

# Chapter 7

## The G-DINA Model Framework



Jimmy de la Torre and Nathan D. Minchen

**Abstract** The development of cognitive diagnosis models (CDMs) has been prolific since the turn of the century; however, they have often been developed in such a way that they lack an overall connective framework. The purpose of this chapter is to review the G-DINA framework. As a general model, it subsumes several simpler and widely-known CDMs; as a general framework, it has also served as the foundation for a variety of model extensions and new methodological developments. We will also discuss associated topics, which include model estimation, Q-matrix validation, computerized adaptive testing, and model selection as they relate to the reviewed models.

### 7.1 Introduction

Cognitive diagnosis models (CDMs) can be viewed as restricted versions of the more general latent class models. In particular, the number of latent classes, as well as their interpretation, are known a priori when CDMs are involved. Further restrictions can be posited regarding how the underlying attributes interact to produce the observed responses. These interactions (or condensation rules; Maris, 1999) include conjunctive, disjunctive, and additive processes (de la Torre, 2011). Assuming a specific underlying process involves the use of a reduced or constrained CDM such the DINA model (Haertel, 1989; Junker & Sijtsma, 2001), DINO model (Templin & Henson, 2006), LLM (Maris, 1999), R-RUM (Hartz, 2002), and A-CDM (de la Torre, 2011). Although more interpretable and requiring smaller sample sizes, reduced models can also lead to poorer model-data fit when they

---

J. de la Torre (✉)

Division of Learning, Development and Diversity, University of Hong Kong, Hong Kong, China  
e-mail: [j.delatorre@hku.hk](mailto:j.delatorre@hku.hk)

N. D. Minchen

Pearson, Bronx, NY, USA  
e-mail: [nathan.minchen@pearson.com](mailto:nathan.minchen@pearson.com)

are incorrectly specified (e.g., Chen & de la Torre, 2013). Notwithstanding their own shortcomings, general or saturated CDMs, such as the G-DINA model (de la Torre, 2011), LCDM (Henson, Templin, & Willse, 2009), and GDM (von Davier, 2008), can be used as an alternative to reduced CDMs to minimize the impact of potential model misspecifications. With the exception of the GDM, which can be specified more generally, the CDMs above are designed for dichotomous attributes and dichotomous responses. It should be noted that when dichotomous attributes and dichotomous responses are involved, the G-DINA model, which is typically written using the identity link function, the LCDM and GDM, which are based on the logit link function, and any saturated CDMs in other link functions (e.g., log) are equivalent to each other. To accommodate a wider range of attribute and response types, extensions of CDMs need to be considered.

An integral component of most, if not all, CDM specifications, general or otherwise, is the Q-matrix (Tatsuoka, 1983). In its typical formulation, a Q-matrix is a  $K \times D$  matrix that identifies the subset of attributes measured by each item, where  $K$  is the number of items and  $D$  the number of attributes measured by the test. The attribute specification for item  $j$  is given in the binary  $D$ -length vector,  $\mathbf{q}_j$ . Correspondingly, the latent variable in CDM is typically a binary  $D$ -length vector,  $\mathbf{a}_l$ , where  $l = 1, \dots, L = 2^D$ , the number of latent classes. As will be shown later, both  $\mathbf{q}_k$  and  $\mathbf{a}_l$  may require some modifications before they can be used in conjunction with CDM extensions.

The valid use of scores derived from CDMs presupposes that the model is adequate for the data. To this end, steps need to be taken to ensure that a discrete latent variable can fit the data, the correct CDMs are employed, and Q-matrix entries are correctly specified. In addition, for greater efficiency, simpler models should be preferred over more complex models whenever appropriate.

Given the large number of CDMs that currently exists, a unifying framework from which these models can be viewed is needed to better understand their unique natures and the extent to which these models relate to each other. Moreover, a coherent framework that permits implementation of various CDM-related procedures can allow for the appropriate use of CDMs to be evaluated more systematically. As will be discussed below, the G-DINA model framework aims to accomplish this two-pronged objective. In addition, the G-DINA as a model can serve as the foundation on which CDM extensions can be built.

## 7.2 The G-DINA Model Framework

### 7.2.1 The G-DINA Model

Without loss of generality, assume that the first  $D_k^*$  attributes are required for item  $k$ , and let  $\mathbf{a}_{lk}^*$  be the  $D_k^*$ -length *reduced* attribute vector,  $l = 1, \dots, 2^{D_k^*}$ , which retains only the attributes required for item  $k$ . The item response function (IRF) of

the G-DINA model is given by

$$g[P(X_k = 1 | \mathbf{a}_{lk}^*)] = \phi_{k0} + \sum_{d=1}^{D_k^*} \phi_{kd} a_{ld} + \sum_{d'=d+1}^{D_k^*} \sum_{d=1}^{D_k^*-1} \phi_{kdd'} a_{ld} a_{ld'} + \dots \\ + \phi_{12\dots D_k^*} \prod_{d=1}^{D_k^*} a_{ld}, \quad (7.1)$$

where  $g[\cdot]$  is either the identity, log, or logit link function,  $\phi_{k0}$  is the intercept,  $\phi_{kd}$  is the main effect due to mastering  $a_d$ , and each of the remaining  $\phi_{k\cdot}$  represent all possible higher-order interaction effects, ranging from two-way to  $D_k^*$ -way. When  $g[\cdot]$  is the logit link, it is equivalent to the LCDM, which has also been shown to be equivalent to a GDM with an extended skill space (von Davier, 2014).

