Psicothema 2020, Vol. 32, No. 1, 122-129
doi: 10.7334/psicothema2019.182

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

# Estimation approaches in cognitive diagnosis modeling when attributes are hierarchically structured

Lokman Akbay[1], and Jimmy de la Torre[2]
[1] Burdur Mehmet Akif Ersoy University, and [2] University of Hong Kong

## Abstract

**Background:** Although research in cognitive psychology suggests refraining from investigating cognitive skills inisolation, many cognitive diagnosis model (CDM) examples do not take hierarchical attribute structures into account. When hierarchical relationships among the attributes are not considered, CDM estimates may be biased. **Method:** The current study, through simulation and real data analyses, examines the impact of different MMLE-EM approaches on the item and person parameter estimates of the G-DINA, DINA and DINO models when attributes have a hierarchical structure. A number of estimation approaches that can result from modifying either the Q-matrix or prior distribution are proposed. Impact of the proposed approaches on item parameter estimation accuracy and attribute classification are investigated. **Results:** For the G-DINA model estimation, the Q-matrix type (i.e, explicit vs. implicit) has greater impact than structuring the prior distribution. Specifically, explicit Q-matrices result in better item parameter recovery and higher correct classification rates. In contrast, structuring the prior distribution is more influential on item and person parameter estimates for the reduced models. When prior distribution is structured, the Q-matrix type has almost no influence on item and person parameter estimates of the DINA and DINO models. **Conclusion:** We can conclude that the Q-matrix type has a significant impact on CDM estimation, especially when the estimating model is G-DINA.

***Keywords:*** CDM, cognitive diagnosis modeling, estimation approaches, Q-matrix.

## Resumen

***Enfoques de estimación en Modelado Diagnóstico Cognitivo cuando los atributos están estructurados jerárquicamente.*** **Antecedentes:** a pesar de que investigación en psicología cognitiva sugiere abstenerse de investigar rasgos cognitivos de forma aislada, muchos de los ejemplos en Modelado Diagnóstico Cognitivo (MDC) no tienen en cuenta la estructura jerárquica de los atributos implicados. Sin embargo, las estimaciones que se hagan con los MDC pueden estar sesgadas cuando no se consideran estas relaciones jerárquicas. **Método:** a través de la simulación y datos reales, el presente estudio estudia el impacto de diferentes enfoques MMLE-EM en los parámetros estimados para los ítems y las personas según los modelos G-DINA, DINA y DINO cuando los atributos tienen una estructura jerárquica. Se proponen una serie de enfoques de estimación que resultan de modificar la Matriz-Q o la distribución previa. Se investiga el impacto de los enfoques propuestos en la precisión en la estimación de los parámetros de los ítems y la clasificación de atributos. **Resultado:** para la estimación del modelo G-DINA, el tipo de Matriz-Q (es decir, explícita vs. implícita) tiene un impacto mayor al de que la distribución previa esté estructurada. Por el contrario, una distribución previa estructurada influye más sobre la estimación de los parámetros de los ítems y las personas en el caso de los modelos reducidos. **Conclusión:** podemos concluir que el tipo de Matriz-Q tiene un impacto significativo en la estimación de MDC, especialmente en el modelo G-DINA.

***Palabras clave:*** MDC, modelado diagnóstico cognitivo, enfoques de estimación, Matriz-Q.

Educational measurement professionals have underscored the appreciable role of cognitive theory in educational testing (e.g., Chipman, Nichols, & Brennan, 1995; Embretson, 1985). Because knowledge, mental processes, and examinees' response strategies define construct representation, Embretson (1983) argued that cognitive theory could improve psychometric practice by guiding the construct representation of a test. Leighton, Gierl, and Hunka (2004) claimed that the cognitive requirements eliciting particular knowledge structures, processes, skills, and strategies, which are referred to as *attributes* (de la Torre, 2009b; de la Torre & Lee, 2010), could be assembled into cognitive models that are then used to develop test items. Assessments that are developed for identifying attribute mastery status of examinees to obtain convincing evidence for diagnostic inferences about examinees' cognitive strengths and weaknesses are referred to as cognitively diagnostic assessments (CDAs: de la Torre & Minchen, 2014). For CDA to impact the testing practice, the role of cognitive theory needs to be well articulated in the test design. However, until quite recently, the impact of cognitive theory on test design has been minimal (Embretson, 1998; National Research Council, 2001). Embretson (1994) has attributed this to the lack of frameworks that use cognitive theory in the test development. Recently, various approaches integrating cognitive theory into psychometric practice have been proposed. These include *the rule space methodology* (Tatsuoka, 1983), *the attribute hierarchy method* (Leighton, Gierl,

Received: June 19, 2019 • Accepted: September 30, 2019
Corresponding author: Lokman AKBAY
Burdur Mehmet Akif Ersoy University
Burdur Mehmet Akif Ersoy Universitesi Egitim Fakultesi A Blok Oda:407
1500 Burdur (Turquía)
e-mail: lokmanakbay@gmail.com

& Hunka, 2004), and *the generalized-DINA model framework* (de la Torre, 2011).

