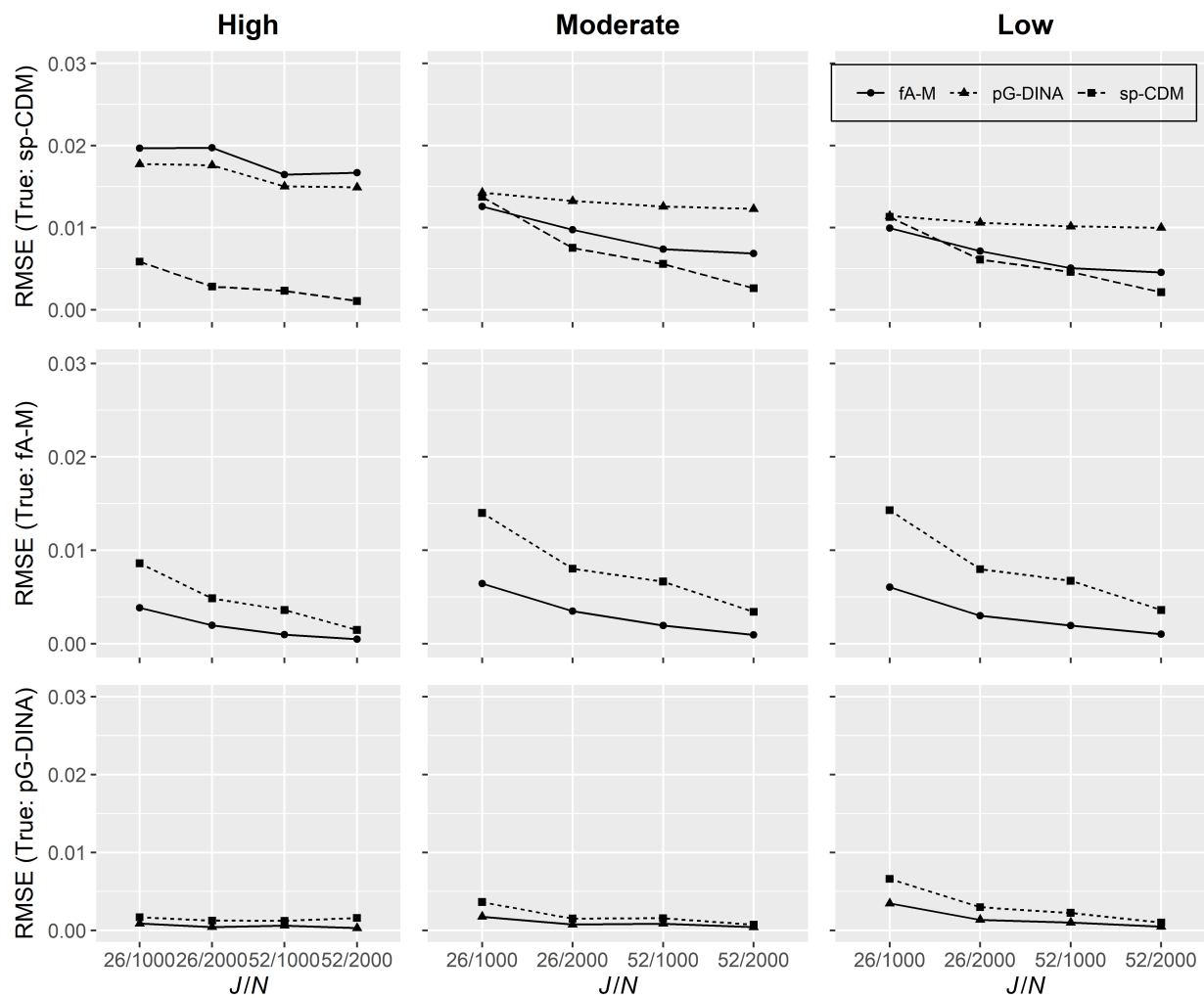
Figure B.2.

Room mean square error (RMSE) in parameter recovery with five attributes

Note: sp-CDM: saturated polytomous cognitive diagnosis models; fA-M: fully additive model for polytomous attributes; pG-DINA: generalized deterministic input, noisy "and" gate model for polytomous attributes with the specific attribute level mastery (SALM) assumption. J: test length; N: sample size

Table B.2.

