Applied Psychological Measurement
http://apm.sagepub.com/

A General Cognitive Diagnosis Model for Expert-Defined Polytomous Attributes

Jinsong Chen and Jimmy de la Torre

Applied Psychological Measurement published online 13 March 2013
DOI: 10.1177/0146621613479818

Article

Applied Psychological Measurement
XX(X) 1–19

DOI: 10.1177/0146621613479818
apm.sagepub.com
SAGE

A General Cognitive Diagnosis Model for Expert-Defined Polytomous Attributes

Jinsong Chen[1] and Jimmy de la Torre[2]

Abstract

Polytomous attributes, particularly those defined as part of the test development process, can provide additional diagnostic information. The present research proposes the polytomous generalized deterministic inputs, noisy, "and" gate (pG-DINA) model to accommodate such attributes. The pG-DINA model allows input from substantive experts to specify attribute levels and is a general model that subsumes various reduced models. In addition to model formulation, the authors evaluate the viability of the proposed model by examining how well the model parameters can be estimated under various conditions, and compare its classification accuracy against that of the conventional G-DINA model with a modified classification rule. A real-data example is used to illustrate the application of the model in practice.

Keywords

CDM, polytomous attributes, expert-defined, Q-matrix, general model

Cognitive diagnosis models (CDMs) are psychometric models that can be used to assess examinees' strengths and weaknesses with respect to a particular set of attributes. Instead of postulating a single proficiency continuum, CDMs construe proficiency as a set of interrelated but separable attributes within a domain. As such, CDMs can provide finer-grained inferences regarding examinees' mastery of those attributes. Over the past decade or so, several CDMs that can be successfully applied across a wide variety of settings have been developed. At present, several CDMs for dichotomous attributes exist, which include the *deterministic inputs, noisy, ''and'' gate* (DINA; Haertel, 1989; Junker & Sijtsma, 2001) model, *deterministic inputs, noisy, ''or'' gate* (DINO; Templin & Henson, 2006) model, additive CDM (*A*-CDM; de la Torre, 2011), the linear logistic model (LLM; Maris, 1999), reduced reparameterized unified model (R-RUM; DiBello, Roussos, & Stout, 2007; Hartz, 2002), the log-linear CDM (LCDM; Henson, Templin, & Willse, 2009), and generalized DINA (G-DINA; de la Torre, 2011) model, to name a few. In contrast, only a few CDMs accommodating polytomous attributes can be found, such as the model based on the ordered-category attribute coding (OCAC; Karelitz,

[1]Sun Yat-Sen University, China
[2]Rutgers, The State University of New Jersey, USA

Corresponding Author:
Jinsong Chen, Department of Psychology, Sun Yat-Sen University, Guangzhou 510275, China.
Email: chenj.sysu@gmail.com

Table 1. Polytomous Attributes in the Eighth-Grade Proportional Reasoning Assessment.

Attributes	Level 1	Level 2
Comparing and ordering of fractions	Students should be able to compare two fractions and determine whether one of these fractions is equal to, less than, or greater than the other.	Students should be able to order three or more fractions.
Constructing ratios and proportions from a situation	Given a problem situation involving ratios, students should be able to construct a single ratio to describe the situation.	Given a proportional situation, students should be able to construct an appropriate proportion.

Note: These attributes were adapted from de la Torre, Lam, Rhoads, and Tjoe (2010). Level 0 is defined as lack of attribute mastery.

2004) framework, R-RUM for polytomous attributes (Templin, 2004), LCDM for polytomous attributes (Templin & Bradshaw, in press), and general diagnostic model (GDM) for polytomous attributes (Haberman, von Davier, & Lee, 2008; von Davier, 2005).

The present research proposes a general CDM to accommodate a special type of polytomous attributes—the expert-defined polytomous attributes. In conventional polytomous attributes, specific attribute levels and their meanings are only derived after the data have been fitted with the model. As such, it is not possible to design items targeting specific levels of the attributes. The authors refer to polytomous attributes with levels inferred as part of the data-fitting process as *data-defined polytomous attributes*. In contrast, with expert-defined polytomous attributes, specific attribute levels are substantively defined prior to the data-fitting process, typically by content experts. An example where a CDM for expert-defined polytomous attributes would be useful can be found in the proportional reasoning assessment for eighth-grade students described by de la Torre, Lam, Rhoads, and Tjoe (2010). The two attributes with three substantively defined attribute levels measured in this assessment are given in Table 1. Similar examples can be found by treating the substandards across different grade levels in the Principles and Standards for School Mathematics (National Council of Teachers of Mathematics [NCTM], 2000) as polytomous attributes, or by classifying students relative to subtopics of different content areas across grade levels in the National Assessment of Educational Progress for mathematics achievement (National Assessment Governing Board, 2008). To provide more accurate diagnostic information at different attribute levels, CDMs that can take into account the definition of each level prior to the model-data fitting would be needed.

From among the different existing CDMs that can accommodate polytomous attributes, the OCAC framework (Karelitz, 2004) is the only model that can capitalize on the additional information afforded by expert-defined polytomous attributes. However, this framework is highly restrictive in terms of both the structure (i.e., joint distribution of the attributes) and measurement (i.e., CDM) models—It assumes that some subsets of the attributes are ordered (i.e., linearly structured), and the CDM to be conjunctive (i.e., the DINA model) with common slip and guessing parameters across all items. In this research, the authors proposed a CDM that builds on the OCAC framework to formulate a parsimonious model for polytomous attributes, and the G-DINA model (de la Torre, 2011) to formulate a general model that subsumes a variety of CDMs. The resulting model is referred to as the polytomous generalized deterministic inputs, noisy, ‘‘and’’ gate (pG-DINA) model, where the *p* is used to denote the polytomous nature of the attributes being handled by the model.

Theoretical Background

Model Formulation

Specific Attribute Level Mastery (SALM) Items. One way to incorporate the substantive definition into the modeling process is to introduce especially designed items targeted to measure specific levels of the attributes. In the most general case, an item involving an M-level polytomous attribute is allowed to divide the examinees into M groups, with each group potentially having its own probability of success. However, to handle such a generality would require a CDM with a large number of parameters, and the complexity of the model exponentially grows as the number of required attributes increases. For example, items that require one, two, and three M-level polytomous attributes would result in saturated CDMs with as many as M, M^2, and M^3 parameters per item, respectively; for reduced CDMs, the number of parameters per item can change depending on the nature of the CDMs. To make CDMs for expert-defined polytomous attributes more manageable, the especially designed items are assumed to distinguish between two types of examinees (i.e., two latent groups): examinees who are *on or above* a specific attribute level and examinees who are *below* this level. Examinees on or above that specific attribute level are assumed to have the necessary attribute mastery level to answer the item correctly; examinees below it do not. Items with this assumption will be referred to as *specific attribute level mastery* (SALM) items. With SALM items, the substantive definition of each attribute level can be incorporated into the modeling process through a modified Q-matrix (Tatsuoka, 1983), as is done in the OCAC framework (Karelitz, 2004). This allows the different levels of attributes to be measurable.