The G-DINA model is considered a saturated CDM because it contains  $2^{D_k^*}$  parameters corresponding to the  $2^{D_k^*}$  latent groups in item  $k$ . As shown by de la Torre (2011), several reduced models can be derived from the G-DINA model by constraining its parameters. The DINA model is equivalent to the G-DINA model with all but the intercept and the highest-order interaction effect set to zero. Its IRF in the G-DINA notation is

$$g[P(X_k = 1 | \mathbf{a}_{lk}^*)] = \phi_{k0} + \phi_{12\dots D_k^*} \prod_{d=1}^{D_k^*} a_{ld}. \quad (7.2)$$

Similarly, the DINO model can be obtained from the G-DINA model using the following constraints:  $\phi_{kd} a_{ld} = -\phi_{kdd'} = \dots = (-1)^{D_k^*+1} \phi_{12\dots D_k^*}$ . Thus, its IRF can be written as

$$g[P(X_k = 1 | \mathbf{a}_{lk}^*)] = \phi_{k0} + \phi_{kd} a_{ld}. \quad (7.3)$$

Finally, when all the interaction effects are set to zero, as in,

$$g[P(X_k = 1 | \mathbf{a}_{lk}^*)] = \phi_{k0} + \sum_{d=1}^{D_k^*} \phi_{kd} a_{ld}, \quad (7.4)$$

the G-DINA model in the identity, log, or logit link is equivalent to the A-CDM, R-RUM, or LLM, respectively. Although the additive property is inherent to a particular link function (e.g., R-RUM is multiplicative when converted to the identity link), Ma, Iaconangelo, and de la Torre (2016) noted the interchangeability of the three additive models for some item parameter combinations. As a whole, recognizing that the G-DINA model subsumes a number of reduced CDMs has important implications in model comparison and model-data fit evaluation.

## 7.3 Model Extensions

### 7.3.1 G-DINA Model for Polytomous Attributes

Although the G-DINA model is a general CDM, it is only so with respect to dichotomous attributes. However, some educational applications may benefit from a finer-grained, and therefore, more instructionally-relevant classification of students. For example, classifying students as having no mastery, basic mastery, and advanced mastery of the skills might be of interest. The middle-school proportional reasoning (PR) assessment described by Tjoe and de la Torre (2013a,b) measures two polytomous attributes, namely, (a) comparing and ordering of fractions, where level 0 represents nonmastery of the attributes, level 1 the ability to compare two fractions, and level 2 the ability to order three or more fractions; and (b) constructing ratios and proportions, where level 0 again represents nonmastery, level 1 the ability to construct a single ratio, and level 2 the ability to construct a proportion, which is made up of two ratios. Such classifications require polytomous attributes.

Define  $\boldsymbol{a}_l = \{a_{ld} \mid a_{ld} \in (0, 1, \dots, M_d)\}$  as the polytomous attribute vector, and again, assume that the first  $D_k^*$  attributes are required for item  $k$ . The reduced attribute vector in this context can be written as  $\boldsymbol{a}_{lk}^* = \{a_{ld}, \dots, a_{lD_k^*}\}$ . When there are no constraints on the model, item  $k$  involves  $M_1 \cdot M_2 \cdots M_{D_k^*}$  latent groups. A saturated CDM for this item would require the same number of parameters, making it too complex to be viable in most practical testing situations. Chen and de la Torre (2013) proposed the polytomous G-DINA (pG-DINA) model as a lower-complexity CDM that can accommodate polytomous attributes. To reduce the number of latent groups, and hence complexity of the corresponding CDM, the pG-DINA model assumes that, for each attribute within an item, an examinee can be classified as either at or below the required attribute level. Examinees on or above the cutoff are assumed to have the necessary attribute mastery level to answer the item correctly, whereas those below it do not. Chen and de la Torre (2013) referred to this as the *specific attribute level mastery* (SALM) assumption. The reduced polytomous attribute vector  $\boldsymbol{a}_{lk}^*$  can be converted to a reduced dichotomous attribute vector  $\boldsymbol{a}_{lk}^*$  as follows:  $\boldsymbol{a}_{lk}^* = \{I(a_{ld} \geq q_{kd})\}$ , for  $d = 1, \dots, D_k^*$ . After the conversion,  $\boldsymbol{a}_{lk}^*$  can be used in the IRF given in (7.1) to model a wide variety of attribute interactions.

