PSYCHOMETRIKA—VOL. 76, NO. 2, 179–199
APRIL 2011
DOI: 10.1007/S11336-011-9207-7

THE GENERALIZED DINA MODEL FRAMEWORK

JIMMY DE LA TORRE

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

The G-DINA (*generalized deterministic inputs, noisy "and" gate*) model is a generalization of the DINA model with more relaxed assumptions. In its saturated form, the G-DINA model is equivalent to other general models for cognitive diagnosis based on alternative link functions. When appropriate constraints are applied, several commonly used cognitive diagnosis models (CDMs) can be shown to be special cases of the general models. In addition to model formulation, the G-DINA model as a general CDM framework includes a component for item-by-item model estimation based on design and weight matrices, and a component for item-by-item model comparison based on the Wald test. The paper illustrates the estimation and application of the G-DINA model as a framework using real and simulated data. It concludes by discussing several potential implications of and relevant issues concerning the proposed framework.

Key words: cognitive diagnosis, DINA, MMLE, parameter estimation, Wald test, model comparison.

1. Overview and Background

Cognitive diagnosis models (CDMs) are latent variable models developed primarily for assessing student mastery and non-mastery on a set of finer-grained skills. In the CDM literature, skills have been generically referred to as attributes, and are represented by the binary vector $\boldsymbol{\alpha}$. Several specific and general CDMs of various formulations have been proposed in the psychometric literature. Examples of specific CDMs include the *deterministic inputs, noisy "and" gate* (DINA; de la Torre, 2009b; Junker & Sijtsma, 2001), and the *reduced reparametrized unified model* (R-RUM; Hartz, 2002; Roussos, DiBello, Stout, Hartz, Henson, & Templin, 2007); examples of general CDMs include the *log-linear CDM* (Henson, Templin, & Willse, 2009), and the *general diagnostic model* (GDM; von Davier, 2005). However, it is not entirely clear whether the different CDM formulations represent different classes of models, or to what extent these models are related to one another. In addition, simultaneously estimating the parameters of the models with disparate formulations and comparing their relative fit at the item level have remained challenging tasks. To address these issues, this paper proposes a framework for (1) relating several CDMs with different formulations, (2) estimating the parameters of multiple CDMs specified within a single test, and (3) comparing the fit of constrained (i.e., specific) and unconstrained (i.e., general) CDMs one item at a time.

The framework will be based on the DINA model, which is one of, if not the simplest, consequently most restrictive, interpretable CDMs available for dichotomously scored test items. (For the purposes of this work, only items scored as right or wrong will be considered in this paper.) The DINA model is a parsimonious model that requires only two parameters for each item regardless of the number of attributes required for the item. It is appropriate when the tasks call for the conjunction of several equally important attributes, and lacking one required attribute for an item is the same as lacking all the required attributes (de la Torre & Douglas, 2004). In the DINA model, item j, $j = 1, \ldots, J$, on a test that measures K attributes partitions the 2^K latent attribute vectors or classes into two latent groups. Given the attribute requirements of item

Requests for reprints should be sent to Jimmy de la Torre, Department of Educational Psychology, Rutgers, The State University of New Jersey, 10 Seminary Place, New Brunswick, NJ 08901, USA. E-mail: j.delatorre@rutgers.edu

j, each of the 2^K attribute vectors is *deterministically* classified as belonging to either group η_{j1}, the group comprised of attribute vectors with all the required attributes for item j, or group η_{j0}, the group comprised of attribute vectors lacking at least one of the required attributes for the same item. With the DINA model, attribute vectors in the same group are assumed to have the same probability of answering the item correctly. Even if the attributes have been correctly identified and specified, this assumption may not always hold for group η_{j0}. That is, because attribute vectors in this group have varying degrees of deficiency with respect to the required attributes, their probabilities of success may not be identical. The paper proposes a generalization of the DINA model, which will be referred to as the *generalized DINA* (G-DINA) model. Specifically, the G-DINA model relaxes the DINA model assumption of equal probability of success for all attribute vectors in group η_{j0}. Moreover, with the use of appropriate design matrices and transformations, the G-DINA model can serve as a general framework for deriving other CDM formulations, estimating some commonly used CDMs, and testing the adequacy of constrained (reduced) models in place of the unconstrained (saturated) model.

The remaining sections of the paper are laid out as follows. The second section introduces the notation used in formulating the G-DINA model, and the role of link functions in formulating various general models for cognitive diagnosis. The third section demonstrates how several familiar CDMs can be represented as constrained cases of the general CDMs. Approaches for estimating the parameters of the constrained and unconstrained models, and a method of testing the adequacy of the fit of constrained models relative to the unconstrained models are given in Sections 4 and 5, respectively. Section 6 illustrates the applicability of the G-DINA model using fraction subtraction and psychiatric data. Lastly, the potential implications of the proposed framework on the practice of cognitive diagnosis modeling, and several relevant issues concerning the G-DINA model in particular and CDMs in general are discussed in Section 7.

2. A Generalization and Link Functions

Like the DINA model, the G-DINA model also requires a $J \times K$ Q-matrix (Tatsuoka, 1983). The element in row j and column k of the Q-matrix, q_{jk}, is equal to 1 if the kth attribute is required to answer item j correctly; otherwise it is equal to zero. However, instead of two groups, the G-DINA model partitions the latent classes into $2^{K_j^*}$ latent groups, where $K_j^* = \sum_{k=1}^{K} q_{jk}$ represents the number of required attributes for item j. For notational convenience but without loss of generality, let the first K_j^* attributes be the required attributes for item j, and $\boldsymbol{\alpha}_{lj}^*$ be the reduced attribute vector whose elements are the required attributes for item j. For example, if only the first two attributes are required for item j, then the attribute vector $\boldsymbol{\alpha}_{lj}$ reduces to $\boldsymbol{\alpha}_{lj}^* = (\alpha_{lj1}, \alpha_{lj2})'$. It would suffice for the purposes of this paper to consider the reduced attribute vector $\boldsymbol{\alpha}_{lj}^* = (\alpha_{lj1}, \ldots, \alpha_{ljK_j^*})'$ in place of the full attribute vector $\boldsymbol{\alpha}_{lj} = (\alpha_{lj1}, \ldots, \alpha_{ljK})'$. Using $\boldsymbol{\alpha}_{lj}^*$ reduces the number of latent groups to be considered for item j from 2^K to $2^{K_j^*}$.

Define the relationship between two attribute vectors $\boldsymbol{\alpha}_{lj}^*$ and $\boldsymbol{\alpha}_{l'j}^*$ as $\boldsymbol{\alpha}_{lj}^* \preccurlyeq \boldsymbol{\alpha}_{l'j}^*$ if and only if $\alpha_{lk} \le \alpha_{l'k}$, for $k = 1, \ldots, K_j^*$. Strict inequality between the attribute vectors is involved (i.e., $\boldsymbol{\alpha}_{lj}^* \prec \boldsymbol{\alpha}_{l'j}^*$) if $\alpha_{lk} < \alpha_{l'k}$ for at least one k. It should be noted that $\boldsymbol{\alpha}_{lj}^* \prec \boldsymbol{\alpha}_{l'j}^*$ implies $\sum_{k=1}^{K_j^*} \alpha_{ljk}^* < \sum_{k=1}^{K_j^*} \alpha_{l'jk}^*$ (i.e., a reduced vector that subsumes another reduced vector will contain more ones). However, its converse is not true. For example, there are fewer ones in the attribute vector $\boldsymbol{\alpha}_l = (001)'$ than in $\boldsymbol{\alpha}_{l'} = (110)'$. However, although $\alpha_{l1} < \alpha_{l'1}$ and $\alpha_{l2} < \alpha_{l'2}$, the inequality between the third elements is in the opposite direction. Hence, $\boldsymbol{\alpha}_l \nprec \boldsymbol{\alpha}_{l'}$. In this paper, when attribute pattern $\boldsymbol{\alpha}_{lj}^*$ is said to have fewer required attributes compared to $\boldsymbol{\alpha}_{l'j}^*$, it is taken to mean that $\boldsymbol{\alpha}_{lj}^* \prec \boldsymbol{\alpha}_{l'j}^*$, unless noted otherwise. Finally, the probability that examinees with attribute pattern

$\boldsymbol{\alpha}^*_{lj}$ will answer item j correctly is denoted by $P(X_j = 1|\boldsymbol{\alpha}^*_{lj}) = P(\boldsymbol{\alpha}^*_{lj})$. Without any constraints, the G-DINA model has $2^{K^*_j}$ parameters for item j, thus affording it greater generality compared to the DINA model whenever $K^*_j > 1$. In its most general formulation, the G-DINA model allows $P(\boldsymbol{\alpha}^*_{lj}) > P(\boldsymbol{\alpha}^*_{l'j})$ for $\boldsymbol{\alpha}^*_{lj} \prec \boldsymbol{\alpha}^*_{l'j}$. That is, examinees with fewer required attributes for an item can have a higher probability of answering the item correctly. For example, this can happen in situations where individuals mastering a subset of the required attributes for the item will pick a particular distractor in a multiple-choice item with a high probability, whereas individuals with no attribute mastery will guess at random; consequently, the probability of success is higher for the latter individuals. However, in many CDM applications, it would be reasonable to impose the constraint $P(\boldsymbol{\alpha}^*_{lj}) \leq P(\boldsymbol{\alpha}^*_{l'j})$ whenever $\boldsymbol{\alpha}^*_{lj} \prec \boldsymbol{\alpha}^*_{l'j}$.