Let q_{jk} denote the element in row j and column k of a $J \times K$ Q-matrix, where J and K represent the numbers of items and attributes, respectively. In the modified Q-matrix, q_{jk} can take on values $0, 1, \ldots, M_k - 1$, where M_k is the number of levels of attribute k. For notational simplicity but without loss of generality, it can be assumed that $M_k = M$ for all attributes. Based on substantive experts' input, $q_{jk} = m$ if mastery at level m or above of attribute k is required to answer the SALM item j correctly. Thus, for a SALM item where $q_{jk} = m$, no distinction is made between mastery levels $m, \ldots, M - 1$; similarly, no distinction is made between mastery levels $0, \ldots, m - 1$. Last, stating that item j does not require any level of mastery of attribute k (i.e., $q_{jk} = 0$) is equivalent to saying that the attribute is not relevant for the item.

Measurement Model. To simplify model formulation, the authors adopt the notation used in the G-DINA model (de la Torre, 2011). First, $K_j^* = \sum_{k=1}^{K} I(q_{jk}>0)$ will be used to denote the number of required attributes for item j, where $I(\cdot)$ is the indicator function. Second, for notational convenience, it is assumed that the first K_j^* attributes are the required attributes for item j. Third, the required attributes for item j can be represented by the reduced vector $\boldsymbol{\alpha}_{lj}^* = (\alpha_{l1}, \ldots, \alpha_{lK_j^*})'$, where $l = 1, \ldots, M^{K_j^*}$ and $M^{K_j^*}$ represent the number of unique attribute patterns. Recall that each element of the attribute, reduced or otherwise, can take on values from 0 to $M-1$. As such, the number of attribute vectors to be considered for item j reduces from M^K to $M^{K_j^*}$.

With the use of SALM items, which only distinguishes between two latent groups for each relevant attribute, the formulation of the CDM can be further simplified. In specifying $q_{jk} = m$ for item j, the M-level attribute α_{lk} can be collapsed into a dichotomous attribute α_{lk}^{**}, where

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

The authors denote $\boldsymbol{\alpha}_{lj}^{**} = (\alpha_{l1}^{**}, \ldots, \alpha_{lK_j^*}^{**})'$ as the *collapsed attribute vector*, where $l = 1, \ldots, 2^{K_j^*}$. With the preceding simplification, they estimate parameters in the space of α_{lj}^{**}. Thus, for an

Table 2. Unique Attribute Patterns for Different Types of Attribute Vectors.

Original $\boldsymbol{\alpha}_{lj}$	Reduced($\boldsymbol{\alpha}_{lj}^{*}$)	Collapsed ($\boldsymbol{\alpha}_{lj}^{**}$)
(0,0,0), (0,0,1), (0,0,2)	(0,0)	(0,0)
(0,1,0), (0,1,1), (0,1,2)	(0,1)	
(1,0,0), (1,0,1), (1,0,2)	(1,0)	(1,0)
(1,1,0), (1,1,0), (1,1,0)	(1,1)	
(2,0,0), (2,0,1), (2,0,2)	(2,0)	
(2,1,0), (2,1,1), (2,1,2)	(2,1)	
(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)
(2,2,0), (2,2,1), (2,2,1)	(2,2)	

item that requires K_j^* attributes, the number of unique α_{lj}^{**} patterns reduces to $2^{K_j^*}$. As an example, assume a test that measures 3 three-level attributes (i.e., $K=3$, $M=3$). From this test, consider a SALM item that requires Levels 1 and 2 of Attributes 1 and 2, respectively; that is, the **q**-vector is (1, 2, 0) with $K_j^*=2$. As shown in Table 2, the original attribute vector $\boldsymbol{\alpha}_{lj}$ will have $M^K=3^3=27$ unique patterns, and the reduced attribute vector $\boldsymbol{\alpha}_{lj}^*$ will have $M^{K_j^*}=3^2=9$ unique patterns, whereas the collapsed attribute vector $\boldsymbol{\alpha}_{lj}^{**}$ will have only $2^{K_j^*}=2^2=4$ unique patterns.

To formulate the pG-DINA model, the probabilities that examinees with collapsed attribute vector $\boldsymbol{\alpha}_{lj}^{**}$ will answer item j correctly are denoted as $P(X_j=1|\boldsymbol{\alpha}_{lj}^{**})=P(\boldsymbol{\alpha}_{lj}^{**})$. The saturated form of the pG-DINA model can be expressed using the following item response function based on the identity link function:

$$P\left(\boldsymbol{\alpha}_{lk}^{**}\right)=\delta_{j0}+\sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{lk}^{**}+\sum_{k'>k}^{K_j^*}\sum_{k=1}^{K_j^*}\delta_{jkk'}\alpha_{lk}^{**}\alpha_{lk'}^{**}+\cdots+\delta_{j1,\ldots,K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lK_j^*}^{**}. \tag{2}$$

The model parameters retain their interpretations as in the G-DINA model: δ_{j0} is the intercept for item j, δ_{jk} is the main effect due to α_{lk}^{**}, $\delta_{jkk'}$ is the interaction effect due to α_{lk}^{**} and $\alpha_{lk'}^{**}$, and $\delta_{j1,\ldots,K_j^*}$ is the interaction effect due to $\alpha_{l1}^{**},\ldots,\alpha_{lK_j^*}^{**}$. As with the original G-DINA model for dichotomous attributes, the pG-DINA model has $2^{K_j^*}$ parameters for an item that requires K_j^* attributes. Accordingly, the total number of item parameters is $\sum_{j=1}^{J}2^{K_j^*}$ in the measurement model. It should be noted that when M = 2, collapsing of the attribute level is not needed, and the pG-DINA model is equivalent to the G-DINA model. Similar saturated models can be obtained using the logit and log link functions, which can be referred to as the *log-odds CDM* and *log CDM* for polytomous attributes, respectively. As noted by de la Torre (2011), the different saturated models will give identical model fits. The fundamental difference between these model lies in that fact that instead of the probability of success, log-odds CDM and log CDM model the logit and log of the probability of success, respectively.

By constraining the parameters of the saturated pG-DINA model, different reduced CDMs similar to those under the G-DINA model can be found. The DINA model for polytomous attributes can be written as

$$P\left(\boldsymbol{\alpha}_{lj}^{**}\right)=\delta_{j0}+\delta_{j1,\ldots,K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lk}^{**}. \tag{3}$$

The A-CDM for polytomous attributes can be written as

$$P\left(\boldsymbol{\alpha}_{lj}^{**}\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk}^{**}. \tag{4}$$

Similarly, LLM and R-RUM for polytomous attributes can be obtained using the additive model under the logit and log link functions, respectively.