In general, the conversion process in the pG-DINA model reduces the number of latent groups to  $2^{D_k^*}$  for item  $k$  regardless of the number of levels of the attributes involved. It should also be noted that the pG-DINA model differs from other polytomous CDMs (e.g., GDM) in that the attribute level required for an item is defined by domain or subject-matter experts a priori, whereas in other CDMs, only the attribute, but not the level, need to be specified. This distinct feature of the pG-DINA model implies a modification of the Q-matrix – instead of only 0 and 1,  $q_{kd} \in (0, 1, \dots, M_d - 1)$ . Using the PR assessment data, Chen and de la Torre (2013) and de la Torre (2015) have shown that the pG-DINA model provides a better fit when compared to the G-DINA model. These results indicate that the pG-DINA model is not only theoretically appealing, but also empirically more appropriate.

### 7.3.2 G-DINA Model for Polytomous Response

Although items that can be scored as either right or wrong (i.e., 1/0) remains the most common item type in large-scale assessments, items that can be scored with ordered polytomous categories are also available. In the CDM literature, it is not uncommon for these scores to be dichotomized and analyzed using existing CDMs for dichotomous response. In recent years, a number of CDMs for ordered polytomous response have been proposed, including the GDM for graded responses (von Davier, 2008), the polytomous LCDM (Hansen, 2013) and the sequential G-DINA (sG-DINA; Ma & de la Torre, 2016) model. Of these, only the sG-DINA model considers the possibility that, within the same item, the subset of attributes being measured can vary from one response category to another.

The sG-DINA model assumes that the problem-solving process is sequential in nature, and allows for different subsets of attributes to be associated with different steps or categories. In the sG-DINA model, the Q-matrix is modified to accommodate  $\mathbf{q}_{kh}$ , the q-vector for category  $h$  of item  $k$ , where  $h = 1, 2, \dots, H_k$ . Note that for ordered polytomous response, 0 is one of the response categories (i.e.,  $X_k = \{h \mid h \in (0, 1, \dots, H_k)\}$ ), but this category does not require a q-vector. Hence, instead of  $K$  rows, the modified Q-matrix contains  $\sum_{k=1}^K H_k$  rows.

We can again assume that the first  $D_k^*$  are the required attributes for category  $h$  of item  $k$ . Conditional on the reduced attribute pattern  $\mathbf{a}_{lh}^*$ , the probability of a correct response to category  $h$  of item  $k$  given the previous step is correctly answered is denoted by

$$S_k(h|\mathbf{a}_{lh}^*) = P(X_{k,h} = 1 \mid X_{k,h-1}, \mathbf{a}_{lh}^*). \quad (7.5)$$

$S_k(h|\mathbf{a}_{lh}^*)$  is referred to as the *processing function* in the item response theory literature (Samejima, 1973). The processing function can be more generally formulated by using various link functions. In doing so, the IRF of the G-DINA model given in (7.1) can be used as the processing function to model a range of attribute interactions associated with the category response. Based on the sG-DINA model, the probability of obtaining a score of  $h$  on item  $k$  is given by

$$P(X_k = 1|\mathbf{a}_{lh}^*) = [1 - S_k(h+1|\mathbf{a}_{lh}^*)] \prod_{h'=1}^h S_k(h'|\mathbf{a}_{lh'}^*), \quad (7.6)$$

where

$$S_k(h|\mathbf{a}_{lh}^*) = \begin{cases} 1, & \text{if } h = 0 \\ 0, & \text{if } h = H_k + 1 \end{cases}.$$

The sG-DINA model is said to be restricted when the attribute-category associations are known. However, for some items, only the attribute-item associations

can be ascertained. For these items, the unrestricted version of the sG-DINA is used, where the same subset of attributes are specified for all categories. Although more general, and therefore more flexible, fitting the unrestricted sG-DINA model when the restrictions are appropriate can lead to suboptimal results. Originally the unrestricted sG-DINA model was designed for ordered responses; however, Ma and de la Torre (2016) have shown that the model can also be used in conjunction with nominal response, and is equivalent to the partial credit DINA model (de la Torre, 2010) and the nominal response diagnostic model (Templin, Rupp, Henson, Jang, & Ahmed, 2008). Finally, as expected, the sG-DINA model performs better than the G-DINA model fitted to dichotomized polytomous data.

### 7.3.3 *G-DINA Model for Continuous Response*

Although a number of CDMs for dichotomous and polytomous responses are available, modeling continuous response in the CDM context is in its infancy. With the proliferation of computer-based testing, perhaps the most obvious and readily-available source of continuous response is latency, or response time. However, other item formats such as placing a mark on a line segment (e.g., Noel, 2014; Noel & Dauvier, 2007) and probability testing (e.g., Ben-Simon, Budescu, & Nevo, 1997) can also yield continuous responses. For illustration purposes, we will use response time to represent continuous response throughout the chapter. As de la Torre and Minchen (2016), Minchen and de la Torre (2018) and Minchen, de la Torre, and Liu (2017) have shown, response time in the CDM context may itself be the work product of interest, or it could be viewed as a type of process data and used in conjunction with response accuracy.

