Cognitive Diagnostic Models With Attribute Hierarchies: Model Estimation With a Restricted Q-Matrix Design

Applied Psychological Measurement
2019, Vol. 43(4) 255–271
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0146621618765721
journals.sagepub.com/home/apm

Dongbo Tu¹, Shiyu Wang², Yan Cai¹, Jeff Douglas³ and Hua-Hua Chang³

Abstract

Attribute hierarchy is a common assumption in the educational context, where the mastery of one attribute is assumed to be a prerequisite to the mastery of another one. The attribute hierarchy can be incorporated through a restricted Q matrix that implies the specified structure. The latent class-based cognitive diagnostic models (CDMs) usually do not assume a hierarchical structure among attributes, which means all profiles of attributes are possible in a population of interest. This study investigates different estimation methods to the classification accuracy for a family of CDMs when they are combined with a restricted Q-matrix design. A simulation study is used to explain the misclassification caused by an unrestricted estimation procedure. The advantages of the restricted estimation procedure utilizing attribute hierarchies for increased classification accuracy are also further illustrated through a real data analysis on a syllogistic reasoning diagnostic assessment. This research can provide guidelines for educational and psychological researchers and practitioners when they use CDMs to analyze the data with a restricted Q-matrix design and make them be aware of the potentially contaminated classification results if ignoring attribute hierarchies.

Keywords

attribute hierarchies, classification accuracy, cognitive diagnostic models, restricted Q matrix

Cognitive diagnostic models (CDMs), also known as diagnostic classification models (DCMs, Rupp, Templin, & Henson, 2010), are psychometric models that allow for the profiling of subjects according to a variety of latent characteristics. The main applications of CDMs have been in the educational context with the aim to provide students with information concerning whether or not they have mastered each of a group of specific skills, that are often generically referred to as attributes. In many cases, such as mathematics learning (Leighton & Gierl, 2007), critical reading (C. Wang & Gierl, 2011), and syllogistic reasoning (Leighton, Gierl, & Hunka, 2004), attributes may be hierarchical, meaning that the mastery of one attribute is prerequisite to the

¹JiangXi Normal University, Nanchang, China

²University of Georgia, Athens, GA, USA

³University of Illinois at Urbana–Champaign, USA

Corresponding Author:

Shiyu Wang, Department of Educational Psychology, University of Georgia, 325N Aderhold Hall, Athens, GA 30602, USA.

Email: swang44@uga.edu

mastery of another one. The rule space model (RSM; K. K. Tatsuoka, 1983, 1985, 1995, 2009) and the attribute hierarchy method (AHM; Leighton et al., 2004) are models which can naturally deal with attribute hierarchies during the classification process. The RSM does not require modeling the hierarchical relations or dependencies, but can incorporate those structures into an adjacency matrix. The AHM is a variation of RSM which assumes a hierarchical ordering of attributes is fundamental to predicting and classifying examinees.

Different from RSM and AHM in which the estimation of attribute profiles mainly depends on a continuous item response model, a series of CDMs that characterize the direct relationship of observed responses to a set of latent categorical attributes have been developed. Most of CDMs use binary attributes, considering examinees to either possess or not possess each attribute. These latent class-based CDMs usually do not assume a hierarchical structure among attributes, meaning all profiles of attributes are possible in a population of interest. Only recently has research considered to model attribute hierarchies and investigate the impact of attribute hierarchies on model fit, and item parameter estimation in latent class-based CDMs. de la Torre, Hong, and Deng (2010) investigated several factors related to the deterministic input noisy “and” gate (DINA) model (Haertel, 1989; Junker & Sijtsma, 2001), item parameters, and classification accuracy under a divergent hierarchical structure. Su, Choi, Lee, Choi, and McAninch (2013) proposed two models which incorporate attribute hierarchies into DINA model and the deterministic input noisy “or” gate (DINO) model (Templin & Henson, 2006). Templin and Bradshaw (2014) introduced a family of hierarchical diagnostic classification models (HDCM) which adapts the log-linear cognitive diagnosis model (LCDM; Henson, Templin, & Willse, 2009) to cases where attribute hierarchies are present, and suggested a generalized likelihood-ratio test to investigate the presence of an attribute hierarchy. One common factor from previous research studies is that they all use an unstructured item-by-attribute Q matrix (K. K. Tatsuoka, 1985) which assumes items do not necessarily reflect the specified attribute hierarchy. Using the unstructured Q matrix, de la Torre et al. (2010) found that the fully Bayesian approach is most accurate when the prior distribution matches the latent class structure for the DINA model. The simulation results from Templin and Bradshaw (2014, Table 2) seem to indicate that the LCDM without restricting the latent class structure can produce the comparable classification results with HDCM that restricts the latent class structure and the corresponding kernel in the item response function.

Different from the unstructured Q-matrix design, Leighton et al. (2004) proposed a restricted Q matrix that reflects the attribute hierarchy and in which, every item is labeled to measure all of its prerequisites. The restricted Q matrix has been implemented in several applications of cognitive diagnostic assessment, and it is interpreted as the *cognitive blueprint* or *cognitive test specification that can be used to develop items that measure the specified attribute hierarchy* (e.g., Gierl, 2007, 2008; C. Wang & Gierl, 2011). If a diagnostic assessment is developed based on a restricted Q matrix, then a natural question to ask is what the appropriate classification procedure is when a latent class-based CDM is used. The motivation to investigate this question is due to the classification results from a cognitive diagnostic assessment that was developed for syllogistic reasoning. Following the hierarchical structure of seven attributes suggested by the Figure 4 in Leighton et al. (2004), 15 questions were developed to reflect the specified attribute hierarchy. The corresponding Q matrix is a restricted structure which only contains items that satisfy the implied attribute hierarchical structure. Both LCDM and HDCM were applied to analyze students' responses, and different from the simulation results of Templin and Bradshaw (2014), these two methods produced very different classification results with attribute profile pattern-wise classification agreement as 6.9%. Especially, the LCDM produced some unreasonable estimates that indicate that the students can master certain skill without mastering its prerequisite. Thus, it is necessary to investigate the factors that affect the classification results when

using a latent class-based CDM with a restricted Q matrix structure. The analysis based on syllogistic reasoning diagnostic assessment is presented in the “Real Data Analysis” section.