Several link functions that are linear in the parameters can be used in specifying general models for cognitive diagnosis. The specific cases of these link functions result in some recognizable CDMs. Three link functions will be considered in this paper, namely, *identity*, *logit*, and *log*. In their saturated forms, all the resulting models have $2^{K^*_j}$ parameters for item j, and provide identical model-data fit. The following notations will be used for each link function.

For the identity link, the original formulation of the G-DINA model based on $P(\boldsymbol{\alpha}^*_{lj})$ can be decomposed into the sum of the effects due the presence of specific attributes and their interactions. Specifically,

$$P\big(\boldsymbol{\alpha}^*_{lj}\big) = \delta_{j0} + \sum_{k=1}^{K^*_j} \delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K^*_j} \sum_{k=1}^{K^*_j-1} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} \ldots + \delta_{j12\cdots K^*_j} \prod_{k=1}^{K^*_j} \alpha_{lk}, \tag{1}$$

where

δ_{j0}	is the intercept for item j;
δ_{jk}	is the main effect due to α_k;
$\delta_{jkk'}$	is the interaction effect due α_k and $\alpha_{k'}$; and
$\delta_{j12\cdots K^*_j}$	is the interaction effect due to $\alpha_1, \ldots, \alpha_{K^*_j}$.

These parameters can be interpreted as follows: δ_0 represents the baseline probability (i.e., probability of a correct response when none of the required attributes is present); δ_k is the change in the probability of a correct response as a result of mastering a single attribute (i.e., α_k); $\delta_{kk'}$, a first-order interaction effect, is the change in the probability of a correct response due to the mastery of both α_k and $\alpha_{k'}$ that is over and above the additive impact of the mastery of the same two attributes; and $\delta_{12\cdots K^*_j}$ represents the change in the probability of a correct response due the mastery of all the required attributes that is over and above the additive impact of the main and lower-order interaction effects. The intercept is always non-negative, the main effects are *typically* non-negative, but the interaction effects can take on any values. The main effects are non-negative if $P(\mathbf{0}_{K^*_j}) \leq P(\boldsymbol{\alpha}^*_{lj})$ for $\sum_{k=1}^{K^*_j} \alpha^*_{ljk} = 1$, where $\mathbf{0}_{K^*_j}$ is the null vector of length K^*_j. This implies that mastering any one of the required attributes corresponds to some increase in an individual's probability of success on the item.

For the logit link,

$$\operatorname{logit}\big[P\big(\boldsymbol{\alpha}^*_{lj}\big)\big] = \lambda_{j0} + \sum_{k=1}^{K^*_j} \lambda_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K^*_j} \sum_{k=1}^{K^*_j-1} \lambda_{jkk'}\alpha_{lk}\alpha_{lk'} \ldots + \lambda_{j12\cdots K^*_j} \prod_{k=1}^{K^*_j} \alpha_{lk}. \tag{2}$$

The logit link results in a general model that can be referred to as the *log-odds CDM*, which is equivalent to the log-linear CDM, and can be viewed as a special case of the GDM.

For the log link,

$$\log P(\boldsymbol{\alpha}_{lj}^*) = \nu_{j0} + \sum_{k=1}^{K_j^*} \nu_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \nu_{jkk'}\alpha_{lk}\alpha_{lk'} \ldots + \nu_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}. \tag{3}$$

The resulting model from the log link function can be referred to as the *log CDM*. Although the three general CDMs have similar formulations, the specifications of these models describe different phenomena: the G-DINA model and logit CDM describe the additive impact of attribute mastery on the probability and logit of the probability of success, respectively, whereas the log CDM describes the multiplicative impact of attribute mastery on the probability of success. Noting this difference is crucial because, as the next sections will show, applying the same constraints (e.g., additive constraint) to the different link functions will result in different reduced models (i.e., models that provide different statistical goodness of fit).

3. Special Cases

3.1. *G-DINA Model: The DINA Model, DINO Model and Additive CDM*

In the DINA model, the item response function (IRF) is given by

$$P(\boldsymbol{\alpha}_{lj}^*) = \begin{cases} g_j & \text{if } \boldsymbol{\alpha}_{lj}^* \prec \mathbf{1}_{K_j^*}, \\ 1 - s_j & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathbf{1}_{K_j^*}$ is a vector of ones and of length K_j^*, g_j is the probability that individuals who lack at least one of the prescribed attributes for item j will guess correctly, and $1 - s_j$ is the probability that individuals who have all the required attributed will not slip and get the item wrong. As Junker and Sijtsma (2001) noted, the terms guessing and slipping are used for mnemonic purposes because they encompass other reasons (e.g., inappropriate Q-matrix specification, use of alternative strategies) why individuals who lack some required attributes may answer the item correctly, whereas individuals who possess all the required attributes may answer the item incorrectly. In the DINA model, except for the attribute vector $\boldsymbol{\alpha}_j^* = \mathbf{1}_{K_j^*}$, the $2^{K_j^*} - 1$ latent groups have identical probability of correctly answering item j. The DINA model can be obtained from the G-DINA model by setting all the parameters, except δ_0 and $\delta_{12\cdots K_j^*}$, to zero. In terms of the G-DINA parameters, $g_j = \delta_{j0}$ and $1 - s_j = \delta_{j0} + \delta_{j12\cdots K_j^*}$. As such, the DINA model has two parameters per item. As can be seen from the model formulation, this model assumes that incremental probability can be expected only when all the required attributes are simultaneously mastered.

For the *deterministic input, noisy "or" gate* (DINO; Templin & Henson, 2006) model, the IRF is given by

$$P(\boldsymbol{\alpha}_{lj}^*) = \begin{cases} g_j' & \text{if } \boldsymbol{\alpha}_{lj}^* = \mathbf{0}_{K_j^*}, \\ 1 - s_j' & \text{otherwise.} \end{cases} \tag{5}$$

In contrast to the DINA model, all the latent groups, except $\mathbf{0}_{K_j^*}$, have the same probability of a correct response to item j. The DINO model can be obtained from the G-DINA model by setting

$$\delta_{jk} = -\delta_{jk'k''} = \cdots = (-1)^{K_j^*+1}\delta_{j12\cdots K_j^*},$$

for $k = 1, \ldots, K_j^*$, $k' = 1, \ldots, K_j^* - 1$, and $k'' > k', \ldots, K_j^*$. That is, in addition to the alternating sign which varies according to the order of interaction, the magnitudes of the main and interaction effects are also constrained to be identical to each other. In contrast to the DINA model, the guessing parameter of the DINO model, g'_j, refers to the probability of a correct response for individuals who have mastered *none* (as opposed to *only a subset*) of the required attributes, and $1 - s'_j$ refers to the probability of not slipping for individuals who have *at least one* (as opposed to *all*) of the required attributes. In terms of the G-DINA parameters, $g'_j = \delta_{j0}$ and $1 - s'_j = \delta_{j0} + \delta_{jk}$. Thus, the DINO model has two parameters per item as well. Incidentally, the DINA and DINO models can also be derived from the logit and log CDMs, albeit on different scales without the appropriate transformations, by applying the same constraints to $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$. For example, Henson, Templin, and Willse (2009) showed that the DINA and DINO models can be obtained from the saturated logs-odds CDM by setting all lower-order interaction terms to zero, and $\lambda_{jk} = -\lambda_{jk'k''} = \cdots = (-1)^{K_j^*+1}\lambda_{j12\cdots K_j^*}$, respectively.