Correctly classified attributes (PCA) and vectors (PCV) (in %) with five attributes and high quality items

| | | Test Length/Sample Size | | | | | | | | | | |
| | | 26/1000 | | | 26/2000 | | | 52/1000 | | | 52/2000 | | |
Gen	EC	1	2	3	1	2	3	1	2	3	1	2	3
sp-CDM	PCA1	86.8	84.1	79.6	87.3	84.4	79.7	94.7	92.2	86.2	94.8	92.0	86.4
	PCA2	80.3	79.1	76.8	81.1	79.9	76.7	90.0	88.2	80.6	90.4	88.1	80.4
	PCA3	87.6	84.8	84.3	88.5	84.9	84.1	95.4	92.8	92.9	95.7	92.7	93.0
	PCA4	78.8	75.6	77.8	80.2	75.6	78.1	88.6	84.4	82.0	89.5	84.0	81.5
	PCA5	77.8	73.2	77.5	79.3	72.8	77.6	87.5	81.3	80.9	88.6	81.4	80.5
	PCV	38.1	30.0	30.7	40.7	30.4	30.8	63.6	50.8	41.6	65.6	50.1	41.1
fA-M	PCA1	87.6	88.0	79.8	88.2	88.3	79.2	94.9	95.0	82.2	95.2	95.3	82.4
	PCA2	80.9	81.1	79.0	81.4	81.8	78.9	89.6	89.8	79.9	90.0	90.0	80.0
	PCA3	84.4	86.1	79.9	85.6	86.7	80.0	93.1	94.0	84.4	93.8	94.1	84.4
	PCA4	77.4	79.9	78.7	78.1	80.3	78.3	85.9	88.8	78.1	87.9	89.0	77.9
	PCA5	78.2	80.8	77.7	79.8	81.5	77.7	87.4	89.8	78.5	89.1	89.9	78.3
	PCV	36.1	39.8	30.0	38.5	41.3	29.8	59.3	64.3	32.3	63.2	65.0	32.4
pG-DINA	PCA1	95.8	95.1	95.9	95.8	95.2	96.0	99.3	99.1	99.3	99.3	99.2	99.4
	PCA2	92.5	92.3	92.9	92.8	92.6	93.1	97.7	97.6	97.8	97.7	97.7	97.8
	PCA3	94.5	93.3	94.9	94.7	93.2	95.0	98.7	98.0	98.8	98.7	98.0	98.9
	PCA4	92.8	84.6	93.3	93.2	83.9	93.5	97.0	96.2	97.2	97.0	96.2	97.2
	PCA5	92.8	79.6	93.3	93.1	77.4	93.5	96.9	95.9	97.1	97.1	96.3	97.3
	PCV	73.1	54.7	74.6	73.9	52.8	75.0	90.1	87.5	90.7	90.3	88.0	91.0

Note. Gen: Generating model; EC: Evaluation criteria; 1: sp-CDM; 2: fA-M; 3: pG-DINA; PCAk: PCA of attribute k

TABLE B.3.

Correctly classified attributes (PCA) and vectors (PCV) (in %) with five attributes and moderate quality items

| | | Test Length/Sample Size | | | | | | | | | | |
| | | 26/1000 | | | 26/2000 | | | 52/1000 | | | 52/2000 | | |
Gen	EC	1	2	3	1	2	3	1	2	3	1	2	3
sp-CDM	PCA1	75.1	75.0	73.4	76.3	76.0	73.3	86.2	85.8	79.5	86.8	85.9	79.2
	PCA2	69.5	67.3	70.0	70.4	67.2	70.1	79.1	77.9	75.2	80.5	78.2	75.3
	PCA3	72.9	74.3	73.4	74.5	75.0	73.5	84.6	84.7	81.1	86.1	84.9	81.0
	PCA4	68.1	69.5	70.1	69.1	69.3	70.3	80.6	79.8	76.4	80.3	79.9	76.6
	PCA5	67.4	67.3	69.6	68.5	67.3	70.1	76.7	76.5	75.5	78.4	76.7	75.3
	PCV	18.5	17.1	17.5	19.2	17.3	18.8	35.1	33.9	27.1	38.5	34.2	27.0
fA-M	PCA1	78.8	80.0	75.1	80.2	80.8	75.2	89.0	89.5	78.5	89.7	89.9	78.5
	PCA2	73.0	73.3	73.2	73.8	74.0	73.5	83.0	83.7	77.8	83.9	84.2	77.7
	PCA3	76.2	79.5	75.7	78.2	80.3	75.8	87.2	89.6	82.2	88.7	89.7	82.3
	PCA4	68.5	71.5	72.3	69.7	72.5	72.4	76.9	81.2	74.9	79.2	82.2	74.8
	PCA5	69.0	71.5	72.4	69.4	72.3	72.4	76.5	80.6	74.4	78.5	81.6	74.2
	PCV	20.7	24.2	21.6	22.4	25.5	21.9	37.8	44.3	27.0	41.6	45.9	27.1
pG-DINA	PCA1	91.1	90.2	91.3	91.4	90.5	91.5	97.6	96.9	97.7	97.7	96.8	97.7
	PCA2	85.8	85.3	86.2	86.1	85.5	86.5	94.3	94.0	94.4	94.4	94.0	94.5
	PCA3	88.0	85.5	88.8	88.5	85.5	89.0	95.9	94.2	96.2	96.1	94.4	96.3
	PCA4	84.5	74.6	85.8	86.0	73.8	86.5	92.0	90.8	92.5	92.4	91.0	92.7
	PCA5	85.0	73.3	86.1	86.3	71.6	86.9	92.0	88.5	92.5	92.3	87.9	92.6
	PCV	51.3	35.8	53.7	53.7	34.7	55.0	75.5	69.2	76.7	76.4	68.9	77.0