The objective of this research is to study the advantages of utilizing attribute hierarchies for increased classification accuracy under a restricted Q-matrix design. The main goal is to investigate different estimation methods to the classification accuracy within the restricted latent class modeling framework. Using several common CDMs as the representatives, misclassification caused by an unrestricted estimation method was discussed within the maximum likelihood estimation (MLE) procedure and the Bayesian estimation procedure. Simulation results also reveal the possible deterioration of the item parameter estimates for the general models due to the unrestricted estimation method. Instead of focusing on one attribute hierarchy structure, four different structures are investigated, with the purpose to generalize the results to different situations. This research can be treated as an empirical investigation for the model identifiability for the restricted latent class model under a restricted Q-matrix design. These results can also provide evidences for theoretical work regarding model identifiability issue in the future.

The rest of the study is organized as follows: Several representatives of the restricted latent class models were briefly introduced and then a set of popular attribute hierarchical structures was reviewed. Next, the advantages of utilizing attribute hierarchies to attribute profile estimation are discussed. This followed a simulation study that further investigates the impacts of several attribute hierarchies to item parameter estimation, model fit, and classification accuracy. A real data analysis is conducted after that to evaluate different attribute profile estimation procedures. Finally, the results and comment on their implications were discussed.

CDMs and Attribute Hierarchies

CDMs

Based on the assumptions about how attributes influence test performance, CDMs can be categorized as noncompensatory or compensatory models. Conjunctive models, the representatives of the noncompensatory models, express the notion that all attributes specified in the Q matrix for an item should be required to answer the item correctly, but allow for slips and guesses in ways that distinguish the models from one another. Unlike noncompensatory models, compensatory models allow one to compensate for what is lacked in some measured skills by having mastered other skills. Encompassing the traditional categories of the reduced CDMs, several general models based on different link functions have been developed to include many of the common compensatory and noncompensatory models. The several representatives of these different types of CDMs were reviewed first.

Consider there are K attributes and J items. Suppose the attribute profile for an examinee is denoted as $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)^T$. Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_J)^T$ be a random response vector of this examinee. The $J \times K$ Q matrix links K attributes with those J items and entry q_{jk} indicates whether item j requires attribute k . Let $\eta(\alpha) = (\eta_1(\alpha), \dots, \eta_J(\alpha))^T$ denotes the corresponding ideal response pattern arising from this attribute profile, where $\eta_j(\alpha) = \prod_{k=1}^K \alpha_k^{q_{jk}}$, describing whether this examinee has mastered all the required attributes for item j .

The DINA model is the simplest conjunctive model with only two parameters for each item. It allows the examinee's response to deviate from the ideal response profile according to slipping parameters, $s_j = P(Y_j = 0 | \eta_j(\alpha) = 1)$, and guessing parameters, $g_j = P(Y_j = 1 | \eta_j(\alpha) = 0)$ for each item. The item response function of the DINA model is

$$P(Y_j = 1 | \alpha) = (1 - s_j)^{\eta_j(\alpha)} g_j^{1 - \eta_j(\alpha)}. \quad (1)$$

The reduced reparameterized unified model (rRUM; Hartz, 2002) is another example of conjunctive model which relaxes the assumptions from DINA and allows slipping and guessing probabilities to vary across items. In this model, the item response function is

$$P(Y_j = 1 | \alpha) = \pi_j^* \prod_{k=1}^K r_{jk}^{*q_{jk}(1-\alpha_k)}, \quad (2)$$

Here, the baseline parameter π_j^* is the probability of answering item j correctly given the examinee possesses all of the required attributes. The parameter r_{jk}^* can be viewed as a penalty parameter for not possessing attribute k for item j , and it is between 0 and 1.

Several general models are proposed based on different link functions and parameterization of kernels, such as the general diagnostic model (GDM; von Davier, 2008), LCDM, and generalized DINA (GDINA; de la Torre, 2011). Here, the GDINA model was introduced as the example. The item response function of the model is

$$P(Y_j = 1 | \alpha_j^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_k + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_k \alpha_{k'} + \cdots + \delta_{j12 \dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_k, \quad (3)$$

where $\alpha_j^* = (\alpha_1, \dots, \alpha_k, \dots, \alpha_{K_j^*})$ is the reduced attribute vector whose elements are the required attributes for item j , K_j^* is the number of attributes measured by item j . $P(Y_j = 1 | \alpha_j^*)$ is the endorsement probability of individuals with the latent attribute profile α_j^* . δ_{j0} is the intercept for item j , δ_{jk} is the main effect due to having attribute k , $\delta_{jkk'}$ is the interaction effect due to having attributes k and k' , $\delta_{j12 \dots K_j^*}$ is the interaction effect due to having attributes 1 up to K_j^* . If there are no interaction effects among attributes, the GDINA model is reduced to the additive CDM (ACDM; de la Torre, 2011), which is an example of the compensatory CDM. Its item response function is

$$P(Y_j = 1 | \alpha_j^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_k. \quad (4)$$