CDAs can be used for formative purposes, and the feedback obtained from analysis of the assessment data can be used to modify teaching and learning activities (DiBello & Stout, 2007). Attention on CDAs has increased as the political changes, such as the *No Child Left Behind Act* (2001), emphasized the need for assessments that are more formative in nature. Thereafter, a number of statistical models, referred to as cognitive diagnosis models (CDMs) or diagnostic classification models (DCMs), to extract diagnostic information from CDAs have been proposed (de la Torre & Minchen, 2014). CDMs involve two components: The interaction of the attributes with each other in the response construction process, and the specification of the attributes needed for each item (Chiu, Douglas, & Li, 2009). In CDMs, a $JxK$ matrix, referred to as the Q-matrix (Tatsuoka, 1983), is used to set the item-by-attribute specifications. The Q-matrix is a binary matrix of $J$ rows and $K$ columns, where $j = 1 \ldots, J,$ represents the test items, and $k = 1, \ldots K$ represents attributes measured by the test. When item $j$ requires examinees to possess attribute $k$ for a successful response, the $q_{jk}$ element of Q-matrix is coded as 1; otherwise, it is coded as 0.

Attributes specified in the Q-matrix may be independent of each other, or they may have a dependent structure. Two types of attribute dependencies exist: higher-order and hierarchical. Attributes are said to follow a higher-order structure when an examinee's mastery probability of each attribute is determined by an overall continuous ability (de la Torre & Douglas, 2004). Such a structure still allows for all the possible attribute patterns, although some patterns are deemed less likely than others. In contrast, under a hierarchical attribute structure, attributes have deterministic prerequisite relationships, where mastery of basic attributes is required for mastering more complex attributes (de la Torre, Hong, & Deng, 2010; Leighton et al., 2004). When this is the case, CDMs need to consider the hierarchical structure to obtain more accurate item and person parameter estimates (Templin & Bradshaw, 2014).

Although several studies investigated the effect of impermissible attribute patterns in CDM estimation when attributes have a hierarchical structure (de la Torre et al., 2010; Templin & Bradshaw, 2014; Tu, Wang, Cai, Douglas, & Chang, 2018), the impact of Q-matrix type (i.e, hierarchically structured and unstructured) has not been investigated in these studies. Therefore, the current study aims to investigate the impact of different CDM estimation approaches, which are based on the constraint or unconstraint status of the Q-matrix and the prior distribution used in the estimation algorithm, on the item parameter estimates and correct attribute classification rates.

## Method

### *Procedure*

When item parameters of a given CDM have been pre-calibrated, the maximum likelihood estimation method can be used for attribute estimation (de la Torre, 2009b). For the binary response data, the likelihood function of latent class $\boldsymbol{\alpha}$ becomes:

$$L(\boldsymbol{X}|\boldsymbol{\alpha}) = \prod_{i=1}^{N}\prod_{j=1}^{J} P(X_{ij}=1|\boldsymbol{\alpha}_i)^{X_{ij}}\left[1-P(X_{ij}=1|\boldsymbol{\alpha}_i)\right]^{1-X_{ij}} \quad (1)$$

where $P(X_{ij} = 1 \mid \boldsymbol{\alpha}_l)$ is the response function of a given CDM, and $X_{ij}$ is observed response of examinee $i$ to item $j$. The $\boldsymbol{\alpha}$ pattern maximizing the likelihood function becomes the estimated attribute profile of the examinee. When item parameters are unknown, marginalized maximum likelihood estimation can be implemented using an expectation-maximization (Dempster, Laird, & Rubin, 1977) algorithm (de la Torre, 2009b). In marginalized maximum likelihood estimation, estimates of the model parameters (structural parameters) are attained by maximizing the likelihood with the attribute class probabilities (incidental parameters , $P(\boldsymbol{\alpha}_l)$, $l = 1, \ldots, 2^K$ ). Given a model response function $P(X_{ij} = 1 \mid \boldsymbol{\alpha}_l)$, the marginalized likelihood is expressed as

$$L(\boldsymbol{X}) = \prod_{i=1}^{N}\prod_{l=1}^{2^K} L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l) \quad , (2)$$

where $p(\boldsymbol{\alpha}_l)$ is the prior probability of attribute class $\boldsymbol{\alpha}_l$. Initial item parameters and attribute class probabilities have to be chosen in the first iteration of the expectation-maximization algorithm. Item parameters and attribute class probabilities are updated in each cycle until convergence is achieved. As an alternative to maximum likelihood estimation, a Bayesian approach can be used for estimation of the examinees' attribute profiles. In this approach, the posterior distribution is obtained using Bayes theorem, and is proportional to the likelihood of the observed data times the prior distribution, as in, $p(\boldsymbol{\alpha} \mid X) \propto L(X \mid \boldsymbol{\alpha})p(\boldsymbol{\alpha})$. Similarly, a Bayesian approach can also be used when the item parameters are unknown to estimate the item parameters and attribute patterns.