Additional interpretable models can be obtained from the G-DINA model. For instance, by setting all the interaction effects to zero, we obtain a model that will be denoted as the *additive CDM* (*A*-CDM), which has the following IRF:

$$P(\boldsymbol{\alpha}_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk}. \tag{6}$$

This model indicates that mastering attribute α_k increases the probability of success on item j by δ_{jk}, and its contribution is independent of the contributions of the other attributes. The *A*-CDM has $K_j^* + 1$ parameters for item j.

3.2. *Logit CDM: The LLM Model*

The IRF of the linear logistic model (LLM; Hagenaars 1990, 1993, Maris, 1999) is given by

$$P(\boldsymbol{\alpha}_{lj}^*) = \frac{\exp(\lambda_{j0} + \sum_{k=1}^{K_j^*} \lambda_{jk}\alpha_{lk})}{1 + \exp(\lambda_{j0} + \sum_{k=1}^{K^*} \lambda_{jk}\alpha_{lk})}. \tag{7}$$

The logit of the probability of a correct response is, therefore,

$$\text{logit}\big[P(\boldsymbol{\alpha}_{lj}^*)\big] = \lambda_{j0} + \sum_{k=1}^{K_j^*} \lambda_{jk}\alpha_{lk}, \tag{8}$$

which is the log-odds CDM without the interaction terms. It can be seen from (8) that the LLM has $K_j^* + 1$ parameters for item j.

3.3. *Log CDM: The Generalized NIDA Model and Reduced RUM*

In the traditional formulation of the *noisy inputs, deterministic "and" gate* (NIDA; Junker & Sijtsma, 2001) model, its IRF can be written as

$$\begin{aligned} P(\boldsymbol{\alpha}_{lj}^*) &= \prod_{k=1}^{K} \big[g_k^{\alpha_{lk}}(1 - s_k)^{(1-\alpha_{lk})}\big]^{q_{jk}} \\ &= \prod_{k=1}^{K_j^*} g_k^{\alpha_{lk}}(1 - s_k)^{(1-\alpha_{lk})}. \end{aligned} \tag{9}$$

In the NIDA model, guessing and slipping occur at the attribute rather than item level. Attribute k can be successfully implemented either by guessing (when $\alpha_{lk}=0$) or not slipping (when $\alpha_{lk}=1$). The probabilities associated with a successful attribute implementation are g_k and $1-s_k$; the probability associated with a correct response for the item is equal to the product of the probabilities of successfully implementing each of the required attributes. In this formulation, the NIDA model is a very restricted model because it assumes the same g_k and s_k for all items. A straightforward extension of the NIDA model is the generalized NIDA (G-NIDA) model where the slip and guessing parameters are allowed to vary across the items. That is,

$$\begin{aligned} P\bigl(\boldsymbol{\alpha}^*_{lj}\bigr) &= \prod_{k=1}^{K^*_j} g_{jk}^{(1-\alpha_{lk})}(1-s_{jk})^{\alpha_{lk}} \\ &= \prod_{k=1}^{K^*_j} g_{jk} \times \prod_{k=1}^{K^*_j} \left(\frac{1-s_{jk}}{g_{jk}}\right)^{\alpha_{lk}}. \end{aligned} \tag{10}$$

Taking the logarithm of (10) gives us

$$\begin{aligned} \log\bigl[P\bigl(\boldsymbol{\alpha}^*_{lj}\bigr)\bigr] &= \sum_{k=1}^{K^*_j} \log(g_{jk}) + \sum_{k=1}^{K^*_j} \alpha_{lk} \log\left(\frac{1-s_{jk}}{g_{jk}}\right) \\ &= v_{j0} + \sum_{k=1}^{K^*_j} v_{jk}\alpha_{lk}, \end{aligned} \tag{11}$$

where $v_{j0}=\sum_{k=1}^{K^*_j}\log(g_{jk})$ and $v_{jk}=\log\bigl[(1-s_{jk})/g_{jk}\bigr]$. This is the log CDM without the interaction terms. The G-NIDA model has K^*_j+1 estimable parameters. The same model was discussed by Maris (1999).

The IRF for the R-RUM, which is based on RUM (Hartz, 2002), is given by

$$P(\boldsymbol{\alpha}_{lj}) = \pi^*_j \prod_{k=1}^{K} r^{*\,q_{jk}\times(1-\alpha_{lk})}_{jk}, \tag{12}$$

which can be written as

$$P\bigl(\boldsymbol{\alpha}^*_{lj}\bigr) = \pi^*_j \prod_{k=1}^{K} r^*_{jk} \times \prod_{k=1}^{K} \left(\frac{1}{r^*_{jk}}\right)^{\alpha_{lk}}. \tag{13}$$

It can be noted that by using $\boldsymbol{\alpha}^*_{lj}$ in place of $\boldsymbol{\alpha}_{lj}$ in (12), q_{jk} becomes implicit in (13). By setting $\pi^*_j\prod_{k=1}^{K} r^*_{jk}=\prod_{k=1}^{K^*_j} g_{jk}$ and $r^*_{jk}=g_{jk}/(1-s_{jk})$, it can be seen that the R-RUM is an alternative way of parametrizing the G-NIDA model. Hence, the R-RUM is also a special case of the log CDM.

Several details need to be noted at this point. First, although the A-CDM, LLM and the G-NIDA are all additive models and have the same number of parameters, they assume different underlying processes, and therefore will not provide identical model-data fit. Specifically, the A-CDM and G-NIDA model are additive models where mastery of an attribute has a *direct* and constant impact on the probability of a correct response—the former has an additive impact whereas the latter has a multiplicative impact. In contrast, the

LLM has a constant additive impact on the *logit* of the probability of a correct response. This makes the interpretations of the A-CDM and G-NIDA model, particularly the former, more straightforward. Second, of the three links, only the logit link includes an explicit normalization step (i.e., $\exp(\cdot)/[1+\exp(\cdot)]$) to ensure that $P(\boldsymbol{\alpha}_{lj}^{*}) \in [0,1]$. In comparison, without appropriate constraints (e.g., $0 \le P(\boldsymbol{\alpha}_{lj}^{*}) \le 1$), some probability estimates based on the reduced models of the G-DINA model and the log CDM may not be contained in the interval $[0,1]$.

4. Estimation

4.1. MMLE and the G-DINA Model

To obtain the parameter estimates of the G-DINA model using marginalized maximum likelihood estimation (MMLE), consider the log-marginalized likelihood of the response data, which can be written as

$$l(\boldsymbol{X}) = \log\bigl[L(\boldsymbol{X})\bigr] = \log \prod_{i=1}^{I} \sum_{l=1}^{L} L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l), \tag{14}$$

where $L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l) = \prod_{j=1}^{J} P(\boldsymbol{\alpha}_{lj})^{X_{ij}}[1-P(\boldsymbol{\alpha}_{lj})]^{(1-X_{ij})}$ is the likelihood of the response vector of examinee i given the attribute vector $\boldsymbol{\alpha}_l$, and $p(\boldsymbol{\alpha}_l)$ is the prior probability of $\boldsymbol{\alpha}_l$. In the G-DINA model the probability of a correct response on item j, $P(\boldsymbol{\alpha}_{lj})$, can be written as $P(\boldsymbol{\alpha}_{lj}^{*})$, where $P(\boldsymbol{\alpha}_{lj}^{*})$ is the reduced attribute vector form of $P(\boldsymbol{\alpha}_{lj})$. Thus, there are $2^{K_j^{*}}$ parameters to be estimated for item j.

By taking the derivative of $l(\boldsymbol{X})$ with respect to $P(\boldsymbol{\alpha}_{lj}^{*})$, and solving for $P(\boldsymbol{\alpha}_{lj}^{*})$, it can be shown after a few algebraic manipulations that the marginal maximum likelihood estimate of $P(\boldsymbol{\alpha}_{lj}^{*})$ is given by

$$\hat{P}\bigl(\boldsymbol{\alpha}_{lj}^{*}\bigr) = \frac{R_{\boldsymbol{\alpha}_{lj}^{*}}}{I_{\boldsymbol{\alpha}_{lj}^{*}}}, \tag{15}$$

where $I_{\boldsymbol{\alpha}_{lj}^{*}} = \sum_{i=1}^{I} p(\boldsymbol{\alpha}_{lj}^{*}|\boldsymbol{X}_i)$ is the number of examinees expected to be in the latent group $\boldsymbol{\alpha}_{lj}^{*}$, $R_{\boldsymbol{\alpha}_{lj}^{*}} = \sum_{i=1}^{I} p(\boldsymbol{\alpha}_{lj}^{*}|\boldsymbol{X}_i)X_{ij}$ is the number of examinees in the latent group $\boldsymbol{\alpha}_{lj}^{*}$ expected to answer item j correctly, and $p(\boldsymbol{\alpha}_{lj}^{*}|\boldsymbol{X}_i)$ represents the posterior probability that examinee i is in the latent group $\boldsymbol{\alpha}_{lj}^{*}$.