Attribute Hierarchies

The attribute hierarchies refer to situations in which the mastery of an attribute is prerequisite to the mastery of another one. Such an assumption is common in the educational context, because the process of teaching generally proceeds sequentially, with each step building upon the last. For example, mathematics encompasses a wide variety of skills and concepts. These skills and concepts are related and often build on one another (Sternberg & Ben-Zeev, 1996). Three common types of attribute hierarchies are linear, divergent, and convergent (Figure 1), which are introduced and discussed by Leighton et al. (2004) and Rupp et al. (2010). The linear attribute hierarchy requires all attributes to be ordered sequentially, and implies that if attribute A1 is not present, then all following attributes will not be present. The convergent structure represents a hierarchy with a convergence branch where two different paths maybe traced from Attribute A1 to Attribute A6. Note that in this structure, one attribute can be a prerequisite of multiple different attributes and an attribute could have many different precursors. The divergent attribute hierarchy refers to different distinct tracks originating from the same single attribute. In the traditional latent class-based CDMs, attributes are assumed independent which can be represented by the fourth structure in Figure 1. Note that independence in this sense is not the same as statistical independence. Though no attribute is a prerequisite for another, the indicators of attribute mastery may be correlated. Another type of attribute hierarchy is a mixed structure where one

Figure 1. Four types of attribute hierarchies and an independent structure.

set of attributes have a structure, the rest attributes have another one, and the attributes between the two sets have an independent structure. The last structure in Figure 1 represents an example, where A1, A2, and A3 are linearly related, A4, A5, and A6 have a divergent structure, and these two sets of attributes are independent.

Classification in Cognitive Diagnostic Modeling With Attribute Hierarchies

In this section, different attribute profile estimators with attribute hierarchies are discussed. A permissible attribute profile was defined as the one which can exist under the specified attribute hierarchy from now on. Correspondingly, an attribute profile that cannot exist under the attribute hierarchy is called an impermissible attribute profile. A set that contains all the permissible/impermissible attribute profiles under a given hierarchical structure is called the permissible/impermissible set, and is denoted as PAS/IPAS. For example, suppose there are two attributes, and Attribute 1 is the prerequisite for Attribute 2, and with this linear relationship, $\alpha^1 = (0, 1)^T$ is an impermissible attribute profile, and $\alpha^2 = (0, 0)^T$, $\alpha^3 = (1, 0)^T$, $\alpha^4 = (1, 1)^T$ are permissible attribute profiles. In this case, $\text{IPAS} = \{(0, 1)^T\}$, $\text{PAS} = \{(1, 0)^T, (1, 1)^T, (0, 0)^T\}$.

The Construction of a Restricted Q Matrix With Attribute Hierarchies

As discussed in “Introduction” section, the authors were interested in a restricted Q-matrix design that can guarantee all the items in the test reflect the corresponding attribute hierarchy.

This is usually the case when a test is designed for a cognitive diagnostic purpose and the Q matrix is developed by content experts before item writing. The restricted Q matrix can be constructed from a reachability matrix (K. K. Tatsuoaka, 1983), which is a $K \times K$ matrix that represents the direct and indirect relationships among attributes, and is referred as R matrix from now on. In this setting, the R matrix is always a lower triangular matrix with all diagonal elements as 1s. The j th element of the i th row in the matrix represents whether attribute j is a direct or indirect prerequisite for attribute i . Following the previous example of the linear relationship of two attributes, the R matrix can be constructed as $\begin{pmatrix} 1, 0 \\ 1, 1 \end{pmatrix}$. Specifically, Q_H was defined as a $J \times K$ matrix that associates the K attributes to J items and reflects the assumed attribute hierarchy satisfying the following conditions: (a) an R matrix is a submatrix of the Q_H matrix and (b) all the transposes of the item-attribute Q vectors in it belong to the permissible attribute set. When constructing a cognitive diagnostic test involving attribute hierarchies, the first step is to construct a Q matrix that is of the form Q_H . From now on, the analysis of the classification results for different attribute profile estimators are based on the prerequisite that the Q matrix of the exam is of a form Q_H .

Attribute Profile Estimation Procedure

The authors, first, started with a simple scenario where the item parameters are pre-calibrated, which are assumed to be known. In this case, the MLE can be used to estimate examinees' attribute profiles. Because of the binary response data, with the assumptions of conditional independence of responses to different items, the likelihood function of α constructed from the chosen model based on J questions is:

$$L(\alpha) = \prod_{j=1}^J P(Y_j = 1 | \alpha)^{Y_j} [1 - P(Y_j = 1 | \alpha)]^{1-Y_j}, \quad (5)$$

where $P(Y_j = 1 | \alpha)$ is the item response function of any chosen CDM, with specific examples given in Equations 1 to 4. When attributes are assumed to be independent, all possible attribute profiles can be candidates for an examinee's attribute profile. The estimator which maximizes (5) over all possible attribute profiles was referred to as the *unrestricted MLE*. However, when a hierarchical structure is assumed for attributes, only the elements in the permissible attribute set are reasonable estimates. In this sense, a *restricted MLE* can be presented as

$$\hat{\alpha}_H = \operatorname{argmax}_{\{\alpha \in \text{PAS}\}} L(\alpha),$$

where $L(\alpha)$ is the likelihood function of the form of Equation 5. The unrestricted and restricted MLE only differ in the attribute profile space; however, this difference will result in different classification results when attributes have a hierarchical structure. A simple example was provided to compare these two estimation procedures.