When attributes do not have any hierarchical structure, all of the $2^K$ attribute patterns are permissible. Conversely, when the attributes are dependent with respect to some hierarchical structure, some attribute patterns are not permissible (de la Torre et al., 2010; Leighton et al., 2004). That is, attribute patterns having an attribute without possessing its prerequisite(s) are not allowed. For example, Figure 1 demonstrates linear, convergent and divergent hierarchies for six attributes (i.e., A1, A2, A3, A4, A5 and A6). The linear hierarchy in the figure implies that A1 must be mastered before mastering A2; A2 must be mastered before mastering A3 and so on. Under the linear structure displayed in Figure 1, the attribute patterns 000000, 100000, 110000, 111000, 111100, 111110, and 111111 are permissible, whereas the remaining 57

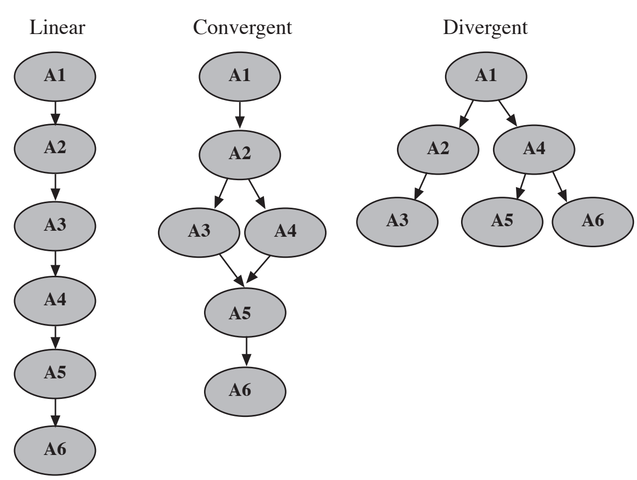


**Figure 1.** Three general types of hierarchies defined by Leighton et al. (2004)

attribute patterns are not. In the convergent structure, A3 and A4 are not prerequisites to one another; however, A1 and A2 are prerequisites for both A3 and A4. Furthermore, mastering either A3 or A4 satisfies the prerequisite for mastering A5. Therefore, in the convergent structure, on top of the permissible attribute patterns in linear structure, five more attribute patterns (i.e., 110100, 110110, 110111, 111010, and 111011) are permissible. The divergent hierarchy allows even more attribute patterns based on the prerequisite relationships among given six attributes. Under these structures, instead of $2^6$ attribute patterns, only seven, 12 and 16 latent classes are permissible under the linear, convergent, and divergent structures, respectively. All attribute patterns allowed by these three structures are given in Table 1. Attribute structure can be incorporated in the expectation-maximization algorithm, specifically, by manipulating the prior distribution. Although we cannot precisely know the distribution of the permissible latent classes, we can assign a prior probability of zero to theoretically impermissible attribute classes by a careful consideration of the hierarchical structure of the attributes. In such cases, imposing zero prior probabilities for the impermissible latent classes yields a *structured prior distribution*.

In cases where attributes have a hierarchical structure test items may or may not explicitly require the more basic attributes (i.e., prerequisites) along with the more complex attributes for a successful response (de la Torre et al., 2010). De la Torre et al. (2010) provided an example where they argue that *taking the derivative* presupposes *knowledge of basic arithmetic operation*, but, an item can be constructed such that it requires ability to differentiate without basic arithmetic operations. Therefore, a Q-matrix for noncompensatory models can be developed by either specifying only explicitly required attributes or specifying both explicitly and implicitly required attributes. For example, when attributes have a linear structure as displayed in Figure 1, the q-vector of an item explicitly requiring only the third out of the six attributes may be represented as 001000 or 111000 based on the explicitly required attribute only or on both the explicitly and implicitly required attributes, respectively.

In this manuscript, Q-matrices that specify only the explicitly required attributes will be referred to as *explicit Q-matrix*, whereas those that specify implicitly required attributes will be referred to as *implicit Q-matrix*. For a disjunctive model, only the most basic attribute is specified in the implicit Q-matrix even though the item also probes attributes that are more complex. For instance, an item explicitly requiring first three attributes under the linear structure given in Figure 1 may be represented as either 100000 by an implicit Q-matrix, or 111000 by an explicit Q-matrix. Impact of implicit and explicit Q-matrices on CDM estimation under hierarchical attribute structures has not been examined. Hence, to fill this gap, the current study investigates the impact of Q-matrix types when attributes follow a hierarchy.

*Simulations*

A simulation study was designed to understand the impact of Q-matrix type, if any, on item parameter estimation and correct attribute classification when the attributes are hierarchical. To accomplish this, explicit and implicit Q-matrices were crossed with the unstructured and structured forms of prior distributions. This resulted in four different approaches that can be employed in CDM estimations. Three general attribute hierarchies (i.e., linear, convergent, and divergent) with six attributes, as defined by Leighton et al. (2004), were considered. The explicit Q-matrix is also given in Table 1. Explicit Q-matrix consisted of items requiring one, two, and three attributes. Although the first 18 items designed to measure each of the six attributes equally, two more items (i.e., item 19 and item 20) were added to have at least two items differentiating adjacent latent classes (e.g., 000000 and 100000, and 111110 and 111111) for implicit Q-matrices.

