# On Comparing Classification Accuracy of Parametric and Non-parametric Cognitive Diagnosis Models

by

Muhammed Veyis Kilincer

A Culminating Project

Submitted to the Faculty

of the

TEACHERS COLLEGE, COLUMBIA UNIVERSITY

in partial fulfillment of the requirements for the

Degree of Master of Education

in

Measurement and Evaluation

May 2016

APPROVED:
Professor Matthew Johnson, Advisor

**Abstract**

Cognitive Diagnosis Models (CDM's), which are designed to diagnose students' mastery level of skills by assigning each individual student into discretely scored latent skill classes, have attracted a great deal of attention. Due to its fruitful contribution in tailoring instructional interventions, various parametric CDM's and their extensions have been introduced by many researchers. Other than parametric CDM's, Chiu and Douglas (2013) introduced an alternative non-parametric diagnosis model that classifies students into latent skill classes based on minimizing the distance between observed and ideal response vectors. In this study, DINA and DINO parametric models as well as several differently specified non-parametric models that are proposed by Chiu and Douglas (2013) were compared in terms of classification accuracy of students into the latent skill categories. The results revealed that non-parametric models that are suitably specified as an alternative to the parametric ones yield to an improvement in classification accuracy in comparison to both DINA and DINO models. As a result, findings suggested that emerging non-parametric diagnosis models can be effectively used in practice instead of parametric ones.

**Keywords**: Non-parametric cognitive diagnosis model, DINA, DINO

## Introduction

Cognitive Diagnosis Models (CDM's), which are designed to diagnose students' mastery level of skills by assigning each individual student into discretely scored latent skill classes, have attracted a great deal of attention, especially, after the "No Child Left Behind" act of 2001 (Tatsuoka, 2009; Huebner, 2010). The initiative emphasizes the importance of enhancing formative assessment tools to adapt instruction to students' specific needs and making educational progress rather than assessing overall educational progress (Tatsuoka, 2009). Moreover, researchers from all levels of education in various domains suggest that utilizing detailed feedback from assessments could be very useful in understanding students' strengths and weaknesses (Bradshaw & Templin, 2014).Because of these, different types of CDM's have been extensively studied and improved in order to determine whether a student has mastered a set of specific skills that are purposed to be measured by a test (Bradshaw & Templin, 2014).

Due to its fruitful contribution in tailoring instructional interventions, different CDM's and their extensions have been introduced by many researchers. Although different CDM's have distinctive features, implementation of these models requires specification of a Q-matrix which maps each item to a set of required skills on a test. That is, Q-matrix specifies a set of skills that need to be possessed by a subject to answer a specific item correctly (Tatsuoka, 1990). In CDM framework, assessment of a subject's mastery level of 'K' attributes, where K = 1, 2… k, on a test is achieved by assigning a binary valued attribute vector to each subject such that $\alpha = (\alpha_1, \ldots, \alpha_k)$; where 1's indicates mastery of the skill and 0's indicates non-mastery. Therefore, assignment of students into latent skill classes is the fundamental purpose of any CDM (DeCarlo, 2011). To achieve this on a test that measures K skills, diagnosis models address minimizing the discrepancy in assigning subjects to proficiency classes among the total of $2^K$

possible skill patterns (Tatsuoka, 1990). In other words, CDM's diagnose the departures from 'ideal response patterns' —"patterns that specifiy which items are passed and failed by someone in the knowledge state" (Embretson & Reise, 2000) — by modeling the relationship between a subject's attribute pattern and the response pattern (Huebner, 2010). Other than various parametric CDM's and several other non-parametric diagnosis models based on clustering algorithm developed to accomplish this purpose, Chiu and Douglas (2013) introduced an alternative non-parametric diagnosis model that classifies students into latent skill classes based on minimizing the distance between observed and ideal response vectors.