Structural Model. It is worth noting that the above simplifications do not apply to the structural model (i.e., the joint distribution of the attributes). For the general or unconstrained structural model, there are $M^K - 1$ parameters, as in, $p(\boldsymbol{\alpha}_l)$, for $l = 1, 2, \ldots, M^K$ with the constraint that $\sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) = 1$. To simplify the joint distribution of the attributes, specific constraints can be imposed on the structural relationships among the attributes. For example, it can be assumed that the attributes have a higher order structure, where $p(\boldsymbol{\alpha}_l|\theta) = \prod_{k=1}^{K} p(\alpha_{lk}|\theta)$ (de la Torre & Douglas, 2004) or an independent structure, where $p(\boldsymbol{\alpha}_l) = \prod_{k=1}^{K} p(\alpha_{lk})$. Alternatively, it can also be assumed that the attributes have a hierarchical structure, where a full mastery of one or more attributes is a prerequisite for the mastery of even the most basic levels of other attributes.

Model Parameter Estimation

With the modified Q-matrix, $\boldsymbol{\alpha}_{lj}^{*}$ can be collapsed into $\boldsymbol{\alpha}_{lj}^{**}$ using Equation 1—this step does not require any estimation. After this, the pG-DINA model is similar to the G-DINA model for dichotomous attributes, except that $\boldsymbol{\alpha}_{lj}^{*}$ will be transformed into $\boldsymbol{\alpha}_{lj}^{**}$ or vice versa whenever necessary. Accordingly, estimation procedures under the G-DINA model and the models it subsume (de la Torre & Chen, 2011) can be used with the pG-DINA model. Specifically, the item parameters for the saturated model, $P(\boldsymbol{\alpha}_{lj}^{**})$, can be estimated using an expectation-maximization (EM) implementation of the marginal maximum likelihood (MML) estimation. As with the G-DINA model, this involves maximizing the likelihood

$$L(\mathbf{X}) = \prod_{i=1}^{N} \sum_{l=1}^{L} L(\mathbf{X}_i|\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l), \tag{5}$$

where N is the sample size, L is the total number of attribute patterns, and $L(\mathbf{X}_i|\boldsymbol{\alpha}_l) = \prod_{j=1}^{J} P(\boldsymbol{\alpha}_{lj}^{**})^{X_{ij}}[1 - P(\boldsymbol{\alpha}_{lj}^{**})]^{(1-X_{ij})}$ is the likelihood of the response vector of examinee i given the attribute vector $\boldsymbol{\alpha}_l$. The prior probability of $\boldsymbol{\alpha}_l$, $p(\boldsymbol{\alpha}_l)$, can be updated to the posterior probability $p(\boldsymbol{\alpha}_l|\mathbf{X})$ during the estimation process using the empirical Bayes's method (Deely & Lindley, 1981). Specifically, the posterior probability of examine i for $\boldsymbol{\alpha}_l$ at the tth iterative estimation process is updated as

$$p(\boldsymbol{\alpha}_l|\mathbf{X}_i)_t = \frac{L(\mathbf{X}_i|\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l|\mathbf{X}_i)_{t-1}}{\sum_{l'=1}^{L} L(\mathbf{X}_i|\boldsymbol{\alpha}_{l'})p(\boldsymbol{\alpha}_{l'}|\mathbf{X}_i)_{t-1}}, \tag{6}$$

where $p(\boldsymbol{\alpha}_l|\mathbf{X}_i)_{t=0} = p(\boldsymbol{\alpha}_l)$. The posterior probability $p(\boldsymbol{\alpha}_l|\mathbf{X})$ is then obtained by averaging the posterior probabilities across all the examinees. The MML estimate of $P(\boldsymbol{\alpha}_{lj}^{**})$ is given by

$$\hat{P}\left(\boldsymbol{\alpha}_{lj}^{**}\right) = \frac{R_{\boldsymbol{\alpha}_{lj}^{**}}}{N_{\boldsymbol{\alpha}_{lj}^{**}}}, \tag{7}$$

where $N_{\boldsymbol{\alpha}_{lj}^{**}}$ is the number of examinees expected to have attribute vector $\boldsymbol{\alpha}_{lj}^{**}$, and $R_{\boldsymbol{\alpha}_{lj}^{**}}$ is the number of examinees with $\boldsymbol{\alpha}_{lj}^{**}$ expected to answer item j correctly. The standard error (SE) of the estimate, $SE[\hat{P}(\boldsymbol{\alpha}_{lj}^{**})]$, can be approximated from the information matrix:

$$\mathbf{I}\left(\mathbf{p}_j\right) = \left\{ \frac{\partial^2 L(\mathbf{X})}{\partial P\left(\boldsymbol{\alpha}_{lj}^{**}\right) \partial \hat{P}\left(\boldsymbol{\alpha}_{l}^{**}\right)} \right\}, \tag{8}$$

where $\mathbf{p}_j = \{P(\boldsymbol{\alpha}_{lj}^{**})\}$. As with the G-DINA model, additional steps can be taken to convert the item parameters of the saturated models to those of the reduced models (de la Torre & Chen, 2011). Finally, for person classification into individual attributes, the marginal modal assignment can be adopted. That is, examinee i will be classified to level m of attribute k if the marginal posterior probability of level m is the largest relative to other marginal posterior probabilities; that is, $p(\alpha_k = m|\mathbf{X}_i) > p(\alpha_k = m'|\mathbf{X}_i)$, where $m \neq m'$.

Simulation Study

Design

To systematically evaluate the viability of the pG-DINA model, a two-part simulation study was designed. The first part examined how well the item parameters can be recovered as a means of verifying whether the proposed model behaved as expected; the second part compared the accuracy of the person classification of the pG-DINA model with that of the G-DINA model modified for polytomous attributes (mG-DINA). The latter was included to demonstrate the advantages, if any, of using the pG-DINA model over a simple dichotomization of the attributes. It should be noted that the only difference between the mG-DINA and G-DINA models is that in the mG-DINA model, all the nonzero entries of the Q-matrix were changed to 1. This is a simple way of converting Q-matrices constructed for polytomous attributes so they can be used in conjunction with models for dichotomous attributes. However, when compared with the pG-DINA model, the person classification rule for the mG-DINA model required some adjustments (see the following details).

For the simulation study, the numbers of attributes and attribute levels were fixed at K = 5 and M = 3, respectively, and the M^K attribute vectors were assumed to be equally likely. Three sample sizes, N = 500, 1,000, and 2,000, and two test lengths, J = 15 and 30, were considered. Two types of Q-matrices with different maximum number of required attributes $(K_j^*)_{\max} = 2$ and 3 were evaluated and they represented relatively less and more complex Q-matrices, respectively. These Q-matrices are given in Tables 3 and 4, and are labeled as Q1 and Q2, respectively. The Q-matrix for J = 15 represented a subset of the Q-matrix for J = 30. In all the four types of Q-matrices, the attributes and the levels within an attribute were specified to occur an equal number of times. In addition, where they existed there was an equal number of one-, two-, and three-attribute items.