The second derivative of the log-marginalized likelihood with respect to $P(\boldsymbol{\alpha}_{lj}^{*})$ and $P(\boldsymbol{\alpha}_{l'j}^{*})$ is equal to

$$\frac{\partial l(\boldsymbol{X})}{\partial P(\boldsymbol{\alpha}_{lj}^{*})P(\boldsymbol{\alpha}_{l'j}^{*})} = -\sum_{i=1}^{I}\biggl[L^{-2}(\boldsymbol{X}_i)\frac{\partial L(\boldsymbol{X}_i)}{\partial P(\boldsymbol{\alpha}_{lj}^{*})}\frac{\partial L(\boldsymbol{X}_i)}{\partial P(\boldsymbol{\alpha}_{l'j}^{*})}\biggr]. \tag{16}$$

By taking the derivatives and simplifying, (16) can be shown to be equal to

$$-\sum_{i=1}^{I}\biggl\{p\bigl(\boldsymbol{\alpha}_{lj}^{*}|\boldsymbol{X}_i\bigr)\frac{X_{ij}-P(\boldsymbol{\alpha}_{lj}^{*})}{P(\boldsymbol{\alpha}_{lj}^{*})[1-P(\boldsymbol{\alpha}_{lj}^{*})]}\biggr\}\biggl\{p\bigl(\boldsymbol{\alpha}_{l'j}^{*}|\boldsymbol{X}_i\bigr)\frac{X_{ij}-P(\boldsymbol{\alpha}_{l'j}^{*})}{P(\boldsymbol{\alpha}_{l'j}^{*})[1-P(\boldsymbol{\alpha}_{l'j}^{*})]}\biggr\}. \tag{17}$$

Using $\hat{P}(\boldsymbol{\alpha}_{lj}^{*})$ and the observed $\boldsymbol{X}$ to evaluate (17), the approximate information matrix for the parameters of item j, $\boldsymbol{I}(\hat{\boldsymbol{P}}_j^{*})$, where $\boldsymbol{P}_j^{*} = \{P(\boldsymbol{\alpha}_{lj}^{*})\}$, can be obtained. The square-root of the lth diagonal element of $\boldsymbol{I}^{-1}(\hat{\boldsymbol{P}}^*_j)$, is an estimate of the standard error of $\hat{P}(\boldsymbol{\alpha}^*_{lj})$, as in, $SE[\hat{P}(\boldsymbol{\alpha}^*_{lj})]$.

The algorithm for estimating the G-DINA model parameters and the corresponding SEs is largely similar to the algorithms in estimating the parameters of the DINA and multiple-choice DINA models described in detail by de la Torre (2009a, 2009b). The parameter estimates of the other saturated CDMs and reduced models can be derived from the $2^{K_j^*}$ G-DINA model parameter estimates. As a last note on this section, readers are referred to von Davier (2005) and von Davier and Yamamoto (2004) for a description and an implementation of the MMLE algorithm for a particular instance of the GDM.

4.2. *Design Matrix*

4.2.1. Saturated Models The design matrix $\boldsymbol{M}_j$ occupies a central role in converting the G-DINA model parameter estimates given in (15) into its components (i.e., $\delta_{j0}, \delta_{jk}, \delta_{jkk'}, \ldots, \delta_{j12\cdots K_j^*}$). It can also be used to derive the parameter estimates of general CDMs based on alternative link functions. Finally, the design matrix can facilitate the estimation of the parameters of a special class of reduced models. The design matrix is of dimension $2^{K_j^*} \times P$, where P is the number of the parameters of the model of interest. Thus, for the saturated model, $P = 2^{K_j^*}$ (i.e., the number of columns is equal to the number of rows). This corresponds to the saturated model because the number of model parameters is equal to the number of latent groups generated by the K_j^* latent attributes.

Let $\boldsymbol{A}_j = \{\alpha_{lk}\}$, a $2^{K_j^*} \times K_j^*$ matrix of the possible combinations of the required attributes for item j, be defined as

$$\boldsymbol{A}_j = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 1 \\ 1 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix}. \tag{18}$$

For the saturated models, the lth row of the saturated design matrix $\boldsymbol{M}_j^{(S)}$ can be generated from the lth row of $\boldsymbol{A}_j$ (i.e., $\boldsymbol{\alpha}^*_l$). It has 1 as its first element, followed by α_{lk}, for $k = 1, \ldots, K_j^*$, then by $\alpha_{lk} \times \alpha_{lk'}$, for $k = 1, \ldots, K_j^* - 1$ and $k' = k + 1, \ldots, K_j^*$, and so forth; the last element of this vector is $\prod_{k=1}^{K_j^*} \alpha_{lk}$. To illustrate, let $K_j^* = 3$. The saturated design matrix is

$$\boldsymbol{M}_{j\,[8\times 8]}^{(S)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{19}$$

Now, define $\boldsymbol{P}_j = \{P(\boldsymbol{\alpha}_{lj}^*)\}$, a vector of length $2^{K_j^*}$ where $P(\boldsymbol{\alpha}_{lj}^*)$ represents the correct response probability of the lth row of $\boldsymbol{A}_j$. To obtain the estimate of $\boldsymbol{\delta}_j = \{\delta_{j0}, \delta_{j1}, \ldots, \delta_{jK_j^*}, \delta_{j12}, \ldots, \delta_{j12\cdots K_j^*}\}'$ given $\hat{\boldsymbol{P}}_j$, find the least-square estimate

$$\hat{\boldsymbol{\delta}}_j = \left(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)}\right)^{-1}\boldsymbol{M}_j^{(S)\prime}\hat{\boldsymbol{P}}_j. \tag{20}$$

To obtain the estimates of $\boldsymbol{\lambda}_j$ and $\boldsymbol{\nu}_j$, find

$$\hat{\boldsymbol{\lambda}}_j = \left(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)}\right)^{-1}\boldsymbol{M}_j^{(S)\prime}\left(\operatorname{logit}\hat{\boldsymbol{P}}_j\right), \tag{21}$$

and

$$\hat{\boldsymbol{\nu}}_j = \left(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)}\right)^{-1}\boldsymbol{M}_j^{(S)\prime}\left(\log\hat{\boldsymbol{P}}_j\right), \tag{22}$$

respectively. As stated earlier, except for the difference in the parameter interpretation, the three formulations provide identical saturated models in terms of their fit to the data.

The standard errors of these models can be computed via the multivariate delta method (Lehmann & Casella, 1998). Specifically, the variance of $f(\hat{\boldsymbol{P}}_j)$ is approximated by

$$\operatorname{Var}\left[f\left(\hat{\boldsymbol{P}}_j\right)\right] \approx \nabla f\left(\hat{\boldsymbol{P}}_j\right)'\operatorname{Var}\left(\hat{\boldsymbol{P}}_j\right)\nabla f\left(\hat{\boldsymbol{P}}_j\right), \tag{23}$$

where

$f(\hat{\boldsymbol{P}}_j)$ is $(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)})^{-1}\boldsymbol{M}_j^{(S)\prime}\hat{\boldsymbol{P}}_j$, $(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)})^{-1}\boldsymbol{M}_j^{(S)\prime}(\operatorname{logit}\hat{\boldsymbol{P}}_j)$, and $(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)})^{-1} \times \boldsymbol{M}_j^{(S)\prime}(\log\hat{\boldsymbol{P}}_j)$ for the G-DINA model, log-odds CDM and log CDM families, respectively;
$\nabla f(\cdot)$ is the gradient of $f(\cdot)$; and
$\operatorname{Var}(\hat{\boldsymbol{P}}_j)$ is the negative of the inverse of the information matrix of $\hat{\boldsymbol{P}}_j$.