The impact of the estimation approaches were studied under various conditions, where two levels of item quality and three generating models (i.e., G-DINA, DINA, and DINO) were employed. The item quality was defined by the item discrimination (i.e., $1 - s - g$) for higher and lower item quality levels. For the higher-quality (HQ) items, the lowest and highest success

*Table 1*
Permissible attribute patterns and the explicit Q-matrix

| Permissible attribute patterns | | | | | | | | | | | | | | | | | | Explicit | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Linear | | | | | | Convergent | | | | | | Divergent | | | | | | Q-matrix | | | | | |
| A1 | A2 | A3 | A4 | A5 | A6 | A1 | A2 | A3 | A4 | A5 | A6 | A1 | A2 | A3 | A4 | A5 | A6 | A1 | A2 | A3 | A4 | A5 | A6 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | | | | | | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| | | | | | | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | | | | | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| | | | | | | | | | | | | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| | | | | | | | | | | | | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| | | | | | | | | | | | | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| | | | | | | | | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| | | | | | | | | | | | | | | | | | | 1 | 0 | 0 | 0 | 1 | 1 |
| | | | | | | | | | | | | | | | | | | 1 | 1 | 0 | 0 | 0 | 1 |
| | | | | | | | | | | | | | | | | | | 1 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 1 |

probabilities (i.e., $P(0)$ and $P(1)$) were generated from $U(0.05, 0.20)$ and $U(0.80, 0.95)$, respectively. For the lower-quality (LQ) items, the lowest and highest success probabilities were drawn from $U(0.15, 0.30)$ and $U(0.70, 0.85)$, respectively. In other words, the slip and guessing parameters to generate the data were drawn from $U(0.05, 0.20)$ and $U(0.15, 0.30)$ for higher and lower item quality conditions, respectively. Attributes were uniformly generated following the linear, convergent, and divergent hierarchies.

Two-levels of item quality, three CDMs, and three general hierarchy types were crossed to form the simulation conditions. In all conditions, the number of items, number of measured attributes and number of examinees were fixed to 20, 6, and 1000, respectively. Moreover, the number of replication for each condition was fixed to 100. To determine the factors and their levels, similar theoretical studies (e.g., de la Torre & Lee, 2010; de la Torre, Hong, & Deng, 2010) and the results of a review study on the empirical CDM applications (i.e., Sessoms & Henson, 2018) were considered. All the factors with varying levels are summarized in Table 2.

For the estimation purposes, only the generating models were fitted to corresponding data to be consistent with the aim of the simulation study, which is to understand what gains, if any, are associated with the different estimation approaches. Although most of the time the true model that is associated with the data may not be known a priori, one may select the best fitting model based on the relative fit statistics, such as AIC (Akaika information criterion) and BIC (Bayesian information criterion). Throughout the study data generation and model estimation conducted in R language and statistical computing environment using R-package GDINA (Ma & de la Torre, 2018). It should be noted that attribute estimation was based on expected a posteriori (EAP) estimator, which is the mean of the posterior distribution. Specifically, EAP is the probability of mastering an attribute and it is converted into 1 or 0 (i.e., mastery and nonmastery) by comparing it to a threshold value such as .50.

To determine the impact of the Q-matrix design on item parameter estimation accuracy and precision, *the absolute bias* and *the root mean squared error (RMSE)* of the estimates across 100 replications were computed. The bias and RMSE for guessing are defined as

$$bias_g = \frac{1}{JxR}\sum_{j=1}^{J}\sum_{r=1}^{R}\left[\hat{g}_{jr} - g_{jr}\right] \tag{3}$$

and

$$RMSE_g = \frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{g}_{jr} - g_{jr}\right)^2}, \tag{4}$$

*Table 2*
Simulation factors

| Type of CDM | Prior distribution | Type of Q-matrix | Type of hierarchy | Item quality |
|---|---|---|---|---|
| G-DINA<br>DINA<br>DINO | Structured<br>Unstructured | Explicit<br>Implicit | Linear<br>Convergent<br>Divergent | Higher Quality<br>Lower Quality |

Note: CDM = cognitive diagnosis model; G-DINA = generalized deterministic input, noisy "and" gate model; DINA = deterministic input, noisy "and" gate model; DINO = deterministic input, noisy "or" gate model

respectively, where $J$ is the number of items; $R$ is the number of replications; $\hat{g}_{jr}$ is the guessing parameter estimate for item $j$ in replication $r$; $g_{jr}$ is the generating guessing parameter for item $j$ in replication $r$. Note that the same formulas can be used for slip parameter, where $g$ is replaced by $s$. Recall that the number of item parameters in the G-DINA model is a function of the number of required attributes, which may be different for the explicit and implicit Q-matrices. Therefore, not all item parameters are directly comparable for different types of estimation approaches employing an explicit or implicit Q-matrix. Therefore, we only considered the recovery of the lowest and highest success probability parameters in computation of item bias and RMSE.