Two generating models were considered in the space of $\boldsymbol{\alpha}_{lj}^{**}$: DINA and A-CDM. Reduced models were considered because they are more interpretable in practice than the saturated models. The generating values of model parameters were set so that $P(\boldsymbol{\alpha}_{lj}^{**})_{\min} = 0.10$ and $P(\boldsymbol{\alpha}_{lj}^{**})_{\max} = 0.90$ for both models. For A-CDM, the impact of δ_{jk} was set such that each required attribute contributed equally to the students' success probability. The Q-matrices and the item parameters followed a symmetric design so that attributes can be discussed interchangeably. A beta prior with Parameters 1 and 2 was used to estimate $P(\boldsymbol{\alpha}_{lj}^{**})$ and stabilize the results. Combining the levels of the different factors, that is, N, J, $(K_j^*)_{\max}$, Q-matrix complexity,

Table 3. Q1 Type of Q-matrices—$(K_j^*)_{max} = 2$.

Item	Attribute				
	α_1	α_2	α_3	α_4	α_5
1[a]	1	0	0	0	0
2[a]	0	1	0	0	0
3[a]	0	0	1	0	0
4[a]	0	0	0	1	0
5[a]	0	0	0	0	1
6[a]	2	0	0	0	0
7[a]	0	2	0	0	0
8[a]	0	0	2	0	0
9[a]	0	0	0	2	0
10[a]	0	0	0	0	2
11[a]	1	2	0	0	0
12[a]	0	1	2	0	0
13[a]	0	0	1	2	0
14[a]	0	0	0	1	2
15[a]	2	0	0	0	1
16	1	1	0	0	0
17	0	1	1	0	0
18	0	0	1	1	0
19	0	0	0	1	1
20	1	0	0	0	1
21	1	0	2	0	0
22	0	1	0	2	0
23	0	0	1	0	2
24	2	0	0	1	0
25	0	2	0	0	1
26	2	2	0	0	0
27	0	2	2	0	0
28	0	0	2	2	0
29	0	0	0	2	2
30	2	0	0	0	2

Note: Items with superscript "a" are used when J = 15.

generating model, resulted in 48 conditions. For each condition, 500 data sets were generated, and each data set was analyzed using the pG-DINA and mG-DINA models. The estimation code used in analyzing the data was custom-written in Ox (Doornik, 2003).

Item Parameter Estimates. This part of the simulation study examined whether the item parameters of different CDMs can be estimated accurately when formulated under the pG-DINA model. The mean estimates of $P(\boldsymbol{\alpha}_{lj}^{**})$ across the replications and corresponding *SE*s were examined. Specifically, the bias of the estimate, the empirical *SE* (i.e., standard deviation [*SD*] across the replications), the computed *SE* (i.e., the root mean squared *SE* across the replications), and the ratios between the *SE*s (i.e., empirical *SE*/computed *SE*) were discussed.

Comparison of Classification Accuracy. This part of the simulation study investigated whether the pG-DINA model, which requires additional input from substantive experts, provided more accurate person classification compared with the mG-DINA model. However, it should be noted that such comparison cannot be carried out directly because, in its traditional application, the G-DINA model, and accordingly the mG-DINA model, provides only dichotomous

Table 4. Q2 Type of Q-Matrices—$(K_j^*)_{max} = 3$.

	Attribute				
Item	α_1	α_2	α_3	α_4	α_5
1[a]	1	0	0	0	0
2[a]	0	1	0	0	0
3[a]	0	0	1	0	0
4[a]	0	0	0	1	0
5[a]	0	0	0	0	1
6[a]	1	2	0	0	0
7[a]	0	1	2	0	0
8[a]	0	0	1	2	0
9[a]	0	0	0	1	2
10[a]	2	0	0	0	1
11[a]	2	2	0	1	0
12[a]	2	1	0	0	2
13[a]	1	0	2	2	0
14[a]	0	2	1	0	2
15[a]	0	0	2	2	1
16	2	0	0	0	0
17	0	2	0	0	0
18	0	0	2	0	0
19	0	0	0	2	0
20	0	0	0	0	2
21	2	0	2	0	0
22	0	2	0	2	0
23	0	0	2	0	2
24	2	0	0	2	0
25	0	2	0	0	2
26	1	0	0	1	1
27	0	1	1	1	0
28	1	1	1	0	0
29	0	1	0	1	1
30	1	0	1	0	1

Note: Items with superscript "a" are used when J = 15.

classification for each attribute. For classification purposes, the marginal posterior probabilities, $p(\alpha_k|X_i)$, obtained using the G-DINA model were converted to 1 when greater than cut point .5, and set to 0 otherwise. This rule can be modified to accommodate polytomous attributes. In this study, the cut points 1/3 and 2/3 were used to convert $p(\alpha_k|X_i)$ into the Attribute Levels 0, 1, and 2 in the mG-DINA model.

The classification accuracy for the individual attribute, CA(α_k), and for the attribute vector, CA(α_l), under the pG-DINA and mG-DINA models was compared. The comparisons were given under two methods of classification: (a) exact classification and (b) weighted classification. For both methods, $\mathrm{CA}(\alpha_k) = \sum_{i=1}^{N} W_{ik}/N$ and $\mathrm{CA}(\boldsymbol{\alpha}_l) = \sum_{i=1}^{N} \prod_{k=1}^{K} W_{ik}/N/K$. For exact classification,

$$W_{ik} = 1 - I(\hat{\alpha}_{ik} = \alpha_{ik}), \tag{9}$$

and for weighted classification,

$$W_{ik} = \begin{cases} 1/2^{|\alpha_{ik} - \hat{\alpha}_{ik}|} & \text{if} |\hat{\alpha}_{ik} - \alpha_{ik}| < M - 1 \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where $I(\cdot)$ is the indicator function, α_{ik} is the generating levels of attribute k for examinee i, $\hat{\alpha}_{ik}$ is its estimate, and M is the attribute levels, which is 3 in this case. The weighted classification takes into account the degree of misclassification, whereas the exact classification does not. For example, a true attribute level of 2 when estimated as 1 is deemed better than when estimated as 0 under the weighted classification method, whereas both estimates are deemed equally incorrect under the exact classification method.

Results

The authors found similar results regardless of whether the generating model was the DINA model or the A-CDM. They also found similar results for the exact and weighted classification methods. Due to space limitations, they present only the results when DINA was the generating model, and the exact classification method was used. Moreover, only the results for some selected items across different K_j^*, N, J, and types of Q-matrices are presented. Because of the symmetry in the design, the results for other items were similar to those in the same conditions. The results in their entirety can be requested from the authors.

Item Parameter Estimates. Table 5 gives the bias of the estimates of selected items. This table shows that in most cases the mean estimates were close to the generating values—the magnitudes of bias were larger than 0.05 in only 7 out of 120 cases. Almost all such biases were associated with relatively small N (i.e., 1,000 or below), small J (i.e., 15), and more complex Q-matrices (i.e., Q2) in both models. As expected, the magnitude of bias tended to decrease with larger N or J, smaller K_j^*, or less complex Q-matrices. As a whole, these findings indicate that the parameters can be accurately estimated across different sample sizes, test lengths, and number of required attributes using the pG-DINA model. The exception was when Q2 and $J = 15$ were used, and the sample size was 1,000 or smaller.