4.2.2. A Special Class of Reduced Models By specifying the appropriate design matrices, the parameters of a special class of reduced models under the identity link can be estimated from $\hat{\boldsymbol{P}}_j$. Let $\boldsymbol{M}^{(r)}$ be the matrix associated with the reduced model r, and $\boldsymbol{M}^{(r-)}$ the last $P-1$ columns of $\boldsymbol{M}^{(r)}$. The reduced model is said to belong to this special class of reduced models under the identity link if $\boldsymbol{M}_j^{(r-)\prime}\boldsymbol{M}_j^{(r-)}$ is a diagonal matrix. Any reduced models with two parameters belong to this special class of models. Examples of two-parameter models are the DINA and DINO models, and their associated design matrices are

$$\boldsymbol{M}_{j\ [2^{K_j^*}\times 2]}^{(r)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \tag{24}$$

and

$$\boldsymbol{M}_{j\ [2^{K_j^*}\times 2]}^{(r)} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{25}$$

respectively. Moreover, the multiple-strategy DINA model proposed by de la Torre and Douglas (2008) is also a two-parameter model belonging to this class. Item 9 of their real-data example can be solved by students who have mastered either α_1 and α_3, or α_1 and α_6. For this item, $K_9^* = 3$ and the corresponding design matrix is

$$\boldsymbol{M}_{9\ [8\times 2]}^{(r)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{26}$$

In addition, models where the probabilities of success are based on the number of attributes mastered, rather than the specific attributes mastered, also belong to this class of reduced models. For example, when $K_j^* = 3$, this model has the constraints $P(100) = P(010) = P(001)$ and $P(110) = P(101) = P(011)$, and the associated design matrix is

$$\boldsymbol{M}_{j\ [8\times 4]}^{(r)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{27}$$

For $\boldsymbol{M}_j^{(r)}$ of reduced models with $P = 2$, say, (24) through (26), it is trivial to show that $\boldsymbol{M}_j^{(r-)\prime}\boldsymbol{M}_j^{(r-)}$ is a diagonal matrix (i.e., $\boldsymbol{M}_j^{(r-)\prime}\boldsymbol{M}_j^{(r-)}$ is a scalar). When $P > 2$, we need to verify that this condition is satisfied. For (27), $\boldsymbol{M}_j^{(r-)\prime}\boldsymbol{M}_j^{(r-)}$ is a diagonal matrix with the elements 3, 3 and 1.

In finding the estimate of $\boldsymbol{\delta}_j$ for models in the special class of constrained G-DINA models, we note that there are more probability estimates than number of parameters to be estimated (i.e.,

$2^{K_j^*} > P$). This implies that combinations of $P(\boldsymbol{\alpha}_{lj}^*)$ will be used to obtain $\hat{\boldsymbol{\delta}}_j$. Unless the latent classes are uniformly distributed, $P(\boldsymbol{\alpha}_{lj}^*)$ needs to be differentially weighted to account for the relative size of the latent class in estimating the parameter. One particular weight that can be used for this purpose is $I_{\boldsymbol{\alpha}_{lj}^*}$ (i.e., the expected number of examinees in latent class l). We can define the diagonal matrix $\boldsymbol{W}_j$ as

$$\boldsymbol{W}_{j\,[2^{K_j^*}\times 2^{K_j^*}]} = \{I_{\boldsymbol{\alpha}_{lj}^*}\}. \tag{28}$$

The estimates of $\boldsymbol{\delta}_j$ for the reduced models can be obtained by finding the weighted least-square estimate

$$\hat{\boldsymbol{\delta}}_j = \left(\boldsymbol{M}_j^{(r)\prime}\boldsymbol{W}_j\boldsymbol{M}_j^{(r)}\right)^{-1}\boldsymbol{M}_j^{(r)\prime}\boldsymbol{W}_j\hat{\boldsymbol{P}}_j. \tag{29}$$

The standard errors of these models can also be computed via the multivariate delta method given in (23), except that $f(\hat{\boldsymbol{P}}_j)$ is defined as $(\boldsymbol{M}_j^{(r)\prime}\boldsymbol{W}_j\boldsymbol{M}_j^{(r)})^{-1}\boldsymbol{M}_j^{(r)\prime}\boldsymbol{W}_j\hat{\boldsymbol{P}}_j$.

It should be noted that the A-CDM, which has the $\boldsymbol{M}_j^{(r)}$ corresponding to the first K_j^*+1 columns of $\boldsymbol{M}_j^{(S)}$, does not belong to this special class of reduced models. The same can be said of the LLM and the G-NIDA model. The parameters of these models and those that do not belong to the special class of reduced models can still be obtained one item at a time given $\boldsymbol{W}_j$ and $\hat{\boldsymbol{P}}_j$ by using various optimization algorithms with the additional constraint that the estimated probabilities are within the interval [0, 1]. Although such an approach is computationally efficient, the asymptotic properties of the estimates derived in this manner remain to be documented.

4.3. $\hat{\boldsymbol{\delta}}_j$, $\hat{\boldsymbol{\lambda}}_j$ *and* $\hat{\boldsymbol{\nu}}_j$ *as MLEs*

4.3.1. Saturated Models To show that $\hat{\boldsymbol{\delta}}_j$, $\hat{\boldsymbol{\lambda}}_j$ and $\hat{\boldsymbol{\nu}}_j$ of the saturated models are MLEs, let $h[\hat{\boldsymbol{P}}_j]$ be $\hat{\boldsymbol{P}}_j$, logit$[\hat{\boldsymbol{P}}_j]$ or log$[\hat{\boldsymbol{P}}_j]$, and $\boldsymbol{\phi}_j$ be $\boldsymbol{\delta}_j$, $\boldsymbol{\lambda}_j$ or $\boldsymbol{\nu}_j$, respectively. It follows that

$$\begin{aligned}
\phi_{j0} &= h\big[P(000\ldots00)\big],\\
\phi_{j0}+\phi_{j1} &= h\big[P(100\ldots00)\big],\\
&\;\;\vdots\\
\phi_{j0}+\phi_{jK_j^*} &= h\big[P(000\ldots01)\big],\\
\phi_{j0}+\phi_{j1}+\phi_{j2}+\phi_{j12} &= h\big[P(110\ldots00)\big],\\
&\;\;\vdots\\
\phi_{j0}+\phi_{jK_j^*-1}+\phi_{jK_j^*}+\phi_{j(K_j^*-1)K_j^*} &= h\big[P(000\ldots11)\big],\\
\phi_{j0}+\phi_{j1}+\phi_{j2}+\phi_{j3}+\phi_{j123} &= h\big[P(111\ldots00)\big],\\
&\;\;\vdots\\
\phi_{j0}+\sum_{k=1}^{K_j^*}\phi_{jk}+\sum_{k'>k}^{K_j^*}\sum_{k=1}^{K_j^*-1}\phi_{jkk'}+\ldots\phi_{j12\ldots K_j^*} &= h\big[P(111\ldots11)\big].
\end{aligned}$$

By isolating each element of $\boldsymbol{\phi}_j$, we get

$$\boldsymbol{\phi}_j = \boldsymbol{L}_j \times h[\hat{\boldsymbol{P}}_j], \tag{30}$$

where $\boldsymbol{L}_j$ is a $2^{K_j^*} \times 2^{K_j^*}$ matrix. The lth row of $\boldsymbol{L}_j$ corresponding to $\boldsymbol{\alpha}_{lj}^*$ has the following elements: $(-1)^{n_{lj}}$, where $n_{lj} = \sum_{k=1}^{K_j^*} \boldsymbol{\alpha}_{lj}^*$, followed by $(-1)^{n_{lj}+1} \times \alpha_{ljk}$, for $k = 1, \ldots, K_j^*$, then by $(-1)^{n_{lj}} \times \alpha_{ljk} \times \alpha_{ljk'}$, for $k = 1, \ldots, K_j^* - 1$, and $k' = k+1, \ldots, K_j^*$, and so forth; the last element of the row is $(-1)^{n_{lj}+1} \times \prod_{k=1}^{K_j^*} \alpha_{ljk}$. Thus, $|\boldsymbol{L}_j| = \boldsymbol{M}_j^{(S)}$. For example, when $K_j^* = 3$,

$$\boldsymbol{L}_{j\,[8\times 8]} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix}. \tag{31}$$

Recall that the estimate of $\boldsymbol{\phi}_j$ is $(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)})^{-1}\boldsymbol{M}_j^{(S)\prime} h[\hat{\boldsymbol{P}}_j]$. But $(\boldsymbol{M}_j^{(S)\prime}\boldsymbol{M}_j^{(S)})^{-1}\boldsymbol{M}_j^{(S)\prime} = \boldsymbol{L}$. Therefore, $\hat{\boldsymbol{\phi}}_j = \boldsymbol{L} \times (h[\hat{\boldsymbol{P}}_j])$. Thus, because $\hat{\boldsymbol{P}}_j$ is an MLE, $\hat{\boldsymbol{\phi}}_j$ is also an MLE due to the invariance property of the MLE. Consequently, the MLE properties (e.g., asymptotic properties) of $\hat{\boldsymbol{P}}_j$ are also applicable to $\hat{\boldsymbol{\phi}}_j$.