The first CDM to handle responses of a strictly continuous type is the continuous DINA (cDINA) model proposed by Minchen et al. (2017). Like the DINA model, the cDINA model involves the same latent variable  $\mathbf{a}_l$ , classifies the examinees into one of two latent groups – those who have the required attributes for the items ( $\eta_{lk} = 1$ ), and those who do not ( $\eta_{lk} = 0$ ). However, instead of a single parameter (i.e., slip or guessing) governing the response of one particular group, the item response of a latent group in the cDINA model is governed by two parameters, representing the mean and standard deviation of the group's, say, response time on item  $k$ . It should also be noted that unlike dichotomous response where examinees in group  $\eta_{lk} = 1$  are expected to score higher, the expected response time of the same examinees can be longer or shorter depending on the context of application. The real data example in Minchen et al. shows that examinees in  $\eta_{lk} = 1$  are more engaged with problems that they are equipped to handle, resulting in longer response times.

Using the cDINA model, the cumulative distribution function for the response  $t_{lk}$  on item  $k$  given  $\mathbf{a}_l$  can be written as

$$P(T_k \leq t | \mathbf{a}_l) = \int_0^t f_{k\eta}(t_k) dt_k, \quad (7.7)$$

where

$$f_{k\eta}(t_k) = \frac{1}{t_k \sqrt{2\pi \sigma_{k\eta}^2}} \exp \left[ -\frac{(\ln t_k - \mu_{k\eta})^2}{2\sigma_{k\eta}^2} \right], \quad (7.8)$$

which is the lognormal distribution with group-specific parameters  $\mu_{k\eta}$  and  $\sigma_{k\eta}$  for  $\eta = 0, 1$ .

The continuous G-DINA (cG-DINA; Minchen & de la Torre, 2018) is a straightforward generalization of the cDINA model. Instead of two latent groups, the cG-DINA model allows for a unique response distribution to be associated with each of the  $2^{D_k^*}$  latent groups; thus it is characterized by  $2^{D_k^*+1}$  parameters. The cumulative distribution of the cG-DINA model for the response  $t_{lk}$  is similar to that in (7.7) with the exception that the lognormal distribution in (7.8) involves  $\mu_{k\eta}$  and  $\sigma_{k\eta}$  for  $\eta = 1, 2, \dots, 2^{D_k^*}$ , and a one-to-one correspondence between  $\eta$  and  $\mathbf{a}_{lk}^*$  can be made.

The cG-DINA model is a saturated model because each of the  $2^{D_k^*}$  latent groups is characterized by a unique parameter set  $(\mu_{k\eta}, \sigma_{k\eta})$ . By imposing the constraints  $\mu_{k1} = \dots = \mu_{k, 2^{D_k^*}-1}$  and  $\sigma_{k1} = \dots = \sigma_{k, 2^{D_k^*}-1}$ , the cDINA model can be easily derived from the cG-DINA model. Similar constraints can be imposed to derive a disjunctive CDM from the cG-DINA model. However, as noted earlier, CDMs for continuous response are in their nascent stages. At present, it is not clear how additive CDMs in this context should be formulated or what constraints on  $\mu_{k\eta}$  and  $\sigma_{k\eta}$  are needed to derive them from the saturated model. Furthermore, the existence of two parameters per latent group raises the possibility that the constrained model for  $\mu_{k\eta}$  may not be the same as that for  $\sigma_{k\eta}$ .

## 7.4 Estimation

An expectation-maximization (EM) implementation of marginalized maximum likelihood estimation (MMLE) can be used to obtain parameter estimates of the CDMs discussed above (e.g., de la Torre, 2009, 2011). Specifically, under the assumption of local independence, the log-marginalized likelihood of the dichotomous response data can be written as

$$\ell(\mathbf{X}) = \log \prod_{n=1}^N \sum_{l=1}^{2^D} P(\mathbf{X}_n | \mathbf{a}_l) p(\mathbf{a}_l), \quad (7.9)$$

where

$$P(\mathbf{X}_n | \mathbf{a}_l) = \prod_{k=1}^K P(X_{nk} = 1 | \mathbf{a}_l)^{X_{nk}} [1 - P(X_{nk} = 1 | \mathbf{a}_l)]^{1-X_{nk}}. \quad (7.10)$$

The MMLE/EM algorithm implements E-step and M-step iteratively item by item until convergence. In particular, the E-step calculates  $N_{a_{lk}^*} = \sum_{n=1}^N P(\mathbf{a}_{lk}^* | \mathbf{X}_n)$ , the expected number of individuals having the attribute pattern  $\mathbf{a}_{lk}^*$ , and  $R_{a_{lk}^*} = \sum_{n=1}^N x_{nk} P(\mathbf{a}_{lk}^* | \mathbf{X}_n)$ , the number of individuals with attribute pattern  $\mathbf{a}_{lk}^*$  expected to answer item  $k$  correctly. Note that  $P(\mathbf{a}_{lk}^* | \mathbf{X}_n)$  is the posterior probability of individual  $n$  having attribute pattern  $\mathbf{a}_{lk}^*$ . In the M-step, as shown in de la Torre (2011), the maximum likelihood estimate of  $P(X_k = 1 | \mathbf{a}_{lk}^*)$  is given by

$$\hat{P}(X_k = 1 | \mathbf{a}_{lk}^*) = \frac{R_{a_{lk}^*}}{N_{a_{lk}^*}}. \quad (7.11)$$

The item parameters  $\phi$  in (7.1) can be derived from (7.11) via the ordinal least-squares approach.