According to Rupp and Templin (2008), various diagnostic models can be categorized with regard to their distinctive characteristics: Models that are compatible with (i) dichotomous versus polytomous observed response variables, (ii) dichotomous versus polytomous latent predictor variables, and (iii) conjunctive (non-compensatory) versus disjunctive (compensatory) models. In the simplest term, conjunctive models assume that an examinee must possess all the required skills by the item to get it correct; whereas, in the disjunctive models, the presence of at least one required skill is enough to answer the item correctly (DeCarlo, 2011; Rupp, Templin & Henson 2010; Chiu & Douglas, 2013). Considering that the purpose of this paper is to compare the classification accuracy of several parametric and non-parametric cognitive diagnosis models on a real data set, one conjunctive (i.e., DINA) and one disjunctive (i.e., DINO) parametric model as well as a non-parametric model —based on proximity of observed response pattern to the ideal response patterns by Chiu and Douglas (2013) — will be briefly explained, in turn, conducted on a real data set.

**Review of Several Cognitive Diagnosis Models**

**DINA**

Sijtsma and Junker (2001) define the item response function of deterministic input, noisy "and" gate (DINA) model as follows: The probability of getting an item correct for the $i$th subject depends on two error probabilities which are slipping and guessing probabilities, and the ideal response pattern ($\eta_{ij}$).

$$P(Y_{ij}=1 \mid \alpha_i) = (1-s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}$$

The slipping probability for the $j$th item ($s_j$) denotes the probability of getting a wrong response on the $j$th item when all necessary attributes are present. Similarly, the guessing probability for the $j$th item ($g_j$) denotes the probability of getting a correct response on the $j$th item when at least one necessary attributes is not possessed.

$$s_j = P(Y_{ij}=0 \mid \eta_{ij}=1) \text{ and } g_j = P(Y_{ij}=1 \mid \eta_{ij}=0)$$

The parameter $\eta_{ij}$—which is sometimes referred as 'latent dichotomous variable' (Junker & Sitjma, 2001) or 'ideal response pattern' (Chiu & Douglas, 2013) — links the latent skill patterns possessed by the $i$th subject, which is $\alpha_{ik} = (\alpha_{i1},\ \alpha_{i2},\ \ldots,\ \alpha_{ik})$, and Q-matrix such that

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$$

It is obvious from the formula of $\eta_{ij}$ that $i$th subject must possess all the required attributes to correctly respond to $j$th item (Sitjma & Junker, 2001). This way of combining latent attribute patterns, which is referred as the 'conjunctive condensation rule' (Rupp &

Templin, 2008), is the distinctive feature of the deterministic part of the DINA model that explains why this model is a conjunctive diagnosis model (DeCarlo, 2011).

**DINO**

The deterministic, noisy "or" gate (DINO) model introduced by Templin and Henson (2006) is the disjunctive analog of the DINA model. In the DINO model, a subject is assumed to possess at least one of the required attribute to answer an item correctly. Therefore, the yielding ideal response pattern of this model is defined as:

$$\omega_{ij} = 1 - \prod_{k=1}^{K} (1 - \alpha_{ik})^{q_{jk}}$$

The item response function of the DINO model is same when the deterministic part of the DINA model, $\eta_{ij}$, is replaced with $\omega_{ij}$ such that

$$P(Y_{ij} = 1 \mid \alpha_i) = g_j^{(1-\omega_{ij})} (1 - s_j)^{\omega_{ij}}$$

Although the interpretation of the slipping and guessing probabilities are slightly different, their formulas are same when $\eta_{ij}$, is replaced with $\omega_{ij}$.

$$s_j = P(Y_{ij} = 0 \mid \omega_{ij} = 1) \text{ and } g_j = P(Y_{ij} = 1 \mid \omega_{ij} = 0)$$

**Non-Parametric Approach as Proximity to Ideal Response Patterns**

Chiu and Douglas (2013) introduce a non-parametric cognitive diagnosis approach that "makes the classification of students to latent attribute classes ($\hat{\alpha}$) by minimizing some measure of the distance over all possible ideal response vectors, and determining the $\alpha$ associated with the nearest ideal response vector". Therefore, $\hat{\alpha}$ is the value of $\alpha$ that produces the smallest distance value between the observed and ideal response vectors. The authors define the ideal response pattern of the model in two ways which are similar to the 'conjunctive condensation rule' of DINA model and 'disjunctive condensation rule' of DINO model. For clarity, I will explain only the non-parametric approach with 'conjunctive condensation rule'. The ideal response pattern of the non-parametric approach is same with the DINA model and formulated as follows:

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$$

Similar to the DINA model, $\eta_i$ is the function of latent $\alpha_i$, and depends only on the Q-matrix. For K latent skills, the model contains all possible ideal response patterns $\eta^{(1)}, \eta^{(2)}, \ldots, \eta^{(K)}$ (i.e., $2^K$ possible values for $\alpha_i$). Then, it minimizes the distance between the observed item response vector and the ideal response vector by the following distance function:

$$d(y_i, \eta^{(m)}), \text{ for } m = 1, 2, \ldots, 2^K$$

At that point, it is important to note that specification of the $d(.)$ function is crucial and should be appropriately chosen to compensate issues related to slipping and guessing as well as several other possible issues. Considering the binary data, 'Hamming distance' which is widely used in clustering analysis to count the number of disagreements between two vectors is a suitable distance function. Although the selection of Hamming distance is suitable with binary data, different specifications of Hamming distance as a distance function is generally required due to the fact that some items on a test could have more response variability than others, also, guessing is not possible on an open-ended item. Considering these practical issues, the authors introduce three different specifications of Hamming distance as a distance functions: (i) Plain

Hamming distance (ii) Weighted Hamming distance, and (iii) Penalized Hamming distance. The 'Plain Hamming' distance function can be formulated similar to Hamming distance as follows:

$$d_h(y,\ \eta) = \sum_{j=1}^{J} | y_j - \eta_j |$$

For the test that includes items with varying response variability, the authors propose the Weighted Hamming distance function so that it puts more weights on terms linked with items having smaller variances. In doing so, Chiu and Douglas (2013) modify the Plain Hamming distance function by weighting with respect to the inverse sample variance, and the resulting Weighted Hamming distance function is:

$d_{wh}(y,\ \eta) = \sum_{j=1}^{J} \frac{1}{\overline{p}_j(1-\overline{p}_j)} | y_j - \eta_j |$, where $\overline{p}_j$ denotes the proportion correct on the $j$ th item.

For the third specification of the approach, the 'Penalized Hamming' distance function, the authors introduce two terms: $w_g$ denotes the weight assigned to a guess, and $w_s$ denotes the weight assigned to a slip. The 'Penalized Hamming' distance becomes equal to 'Plain Hamming' distance when $w_g = w_s = 1$. The authors suggest that the distance function can be set to assign more weight to guesses when $g < s$, and more weight to slips when $g > s$. Besides, these two penalizing terms can be specified for each item if needed. The resulting distance function is:

$$d_{gs}(y,\ \eta) = \sum_{j=1}^{J} w_g I[y_j = 1] | y_j - \eta_j | + \sum_{j=1}^{J} w_s I[y_j = 0] | y_j - \eta_j |$$

The Penalized Hamming distance function can be appropriately specified for the type of the test. For example, one can penalize guesses for open-ended subject responses because guessing is impossible, in turn, penalizing guessing will yield to an improvement in terms of disagreements between the observed and ideal response vectors. However, for a multiple-choice test, guessing should not be penalized as much as a disagreement because $y_j = 0$ when $\eta_j = 1$ is sometimes considered as a slip.

These non-parametric diagnostic methods depend on ideal response patterns, and therefore they are not directly using item parameters of any CDM such as slipping and guessing parameters. Due to this advantage of non-parametric model, it is computationally efficient and can be effectively used to utilize classify subjects into latent attribute classes with any sample size (Chiu & Douglas, 2013).