The correct attribute classification rates at the individual-attribute level (i.e., *correct attribute classification rate; CAC*) and at the attribute-vector level (i.e., *correct vector classification rate; CVC*) were also investigated. The CAC and CVC can be computed using the formulae

$$CAC_k = \sum_{r=1}^{R}\sum_{i=1}^{N}\frac{I\left[\hat{a}_{ik}^{r} - a_{ik}^{r}\right]}{NR} \tag{5}$$

and

$$CVC = \sum_{r=1}^{R}\sum_{i=1}^{N}\frac{I\left[\hat{\boldsymbol{a}}_{i}^{r} - \boldsymbol{a}_{i}^{r}\right]}{NR}, \tag{6}$$

respectively, where $N$ is the total number of examinees; $I$ is the indicator function; $a^{r}_{ik}$ is true mastery status of examinee $i$ for attribute $k$ in replication $r$; $\hat{a}^{r}_{ik}$ is the EAP estimate of examinee $i$ for attribute $k$ in replication $r$; $\boldsymbol{\alpha}^{r}_{i}$ is true attribute pattern of examinee $i$ in replication $r$ ; and $\hat{\boldsymbol{\alpha}}^{r}_{i}$ is the estimated attribute pattern for the same examinee in the same replication.

## Results

### *Simulation results*

#### *Results on Model Parameter Estimation*

The absolute bias and RMSE of model parameter estimates obtained from the four estimation approaches are provided in the upper and lower panels of Table 3, respectively. The observed bias and RMSE displayed similar patterns across the higher and lower item quality conditions. Although the higher quality items resulted in smaller bias and RMSE, we discussed only the results based on lower quality items as the differences produced by the estimation approaches are magnified under this condition. For the G-DINA model, in comparison to use of explicit Q-matrix, using an implicit Q-matrix resulted in elevated bias and RMSE. In other words, the use of explicit Q-matrix better recovered these lowest and highest success probability parameters. Recoveries of these parameters were even better when the estimation involved a structured prior distribution. Therefore, under the G-DINA model, model parameter estimates were best when an explicit Q-matrix was used along with a hierarchically structured prior distribution.

The bias and RMSE results for the DINA and DINO models are also given in Table 3. Although all four estimation approaches had good recovery of the item parameters; the best results (i.e., the lowest bias and RMSE) were obtained when a structured prior distribution was employed in the estimation. When the prior was structured, the Q-matrix type had no impact on attribute classification (i.e., results were almost identical under all conditions). However, when prior distribution was unstructured, use of implicit Q-matrix produced the lowest bias and RMSE. The bias and RMSE levels were quite similar to the ones obtained with structured prior distribution. Slightly elevated bias and RMSE were observed when the explicit Q-matrix was used and all latent classes are considered in the estimation. These elevated bias and RMSE were observed especially on the guessing parameter when the DINA model was fitted; and on the slip parameter when the DINO was fitted.

In comparison to the parameter recovery of the general model, we observed that all four estimation approaches resulted in smaller bias and RMSE for the reduced models. The difference in the performance of the estimation approaches may be attributed to the different model complexities of the CDMs (e.g., the GDINA model complexity is increased by the number of attributes required for the item). Moreover, the impact of implicit and explicit Q-matrices were different for the general and reduced models. In the general model, employment of implicit Q-matrix resulted much better item parameter recovery; whereas improved parameter estimation observed with explicit Q-matrix in the estimation of reduced models.

### *Results on Attribute Estimation*

The CAC and CVC rates of the models by the estimation approaches are documented in Table 4. Results under lower and higher item quality conditions displayed similar patterns; however, we only discussed the higher item quality conditions to emphasize the importance of Q-matrix type even when items in a test are of very good quality. The overall pattern indicates that higher CAC and CVC rates were obtained with the explicit Q-matrix use when the generating and fitted model was the G-DINA model. Under the G-DINA model, the correct classification rates were the highest when explicit Q-matrix was used along with a structured prior distribution; and the lowest attribute and vector classification rates were obtained when the implicit Q-matrix was used along with an unstructured prior distribution. For example, for the G-DINA model, under the linear hierarchy, CVC rates of the explicit and implicit Q-matrices were .841 and .110, respectively, when an unstructured prior distribution was used. These rates went up to .865 and .588 when a structured prior distribution was employed in the estimation.

The DINA and DINO model attribute classification results showed that under the both models the highest CAC and CVC rates were observed when impermissible attribute patterns were excluded from the estimation via a structured prior distribution. Under the structured prior conditions, correct attribute and vector classification rates of the two types of Q-matrices were almost identical. In contrast, when an unstructured prior distribution was employed, explicit Q-matrix produced higher CAC and CVC rates when generating and the fitted model was the DINA. For instance, when all attribute patterns were permissible, the observed differences in the CVC rates in using explicit and implicit Q-matrix were 1.2%, 3.7%, and 2.5% for the linear, convergent, and divergent cases, respectively. For the unstructured prior distribution conditions, in contrast to the DINA model cases, implicit Q-matrix yielded higher CAC and CVC rates when the generating and fitted model