For the DINA and DINO models, the  $2^{D_k^*}$  latent groups are further partitioned into two non-overlapping groups  $\eta_{k0}$  and  $\eta_{k1}$ , where individuals in the former and latter groups are expected to answer item  $k$  incorrectly and correctly, respectively. The maximum likelihood estimate of the probability of success for individuals in group  $\eta_{ku}$  where  $u \in (0, 1)$  is

$$\hat{P}(X_k = 1 | \eta_{ku}) = \frac{\sum_{\mathbf{a}_{lk}^* \in \eta_{ku}} R_{a_{lk}^*}}{\sum_{\mathbf{a}_{lk}^* \in \eta_{ku}} N_{a_{lk}^*}}. \quad (7.12)$$

For A-CDM, LLM and R-RUM, the maximum likelihood estimate can be found using various optimization functions based on  $R_{a_{lk}^*}$  and  $N_{a_{lk}^*}$ . The parameters of the pG-DINA model can be estimated as in the G-DINA model after converting  $\mathbf{a}_{lk}^*$  to reduced dichotomous attribute vector  $\mathbf{a}_{lk}^*$ . For the sG-DINA model, the following objective function is maximized in the M-step,

$$\ell_k = \sum_{l=1}^{2^{D_k^*}} \sum_{h=0}^{H_k} R_{a_{lkh}^*} \log [P(X_k = h | \mathbf{a}_{lkh}^*)],$$

where  $R_{a_{lkh}^*} = \sum_{n=1}^N I(x_{nk} = h) P(\mathbf{a}_{lkh}^* | \mathbf{X}_n)$  is the number of individuals with attribute pattern  $\mathbf{a}_{lkh}^*$  expected to obtain a score of  $h$  on item  $k$ . Note that the EM algorithm for estimating the sG-DINA model can also be implemented at the category level after transforming the polytomous data to dichotomous data with missing values using the mapping matrix (Ma, 2018).

For the cG-DINA model, the conditional likelihood given in (7.10) can be written as

$$P(\mathbf{t}_n | \mathbf{a}_l) = \prod_{k=1}^K f_j(t_{nk} | \mathbf{a}_l). \quad (7.13)$$



Following several steps of derivation, the maximum likelihood estimates of  $\mu_{k\eta}$  and  $\sigma_{k\eta}^2$  can be shown to be equal to

$$\hat{\mu}_{k\eta} = \sum_{n=1}^N p^*(\mathbf{a}_{lk}|\mathbf{t}_n) \log t_{nk}, \tag{7.14}$$

and

$$\hat{\sigma}_{k\eta}^2 = \sum_{n=1}^N p^*(\mathbf{a}_{lk}|\mathbf{t}_n) (\log t_{nk} - \hat{\mu}_{k\eta})^2, \tag{7.15}$$

respectively, where  $p^*(\mathbf{a}_{lk}|\mathbf{t}_n)$  is the posterior probability (normalized across the  $N$  examinees) of examinee  $n$  being in the reduced attribute pattern  $\mathbf{a}_{lk}$ .

Unlike traditional IRT, where the prior ability distribution can be reasonably specified, for example, using  $N(0, 1)$ , the multinomial attribute distribution  $p(\mathbf{a}_l)$  in CDM cannot be readily determined a priori. A convenient way of specifying  $p(\mathbf{a}_l)$  is to employ the empirical Bayes estimate. In particular, we let  $p^{(c+1)}(\mathbf{a}_l)$ , the prior distribution at iteration  $c + 1$ , be equal to the  $p^{(c)}(\mathbf{a}_l | \mathbf{X})$ , the posterior distribution at iteration  $c$ . It should be noted that in the CDM context, estimation of the item response model can impact the joint attribute distribution estimate, and vice versa. Therefore, in situations where the impact of model misspecification on item parameter estimates needs to be isolated, one can use the G-DINA model to arrive at the correct attribute distribution estimate in the first step, and, fixing the attribute distribution, use the EM algorithm to obtain the item parameter estimates of the reduced model in the second step.

## 7.5 G-DINA Model-Based Methodologies

### 7.5.1 Q-Matrix Validation

In typical CDM applications, Q-matrices are built by subject-matter experts. In addition to subjective judgments, experts may not reach complete agreement on each of the Q-matrix entries. For these reasons, the correctness of the entire Q-matrix cannot be guaranteed. To address this issue, statistical procedures, referred to in the literature as *empirical Q-matrix validation* methods, have been proposed.