Example. Considering the example of using the DINA model for classification. Suppose there are $K = 3$ attributes and $J = 10$ questions. The attributes are assumed to have a linear

relationship described by Figure 1. The corresponding R matrix is $\begin{pmatrix} 1, 0, 0 \\ 1, 1, 0 \\ 1, 1, 1 \end{pmatrix}$. The Q_H

matrix for this test is constructed by replicating this R matrix 3 times and adding the row vector (1, 1, 0) as the last row. Suppose the true attribute profile for an examinee is

$\alpha^0 = (1, 0, 0)^T$. Let us consider an impermissible attribute profile $\alpha^1 = (1, 0, 1)^T$. It is easy to verify that the ideal response profiles determined by these two attribute profiles are exactly the same, that is $\eta_j(\alpha^0) = \eta_j(\alpha^1), j = 1, 2, \dots, 10$. If the DINA model is the true data-generating model, because α^0 and α^1 have the same ideal response profile, then the two patterns will result in the same item response function. This will make these two attribute profiles have the exactly same likelihoods, that is $L(\alpha^0) = L(\alpha^1)$. In this case, if α^1 is not excluded from the attribute profile estimation procedure, then it will have the same likelihood as α^0 . Such a result is usually called the nonidentifiability of attribute profiles (Zhang, DeCarlo, & Ying, 2013).

From the example, it can be noted that, if attribute hierarchies exist, two attribute profiles having the same ideal response profile will have a high chance to result in the same or similar likelihood values. In fact, the phenomenon that two different attribute profiles result in the same ideal response profile is usually interpreted as these two attribute profiles cannot be separated, and they are in an equivalence class (Chiu, Douglas, & Li, 2009; K. K. Tatsuoaka, 1995, 2009; S. Wang, 2017; S. Wang & Douglas, 2015; Zhang, 2014). If the DINA model is used, the attributes in the same equivalence class cannot be separated from each other because they all have the same likelihood. For the other more general models, such as rRUM, ACDM, and GDINA, attribute profiles in the same equivalence class will have very similar likelihood values, as the item response function is not completely determined only by the ideal response profile, but also by the specific form of the attribute profile. In this case, the unrestricted MLE cannot distinguish one from another and could lead to estimates in the impermissible attribute profiles.

The Bayesian estimation approach is an alternative way to estimate attribute profile. When such a method is used, the classification is conducted based on the posterior distribution $p(\alpha|Y) \propto L(Y|\alpha)p(\alpha)$, where $p(\alpha)$ denotes the prior distribution for the population attribute profiles and $L(Y|\alpha)$ is the likelihood function constructed from response data. In this case, for two attribute profiles in the same equivalence class, say α^1 and α^2 , $\frac{p(\alpha^1|Y)}{p(\alpha^2|Y)} = \frac{L(Y|\alpha^1)p(\alpha^1)}{L(Y|\alpha^2)p(\alpha^2)}$. The prior distribution for the attribute profile is usually assumed from a uniform discrete distribution over all possible 2^K attribute profiles (e.g., Huebner & Wang, 2011) if attribute hierarchies do not exist. In this case, the posterior probability depends on the likelihood function according to $\frac{p(\alpha^1|Y)}{p(\alpha^2|Y)} = \frac{L(Y|\alpha^1)}{L(Y|\alpha^2)}$. Then the attribute profile within an equivalence class will have exactly the same or similar posterior probability and cannot be separated from each other. Such indistinguishable Bayesian classifiers have been discussed by Groß and George (2014), in which they use the DINA model as an example. If the prior distribution does not assume each attribute profile has an equal probability over all possible values, then the posterior will be a reflection of the prior. Within an equivalence class, the posteriors are proportional to priors. The classification based only on the prior may be invalid because no additional information from the data is used. To avoid such an invalid classification, the prior for each impermissible attribute profiles should be set to zero.

When the item parameters and attribute profiles are both unknown, the joint MLE, the marginal MLE (MMLE; e.g., de la Torre, 2009), or Bayesian estimation methods (Hartz, 2002) can be used to estimate them together. Those estimation methods are usually implemented either by the expectation maximization (EM) algorithm or Markov chain Monte Carlo (MCMC) methods based on different model structures. No matter what algorithm is used, the item parameters and latent attribute profile are estimated through an iterative process, and the likelihood function always acts as an important role in these two types of parameter estimations. In this case, it is still important to restrict the latent attribute profile space in the specific hierarchical structure.

This will help avoid the inseparable attribute profiles situation during the latent profile estimation step and also improve the item parameter estimation. The following simulation study was used to support the argument about the item parameter estimation and attribute profile classification in this scenario.

Simulation Study

In this section, a simulation study is conducted to investigate the advantage of incorporating attribute hierarchies to the model fit, item parameter estimation, and classification accuracy when both item parameters and attribute profiles are unknown. Four models, the DINA, the rRUM, the ACDM, and the GDINA are used to generate response data. Two MMLE estimation procedures are evaluated: one considers only the permissible attribute patterns, and the other considers 2^K attribute profiles. To simplify the argument, the two estimation procedures were referred as restricted estimation and unrestricted estimation, respectively. The CDM R Package (George, Robitzsch Kiefer, Groß, & Ünlü, 2016) is used to estimate parameters and this package implemented the MMLE through the EM algorithm. The sample size was 1,000 examinees for each of 100 replications. Item parameters for the four models were simulated based on $1 - P(X_{ij} = 1 | \alpha_{ij}^* = \mathbf{1}) = P(X_{ij} = 1 | \alpha_{ij}^* = \mathbf{0}) = 0.1$, as that suggested by Ma and de la Torre (2016) to represent items with high quality. To focus on the comparison of the two estimation procedures with different attribute hierarchies, the test length was fixed as 30 items and the number of attributes as 5. Four types of attribute hierarchies are considered in this study: linear, convergent, divergent, and mixed structures (see Figure 1). Note that the mixed structure now consists of two sets of linear structure, one has a linear structure for Attributes 1 to 3, and the other has a linear structure for Attributes 4 to 5. These two sets of attributes are independent. The true latent class structure for each data was simulated from a uniform discrete distribution on all the permissible attribute profiles under the assumed attribute hierarchy.