The empirical *SE*s of estimates are given in Table 6. Empirical *SE* indicates the sampling variability of the estimates, and accordingly, a larger value suggests more unstable estimates. The largest values were associated with the Q-matrix Q2 and $J = 15$ for both models. Empirical *SE*s tended to be larger for the A-CDM compared with the DINA model under similar conditions (A-CDM results not presented). This was a reasonable finding due to higher number of model parameters and larger magnitude of true generating parameters in the A-CDM. The empirical *SE*s tended to decrease in both models when N or J was larger, K_j^* was smaller, or the Q-matrix was less complex.

Table 7 shows the ratio of the empirical *SE* and the computed *SE*. When the computed *SE* accurately reflected the sampling variability of the estimates, the ratio was close to 1. As the results show, the ratios were close to 1 in most cases—with the values between 2/3 and 3/2 in 110 out of 120 cases. Almost all ratios smaller than 2/3 or larger than 3/2 were associated with the Q2 and $J = 15$. In most cases, the ratios were larger than 1, suggesting that the computed *SE* tended to be smaller (i.e., less conservative) relative to the sampling variability. As expected, the ratios tended to be closer to 1 when n or J was larger, or the Q-matrix was less complex. However, a clear pattern with respect to the effect of K_j^* on the ratios cannot be discerned. Overall, these findings indicate that the estimates were stable across replications, and the computed *SE*s accurately reflected the true sampling variability across different sample sizes, test lengths, and number of required attributes, except when the Q-matrix was Q2 and $J = 15$.

Comparison of Classification Accuracy. Table 8 gives the mean of the classification accuracy CA(α_k) and CA(α_l), and the related *SD* across all replications, when pG-DINA was the fitted model using the exact classification method. For CA(α_k), the mean was between 85% and 93% across all conditions except for Q2 and $J = 15$, where the value was between 75% and 78%.

Table 5. Biases of Mean Estimates of Selected Item for the DINA Model.

				$P(\boldsymbol{\alpha}_{lj}^{**})$							
Type of Q-matrix	J	N	K_j^*	0 00 000	1 01 001	10 010	11 011	100	101	110	111
Q1	15	500	1	−0.02	0.01						
			2	0.01	0.04	0.00	−0.02				
		1,000	1	−0.01	0.00						
			2	0.00	0.02	0.00	−0.01				
		2,000	1	0.01	0.00						
			2	0.00	0.01	0.00	0.00				
	30	500	1	0.00	0.00						
			2	0.01	0.02	0.00	−0.01				
		1,000	1	0.00	0.00						
			2	0.00	0.01	0.00	0.00				
		2,000	1	0.00	0.00						
			2	0.00	0.00	0.00	0.00				
Q2	15	500	1	−0.02	0.02						
			2	0.00	*0.14*	0.01	0.00				
			3	0.01	0.00	**0.06**	0.05	0.04	0.03	*0.13*	−**0.06**
		1,000	1	−0.01	0.02						
			2	0.00	**0.08**	0.00	0.00				
			3	0.00	0.00	0.03	0.03	0.02	0.02	**0.07**	−0.04
		2,000	1	0.01	0.01						
			2	0.00	0.03	0.00	0.00				
			3	0.00	0.00	0.01	0.01	0.01	0.00	0.04	−0.02
	30	500	1	0.00	0.00						
			2	0.01	0.02	0.00	−0.01				
			3	0.01	0.00	0.02	0.01	0.03	0.01	**0.06**	−0.03
		1,000	1	0.00	0.00						
			2	0.00	0.01	0.00	−0.01				
			3	0.01	0.00	0.01	0.01	0.01	0.01	0.03	−0.02
		2,000	1	0.00	0.00						
			2	0.00	0.01	0.00	0.00				
			3	0.00	0.00	0.01	0.00	0.00	0.00	0.02	−0.01

Note: Biases given in boldface are larger than 0.05, and biases given in italics are larger than 0.1.

For CA(α_l), the mean was between 45% and 70% across all conditions except for Q2 and J = 15, where the value was only between 27% and 30%. The mean was always better for the DINA model than for the A-CDM (again, A-CDM results not presented), but in most cases, the difference was within 5% with respect to an individual attribute, α_k. Although the accuracy tended to increase as the sample size increased, the increment was relatively small—within 2% for CA(α_k) and 4% for CA(α_l) from N = 500 to 2,000. When cases under Condition Q2 and J = 15 were excluded, the improvement of CA(α_k) from J = 15 to 30 was within 8%, and the improvement from Q2 to Q1 was within 2%. The *SD*s were small, suggesting stable estimates across replications. As a whole, these findings indicate that, when cases of Q2 and J = 15 were excluded, the classification accuracy for the individual attributes was high and stable across different sample sizes, test length, and Q-matrices, whereas the classification accuracy for the attribute vector can be considered moderate.

When the weighted classification method was used (tabulated results not presented), the authors found similar results, except that accuracy was considerably higher compared with the

Table 6. Empirical *SE*s of Estimates of Selected Item for the DINA Model.

				$P(\boldsymbol{\alpha}_{lj}^{**})$							
Type of Q-matrix	J	N	K_j^*	0 00 000	1 01 001	10 010	11 011	100	101	110	111
Q1	15	500	1	0.04	0.03						
			2	0.04	0.06	0.03	0.05				
		1,000	1	0.04	0.02						
			2	0.03	0.05	0.02	0.04				
		2,000	1	0.03	0.01						
			2	0.02	0.04	0.02	0.03				
	30	500	1	0.03	0.02						
			2	0.03	0.05	0.02	0.03				
		1,000	1	0.02	0.01						
			2	0.02	0.03	0.02	0.02				
		2,000	1	0.02	0.01						
			2	0.02	0.02	0.01	0.02				
Q2	15	500	1	0.05	0.03						
			2	0.05	**0.18**	0.04	0.05				
			3	0.06	0.05	**0.12**	0.09	0.09	0.07	**0.15**	0.08
		1,000	1	0.04	0.03						
			2	0.04	**0.13**	0.03	0.04				
			3	0.05	0.04	0.09	0.07	0.06	0.05	0.09	0.07
		2,000	1	0.04	0.02						
			2	0.03	0.08	0.02	0.04				
			3	0.03	0.03	0.06	0.05	0.04	0.03	0.07	0.05
	30	500	1	0.03	0.02						
			2	0.03	0.05	0.02	0.04				
			3	0.04	0.03	0.05	0.04	0.06	0.04	0.08	0.06
		1,000	1	0.02	0.01						
			2	0.02	0.03	0.02	0.03				
			3	0.03	0.02	0.04	0.03	0.04	0.03	0.06	0.04
		2,000	1	0.02	0.01						
			2	0.02	0.03	0.01	0.02				
			3	0.02	0.01	0.03	0.02	0.03	0.02	0.04	0.03

Note: *SE*s = standard errors; Values given in boldface are larger than 0.1.

exact method. This was expected as the weighted method is less stringent. The increment can be up to 10% and 20% for CA(α_k) and CA(α_l), respectively.