De la Torre and Chiu (2016) proposed the G-DINA model discrimination index (GDI) for an item with any q-vector. For simplicity of notation, let us assume again that the first  $D_k^*$  attributes are required for item  $k$ . The GDI is defined as

$$\varsigma_{1:D_k^*}^2 = \sum_{l=0}^{2^{D_k^*}} p(\mathbf{a}_l^*) \left[ P(X_k = 1 | \mathbf{a}_l^*) - \bar{p}_k \right]^2 \tag{7.16}$$

where  $p(\mathbf{a}_l^*)$  is relative size of the reduced attribute pattern  $\mathbf{a}_l^*$ , and  $\bar{p}_k$  is the mean success probability on item  $k$ . As can be seen from (7.16), the GDI is simply the variance of the success probabilities given a particular q-vector. For each item,  $2^D - 1$  q-vectors can be specified, each corresponding to one GDI. De la Torre and Chiu (2016) defined a q-vector that results in the maximum  $\zeta_k^2$  as an *appropriate* q-vector to item  $k$ . Of the appropriate q-vectors, the q-vector with the minimum number of attributes specified is deemed *correct*.

The GDI serves as the basis of the EM-based data-driven algorithm (de la Torre & Chiu, 2016) developed to validate the expert-based provisional Q-matrix. Compared to other data-driven Q-matrix validation methods that are designed for specific CDMs (e.g., the  $\delta$ -method for the DINA model; de la Torre, 2008), the GDI is based on a general model so it can be used with any reduced CDMs the G-DINA model subsumes. In practice, the inequality established by de la Torre and Chiu (2016) may not hold due to potential misspecifications in the provisional Q-matrix as well as noise in the data. As a matter of fact, the maximum  $\zeta_k^2$  is always achieved when  $\mathbf{q}_k = \mathbf{1}$ , which, more often than not, is an overspecification. To address this issue, they recommended examining the proportion of variance accounted for a particular q-vector relative to the maximum  $\zeta_k^2$ , and suggested selecting the simplest q-vector from a set of q-vectors with GDIs above a particular cutoff (e.g.,  $\zeta^2 > 0.95$ ). Although it has been shown that the GDI-based procedure can be a reliable method of empirically validating a provisional Q-matrix, particularly when high quality items are involved, determining a single cutoff that is optimal across a variety of conditions remains a challenge. To minimize dependence on a single cutoff and to allow for quantitative and graphical information to be combined in determining the correct q-vector for an item, de la Torre and Ma (2016) proposed the use of the *mesa plot*. The mesa plot displays the GDIs of different q-vectors in ascending order. Instead of a single recommendation, a number of q-vectors in the vicinity where the plot plateaus or forms a tabletop are suggested from which the correct q-vector can be selected.

## 7.5.2 Cognitive Diagnosis Computerized Adaptive Testing

As in traditional IRT, computerized adaptive testing can also be used to improve test efficiency (i.e., shorter test or greater accuracy) in the CDM context by administering items that are tailored to an examinee's most current attribute estimate. However, due to the discrete and multidimensional nature of the attributes, the method for determining the optimal item in cognitive diagnosis computerized adaptive testing (CD-CAT) differs.

Kaplan, de la Torre, and Barrada (2015) used the GDI as an item selection index for CD-CAT. Specifically, for examinee  $n$ , the GDI for item  $k$  at time  $t$  (i.e., after  $t$  items have been administered) is computed as

$$\zeta_k^{2(t)} = \sum_{l=0}^{2^{D_k^*}} p(\mathbf{a}_l | \mathbf{X}_n^{(t)}) \left[ P(X_k = 1 | \mathbf{a}_l^*) - \bar{p}_{nk}^{(t)} \right]^2, \quad (7.17)$$

where  $p(\mathbf{a}_l | \mathbf{X}_n^{(t)})$  is posterior probability of  $\mathbf{a}_l^*$  at time  $t$ ,  $P(X_k = 1 | \mathbf{a}_l^*)$  is the time-invariant success probability on item  $k$  given  $\mathbf{a}_l^*$ , and  $\bar{p}_{nk}^{(t)}$  current overall item difficulty. Note that, as a CD-CAT item selection index, (7.17) is a function of  $p(\mathbf{a}_l | \mathbf{X}_n^{(t)})$ , which changes over time. The item with the largest  $\zeta_k^{2(t)}$  is deemed most informative, and hence administered at time  $t + 1$ .

To examine the viability of the GDI as a CD-CAT item selection index, Kaplan et al. (2015) compared it with the posterior-weighted Kullback-Leibler (PWKL; Cheng, 2009) index, as well as the doubly-posterior-weighted modified PWKL (MPWKL) index, which they also introduced. They found that the GDI and MPWKL outperformed the PWKL when the reduced model is either the DINA or DINO model, but not when it is the A-CDM. In addition, although GDI and MPWKL performed similarly in terms of correct classification rate or average test length, the former was deemed more efficient in that it only required a fraction of the time to be implemented.

### 7.5.3 Item-Level Model Comparison

Given the variety of CDMs currently available, it is not obvious how the choice between these models can be made in practice. Previously, researchers assume a particular underlying process (e.g., conjunctive, additive) to fit a particular CDM (i.e., DINA model, R-RUM) to the data. With the availability of general models, fitting CDMs with less restrictive assumptions has been advocated. However, recent analyses of real data show that different items may require different types of CDMs, both reduced and saturated. These findings imply that a single reduced CDM would likely not provide a sufficient model-data fit. Moreover, even if a general model may provide an adequate fit assuming CDMs are appropriate, the parsimony principle (Beck, 1943) dictates that the simplest set of models that can provide equally good fit to the data be chosen. These findings also imply that using a test-level comparison using, say, Akaike (1973) or Bayesian information criterion, or the likelihood ratio test to choose *en masse* from among the CDMs that have been specified a priori may not lead to the selection of the optimal CDMs for the data.