**ECPE Data**

The Examination for the Certificate of Proficiency in English (ECPE) data set used in the analysis, for N = 2922 participants. The ECPE data comprises 28 multiple-choice items that are devised to measure the of use English skills as a second language. These 28 items are associated to 3 latent skills: knowledge of (i) morphosyntactic rules, (ii) cohesive rules, and (iii) lexical rules as shown in Table 1 (Templin & Hoffman, 2013; Templin & Bradshaw, 2014). The ECPE data and Q-matrix is available in the 'CDM' package (Robitzsch, Kiefer, George, & Ünlü, 2014) in R (R Core Team, 2012)[1].

**Methods**

In order to compare the classification accuracy of parametric and non-parametric diagnosis models, two parametric diagnosis models and non-parametric diagnosis approach with 6 different specifications were used. For the parametric models, the DINA model as in Sijtsma

[1] The ECPE data in 'CDM' package is retrieved from the following link: http://jonathantemplin.com/dcm-workshop-spring-2012-ncme/

and Junker (2001) and the DINO model as in Templin and Henson (2006) conducted on the ECPE data by using ‘gdina’ function with default settings in the ‘CDM’ package (Robitzsch et al., 2014) in R (R Core Team, 2012). By default, this function implements Expectation Maximization (EM) algorithm in estimating diagnostic model parameters (De la Torre, 2011 as cited in Robitzsch et al., 2014). It also enables implementing various parametric CDM’s and their extensions by specification of necessary arguments such as link function and condensation rule. For the analysis, the DINA and the DINO model explained throughout this paper are fitted by using the ‘identity’ link function introduced in De la Torre (2011). The ‘identity’ link function in Generalized-DINA (G-DINA) framework is considered such that probability of a subject answering an item correctly is determined by the presence of specific latent attributes and their interactions. Therefore, the DINA model can be estimated by setting all interaction terms to zero through the ‘identity’ link of G-DINA; whereas, all main effects and interactions should be constrained to be identical to each other for the DINO model (De la Torre, 2011).

*Table 1. Q-matrix for the ECPE test.*

| **Item** | **Skill 1** | **Skill 2** | **Skill 3** | **Item** | **Skill 1** | **Skill 2** | **Skill 3** |
|---|---|---|---|---|---|---|---|
| **1** | 1 | 1 | 0 | **15** | 0 | 0 | 1 |
| **2** | 0 | 1 | 0 | **16** | 1 | 0 | 1 |
| **3** | 1 | 0 | 1 | **17** | 0 | 1 | 1 |
| **4** | 0 | 0 | 1 | **18** | 0 | 0 | 1 |
| **5** | 0 | 0 | 1 | **19** | 0 | 0 | 1 |
| **6** | 0 | 0 | 1 | **20** | 1 | 0 | 1 |
| **7** | 1 | 0 | 1 | **21** | 1 | 0 | 1 |
| **8** | 0 | 1 | 0 | **22** | 0 | 0 | 1 |
| **9** | 0 | 0 | 1 | **23** | 0 | 1 | 0 |
| **10** | 1 | 0 | 0 | **24** | 0 | 1 | 0 |
| **11** | 1 | 0 | 1 | **25** | 1 | 0 | 0 |
| **12** | 1 | 0 | 1 | **26** | 0 | 0 | 1 |
| **13** | 1 | 0 | 0 | **27** | 1 | 0 | 0 |
| **14** | 1 | 0 | 0 | **28** | 0 | 0 | 1 |

*Note. skill 1: morphosyntactic rules, skill 2: cohesive rules, Skill 3: lexical rules.*

For the non-parametric approach, the methods explained in Chiu and Douglas (2013) was adopted, and 6 different models were conducted on the ECPE data set by using ‘AlphaNP’ function with default settings in the ‘NPCD’ package (Zheng & Chiu, 2015) in R (R Core Team, 2012). These 6 models were specified according to 2 conditions for the condensation rule (i.e., conjunctive and disjunctive condensation rule), and 3 conditions for selection of the distance function that minimizes the distance between the observed and ideal response vectors (i.e., ‘Plain’, ‘Weighted’, and ‘Penalized’ Hamming distance). Four models that include either the Plain or the Weighted Hamming distance were fit with default settings. For the two Penalized Hamming distance models —conjunctive Penalized Hamming distance model and disjunctive Penalized Hamming distance model— more weight assigned to slips because guess parameters are larger than slips as presented in Table 2. Therefore, several slip weights, $\omega_s = 2,3,4$ , were considered while keeping guessing weight constant, $\omega_g = 1$ .