Note. Gen: Generating model; EC: Evaluation criteria; 1: sp-CDM; 2: fA-M; 3: pG-DINA; PCAk: PCA of attribute k

Table B.4.

Correctly classified attributes (PCA) and vectors (PCV) (in %) with five attributes and low quality items

| | | Test Length/Sample Size | | | | | | | | | | |
| | | 26/1000 | | | 26/2000 | | | 52/1000 | | | 52/2000 | | |
Gen	EC	1	2	3	1	2	3	1	2	3	1	2	3
sp-CDM	PCA1	72.8	72.7	71.7	74.1	73.7	71.8	84.2	83.6	78.5	84.7	83.8	78.3
	PCA2	67.9	65.0	68.2	68.9	65.2	68.2	76.9	75.5	73.5	77.9	75.9	73.5
	PCA3	70.4	71.6	71.2	71.8	72.3	71.5	82.0	82.3	79.0	83.4	82.6	79.1
	PCA4	65.5	66.8	67.9	66.8	67.4	68.2	78.6	77.0	74.6	77.1	77.5	74.5
	PCA5	65.2	65.5	67.3	66.1	65.5	67.9	75.7	75.6	73.8	76.1	76.2	74.0
	PCV	16.1	14.5	16.0	16.6	15.0	16.4	31.0	30.1	24.6	32.9	30.7	24.6
fA-M	PCA1	74.2	75.7	72.5	75.5	76.4	72.6	85.4	86.1	78.9	86.1	86.4	79.0
	PCA2	66.9	69.5	68.1	68.0	68.0	68.3	75.9	76.7	73.8	77.2	77.7	74.0
	PCA3	69.3	73.0	70.4	71.0	73.8	70.9	80.5	83.7	77.7	82.3	84.0	77.8
	PCA4	64.3	67.4	67.3	65.7	68.2	67.6	72.6	76.3	72.8	74.4	77.1	72.9
	PCA5	62.1	68.6	66.5	64.0	67.1	67.1	69.7	73.7	70.7	71.3	74.3	70.8
	PCV	13.8	16.8	15.8	15.4	17.9	16.1	26.4	31.6	23.2	29.3	32.8	23.2
pG-DINA	PCA1	80.2	77.2	80.8	80.7	77.4	81.0	90.2	87.2	90.5	90.7	87.5	90.8
	PCA2	76.5	70.5	77.0	76.8	70.0	77.2	85.5	82.6	85.8	85.8	82.7	85.9
	PCA3	77.7	75.4	79.3	79.0	75.4	79.9	88.8	85.6	89.6	89.4	85.7	89.8
	PCA4	72.5	69.7	74.9	74.2	69.2	75.7	83.0	81.4	84.0	83.7	81.7	84.4
	PCA5	73.0	68.2	75.2	74.5	67.5	75.8	83.0	78.2	84.0	83.8	76.8	84.3
	PCV	27.9	19.1	30.6	30.0	18.6	31.7	49.6	39.4	51.6	51.3	39.0	52.4