To determine empirically (i.e., *post hoc*) the most appropriate CDM for each item, de la Torre (2011) developed an item-level model selection method using the Wald test. Assuming the Q-matrix has been validated, the Wald test can be used to determine whether one or more reduced CDMs can be used in place of the saturated CDM. For item  $k$ , the Wald statistic for comparing the reduced CDM  $\varrho$  against the saturated model is defined as

$$W_{k\varrho} = [\mathbf{R}_{k\varrho} \times g(\mathbf{P}_k)]' [\mathbf{R}_{k\varrho} \times \text{Var}[g(\mathbf{P}_k)] \times \mathbf{R}'_{k\varrho}] [\mathbf{R}_{k\varrho} \times g(\mathbf{P}_k)], \quad (7.18)$$

where  $g(\mathbf{P}_k)$  is  $g[P(X_k = 1 | \mathbf{a}_{lk}^*)]$ ,  $\text{Var}[g(\mathbf{P}_k)]$  is the corresponding variance matrix, and  $\mathbf{R}_{k\varrho}$  is the restriction matrix associated with the reduced CDM  $\varrho$ . The

restriction matrix  $\mathbf{R}_{k_Q}$  is of size  $(2^{D_k^*} - p) \times 2^{D_k^*}$ , where  $p$  is the number of parameters in model  $Q$ . Below are examples of  $\mathbf{R}$  for the (1) DINA model, (2) DINO model, and (3) additive models when  $D_k^* = 2$ :

$$\mathbf{R}^{(1)} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}, \quad \mathbf{R}^{(2)} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \text{and} \quad \mathbf{R}^{(3)} = [1 \ -1 \ -1 \ 1].$$

The Wald statistic  $W_{k_Q}$  is assumed to be asymptotically  $\chi^2$ -distributed with  $(2^{D_k^*} - p)$  degrees of freedom. It should be noted that using the Wald test for the purpose of evaluating the appropriateness of reduced CDMs is only meaningful when  $D_k^* \geq 2$ .

With a sufficiently large sample size and reasonable item quality, the Wald test has acceptable Type I error and power across various reduced models (de la Torre & Lee, 2013; Ma & de la Torre, 2016). Furthermore, in comparing the fit of CDMs selected via the Wald test against that of the G-DINA model, Ma, Iaconangelo, and de la Torre (2016) found using simulated and real data that the former provided higher correct classification rate than the latter, particularly when lower item qualities and smaller sample sizes are involved. More recently, de la Torre and Ma (2017) have shown that performing the Wald test is a necessary step to accurately evaluate whether or not a test can potentially identify all the possible attribute patterns. An evaluation of the expected item response profiles derived from fitting a saturated model without considering the appropriateness of reduced CDMs can lead to incorrect conclusions about the identifiability, or lack thereof, of the attribute patterns. Lastly, the use of the Wald test in the CDM context extends beyond item-level model comparison – it has also been used to evaluate differential item functioning (e.g., Hou & Terzi, 2017).

## 7.6 Discussion

This chapter presents the G-DINA model as framework for conducting analysis in the CDM context. As a general model and with appropriate link functions, the G-DINA model can be shown to subsume a number of familiar reduced CDMs in the literature. With it as the base model, the G-DINA model can be extended in various directions to address a wider range of practical testing situation needs. As a framework, the G-DINA model provides a coherent environment where CDM-related procedures can be developed and implemented. Thus far, the CDM-based methodologies that have been developed are largely applicable to CDM for dichotomous responses and attributes. To further improve the practicability of CDMs, these methodologies should be expanded to also apply to other CDM types.

The surge in the development of CDMs and related methodologies in recent years is without a doubt a positive development in this field. However, using these models and methodologies systematically and integratively can be daunting, particularly to many applied researchers. If any suggestions could be proffered regarding this

matter, they would be as follows. First, validate the Q-matrix specification. To do so without conflating Q-matrix misspecifications with potential CDM misspecifications, fit the G-DINA model. Second, check whether reduced CDMs can be used in place of the G-DINA model for items where  $D_k^* \geq 2$ . More likely than not, this would result in different items retaining different CDMs. Third, recalibrate the data using the CDMs selected in the previous step to update the estimates of the item parameters and attribute distributions. These are the estimates that one can use in estimating the examinees' attribute patterns. Optionally, in some applications, one can also consider whether the attribute distribution, which is typically estimated in saturated form (i.e., without constraints), can be simplified. An alternative is to specify the attribute distribution using a higher-order formulation (de la Torre & Douglas, 2004). As a final step, evaluate the absolute fit (i.e., goodness-of-fit) of the model to the data. One way this can be accomplished is by comparing the expected and observed moments, particularly the correlation and log-odds ratio, of each item pair (Chen, de la Torre, & Zhang, 2013; de la Torre & Douglas, 2008).