4.3.2. Special Class of Reduced Models For the special class of reduced models under the identity link, the $2^{K_j^*}$ latent groups for item j are partitioned into P non-overlapping latent groups g_{jp}, $p = 0, \ldots, P-1$. The maximum likelihood estimate of the probability of success for individuals in group g_{jp} is equal to

$$\hat{P}(g_{jp}) = \frac{\sum_{\boldsymbol{\alpha}_{lj}^* \in g_{jp}} R_{\boldsymbol{\alpha}_{lj}^*}}{\sum_{\boldsymbol{\alpha}_{lj}^* \in g_{jp}} I_{\boldsymbol{\alpha}_{lj}^*}} = \frac{\sum_{\boldsymbol{\alpha}_{lj}^* \in g_{jp}} I_{\boldsymbol{\alpha}_{lj}^*} \hat{P}(\boldsymbol{\alpha}_{lj}^*)}{\sum_{\boldsymbol{\alpha}_{lj}^* \in g_{jp}} I_{\boldsymbol{\alpha}_{lj}^*}}. \tag{32}$$

By setting $\hat{\delta}_{j0} = \hat{P}(g_{j0})$, and $\hat{\delta}_{jp} = \hat{P}(g_{jp}) - \hat{P}(g_{j0})$ for $p = 1, \ldots, p-1$, we find the estimate of $\hat{\boldsymbol{\delta}}_j$. This estimate is equivalent to the estimate obtained using (29). Thus, $\hat{\boldsymbol{\delta}}_j$ for the special class of reduced models under the identity link is also a maximum likelihood estimate.

5. Model Comparison

5.1. The Wald Test

Assuming that the relevant attributes have been identified, and the Q-matrix correctly specified, the saturated model will give the best model-data fit. To examine whether one of the reduced models, interpretable or otherwise, can be used in place of the saturated model, the fit of the saturated and reduced models can be statistically compared.

One method of testing the adequacy of a reduced model is to use the *Wald test*. For item j and reduced model p, this test requires setting up $\boldsymbol{R}_{jp}$, a $(2^{K_j} - p) \times 2^{K_j}$ matrix of restrictions.

For example, the $\boldsymbol{R}_{jp}$ matrices for the DINA model and A-CDM when $K_j = 3$ are

$$\boldsymbol{R}_{jp6\times 8} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix}, \tag{33}$$

and

$$\boldsymbol{R}_{jp4\times 8} = \begin{pmatrix} 1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix}, \tag{34}$$

respectively. The first matrix constrains all $P(\boldsymbol{\alpha}^*_{lj})$ except $P(\mathbf{1}_{K^*_j})$ to be identical. In contrast, the second matrix sets $\delta_{12} = \delta_{13} = \delta_{23} = \delta_{123} = 0$; the same matrix can also be used to set the interaction terms of the log and logit link functions to zero. The Wald statistic W is computed as

$$W = \left[\boldsymbol{R}_{jp} \times f(\boldsymbol{P}_j)\right]' \left\{\boldsymbol{R}_{jp} \times \mathrm{Var}\left[f(\boldsymbol{P}_j)\right] \times \boldsymbol{R}'_{jp}\right\}^{-1} \left[\boldsymbol{R}_{jp} \times f(\boldsymbol{P}_j)\right], \tag{35}$$

and is asymptotically $\chi^2_{2^{K^*_j} - p}$ under the null hypothesis that $\boldsymbol{R}_{jp} \times f(\boldsymbol{P}_j) = \mathbf{0}$. It should be noted that the proposed Wald test for comparing saturated and reduced models does not require estimation of the reduced models—finding $f(\boldsymbol{P}_j)$, $\mathrm{Var}[f(\boldsymbol{P}_j)]$, and $\boldsymbol{R}_{jp}$ is sufficient for its implementation.

5.2. *An Example*

This example serves as an illustration on how the Wald test can be used with the G-DINA model. The example focuses on the Type I error of the Wald test when the true model is the A-CDM, and the power of the test to detect when the A-CDM is an inadequate model for the data (i.e., the true model is either the DINA or DINO model). For this example, the sample size, number of items, and number of attributes were fixed to $I = 2000$, $J = 30$ and $K = 5$, respectively. As the Q-matrix for this example shows (see Table 1), the test is composed of three item types: one-attribute, two-attribute, and three-attribute items, and ten items of each type were used. For the A-CDM, the parameters for the one-attribute items were $\delta_0 = 0.20$ and $\delta_1 = 0.60$; the parameters for the two-attribute items were $\delta_0 = 0.20$ and $\delta_k = 0.30$, for $k = 1, 2$; and the parameters for the three-attribute items were $\delta_0 = 0.20$ and $\delta_k = 0.20$, for $k = 1, 2, 3$. For the DINA and DINO models, the guessing and slip parameters were set to 0.20. These values were selected so that the minimum and maximum probabilities of success were equal to 0.20 and 0.80, respectively, across all models and item types.

For each of the three models (i.e., A-CDM, DINA and DINO models), 1,000 data sets were generated. The Type I error rate and power of the Wald test for the A-CDM were examined using the significance levels 0.01, 0.05, and 0.10. At the 0.01 significance level, the mean rejection rates of the Wald test for the two- and three-attribute items were both 0.01; at the 0.05 significance level, these rejection rates were 0.05 and 0.06; and at the 0.10 significance level, these rejection rates were both 0.11. The results indicated that the Type I error rates of the Wald test for the A-CDM were close, if not equal to the nominal significance levels. At all significance levels considered, the Wald test consistently rejected the adequacy of the A-CDM when the data were generated using the DINA or the DINO model. Thus, for these particular simulation setup, the Wald test has a power of 1.0, and this was achieved without inflating the Type I error.

TABLE 1.
Simulation study Q-matrix.

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

6. Real Data Illustrations

6.1. Fraction Subtraction Data

The data for this illustration were a subset of the data originally described and used by Tatsuoka (1990), and more recently by Tatsuoka (2002, 2005), and de la Torre and Douglas (2004, 2008). These data were responses of 536 middle school students to 12 fraction subtraction problems involving four of the eight original attributes. The attributes are (1) performing basic fraction subtraction operation, (2) simplifying/reducing, (3) separating whole number from fraction, and (4) borrowing one from whole number to fraction. The fraction subtraction items and the Q-matrix are given in Table 2. Based on this Q-matrix, there were $\sum_{j=1}^{12} 2^{K_j^*} = 106$ parameters to be estimated when these data were analyzed using the G-DINA model. Due to the sample size of the data for this illustration, $P(\boldsymbol{\alpha}_{lj}^*)$ was constrained to be less than or equal to $P(\boldsymbol{\alpha}_{l'j}^*)$ when $\boldsymbol{\alpha}_{lj}^* \prec \boldsymbol{\alpha}_{l'j}^*$ after initial parameter estimates were obtained. In estimating the model parameters, the MMLE algorithm, which was written in Ox (Doornik, 2003) and implemented using a 3-GHz Pentium 4 computer with a convergence criterion of 0.001, took under 16 seconds to estimate the model parameters.

TABLE 2.
Q-matrix for the fraction subtraction data.

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

Given in Table 3 are the parameter estimates of the G-DINA model based on the fraction subtraction data. The table entries represent the probabilities of success on the items for specific attribute mastery patterns. As noted earlier, the number of parameters for each item is a function of the number of required attributes for the item. For example, Items 2, 5, 6, 11 and 12 each had 16 parameters, whereas Items 1 and 3 only had two parameters each. The specific reduced attribute patterns corresponding to the parameter estimates are also given on the top rows of the table. It should be noted that even when $K_j^* = K_{j'}^*$, the attributes represented in the reduced attribute vector may not be the same set of attributes except when $K_j^* = K$. For instance, Items 7 and 8 both require two attributes, but the attributes required for these two items are not the same (i.e., α_1 and α_2 are required for Item 7, whereas α_1 and α_3 are required for Item 8). One caveat needs to be articulated at this point: Due to the relatively small number of examinees who took the fraction subtraction test, the results below should be interpreted with caution.