The Q_H Matrices Under Different Attribute Hierarchies

The test Q_H matrices under each of the four attribute hierarchies were developed as follows: All the possible permissible attribute profiles were first constructed under each structure. The transpose of those permissible attribute vectors form a base Q matrix and in which they are the row vectors. These base Q matrices of the four types of attribute structure are provided in the online-support document. The Q_H matrices in different hierarchical structures were generated based on the corresponding base Q matrices. Specifically, when $K = 5$, the number of permissible attribute patterns for linear, convergent, divergent, and mixed structures were 5, 6, 9, and 11, respectively. The Q_H matrices for the tests with 30 items were constructed by replicating the corresponding base Q matrices 6, 5, 3, 2 times for those four structures. For the Q_H matrices with divergent and mixed structure, the last three and last eight item-attributes were randomly selected from their corresponding permissible attribute profiles, respectively.

Evaluation Criteria

The two estimation methods are evaluated through model fit, item parameter estimation and classification accuracy. The Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) are used to compare the model fit of the restricted estimation and the unrestricted estimation procedure. These two indexes are defined as $AIC = -2 \times \log(L) + 2d$, $BIC = -2 \times \log(L) + d \times \log(N)$. Here, L is the likelihood based on

MLE, d refers to the number of parameters under the assumed model, and N is sample size. In the unrestricted estimation procedure, $d = (2^K - 1) + P$, whereas for the restricted estimation procedure, $d = (M - 1) + P$. Here, P is the number of item parameters and M is the cardinality of the set of the permissible attribute profiles. The smaller value of AIC or BIC indicates a better model fit. Due to multiple replications in this study, the mean AIC and BIC indexes are calculated by $\text{MAIC} = \frac{1}{R} \sum_{r=1}^R \text{AIC}_r$ and $\text{MBIC} = \frac{1}{R} \sum_{r=1}^R \text{BIC}_r$. Here, R is the number of simulation replications, and in this case, $R = 100$. The item parameter estimation is evaluated based on the root mean square error (RMSE), which is defined as,

$$\text{RMSE} = \sqrt{\frac{\sum_{r=1}^R \sum_{c=1}^C \sum_{j=1}^J \left[P^{(r)}(X_j = 1 | \alpha_c) - \hat{P}^{(r)}(X_j = 1 | \alpha_c) \right]^2}{J \times C \times R}},$$

where J , K , C , and R denote the number of items, attributes, permissible attribute pattern, and replications, respectively. $P^{(r)}(X_j = 1 | \alpha_c)$ and $\hat{P}^{(r)}(X_j = 1 | \alpha_c)$ represent the true and estimated correct response probability of item j for individuals with attribute pattern α_c in the r th replication. Note that the item parameters for all the CDMs are the reparameterization of the correct response probability under a given latent attribute profile (S. Wang, 2017); thus, the RMSE defined above can be viewed as the criterion for item parameter estimation recovery rate across all items and all replications. The classification accuracy was evaluated by two indexes. The first index is the pattern-wise agreement rate (PAR) which reflects the agreement between estimated attribute profiles and the true attribute profiles. PAR is defined as $\text{PAR} = \frac{1}{N} \sum_{i=1}^N I\{\alpha_i = \hat{\alpha}_i\}$, that is the proportion of accurately estimated attribute profiles. The second is defined as the impermissible attribute pattern rate (IPR). The impermissible attribute pattern rate is defined as $\text{IPR} = m/N$, where m is the number of impermissible attribute profile estimations and N is the sample size. As discussed before, some attribute patterns are impermissible under the specified attribute hierarchy, so the lower the IPR, the higher the classification accuracy of the estimation method.

Results

The overall model fit of four CDMs in different simulation conditions are presented in Figure 2. An improvement of model fit can be observed by using the restricted estimation in each simulation condition. The RMSEs of the item parameters from the two estimation procedures across the four CDMs are documented in Figure 3. For the DINA model, the two estimation methods both have good recovery of item parameters. The restricted method produces a slightly better result for the rRUM and the ACDM, and improved a lot for the GDINA, compared with the unrestricted estimation method. The different performance of the two estimation methods on the item parameter estimation across these four models may due to different model complexities. The DINA model only has two types of correct response probability for each item, no matter whether the latent attribute profile space is restricted or not. The GDINA, as a representative of the saturate CDM, has $2^{K_j^*}$ item parameters for item j when the attribute profile space is not restricted to the specific structure. In the case where the attribute hierarchy does exist, the number of correct response probabilities of a given item under the GDINA model will be smaller than $2^{K_j^*}$. However, the unrestricted method still produces $2^{K_j^*}$ possible estimates of the correct response probabilities and this might contaminate the estimation on the correct response probability within the permissible attribute profiles. The model complexities of rRUM and the

Figure 2. Model-fit indexes.

Note. DINA-I and DINA-H represented the unrestricted estimation procedure and restricted estimation procedure when using DINA model, respectively. The similar explanations are for the rRUM-I versus rRUM-H, ACDM-I versus ACDM-H, and GDINA-I versus GDINA-H. DINA = deterministic input noisy "and" gate; rRUM = reduced reparameterized unified model; AIC = Akaike information criterion; BIC = Bayesian information criterion; ACDM = additive cognitive diagnosis modeling; GDINA = generalized DINA; MAIC = mean Akaike information criterion; MBIC = mean Bayesian information criterion.

Figure 3. RMSEs of item parameters.

Note. DINA-I and DINA-H represented the unrestricted estimation procedure and restricted estimation procedure when using DINA model, respectively. The similar explanations are for the rRUM-I versus rRUM-H, ACDM-I versus ACDM-H, and GDINA-I versus GDINA-H. RMSE = root mean square error; DINA = deterministic input, noisy "and" gate; rRUM = reduced reparameterized unified model; ACDM = additive cognitive diagnosis modeling; GDINA = generalized DINA.