As a last word, we should be cognizant that, despite the numerous developments pertaining to CDM and related methodologies in the last two decades, these advances represent but one side of the coin. To fully take advantage of the potential of CDMs, we should also focus our attention on the other side of the same coin, which is developing cognitively diagnostic assessments (CDAs; de la Torre & Minchen, 2014). In particular, we need to develop diagnostic assessments from the ground up using a CDM framework. On one hand, without the appropriate data, psychometric tools no matter how sophisticated cannot produce the rich information needed for precise diagnosis of student needs. On the other hand, without the appropriate psychometric tools, information no matter how rich cannot be properly extracted and utilized. Thus, for cognitive diagnosis modeling to break new ground in the near future, CDMs and CDAs must be used hand in hand.

## References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.
- Beck, L. W. (1943). The principle of parsimony in empirical science. *The Journal of Philosophy*, *40*, 617–633.
- Ben-Simon, A., Budescu, D. V., & Nevo, B. A. (1997). Comparative study of measures of partial knowledge in multiple choice tests. *Applied Psychological Measurement*, *21*, 65–88.
- Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement*, *37*, 419–437.
- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, *50*, 123–140.
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, *74*, 619–632.
- 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. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115–130.
- de la Torre, J. (2010, July). The partial-credit DINA model. Paper Presented at the International Meeting of the Psychometric Society, Athens, GA.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199.
- de la Torre, J. (2015, December). *Cognitively diagnostic assessment and cognitive diagnosis modeling: An example from start to finish*. Invited Presentation at the Global Chinese Conference on Educational Information and Assessment and Chinese Association of Psychological Testing Annual Conference, Taichung, Taiwan.
- de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, *81*, 253–273.
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models 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.
- de la Torre, J., & Lee, Y.-S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, *50*, 355–373.
- de la Torre, J., & Ma, W. (2016, August). *Cognitive diagnosis modeling: A general framework approach and its implementation in R*. A Short Course at the Fourth Conference on Statistical Methods in Psychometrics, Columbia University, New York.
- de la Torre, J., & Ma, W. (2017, November). *Do I complete Q?* Invited Presentation at the Fifth Conference on the Statistical Methods in Psychometrics, Department of Statistics, Columbia University, New York.
- de la Torre, J., & Minchen, N. D. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, *20*, 89–97.
- de la Torre, J., & Minchen, N. D. (2016, May). *Modeling response time in cognitive diagnosis*. Invited Presentation at the Graduate Institute of Educational Measurement and Statistics Colloquium, National Taichung University of Education, Taiwan.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 301–321.
- Hansen, M. (2013). *Hierarchical item response models for cognitive diagnosis*. (Unpublished doctoral dissertation). Los Angeles: University of California.
- 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.
- Hou, L., & Terzi, R. (2017, April). *Examining DIF in the proportional reasoning test using various Wald test formulations*. Paper Presented at the Meeting of National Council on Measurement in Education, San Antonio, TX.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.
- Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*(3), 167–188.
- Ma, W. (2018). A diagnostic tree model for polytomous responses with multiple strategies. *British Journal of Mathematical and Statistical Psychology*. Advanced online publication. <https://doi.org/10.1111/bmsp.12137>
- Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, *69*, 253–275.
- Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, *40*, 200–217.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187–212.

- Minchen, N. D., & de la Torre, J. (2018). A general cognitive diagnosis model for continuous-response data. *Measurement: Interdisciplinary research and perspective*, *16*, 30–44.
- Minchen, N. D., de la Torre, J., & Liu, Y. (2017). A cognitive diagnosis model for continuous response. *Journal of Educational and Behavioral Statistics*, *42*, 651–677.
- Noel, Y. (2014). A beta unfolding model for continuous bounded responses. *Psychometrika*, *79*, 647–674.
- Noel, Y., & Dauvier, B. (2007). A beta item response model for continuous bounded responses. *Applied Psychological Measurement*, *31*, 47–73.
- Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika*, *38*, 203–219.
- 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. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287–305.
- Templin, J. L., Rupp, A., Henson, R., Jang, E., & Ahmed, M. (2008, March). *Cognitive diagnosis models for nominal response data*. Paper Presented at the Annual Meeting of the National Council on Measurement in Education, New York.
- Tjoe, H., & de la Torre, J. (2013a). Designing cognitively-based proportional reasoning problems as an application of modern psychological measurement models. *Journal of Mathematics Education*, *6*, 17–22.
- Tjoe, H., & de la Torre, J. (2013b). The identification and validation process of proportional reasoning attributes: An application of a cognitive diagnosis modeling framework. *Mathematics Education Research Journal*, *26*, 237–255.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307.
- von Davier, M. (2014). *The log-linear cognitive diagnostic model (LCDM) as a special case of the general diagnostic model (GDM)* (Research Report No. RR-14-40). Princeton, NJ: Educational Testing Service. Retrieved from <https://doi.org/10.1002/ets2.12043>