*Table 3*
Item parameter bias and RMSE under lower item quality conditions

| | | | Guessing parameter | | | | Slip parameter | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Explicit Q | | Implicit Q | | Explicit Q | | Implicit Q | |
| Statistic | Model | Hierarchy | Unstr. | Struc. | Unstr. | Struc. | Unstr. | Struc. | Unstr. | Struc. |
| bias | G-DINA | Linear | .030 | .025 | .174 | .061 | .032 | .025 | .161 | .055 |
| | | Convergent | .032 | .027 | .173 | .073 | .033 | .027 | .159 | .060 |
| | | Divergent | .049 | .042 | .148 | .074 | .034 | .028 | .082 | .043 |
| | DINA | Linear | .030 | .019 | .019 | .019 | .023 | .023 | .023 | .023 |
| | | Convergent | .028 | .020 | .021 | .020 | .024 | .024 | .024 | .024 |
| | | Divergent | .033 | .027 | .028 | .027 | .025 | .024 | .024 | .024 |
| | DINO | Linear | .022 | .022 | .024 | .022 | .029 | .021 | .024 | .021 |
| | | Convergent | .025 | .024 | .026 | .024 | .029 | .021 | .027 | .021 |
| | | Divergent | .038 | .036 | .040 | .037 | .026 | .020 | .022 | .020 |
| RMSE | G-DINA | Linear | .040 | .031 | .186 | .090 | .041 | .032 | .172 | .071 |
| | | Convergent | .041 | .034 | .187 | .103 | .043 | .034 | .171 | .077 |
| | | Divergent | .063 | .052 | .171 | .098 | .043 | .037 | .105 | .057 |
| | DINA | Linear | .036 | .024 | .024 | .024 | .029 | .029 | .029 | .029 |
| | | Convergent | .035 | .026 | .026 | .025 | .030 | .029 | .029 | .029 |
| | | Divergent | .040 | .034 | .034 | .033 | .031 | .030 | .030 | .030 |
| | DINO | Linear | .028 | .027 | .031 | .027 | .036 | .027 | .030 | .027 |
| | | Convergent | .030 | .030 | .032 | .030 | .035 | .026 | .033 | .026 |
| | | Divergent | .047 | .045 | .049 | .045 | .032 | .026 | .027 | .025 |

Note: Unstr. = Unstructured prior distribution; Struc. = Structured prior distribution

*Table 4*
Correct attribute and vector classification rates under higher item quality conditions

| | | | | Correct attribute classification rates | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDM** | **Hierarchy** | **Q** | **Prior** | **A1** | **A2** | **A3** | **A4** | **A5** | **A6** | **CVC** |
| G-DINA | Linear | Explicit | Unstr. | .983 | .971 | .968 | .967 | .959 | .974 | .841 |
| | | | Struc. | .986 | .976 | .974 | .972 | .966 | .978 | .865 |
| | | Implicit | Unstr. | .965 | .887 | .773 | .695 | .698 | .600 | .110 |
| | | | Struc. | .983 | .945 | .899 | .861 | .883 | .958 | .588 |
| | Convergent | Explicit | Unstr. | .983 | .969 | .948 | .954 | .966 | .976 | .824 |
| | | | Struc. | .986 | .974 | .951 | .957 | .972 | .980 | .841 |
| | | Implicit | Unstr. | .953 | .882 | .851 | .832 | .681 | .605 | .219 |
| | | | Struc. | .985 | .968 | .924 | .931 | .925 | .960 | .747 |
| | Divergent | Explicit | Unstr. | .987 | .940 | .927 | .958 | .922 | .950 | .732 |
| | | | Struc. | .993 | .942 | .933 | .966 | .925 | .951 | .750 |
| | | Implicit | Unstr. | .975 | .933 | .846 | .951 | .890 | .910 | .614 |
| | | | Struc. | .992 | .940 | .927 | .963 | .921 | .946 | .738 |
| DINA | Linear | Explicit | Unstr. | .966 | .960 | .975 | .974 | .963 | .995 | .856 |
| | | | Struc. | .967 | .965 | .984 | .984 | .984 | .997 | .894 |
| | | Implicit | Unstr. | .967 | .954 | .973 | .965 | .963 | .993 | .844 |
| | | | Struc. | .967 | .965 | .984 | .984 | .984 | .997 | .894 |
| | Convergent | Explicit | Unstr. | .966 | .957 | .978 | .943 | .973 | .996 | .843 |
| | | | Struc. | .966 | .961 | .982 | .950 | .987 | .998 | .868 |
| | | Implicit | Unstr. | .966 | .950 | .973 | .922 | .964 | .994 | .806 |
| | | | Struc. | .967 | .962 | .982 | .950 | .987 | .998 | .869 |
| | Divergent | Explicit | Unstr. | .981 | .941 | .981 | .947 | .948 | .976 | .816 |
| | | | Struc. | .983 | .943 | .984 | .951 | .953 | .977 | .829 |
| | | Implicit | Unstr. | .983 | .938 | .977 | .945 | .931 | .970 | .791 |
| | | | Struc. | .983 | .943 | .984 | .951 | .953 | .977 | .829 |
| DINO | Linear | Explicit | Unstr. | .996 | .965 | .972 | .975 | .957 | .965 | .854 |
| | | | Struc. | .997 | .986 | .985 | .983 | .962 | .966 | .893 |
| | | Implicit | Unstr. | .997 | .985 | .984 | .982 | .959 | .963 | .884 |
| | | | Struc. | .997 | .986 | .985 | .983 | .962 | .966 | .892 |
| | Convergent | Explicit | Unstr. | .996 | .974 | .942 | .977 | .959 | .966 | .844 |
| | | | Struc. | .998 | .986 | .950 | .980 | .963 | .967 | .868 |
| | | Implicit | Unstr. | .998 | .985 | .941 | .979 | .962 | .964 | .854 |
| | | | Struc. | .998 | .986 | .950 | .980 | .963 | .967 | .868 |
| | Divergent | Explicit | Unstr. | .996 | .912 | .921 | .981 | .907 | .937 | .718 |
| | | | Struc. | .998 | .931 | .921 | .986 | .908 | .937 | .738 |
| | | Implicit | Unstr. | .999 | .930 | .916 | .986 | .906 | .937 | .730 |
| | | | Struc. | .999 | .931 | .921 | .986 | .908 | .937 | .737 |