Figure 4. The IPRs and PARs of four models.

Note. DINA-I and DINA-H represented the unrestricted estimation procedure and restricted estimation procedure when using DINA model, respectively. The similar explanations are for the rRUM-I versus rRUM-H, ACDM-I versus ACDM-H, and GDINA-I versus GDINA-H. IPR = impermissible attribute pattern rate; PAR = pattern-wise agreement rate; DINA = deterministic input noisy “and” gate; rRUM = reduced reparameterized unified model; ACDM = additive cognitive diagnosis modeling; GDINA = generalized DINA.

ACDM are between the DINA and the GDINA, so the performances of the two estimation methods on these two models are between them too.

Finally, Figure 4 documents the IPRs and PARs of four models across different attribute hierarchies. The overall trend indicates that the PARs from the restricted estimation approach are always higher than those from the unrestricted estimation method under each of the four attribute structures. Compared with the PARs from the restricted estimation procedure, the PARs from the unrestricted estimation procedure decrease around 35% to 50%, 30% to 42%, 34% to 51%, and 40% to 61% for the DINA, rRUM, ACDM, and GDINA, respectively. The IPRs for these four models due to the unrestricted estimation procedure are around 42.6% to 47.8%. As for the restricted MLE, it always produces 0% impermissible attribute pattern rates due to the estimation procedure naturally exclude the impermissible attribute patterns. For the DINA, rRUM, and ACDM model, although the item parameter estimation performs well, some impermissible attribute patterns can have the exactly same or similar likelihood value with that from a permissible attribute pattern, and the unrestricted estimation procedure cannot distinguish them from each other. For the GDINA model, when the unrestricted estimation method is used, the item parameter recovery is relatively low among the four models and this can further contaminate the likelihoods used for the attribute profile classification. It may due to this reason to produce the largest misclassification rate and largest IPR among the four models.

Real Data Analysis

In this section, the performance of the restricted attribute profile estimation was compared with the unrestricted one based on a real data application. This data set contains responses to 15 items involving syllogistic reasoning from 769 undergraduate students in two universities of China. This is the motivating example presented in the “Introduction” section. Leighton et al. (2004) applied AHM to model syllogistic reasoning where they suggested a divergent attribute hierarchy can be well constructed based on Philip Johnson-Laird’s theory of mental models (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). The details of these mental models can be referred in Leighton et al. (2004). The items in this test are designed based on the categorical syllogisms, and three examples are presented in Figure 5 below.

<div>All A are B All B are C ? All A are C</div>	<div>All A are not B All C are B ? All C are not A</div>	<div>Some A are B No B are C ? Some A are not C</div>
--	--	---

Figure 5. Three examples of syllogistic reasoning item.

Table 1. Permissible Attribute Profiles Based on the Divergent Attribute Hierarchy in Figure 4 of Leighton, Gierl, and Hunka (2004).

ID	Attribute profile
1	0000000
2	1000000
3	1100000
4	1110000
5	1101000
6	1101100
7	1101010
8	1101011
9	1111000
10	1111100
11	1111010
12	1101110
13	1111011
14	1101111
15	1111110
16	1111111

The hierarchical structure of seven attributes is suggested by the Figure 4 in Leighton et al. (2004), which is also presented in the online-support document. Based on this attribute hierarchy, there are only 16 types of permissible attribute profiles which are listed in Table 1. The Q_H matrix of the 15 items was constructed from all the permissible attribute profiles except the first one whose elements are all zeros. Each selected permissible attribute pattern represents a vector in the Q_H matrix.

In this analysis, the DINA, the rRUM, the ACDM, the GDINA, as well as the LCDM and the HDCM are used to do classification. Each model is associated with two types of attribute profile estimation procedures: the restricted estimation procedure that restricts the attribute profile space to the permissible attribute profile set, and unrestricted one that considers 2^7 attribute profiles. The HDCM is defined to be paired with the restriction estimation in the original article (Templin & Bradshaw, 2014), and here, the unrestricted estimation procedure was also used to further investigate it. The model parameters and the attribute patterns were estimated through a MMLE/EM algorithm as those in the simulation study. Note that the purpose is not to compare different models, but to investigate the performance of the restricted estimation procedure and unrestricted estimation procedure across different models.

Table 2. Model-Fit Indices for Syllogistic Reasoning Data.

Models	Estimation procedure	Model fit	
		AIC	BIC
DINA	DINA-I	11,217	11,946
	DINA-H	10,993	11,202
rRUM	rRUM-I	10,551	11,517
	rRUM-H	10,981	11,427
ACDM	ACDM-I	10,956	12,507
	ACDM-H	10,785	11,366
LCDM	LCDM-I	12,077	15,073
	LCDM-H	12,413	14,889
GDINA	GDINA-I	10,591	13,587
	GDINA-H	10,653	13,129
HDCM	HDCM-I	11,289	12,306
	HDCM-H	11,721	12,218

Note. DINA-I and DINA-H represented the unrestricted estimation procedure and restricted estimation procedure when using DINA model, respectively. The similar explanations are for the rRUM-I versus rRUM-H, ACDM-I versus ACDM-H, LCDM-I versus LCDM-H, and GDINA-I versus GDINA-H. AIC = Akaike information criterion; BIC = Bayesian information criterion; DINA = deterministic input noisy “and” gate; rRUM = reduced reparameterized unified model; ACDM = additive cognitive diagnosis modeling; LCDM = log-linear cognitive diagnosis model; GDINA = generalized DINA; HDCM = hierarchical diagnostic classification models.