For items requiring a single attribute (i.e., Items 1 and 3), $P(0)$ and $P(1)$ can be interpreted as the g and $1-s$ of the DINA model, respectively. For example, examinees who have not mastered *basic fraction subtraction* have a 12% chance of guessing and correctly answering Item 3; in comparison, those who have mastered this attribute will have a 95% chance of answering the item correctly. For the remaining items (i.e., items requiring multiple attributes), only a minority of the items, Items 5 and 10, appear to satisfy the conjunctive assumption of the DINA model that students who lack at least one of the required attribute for the item have the same probability of success; the remaining eight items clearly do not satisfy this assumption. For these two items, $P(\boldsymbol{\alpha}_{lj}^*)$ are very similar whenever $\boldsymbol{\alpha}_{lj}^* \nless \mathbf{1}$; in addition, a dramatically higher probability of success can be observed for examinees who mastered all the required attributes for the items.

In contrast to items that satisfy the DINA model assumption, estimates for Items 4, 7, 8 and 9, all of which require two attributes, show that examinees who mastered one of the two required attributes for these items can have higher probabilities of success compared to individuals who mastered none of the attributes. These results also show that the improvement depends on which attributes were mastered. For these four items, mastering *basic fraction subtraction* (which rep-

TABLE 3.
Results of the fraction subtraction data analysis.

Item	Attribute Pattern														
	0	1													
	00	10	01	11											
	000	100	010	001	110	101	011	111							
	0000	1000	0100	0010	1000	1100	1010	1001	0110	0101	0011	1110	1101	1011	0111	1111
1	0.00	0.71														
2	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.23	0.23	0.23	0.23	0.23	0.74	0.23	0.89
3	0.12	0.95														
4	0.33	0.47	0.33	0.77												
5	0.01	0.03	0.01	0.01	0.01	0.03	0.04	0.05	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.81
6	0.04	0.05	0.04	0.04	0.04	0.05	0.05	0.10	0.05	0.05	0.05	0.10	0.10	0.81	0.10	0.94
7	0.11	0.59	0.11	0.97												
8	0.04	0.42	0.04	0.97												
9	0.07	0.43	0.07	0.92												
10	0.01	0.05	0.01	0.01	0.05	0.07	0.05	0.88								
11	0.00	0.00	0.00	0.00	0.00	0.27	0.18	0.00	0.00	0.00	0.00	0.27	0.41	0.81	0.27	0.85
12	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.16	0.37	0.02	0.85

TABLE 4.
Summary of the Q-matrix for the MCMI-III data.

Specification type	Scale: Anxiety	Somatoform	Thought disorder	Major depression
Single	9	2	11	6
Joint	5	10	6	12
Total	14	12	17	18

resents the simplest attribute) by itself resulted in observable improvements, whereas mastering either *simplifying/reducing* or *separating whole number from fraction* alone did not. A dramatic increase in the probability of success on these items can be achieved by mastering both required attributes.

For Items 2, 6 and 11, mastering three of the four required attributes, namely, *performing basic fraction subtraction operation*, *separating whole number from fraction*, and *borrowing one from whole number to fraction*, provided a markedly higher probability of success compared to any other subsets of the attributes. Additional improvements can be achieved by mastering the remaining attribute, *simplifying/reducing*. Finally, although Item 12 does not come close to satisfying the assumption of the DINA model in that the probability of success on the item is highly variable when $\boldsymbol{\alpha}_{lj}^{*} \not\prec \mathbf{1}$, results indicate that mastery of all the required attributes was necessary for examinees to perform well on this item.

6.2. *MCMI-III Data*

Applications of models for cognitive diagnosis extend beyond problems in the educational settings. Examples of this type of applications include determining the cognitive deficit profiles of patients with schizophrenia (Jaeger, Tatsuoka, & Berns, 2003), and evaluation of individuals with gambling pathology (Templin & Henson, 2006). The following brief example using the Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, Millon, Davis, & Grossman, 2006) further illustrates the generality and portability of CDMs as statistical tools for diagnostic purposes.

The data were based on the Dutch language version of the MCMI-III (Rossi, Sloore, & Derksen, 2008), and the current sample has also been reported in Rossi, van der Ark, and Sloore (2007) and in Rossi, Elklit, and Simonsen (2010). Responses came from 1210 Belgian subjects, 39% of which were female. Sixty-eight percent of the subjects were clinical patients; the remainder were inmates. Of the MCMI-III data, four scales with a total of 44 items were analyzed using the unconstrained G-DINA model. The four scales, which can be construed as the attributes, are: Anxiety (Scale A), Somatoform (Scale H), Thought Disorder (Scale SS), and Major Depression (Scale CC). Table 4 summarizes the Q-matrix for these data and gives the number of times each attribute was specified singly or jointly with other attributes. Of the 44 items, 28 have single-attribute specifications, 15 have two-attribute specifications, and one item has a three-attribute specification. Overall, the four attributes were specified 12 to 18 times each. Using the same estimation code, computer system and convergence criterion as the fraction subtraction data, it took under 20 seconds to obtain the model parameter estimates.

For illustration purposes, the item, *I can't seem to sleep, and wake up just as tired as when I went to bed*, which measures Somatoform and Major Depression, is discussed. Results show that individuals with neither Somatoform or Major Depression have a 5% chance of endorsing the item; individuals with Somatoform or Major Depression have a 49% or 32% chance of endorsing the statement, respectively; finally, individuals who have both Somatoform and Major Depression have a 74% chance of endorsing the item. The results indicate that individuals with

Somatoform are more likely *to have difficulty sleeping and to wake up tired* than those with Major Depression. In addition, the results also indicate that a disjunctive model (e.g., DINO model) is not appropriate for this item because the rate of endorsement among individuals with both disorders is higher compared to individuals with only one of the two disorders. If any, the results show that the two disorders have an additive effect on the probability of item endorsement. Based on the additive model under the identity link, $P(11) = \delta_{00} + \delta_{10} + \delta_{01}$, which is equal to $P(10) + P(01) - P(00)$. The expected probability for $P(11)$ of 0.76 under the additive model is close to the estimated probability of 0.74 under the saturated model.

7. Summary and Discussion

The paper proposes the G-DINA model as a generalization of the DINA model with more relaxed assumptions. As a general model, the G-DINA model is an interpretable model based on the identity link function and represents one of the many alternative general CDM formulations. When appropriate constraints are applied, several commonly used CDMs are shown to be special cases of these general models. In their saturated forms, the general models provide identical model-data fit and therefore are equivalent models; however, the additive cases of the general models, despite their identical formulation and number of parameters, do not provide the same fit, and thus, cannot be used interchangeably.

The G-DINA model proposed in this paper is a general CDM framework that includes components for item-by-item model estimation and hypothesis testing. The parameters of the G-DINA model can be estimated using the MMLE algorithm. From these estimates, parameters of several saturated models and a special class of reduced models can be estimated using least-square estimation and appropriate design and weight matrices. The corresponding SEs can be obtained from the G-DINA model SEs via the multivariate delta method. Parameters of models outside this special class of reduced models can also be obtained item by item using the initial G-DINA model parameter estimates and the weight matrix. In addition to parameter estimation, the proposed framework also allows for statistical comparison of saturated and reduced models. The Wald test can be used to empirically determine the adequacy of the reduced models relative to the saturated model for each of the multi-attribute items.

The G-DINA model framework as described in this paper can extend the flexibility and practical usefulness of cognitive diagnosis modeling. For one, parameter estimation can be carried out faster using the more efficient MMLE algorithm rather than the computer-intensive Markov chain Monte Carlo algorithm. In addition, different reduced models can be fitted without the necessity of estimating the parameters from the original response data. By estimating the reduced model parameters one item at a time, fitting all the possible combinations of feasible reduced models becomes possible. Instead of J^{*m} number of estimations involving the entire response data (J^* and m are the numbers of multi-attribute items and feasible reduced models), only $J^* \times m$ number of estimations involving item-specific parameters are necessary under this framework. From the collection of possible reduced models, the most appropriate reduced model in the statistical sense can be determined for each item. Such empirically derived models can be compared with the researchers' theoretically based expectations.

The estimation and hypothesis-testing components of the G-DINA model framework can have several implications in the practice of cognitive diagnosis modeling. First, researchers who cannot specify the reduced CDM a priori can still obtain CDM parameter estimates based on the general formulation with the final model specification determined *a posteriori*. Second, researchers who can posit the underlying process for a subset or all of the items can verify their hypothesis empirically. Third, with the possibility of testing the reduced models one item at a time, it is also possible to construct a test where multiple underlying processes (i.e., CDMs) are involved at the same time. Fourth, as the fraction subtraction data example illustrates, barring any developmental constraints, the differential impact of attribute mastery may be informative in determining the sequence by which students must acquire the different attributes to optimize their performance in an assessment.