Note: Q = Q-matrix; Prior = Prior distribution; A1-A6 = Measured attributes; CVC = Correct vector classification rate; Unstr. = Unstructured prior distribution; Struc. = Structured prior distribution

was the DINO. For instance, when unstructured prior distribution was involved; the observed differences in the CVC rates of implicit and explicit Q-matrix use are 3.0%, 1.0%, and 1.2% for the linear, convergent, and divergent cases, respectively.

In general, the simulation results indicated that impact of Q-matrix type on the estimation of the G-DINA model was much higher in comparison to its impact on the reduced models. For the G-DINA model estimation, the impact of Q-matrix type was higher than impact of prior distribution type. In contrast, type of prior distribution was more influential on item and person parameter estimations for the reduced models. When prior distribution was structured, the Q-matrix type had almost no influence on item and person parameter estimation.

*Real Data Analysis*

We also conducted a numerical analysis to understand how the estimation approaches perform when real data involved. The analyzed data consisted of 2,922 examinees' binary responses to 28 items in the grammar section of the Examination for the Certificate of Proficiency in English (ECPE), which was developed and administered by the University of Michigan English Language Institute in 2003. The dataset and the explicit Q-matrix are available in the 'CDM' package (Robitzsch, Kiefer, George, & Uenlue, 2019) in R software environment. The dataset have been analyzed in several studies (e.g., Chiu et al., 2009; Henson & Templin, 2007; Templin & Bradshaw, 2014). Templin and Bradshaw (2014) reported a linear hierarchy among these three attributes measured by the ECPE grammar test. Based on this described linear hierarchy, Lexical rules (A1) is a prerequisite attribute to Cohesive rules (A2), which in turn is a prerequisite attribute to Morphosyntactic rules (A3).

The G-DINA model was used in the analysis for examinee classification. Model fit results obtained via all four estimation approaches are presented in Table 5. When we compare the model fit statistics resulted from employment of explicit Q-matrix, AIC, BIC, and the likelihood ratio test (LRT: $\chi^2 = 22.69$, $df = 4$, $p - value = .0001$) indicated that the G-DINA model allowing all latent classes better fits to the data. Likewise, under the implicit Q-matrix, the G-DINA model better fitted to the data when all latent classes were permissible (LRT: $\chi^2 = 115.92$, $df = 4$, $p - value = .0001$). Therefore, regardless of the Q-matrix type, model selection results show that a model constrained by the linear hierarchy does not fit to the data as well as the unconstrained model.

When the G-DINA model let all latent classes in the estimation, implicit and explicit Q-matrix classifications were compared to determine the attribute classification agreements between the two approaches. Individual attribute level and attribute vector level agreements were computed to be .966, .466, .836, and .356 for the three attributes and attribute vector, respectively. When an unstructured prior distribution is employed, as shown in the simulation study earlier, when the G-DINA model is fitted the explicit Q-matrix provides higher correct classification rates than its implicit counterpart does. The largest disagreement is on the second attribute, which may also be due to the fact that implicit Q-matrix follow a linear hierarchy that might not be accurate. Furthermore, in their study, Tu et al. demonstrated that when attributes are hierarchical, estimation approach excluding impermissible attributes results in good data-model fit when fitting model is the G-DINA (2018). However, in our analysis of the ECPE data, fit of the G-DINA model with unstructured prior distribution yielded better data-model fit (i.e., lower AIC and BIC). In the light of this information, we can argue that either the Q-matrix or the specified hierarchical structure might be inaccurate so that it requires further investigation.