Result

Table 2 summarizes the model fit from different models in terms of the AIC and BIC. Based on the BIC index, the model fit is improved when using the restriction estimation procedure for each of the models, especially for the more general model, while based on the AIC index there is a mixed effect across different models. This is reasonable as the restriction procedure reduces the number of attribute profiles as well as the corresponding item parameters in the kernel of the general model, thus reduced the corresponding BIC. The HDCM-H has a better fit than the LCDM-H, which is consistent with the finding from Templin and Bradshaw (2014), as it further reduces the number of item parameters due to the restricted kernel form.

Note that only 16 out of 128 attribute patterns are permissible under the specified attribute hierarchy. The unrestricted estimation procedure can produce many impermissible attribute profile estimations, even if it might have a better model fit in terms of AIC for some models. Specifically, the impermissible attribute rates of the unrestricted estimation procedure based on the DINA, the rRUM, the ACDM, the GDINA, the LCDM, and the HDCM are 40.7%, 61.1%, 56.2%, 76.9%, 66.1%, and 89.5%, respectively. For example, when using the unrestricted estimation procedure, 2.6% examinees’ attribute profiles were estimated as (1000111) by the DINA model, 3.6% examinees’ attribute profiles were estimated as (0100100) by the rRUM, 3.1% examinees’ attribute profiles were estimated as (0100100) by the ACDM, 9.9% examinees’ attribute profiles were estimated as (0100110) by the LCDM, 6.9% examinees’ attribute profiles were estimated as (0000011) by the GDINA model, and 6.8% examinees’ attribute profiles were estimated as (10110000) by the HDCM model. According to the assumed attribute hierarchy, Attributes 1 and 2 were the precursor attributes of all others; Attribute 4 was the precursor of Attribute 5, 6, and 7; and Attribute 6 was the precursor attribute of 7. Such estimated impermissible attribute profiles indicate examinees could master a difficult skill but slip on an easy skill. Finally, the estimated mastery proportion of each attribute, which is calculated as the ratio of the number of examinees who were masters of that attribute to the total number of examinees,

Table 3. The Estimated Master Proportion of Each Attribute.

Model	Attributes						
	1	2	3	4	5	6	7
DINA							
DINA-I	0.906	0.914	0.598	0.771	0.502	0.575	0.581
DINA-H	0.953	0.910	0.685	0.815	0.519	0.588	0.580
rRUM							
rRUM-I	0.521	0.579	0.641	0.449	0.397	0.334	0.230
rRUM-H	0.860	0.787	0.459	0.542	0.407	0.380	0.313
ACDM							
ACDM-I	0.485	0.391	0.406	0.324	0.355	0.303	0.100
ACDM-H	0.714	0.568	0.287	0.391	0.282	0.225	0.189
LCDM							
LCDM-I	0.558	0.547	0.511	0.417	0.547	0.360	0.241
LCDM-H	0.834	0.223	0.121	0.376	0.120	0.100	0.068
GDINA							
GDINA-I	0.365	0.506	0.398	0.397	0.451	0.324	0.555
GDINA-H	0.922	0.886	0.598	0.685	0.545	0.566	0.542
HDCM							
HDCM-I	0.745	0.365	0.433	0.537	0.610	0.553	0.615
HDCM-H	0.870	0.683	0.485	0.579	0.345	0.319	0.183

Note. DINA-I and DINA-H represented the unrestricted estimation procedure and restricted estimation procedure when using DINA model, respectively. The similar explanations are for the rRUM-I versus rRUM-H, ACDM-I versus ACDM-H, LCDM-I versus LCDM-H, GDINA-I versus GDINA-H, and HDCM-I versus HDCM-H. DINA = deterministic input noisy “and” gate; rRUM = reduced reparameterized unified model; ACDM = additive cognitive diagnosis modeling; LCDM = log-linear cognitive diagnosis model; GDINA = generalized DINA; HDCM = hierarchical diagnostic classification models.

was summarized in Table 3. Again, the restricted estimation method produces more reasonable results than the unrestricted estimation procedure. For example, when using the unrestricted estimation procedure, more examinees were diagnosed to master Attribute 2 than 1, or Attribute 7 than 6 under the DINA model; more examinees were diagnosed to master Attribute 2 than 1, or Attribute 3 than 2 under the rRUM model; more examines were diagnosed to master Attribute 3 than 2, or Attribute 5 than 4 under the ACDM model; more examines were diagnosed to master Attribute 5 than 4 under the LCDM model; more examines were diagnosed to master Attribute 2 than 1, or Attribute 5 than 4, or Attribute 7 than 6 under the GDINA model; more examinees were diagnosed to master Attribute 3 and 4 than 2, or Attribute 7 than 6 under the HDCM model.

Conclusion

Attribute hierarchies describe the interdependency among cognitive attributes. A hierarchical order implies that the higher order attribute can only be mastered if a lower order attribute is mastered. When attribute hierarchies exist, a restricted Q-matrix design which reflects the specified attribute structure can be developed before item writing. When the items are written according to such a Q matrix structure, the traditional nonhierarchical CDMs that assume all profiles of attributes are present in a population of interest may result in many equivalence classes. All the attribute profiles in the same equivalence class have the same ideal response profile, and

they can result in the exactly same or similar likelihood values. This could lead to misclassification of attribute profiles due to the nonidentifiability problem.