Note. Gen: Generating model; EC: Evaluation criteria; 1: sp-CDM; 2: fA-M; 3: pG-DINA; PCAk: PCA of attribute k
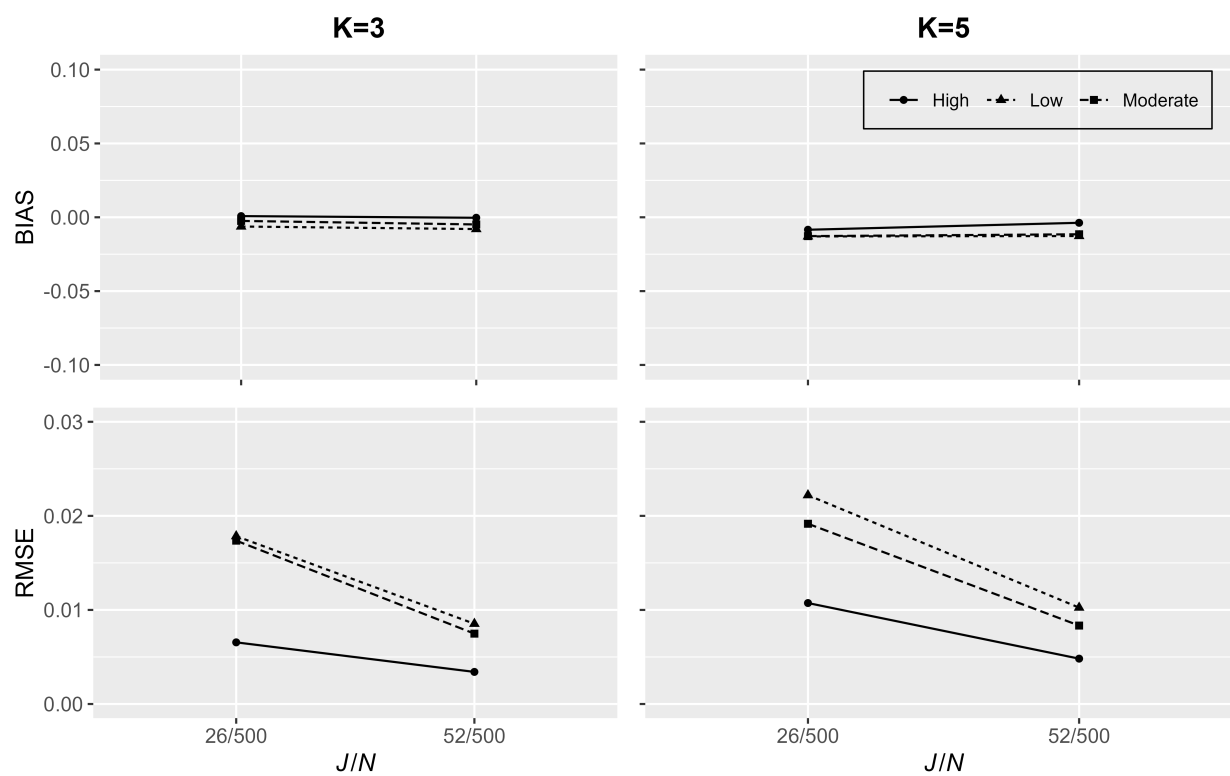
Figure B.3.

Results of parameter recovery for the sp-CDM with $N = 500$

Note: sp-CDM: saturated polytomous cognitive diagnosis models; J: test length; N: sample size

TABLE B.5.

Correctly classified attributes (PCA) and vectors (PCV) (in %) for the sp-CDM with $N = 500$

| | K=3 | | | | | | K=5 | | | | | |
| | 26/500 | | | 52/500 | | | 26/500 | | | 52/500 | | |
EC	H	M	L	H	M	L	H	M	L	H	M	L
PCA1	89.3	76.0	70.1	96.1	87.8	82.0	85.7	73.5	71.2	94.2	84.8	82.7
PCA2	87.6	73.2	66.9	95.3	84.0	77.7	79.1	68.6	67.1	89.4	78.0	75.8
PCA3	86.8	72.5	64.2	94.9	84.6	74.4	86.6	70.7	69.2	94.7	82.0	79.8
PCA4	–	–	–	–	–	–	77.8	66.9	63.5	87.2	76.7	73.7
PCA5	–	–	–	–	–	–	76.6	66.2	63.2	86.2	75.5	72.6
PCV	68.5	40.6	30.1	87.2	62.8	47.5	35.2	15.7	13.4	60.1	31.4	26.7

Note. sp-CDM: saturated polytomous cognitive diagnosis models; EC: Evaluation criteria; H: High quality items; M: Moderate quality items; L: Low quality items; PCAk: PCA of attribute k