*Table 5*

Model fits

| | **Unstructured prior distribution** | **Structured prior distribution** |
|---|---|---|
| Explicit Q-matrix | AIC = 85642.40 | AIC = 85665.09 |
| | BIC = **86126.79** | BIC = 86149.47 |
| | -2LL = 85480.40 | -2LL = 85503.09 |
| Implicit Q-matrix | AIC = **85407.82** | AIC = 85523.74 |
| | BIC = 86310.81 | BIC = 86426.72 |
| | -2LL = 85105.82 | -2LL = 85221.74 |

Note: AIC = Akaika information criterion; BIC = Bayesian information criterion; and -2LL = Deviance (-2 times loglikelihood)

## Discussion

CDMs are useful tools that provide fine-grained information on examinees' strengths and weaknesses, which in turn can be used to inform classroom instructions and learning. Although CDAs primarily serve for formative assessments in low-stakes contexts, we cannot discount their potential use in high-stakes testing. In such cases, the use of an estimation approach that results in more accurate item and person parameter estimates, even if the improvement is slight, might be vital. When attributes follow a hierarchical structure, more accurate item parameter estimation and examinee classification may be achieved by structuring either the Q-matrix or prior distribution in the model estimation procedure. Then, potential contribution of the distinct estimation approaches need to be considered when attributes follow a hierarchical structure.

This study was designed to understand the impact of an implicit Q-matrix on item parameter estimation and examinee classification rate when attributes follow a hierarchy. Study results indicated that employment of explicit rather than implicit Q-matrix in the G-DINA model estimation yielded significant increase in item parameter estimation accuracy and correct attribute classification rate. Results also indicated that, for the DINA and DINO models, both the implicit and explicit versions of the Q-matrix yielded almost identical item parameter estimates and examinee classifications when the prior distribution was already structured. However, under unstructured attribute conditions, some differences on item and person parameters emerged due to Q-matrix type. When the prior distribution is not structured, using an explicit Q-matrix improves the classification accuracy of the DINA model; while using an implicit Q-matrix increases the correct classification rates of the DINO model.

When the prior distribution or Q-matrix is structured, hierarchy is assumed to be known. However, in practice, hierarchical structure among the attributes may not always be well established. Thus, incorrect specification of the hierarchical attribute structure may adversely impact model estimations. In such cases, item parameter estimates and examinee classifications may be adversely affected by structured prior distribution or Q-matrix. Therefore, correctly identifying hierarchical structure is of vital importance. To this end, development of statistical methods to validate expert-based hierarchical structures can be a potential future research direction. Also, it would be interesting to see the impact of unbalanced, misspecified, and incomplete (i.e., not all possible single-attribute items included) Q-matrices on model estimation approaches.

## References


Chipman, S. F., Nichols, P. D., & Brennan, R. L. (1995). Introduction. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 1-18). Hillsdale, NJ: Erlbaum.

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

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement, 33,* 163-183. https://doi.org/10.1177/0146621608320523

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

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

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

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement, 47,* 227-249. https://doi.org/10.1111/j.1745-3984.2010.00110.x

de la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement, 47,* 115-127. https://doi.org/10.1111/j.1745-3984.2009.00102.x

de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicologia Educative, 20,* 89-97. https://doi.org/10.1016/j.pse.2014.11.001

DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement, 44,* 285-291. https://doi.org/10.1111/j.1745-3984.2007.00039.x

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. New York, NY: Springer-Verlag.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179-197. http://dx.doi.org/10.1037/0033-2909.93.1.179

Embretson, S .E. (Ed.) (1985). *Test design: Developments in psychology and psychometrics*. Orlando, FL: Academic Press.

Embretson, S. E. (1994). Applications of cognitive design systems to test development. In C. R. Reynolds (Ed.), *Cognitive assessment: A multidisciplinary perspective* (pp. 107-135). New York: Plenum Press.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3,* 380-396. http://dx.doi.org/10.1037/1082-989X.3.3.380

Henson, R. A., & Templin, J. (2007, April). *Large-scale language assessment using cognitive diagnosis models*. Paper presented the meeting of the National Council on Measurement in Education, Chicago, IL.

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. https://doi.org/10.1177/01466210122032064

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement, 41,* 205-237. https://doi.org/10.1111/j.1745-3984.2004.tb01163.x

Ma, W., & de la Torre, J. (2018). GDINA: The Generalized DINA Model Framework [Computer software]. Retrieved from https://CRAN.R-project.org/package= GDINA

Maris, E. (1995). Psychometric latent response models. *Psychometrika, 60,* 523-547. https://doi.org/10.1007/BF02294327

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64,* 187-212. https://doi.org/10.1007/BF02294535

No Child Left Behind Act of 2001, Pub. L. No. 1-7-110 (2001).

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2019). CDM: Cognitive diagnosis modeling [Computer software]. Retrieved from https://CRAN.Rproject.org/package=CDM

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement, 6,* 219-262. https://doi.org/10.1080/15366360802490866

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20,* 345-354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Team, R. C. (2015). R: A Language and Environment for Statistical Computing [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Templin, J. L., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika, 79,* 317-339. https://doi.org/10.1007/s11336-013-9362-0

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

Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H-H (2018). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q-matrix design. *Applied Psychological Measurement*. Advance online publication. doi:10.1177/0146621618765721