Table 9 presents the improvement in classification accuracy for individual attribute and the attribute vectors gained by moving from the mG-DINA model to the pG-DINA model using the exact classification method. The improvements were dramatic particularly when Q2 and J = 15 were excluded, ranging from 15% to 26% for individual attributes and from 27% to 57% for the attribute vector. The improvements were considerably larger when the generating model was DINA compared with the A-CDM, or when the test length was longer (for Q1 only); however, the improvements were largely independent of the sample size or type of Q-matrix (when J = 30). Overall, these findings indicate that the improvements in classification accuracy from the pG-DINA to mG-DINA model can range from considerable to dramatic, except for Q2 and J = 15, and improvements for attribute vector were especially prominent.

Table 7. Ratios of Empirical *SE*s Over Computed *SE*s of Selected Item for the DINA Model.

				$P(\boldsymbol{\alpha}_{lj}^{**})$							
				0	1						
				00	01	10	11				
Type of Q-matrix	J	N	K_j^*	000	001	010	011	100	101	110	111
Q1	15	500	1	0.84	1.26						
			2	0.99	0.93	1.30	1.18				
		1,000	1	1.06	1.23						
			2	1.03	1.01	1.32	1.20				
		2,000	1	1.27	1.16						
			2	1.01	1.11	1.37	1.31				
	30	500	1	1.04	1.10						
			2	0.92	0.89	1.08	1.00				
		1,000	1	1.09	1.07						
			2	1.01	0.94	1.02	1.00				
		2,000	1	1.05	1.02						
			2	0.96	0.96	1.02	1.01				
Q2	15	500	1	0.87	1.32						
			2	1.43	**1.92**	1.33	0.90				
			3	1.30	**1.61**	1.10	1.38	0.90	1.16	0.81	0.70
		1,000	1	1.09	**1.51**						
			2	**1.55**	*2.10*	1.34	1.06				
			3	1.35	**1.68**	1.37	**1.62**	1.05	1.32	0.83	0.85
		2,000	1	1.22	1.37						
			2	1.49	**1.83**	1.40	1.21				
			3	1.27	1.48	1.29	**1.51**	1.07	1.26	0.92	0.99
	30	500	1	1.10	1.09						
			2	0.99	0.92	1.04	1.03				
			3	0.89	1.00	0.73	0.94	0.77	0.94	**0.59**	0.74
		1,000	1	1.08	1.10						
			2	0.98	0.93	1.08	1.10				
			3	1.01	1.00	0.92	0.95	0.89	0.97	0.74	0.92
		2,000	1	1.07	1.03						
			2	1.00	1.00	1.06	1.03				
			3	0.94	1.03	0.95	0.97	0.94	1.03	0.88	0.97

Note: *SE*s = standard errors; Ratios given in boldface are smaller than 2/3 or larger than 1.5, and ratios given in italics are smaller than 0.5 or larger than 2.0.

When the weighted classification method was used (tabulated results not presented), the improvement was less dramatic compared with the exact method, but still considerable. The improvements ranged up to 13% for individual attributes and up to 44% for the attribute vector.

Real-Data Example

In the introduction of this article, the authors presented two polytomous attributes in a proportional reasoning assessment for the eighth-grade students (de la Torre et al., 2010). In addition, two dichotomous attributes were measured by the assessment: *identifying a multiplicative relationship between sets of quantities* and *applying algorithms to solve a proportional reasoning problem*. A set of 15 items had been developed by content experts to provide diagnostic information relative to the four attributes. The assessment was completed by 393 students using a multiple-choice format. Due to the small sample size, a beta prior with parameters 0.5 and 1

Table 8. Classification Accuracy Using pG-DINA, Exact Classification.

Type of Q-matrix	J	N	Accuracy α_1	α_2	α_3	α_4	α_5	α_l[a]	SD α_1	α_2	α_3	α_4	α_5	α_l[a]
Q1	15	500	0.85	0.85	0.85	0.85	0.85	0.45	0.02	0.02	0.02	0.02	0.02	0.02
		1,000	0.85	0.85	0.85	0.85	0.85	0.46	0.01	0.01	0.01	0.01	0.01	0.02
		2,000	0.85	0.85	0.85	0.85	0.85	0.46	0.01	0.01	0.01	0.01	0.01	0.01
	30	500	0.92	0.92	0.92	0.92	0.92	0.68	0.01	0.01	0.01	0.01	0.01	0.02
		1,000	0.92	0.92	0.92	0.92	0.92	0.69	0.01	0.01	0.01	0.01	0.01	0.01
		2,000	0.93	0.92	0.93	0.93	0.93	0.70	0.01	0.01	0.01	0.01	0.01	0.01
Q2	15	500	0.76	0.76	0.77	0.75	0.77	0.27	0.02	0.02	0.02	0.02	0.02	0.02
		1,000	0.77	0.77	0.78	0.76	0.78	0.28	0.01	0.02	0.01	0.02	0.02	0.02
		2,000	0.78	0.77	0.78	0.76	0.78	0.30	0.01	0.01	0.01	0.01	0.01	0.01
	30	500	0.90	0.90	0.90	0.90	0.90	0.63	0.02	0.01	0.01	0.01	0.01	0.02
		1,000	0.91	0.90	0.91	0.91	0.90	0.64	0.01	0.01	0.01	0.01	0.01	0.02
		2,000	0.91	0.91	0.91	0.91	0.91	0.65	0.01	0.01	0.01	0.01	0.01	0.01

Note: SD = standard deviation.
For the DINA model only.
[a]Attribute vector.

Table 9. Improvement of Classification Accuracy (pG-DINA–mG-DINA), Exact Classification.

J	N	Q1 α_1	α_2	α_3	α_4	α_5	$\boldsymbol{\alpha}$[a]	Q2 α_1	α_2	α_3	α_4	α_5	$\boldsymbol{\alpha}$[a]
15	500	0.21	0.22	0.21	0.21	0.21	0.35	0.13	0.13	0.19	0.12	0.14	0.18
	1,000	0.22	0.22	0.22	0.22	0.22	0.36	0.15	0.16	0.21	0.10	0.14	0.19
	2,000	0.22	0.22	0.22	0.22	0.22	0.36	0.11	0.11	0.22	0.10	0.12	0.19
30	500	0.25	0.25	0.25	0.26	0.25	0.56	0.25	0.24	0.25	0.24	0.24	0.51
	1,000	0.26	0.26	0.26	0.26	0.26	0.57	0.25	0.24	0.25	0.24	0.25	0.53
	2,000	0.26	0.26	0.26	0.26	0.26	0.57	0.25	0.24	0.26	0.24	0.25	0.53

Note: mG-DINA = G-DINA model modified for polytomous attributes.
For the DINA model only.
[a]Attribute vector.

was used to estimate the item parameters. As shown in Table 10, the Q-matrix consists of 15 items with two polytomous attributes, α_1 and α_2, and two dichotomous attributes, α_3 and α_4. This item-attribute-sample size configuration was comparable with the worst condition in this simulation study. The definitions of polytomous attribute levels can be found in Table 1. This Q-matrix was used with the pG-DINA model, and the Q-matrix and model combination was called M1. An alternative way of analyzing the data was to convert each of the three-level polytomous attributes into two dichotomous attributes, with the first attribute corresponding to the mastery or nonmastery of Level 1 and the second to the mastery and nonmastery of Level 2. A new Q-matrix with six dichotomous attributes was created and used with the G-DINA model. This Q-matrix and model combination was called M2, and was compared with M1 in the following analyses.