Detailed analysis was provided about the advantages of using a restricted estimation procedure that limits the latent class structure within the permissible attribute patterns and evaluates them on four attribute hierarchies for the latent class-based CDM through a simulation study and a real data analysis. The results are obtained under two conditions: (a) a restricted Q-matrix design is used in the diagnostic assessment and (b) a latent class-based CDM is appropriate for analyzing the response data. With these two design conditions, the attribute profile estimation should be restricted in the set which contains all permissible attribute profiles under the specified hierarchical structure, no matter using the frequentist estimation procedure or Bayesian estimation procedure. If the EM algorithm is used, as suggested by de la Torre (2012), this can be achieved by either setting the population probability of impermissible attribute profiles as zero or dropping the impermissible attribute profiles. The simulation study and the real data example show the consequence of getting undesirable classification results if the attribute profile estimation was not restricted in the permissible attribute profile set. By using the restricted estimation method, the number of latent classes can be reduced from the full space when attribute hierarchies are assumed to exist; this can in turn to improve the model fit and item parameter estimation as indicated by the simulation study. At the same time, because only the permissible attribute profiles are considered, the number of latent classes can be reduced from 2^K . The sample size needed to obtain stable parameter estimates from CDM calibrations will decrease as well, and this can result in faster computing time especially when there are a large number of attributes and items (Su et al., 2013).

The different results of classification accuracy from the HDCM and LCDM (Templin & Bradshaw, 2014) may be due to the different modeling of the attribute hierarchy. For the HDCM with a balanced or unstructured Q matrix, the attribute hierarchy is reflected through the restricted kernel of the item response function (Templin & Bradshaw, 2014, Section 7, p. 325). This parameterization is only suitable for several general models, such as LCDM, GDM, and GDINA. However, for the simple model, the DINA model, the attribute hierarchy cannot be reflected through it. In this setup, the attribute hierarchy is reflected through the restricted Q matrix structure. Although the item response function itself does not change, the restricted Q matrix together with the restricted latent class structure reflects the specified hierarchical structure. Within this framework, no matter the measurement model is a parsimonious model, such as DINA or a general model, such as LCDM, the attribute hierarchical structure can be reflected through the restricted Q matrix. The study assumes the attribute hierarchical structure is reflected through the Q matrix, and investigates the factors which might affect the classification accuracy. Both the simulation study and the real data application implies that it is important to restrict the latent class structure within the permissible attribute profile spaces no matter which type of CDM is used. Especially when using the restricted estimation procedure, the parsimonious model, such as DINA, can also produce the reasonable and good classification results under attribute hierarchy, as indicated by the real data analysis.

The current study can be treated as an empirical study to investigate the model identifiability with a restricted Q-matrix design. The research work of model identifiability for CDMs under an unstructured Q-matrix design (Xu, 2017; Xu & Zhang, 2016) has been well addressed; however, there has not much work been done to investigate the estimation of CDMs with a restricted Q-matrix design. The current study assumes the Q matrix is correctly specified and contains a reachability matrix to reflect the attribute structure. There might be different ways to construct the Q matrix given a specified attribute structure. It is important to investigate the impact of different Q-matrix designs to the classification accuracy when combined with different types of CDMs, which hopefully can be rigorously generalized to

the discussion of the model identifiability issues when the attribute hierarchy is specified before test development.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (31660278, 31760288).

Supplemental Material

Supplementary material is available for this article online.

Reference

- Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automated Control*, 19, 716-723.
- Chiu, C. Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74, 633-665.
- 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. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J. (2012, April). *Cognitive diagnosis modeling: A general framework approach* (Session5: Estimation of CDMs). Training session provided at the annual meeting of the National Council of Measurement Research, Vancouver, Canada.
- 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.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, 74(2), 1-24.
- Gierl, M. J. (2007). Making diagnostic inferences about cognitive attributes using the rule space model and attribute hierarchy method. *Journal of Educational Measurement*, 44, 325-340.
- Gierl, M. J. (2008). Using the attribute hierarchy method to identify and interpret cognitive skills that produce group differences. *Journal of Educational Measurement*, 4, 565-589.
- Groß, J., & George, A. C. (2014). On permissible attribute classes in noncompensatory Cognitive Diagnosis Models. *Methodology*, 3, 100-107.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 333-352.
- Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
- Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.
- Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, 71, 407-419.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, 16, 1-61.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum.

- 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.
- Leighton, J. P., & Gierl, M. J. (2007). *Cognitive diagnostic assessment for education: Theory and practice*. Cambridge, UK: Cambridge University Press.
- Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, 41, 205-236.
- 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.
- Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic assessment: Theory, methods, and applications*. New York, NY: Guilford Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Sternberg, R. J. & Ben-Zeev, T. (Eds.) (1996). *The nature of mathematical thinking*. Hillsdale, NJ: Erlbaum.
- Su, Y., Choi, K., Lee, W., Choi, T., & McAninch, M. (2013). *Hierarchical cognitive diagnostic analysis for TIMSS2003 mathematics* (CASMA Research Report No. 35). Iowa City: Center for Advanced Studies in Measurement and Assessment, University of Iowa.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational and Behavioral Statistics*, 10, 55-73.
- Tatsuoka, K. K. (1995). Architecture of knowledge structure and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327-361). Hillsdale, NJ: Lawrence Erlbaum.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction of the rule space method*. New York, NY: Routledge.
- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79, 317-339.
- Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnostic models. *Psychological Methods*, 11, 287-305.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287-307.
- Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, 48, 165-118.
- Wang, S. (2017). Two-stage maximum likelihood estimation in the misspecified restricted latent class model. *British Journal of Mathematical and Statistical Psychology*. Advance online publication. doi:10.1111/bmsp.12119
- Wang, S., & Douglas, J. (2015). Consistency of nonparametric classification in cognitive diagnosis. *Psychometrika*, 80, 85-100.
- Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, 45, 675-707.
- Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81, 625-649.
- Zhang, S. S. (2014). *Statistical inference and experimental design for Q-matrix based Cognitive Diagnosis Models* (Doctoral dissertation). Columbia University, New York, NY.
- Zhang, S. S., DeCarlo, L. T., & Ying, Z. (2013). *Non-identifiability, equivalence classes, and attribute-specific classification in Q-matrix based Cognitive Diagnosis Models* (arXiv:1303.0426). Retrieved from <https://arxiv.org/pdf/1303.0426.pdf>