To the extent that both frameworks decompose item parameters into more *basic parameters*, the G-DINA model is similar to the linear logistic test model (LLTM; Fischer, 1973). However, several notable differences between the two frameworks exist. First, the decomposition of the parameters simultaneously involves all the items in LLTM, whereas this decomposition involves one item at a time in the G-DINA model. Second, because sufficient statistics are available in LLTM but not in the G-DINA model, parameters can be estimated in the former using conditional maximum likelihood estimation, whereas the latter requires MMLE. Finally, estimation of the basic parameters in LLTM involves a single step, whereas a two-step procedure is implemented in the G-DINA model. In this regard, the two-step procedure described in this paper is more similar to that of Scheiblechner (1972, as cited in Fischer, 1997), where least-square regression is used in deriving the basic parameter estimates from the item difficulty parameter estimates. In closing this paragraph, it should be noted that although unidimensional item response models such as the LLTM or the three-parameter logistic model have been used for diagnostic purposes, some researchers (e.g., Stout, 2007; von Davier, 2009) do not consider them as diagnostic models in terms of both their original intention and capability.

For the G-DINA model framework to be deemed practically viable, additional work needs to be carried out to better understand its properties. For one, it is unclear at this point under what circumstances the proposed method can be expected to provide accurate parameter estimates and SEs. Sample size and attribute distribution are particularly salient for the G-DINA model because, as de la Torre (2009a) has shown, the stability of proportion-based parameter estimates can in part be traced to the number of observations used in estimating these parameters. For another, the Wald test statistic for the G-DINA model is assumed to be asymptotically distributed. However, it is not clear what sample size is deemed reasonable to provide acceptable approximations, and whether the properties of this test as a function of sample size will vary with other factors, such as the type of reduced model and size of K_j^*. Another legitimate concern about the proposed test is related to the possibility of inflating the Type I error due to the high degree of complexity of the saturated model and the large number of comparisons to be made. It would be instructive to examine whether indices that account for both the complexity of the models and their goodness of fit, such as Akaike's information criterion (Akaike, 1973), and measures that adjust for multiple comparisons at the item or test level, such as the Bonferroni correction, can be used to address this concern.

As noted by one reviewer, the estimation time given in this paper can be considered conservative. This is because the estimation code was written by a nonprofessional programmer in a higher-level programming language (i.e., Ox). Additional efficiency can be gained if the estimation code can be written by a more experienced programmer using a lower-level language. Such code can make the implementation of the G-DINA model with larger data sets more practicable.

The G-DINA model framework allows for reduced models to be estimated one item at a time. This can be a very efficient approach when several reduced models are being considered simultaneously. However, such an approach can have its drawbacks. For one, the framework cannot be used when constraints need to be imposed across items (e.g., common parameter estimates for items with the same attribute specification). This limitation is true of the current proposed MMLE algorithm for the saturated G-DINA model as well. For another, although the estimation method within the framework can provided maximum likelihood estimates for the saturated models and a special class of reduced models, the same cannot be said of estimates for reduced models that do not belong to this special class of models. Additional work needs to be done to document and better understand the statistical properties of these estimates.

Its current limitations notwithstanding, the generality of the G-DINA model can open multiple possibilities. It would be interesting to examine whether the greater flexibility afforded by the general formulation of the model can be exploited to address other important issues in cognitive diagnosis. For example, one can investigate whether the procedure of empirically validating the Q-matrix for the DINA model (de la Torre, 2008) can be extended to the G-DINA model as a tool for examining the appropriateness of Q-matrix specifications across different types of CDMs.

As with other CDMs, employing a saturated formulation of the joint distribution of the attributes when the number of attributes is large (e.g., $K \geq 20$) may not be feasible given the current computer memory limitations, and the sample size needed to calibrate such a test. As an alternative, whenever reasonable, the joint distribution of the attributes can be simplified by conditioning the attributes on a higher-order latent trait (de la Torre & Douglas, 2004, 2008). This joint distribution can also be simplified by imposing structures on the attributes (e.g., Leighton, Gierl, & Hunka, 2004) to reduce the number of attribute vectors to be considered.

It is important to recognize that the proposed framework represents only the psychometric component of cognitive diagnosis modeling—the other component is represented by cognitively diagnostic assessments with which the framework can be used. It is, therefore, of paramount importance that necessary resources be invested to develop assessments that can support diagnostic inferences. To this end, the contributions of experts from other relevant disciplines such as cognitive and learning sciences, subject domains, and didactics, cannot be overemphasized across the various stages of cognitive diagnosis modeling. Thus, collaborative effort between individuals with diverse expertise is needed in defining what attributes to measure, designing which tasks will best measure the attributes, selecting psychometric tools most appropriate for data analysis, and interpreting the test scores and prescribing apposite courses of action to take.

Acknowledgements

This research was supported by National Science Foundation Grant DRL-0744486. The author would also like to thank Kikumi Tatsuoka and Gina Rossi for access to their data.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N., & Csaki, F. (Eds.) *Proceedings of the second international symposium on information theory* (pp. 267–281). Budapest: Akad. Kiado.

de la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: development and applications. *Journal of Educational Measurement*, *45*, 343–362.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively-based multiple-choice options. *Applied Psychological Measurement*, *33*, 163–183.

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

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

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

Doornik, J.A. (2003). *Object-oriented matrix programming using Ox (version 3.1) [Computer software]*. London: Timberlake Consultants Press.

Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*, 359–374.

Fischer, G.H. (1997). Unidimensional linear logistic Rasch models. In van der Linden, W., & Hambleton, R.K. (Eds.), *Handbook of modern item response theory* (pp. 225–244). New York: Springer.

Hagenaars, J.A. (1990). *Categorical longitudinal data: loglinear panel, trend, and cohort analysis*. Thousand Oaks: Sage.

Hagenaars, J.A. (1993). *Loglinear models with latent variables*. Thousand Oaks: Sage.

Hartz, S.M. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: blending theory with practicality*. Unpublished doctoral dissertation.

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210.
Jaeger, J., Tatsuoka, C., & Berns, S. (2003). Innovative methods for extracting valid cognitive deficit profiles from NP test data in schizophrenia. *Schizophrenia Research*, *60*, 140–140.
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.
Lehmann, E.L., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York: Springer.
Leighton, J.P., Gierl, M.J., & Hunka, S. (2004). The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*, 205–236.
Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187–212.
Millon, T., Millon, C., Davis, R., & Grossman, S. (2006). *MCMI-III manual* (3rd ed.). Minneapolis: Pearson Assessments.
Rossi, G., Elklit, A., & Simonsen, E. (2010). Empirical evidence for a four factor framework of personality disorder organization: multigroup confirmatory factor analyses of the Millon Clinical Multiaxial Inventory—III personality disorder scales across Belgian and Danish data samples. *Journal of Personality Disorders*, *24*, 128–150.
Rossi, G., Sloore, H., & Derksen, J. (2008). The adaptation of the MCMI-III in two non-English-speaking countries: state of the art of the Dutch language version. In Millon, T., & Bloom, C. (Eds.), *The Millon inventories: a practitioner's guide to personalized clinical assessment* (2nd ed., pp. 369–386). New York: Guilford.
Rossi, G., van der Ark, L.A., & Sloore, H. (2007). Factor analysis of the Dutch language version of the MCMI-III. *Journal of Personality Assessment*, *88*, 144–157.
Roussos, L.A., DiBello, L.V., Stout, W., Hartz, S.M., Henson, R.A., & Templin, J.L. (2007). The fusion model skills diagnosis system. In Leighton, J.P., & Gierl, M.J. (Eds.), *Cognitively diagnostic assessment for education: theory and applications* (pp. 275–318). Cambridge: Cambridge University Press.
Stout, W. (2007). Skills diagnosis using IRT-Based continuous latent trait models. *Journal of Educational Measurement*, *44*, 313–324.
Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *51*, 337–350.
Tatsuoka, C. (2005). Corrigendum: data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *54*, 465–467.
Tatsuoka, K. (1983). Rule space: an approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345–354.
Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In Frederiksen, N., Glaser, R., Lesgold, A., & Safto, M. (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453–488). Hillsdale: Erlbaum.
Templin, J., & Henson, R. (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 RR-05-16). Princeton: Educational Testing Service.
von Davier, M. (2009). Some notes on the reinvention of latent structure models as diagnostic classification models. *Measurement*, *7*, 67–74.
von Davier, M., & Yamamoto, K. (2004, October). *A class of models for cognitive diagnosis*. Paper presented at the 4th Spearman Conference, Philadelphia, PA.

Manuscript Received: 24 DEC 2008
Final Version Received: 23 SEP 2010
Published Online Date: 15 MAR 2011