Table 11 presents the results of fitting M1 and M2 to the proportional reasoning data. As the results show, the -2 log likelihood of M2 was slightly better. This was expected because M2 was more general in both the structural and measurement models. After taking into account the

Table 10. Q-Matrix of the Proportional Reasoning Data.

Item	α_1	α_2	α_3	α_4
1	0	1	0	0
2	0	0	0	1
3	0	2	0	1
4	0	0	1	0
5	2	0	0	0
6	0	1	0	0
7	0	2	0	0
8	2	0	0	0
9	1	1	0	0
10	1	1	0	0
11	0	0	1	1
12	1	1	0	0
13	2	0	0	0
14	2	1	0	0
15	2	1	0	0

Note: α_1 = comparing and ordering of fractions; α_2 = constructing ratios and proportions from a situation; α_3 = identifying a multiplicative relationship between sets of quantities; α_4 = applying algorithms to solve a proportional reasoning problem.

Table 11. M1 and M2 Model Fit to the Proportional Reasoning Data.

Model	−2LL	AIC	BIC	NP
M1	6,489	6,647	6,961	79
M2	6,483	6,697	7,122	107

Note: −2LL = −2 log likelihood; AIC = Akaike's information criterion; BIC = Bayesian information criterion; NP = number of parameters.

number of model parameters, the Akaike's information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarzer, 1976) indicated that M1 had better fit. The results discussed next pertain to those of M1.

Table 12 shows the proportion of examinees at different levels of the attributes. At least, 50% of the examinees have some mastery of each attribute, with α_1 being the lowest at 50% and α_2 being the highest at 66%. However, of the four attributes, α_4 has the highest full mastery at 63%. The majority (more than 80%) of the examinees with at least partial masteries were at Level 2 of the polytomous attributes.

Table 13 gives the item parameter estimates (i.e., probability of success for reduced attribute patterns) and the corresponding *SE*s based on M1. Most of the items provided adequate diagnostic information. For 12 of the 15 items, the difference in the probabilities of success between examinees who have all the required attribute levels and those who have none, as in, $P(\mathbf{1}) - P(\mathbf{0})$, was at least 0.25, with a mean difference of 0.49. Item 1 provided the lowest diagnostic information, with $P(\mathbf{1}) - P(\mathbf{0})$ being only 0.11. This item had a very high guessing parameter of 0.81, which could indicate attribute misspecification. To a lesser extent, the same can be said of Item 13. Items 8, 11, and 14 had relatively low $P(\mathbf{1})$ (i.e., less than 0.50), which could indicate that the attributes or attribute levels had been underspecified. Finally, for Items 3, 9, and 11,

Table 12. Proportion of Examinees in Different Levels of the Proportional Reasoning Attributes.

α_1		α_2		α_3	α_4
L1	L2	L1	L2	L1	L1
0.07	0.43	0.11	0.55	0.54	0.63

Note: α_1 = comparing and ordering of fractions; α_2 = constructing ratios and proportions from a situation; α_3 = identifying a multiplicative relationship between sets of quantities; α_4 = applying algorithms to solve a proportional reasoning problem.

Table 13. Item Parameters ($P(\boldsymbol{\alpha}_{lj}^{**})$) for M1.

	Estimate				*SE*			
Item	0 00	1 01	10	11	0 00	1 01	10	11
1	0.81	0.92			0.04	0.02		
2	0.02	0.94			0.04	0.02		
3	0.28	0.86	0.21	0.95	0.05	0.05	0.11	0.02
4	0.40	0.75			0.04	0.04		
5	0.12	0.82			0.03	0.04		
6	0.38	0.92			0.05	0.02		
7	0.36	0.97			0.04	0.02		
8	0.16	0.41			0.03	0.04		
9	0.34	0.64	0.33	0.91	0.06	0.08	0.17	0.03
10	0.50	0.55	0.77	0.88	0.06	0.08	0.16	0.03
11	0.26	0.04	0.21	0.46	0.05	0.06	0.10	0.04
12	0.15	0.21	0.84	0.70	0.05	0.08	0.25	0.04
13	0.69	0.96			0.03	0.02		
14	0.22	0.26	0.34	0.49	0.04	0.06	0.20	0.04
15	0.24	0.36	0.70	0.80	0.05	0.07	0.20	0.04

Note: *SE* = standard error.

$P(\mathbf{0}) > P(10)$ or $P(01)$, and for Item 12, $P(10) > P(\mathbf{1})$. However, most of these differences were well within the *SE*s of the estimates, particularly when one considers the size of *SE*s of $P(10)$. Moreover, in multiple-choice items, it is possible for individuals who lack all the required attribute levels and guess at random to have a higher probability of success than those who master a subset of the required attribute levels.

To evaluate the validity of inferences from the proposed model, absolute fit statistics for dichotomous attributes by Chen, de la Torre, and Zhang (in press) can be readily extended to polytomous attributes. Specifically, the residual between the observed and predicted Fisher-transformed correlation of item pairs (referred to as r) and the residual between the observed and predicted log-odds ratios of item pairs (referred to as l) are evaluated. Denote $\mathbf{X}_j$ and $\tilde{\mathbf{X}}_j$ as the observed and predicted response vector for item j, respectively. For item j,

$$r_{jj'} = \left| Z\left[\mathrm{Corr}\left(\mathbf{X}_j, \mathbf{X}_{j'}\right)\right] - Z\left[\mathrm{Corr}\left(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_{j'}\right)\right] \right|, \tag{11}$$

$$l_{jj'} = \left| \log\left(\frac{N_{11}N_{00}}{N_{01}/N_{10}}\right) - \log\left(\frac{\tilde{N}_{11}\tilde{N}_{00}}{\tilde{N}_{01}/\tilde{N}_{10}}\right) \right|, \tag{12}$$

where $\tilde{N}$ is the sample size, generally large, used to generate the predicted response patterns, $j \neq j'$; $Z[\mathrm{Corr}(\cdot)]$ is the Fisher transformation of Pearson's correlation; and $N_{yy'}$ and $\tilde{N}_{yy'}$ are the number of observed and predicted examinees, respectively, who scored y on item j and y' on item j'. When the model fits the data adequately, both residuals should approach 0. The approximate *SE*s of r and l can be computed as

$$SE\left[r_{jj'}\right] = [N - 3]^{1/2}, \tag{13}$$

$$SE\left[l_{jj'}\right] = \left[\frac{\tilde{N}\left(1/\tilde{N}_{11} + 1/\tilde{N}_{00} + 1/\tilde{N}_{01} + 1/\tilde{N}_{10}\right)}{N}\right]^{1/2}. \tag{14}$$

With these *SE*s, the z scores of r and l can be derived to test their statistical significance. Chen et al. (in press) found similar performance for both statistics: They both have very high statistical power and conservative Type I error in most conditions after the Bonferroni correction.

For the preceding proportional reasoning data fitted with the pG-DINA model, the authors found that the z scores of the maximum r and l statistics are 2.88 and 2.47, respectively. Both are well below the critical z score of 3.30 at a significance level of 0.1 with the Bonferroni correction accounting for $15 \times 14 = 210$ possible tests for the different item pairs. These results indicate that the pG-DINA model provided an adequate fit to the proportional reasoning data.

Discussion

Expert-defined polytomous attributes can increase the practical usefulness of cognitively diagnostic assessments. CDMs with sufficient generality to accommodate such attributes across a wide range of settings are needed, and this research proposed a general model for such a purpose. The pG-DINA model allows input from substantive experts to specify attribute levels using SALM items. Parameters for the saturated model can be estimated using an EM implementation of the MML algorithm. Through a simulation study, the authors showed that, except for one Q-matrix, not only can item parameters be accurately estimated but also that the estimates of item parameters were stable, and the estimates of the *SE*s were accurate across various conditions. Moreover, the classification accuracy for individual attributes was relatively high and stable, whereas the accuracy for the attribute vector was moderate. The accuracy can be increased considerably if the weighted classification is adopted. The improvements of accuracy from the pG-DINA to mG-DINA model ranged from considerable to dramatic, with the improvements for the attribute vector especially prominent.

The authors also used empirical data to evaluate the pG-DINA model, and demonstrated that, for the data at hand, using expert-defined polytomous attributes was better than converting the attributes into multiple dichotomous attributes. Although the items developed for the test were generally diagnostic, the item parameter estimates indicate that some attributes might have been misspecified. Additional data are needed to examine whether such a pattern persists when a larger sample size is involved.

The poor results in the simulation study were largely confined to the conditions that involved Q2 and $J = 15$, and were independent of different sample sizes or generating models. The authors conjecture that that the problem is related to the structure of the Q-matrix. Specifically, several attribute patterns (e.g., 22220 and 22221) are not distinguishable using this Q-matrix. In the other three Q-matrices, one-attribute items at Levels 1 and 2 exist for each attribute.

Accordingly, all attribute patterns are distinguishable based on combinations of these items. For Q2 and $J = 15$, however, any one-attribute item at Level 2 was missing for all attributes, and the combinations of two- or three-attribute items were not enough to distinguish attribute patterns that cannot be completely distinguished using one-attribute items at Level 2. Accordingly, the model could have difficulty in correctly distinguishing examinees with those attribute patterns. As the results showed, this problem extended to the lack of accuracy of the item parameter estimates. More research is needed to determine how the modified Q-matrix for the pG-DINA model needs to be constructed to yield more accurate item parameter estimates and attribute classifications.

In addition to a better understanding of how the modified Q-matrix can be constructed, additional work is needed to better understand the proposed model and broaden the generalizability of the current findings. First, the authors only studied the saturated form of the pG-DINA model. In practice, reduced models (e.g., DINA or additive models for polytomous attributes), when they fit the data, are preferable because of their interpretability. Accordingly, their properties in conjunction with polytomous attributes are worth evaluating in future works. Second, both the generating and fitted models in the simulation were saturated in terms of the attribute distribution. In practice, however, attributes with different structures (e.g., higher order, hierarchical structure) can be encountered, and thus, it would be useful to evaluate how the proposed model can perform under varying attribute structures. Third, the SALM assumption was imposed on all items that involved polytomous attributes, which might be too strict for some item-attribute combinations. For instance, it is likely that a multiple-attribute item meets the SALM assumption for some attributes (i.e., with two latent groups of success and failure), but fails to meet the assumption for the other attributes (i.e., with more than two latent groups). For this reason, developing a method that verifies the reasonableness of this assumption would be desirable. Similarly, correctly specifying all the entries of the modified Q-matrix can be challenging for content experts. Extending the empirical Q-matrix validation procedure by de la Torre and Chiu (2010) to polytomous attributes could help address misspecification issues in the modified Q-matrix.

As practical settings with expert-defined polytomous attributes emerge more frequently, researchers will encounter both opportunities and challenges in using the pG-DINA model. As a general model, it has the ability to accommodate various CDMs to meet a wide range of requirements from applied researchers, with a few modifications at most. It is possible, in the context of expert-defined polytomous attributes, to also address issues that have already been addressed in the context of dichotomous attributes (e.g., higher order latent traits, multiple strategy). The proposed model also provides an opportunity to incorporate greater input from substantive experts, thereby making the model more relevant in practice. However, the variety of models it can accommodate and the greater expert contribution under the model can also be a challenge because they imply that a more extensive collaboration across disciplines is necessary. The collaboration between psychometric and substantive areas must take place across the different phases of the cognitive diagnostic assessment development, including attribute and attribute-level definition, item development, and various types of validation. Without a deliberate effort to bring together experts from different fields that can collaborate closely, the goal of developing assessments with diagnostic potential to improve classroom instruction and student learning may remain elusive.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Science Foundation Grant DRL-0744486.

References

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automated Control*, *19*, 716-723.

Chen, J., de la Torre, J., & Zhang, Z. (in press). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*.

Deely, J. J., & Lindley, D. V. (1981). Bayes Empirical Bayes. *Journal of the American Statistical Association*, *76*, 833-841.

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

de la Torre, J., & Chen, J. (2011, April). *Estimating different reduced cognitive diagnosis models using a general framework*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

de la Torre, J., & Chiu, C.-Y. (2010, May). *A general method of empirical Q-matrix validation using the G-DINA model discrimination index*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

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

de la Torre, J., Lam, D., Rhoads, K., & Tjoe, H. (2010, May). *Measuring grade 8 proportional reasoning: The process of attribute identification and task development and validation*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

DiBello, L., Roussos, L., & Stout, W. (2007). Cognitive diagnosis Part I. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics (Vol. 26): Psychometrics* (pp. 979-1030). Amsterdam, Netherlands: Elsevier.

Doornik, J. A. (2003). Object-Oriented Matrix Programming using Ox (Version 3.1) [Computer software]. London, England: Timberlake Consultants Press.

Haberman, S. J., von Davier, M., & Lee, Y.-H. (2008). *Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions* (ETS Research Report no. RR-05–16). Princeton, NJ: Educational Testing Service.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 333-352.

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

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

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

Karelitz, T. M. (2004). *Ordered category attribute coding framework for cognitive assessments* (Unpublished doctoral dissertation). University of Illinois at Urbana–Champaign.

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

National Assessment Governing Board. (2008). *Mathematics framework for the 2009 national assessment of educational progress*. Washington, DC: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Schwarzer, G. (1976). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345-354.

Templin, J. (2004). *Generalized linear proficiency models for cognitive diagnosis* (Unpublished doctoral dissertation). University of Illinois at Urbana–Champaign.

Templin, J., & Bradshaw, L. (in press). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*.

Templin, J., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287-305.

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (ETS Research Report no. RR-05-16). Princeton, NJ: Educational Testing Service.