**Results & Conclusion**

Classification accuracy of the non-parametric models and the parametric models are determined according to two agreement indices: “pattern-wise agreement rate (PAR) which represents the proportion of agreement between the attribute patterns of the two models compared, and attribute-wise agreement rate (AAR) which represents the proportion of agreement between the individual attributes of the two models compared” (Chiu & Douglas, 2013). The mathematical notation of the agreement indices are:

$$PAR = \sum_{i=1}^{N} \frac{I[\widehat{\alpha}_i - \alpha_i]}{N} \qquad AAR = \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{I[\widehat{\alpha}_{ik} = \alpha_{ik}]}{NK}$$

*Table 2. Guess and slip parameters for the DINA and the DINO moedel*

| | **DINA** | | | | **DINO** | | | |
|---|---|---|---|---|---|---|---|---|
| **Item** | **Guess** | **SE (g)** | **Slip** | **SE (s)** | **Guess** | **SE (g)** | **Slip** | **SE (s)** |
| **1** | 0.7035 | 0.0123 | 0.0885 | 0.0089 | 0.6570 | 0.0165 | 0.1057 | 0.0076 |
| **2** | 0.7141 | 0.0175 | 0.1045 | 0.0073 | 0.7263 | 0.0153 | 0.0991 | 0.0074 |
| **3** | 0.4377 | 0.0135 | 0.2670 | 0.0129 | 0.4250 | 0.0141 | 0.2966 | 0.0130 |
| **4** | 0.4797 | 0.0158 | 0.1625 | 0.0099 | 0.5371 | 0.0141 | 0.1507 | 0.0105 |
| **5** | 0.7630 | 0.0133 | 0.0405 | 0.0053 | 0.7995 | 0.0112 | 0.0383 | 0.0056 |
| **6** | 0.7168 | 0.0141 | 0.0667 | 0.0068 | 0.7520 | 0.0121 | 0.0599 | 0.0071 |
| **7** | 0.5424 | 0.0135 | 0.0852 | 0.0087 | 0.5147 | 0.0142 | 0.1132 | 0.0095 |
| **8** | 0.7918 | 0.0152 | 0.0423 | 0.0049 | 0.8040 | 0.0134 | 0.0381 | 0.0045 |
| **9** | 0.5336 | 0.0157 | 0.1993 | 0.0106 | 0.5662 | 0.0140 | 0.1818 | 0.0111 |
| **10** | 0.4786 | 0.0139 | 0.1637 | 0.0110 | 0.4900 | 0.0134 | 0.1374 | 0.0120 |
| **11** | 0.5553 | 0.0134 | 0.0999 | 0.0092 | 0.5258 | 0.0142 | 0.1227 | 0.0097 |
| **12** | 0.1933 | 0.0113 | 0.3066 | 0.0136 | 0.1629 | 0.0104 | 0.3497 | 0.0135 |
| **13** | 0.6302 | 0.0133 | 0.1224 | 0.0096 | 0.6389 | 0.0127 | 0.1052 | 0.0105 |
| **14** | 0.5141 | 0.0139 | 0.2130 | 0.0119 | 0.5214 | 0.0134 | 0.1913 | 0.0130 |
| **15** | 0.7482 | 0.0135 | 0.0400 | 0.0054 | 0.7871 | 0.0114 | 0.0372 | 0.0057 |
| **16** | 0.5485 | 0.0134 | 0.1268 | 0.0101 | 0.5279 | 0.0142 | 0.1541 | 0.0105 |
| **17** | 0.8135 | 0.0117 | 0.0597 | 0.0067 | 0.7956 | 0.0136 | 0.0635 | 0.0062 |
| **18** | 0.7288 | 0.0139 | 0.0862 | 0.0075 | 0.7602 | 0.0120 | 0.0815 | 0.0080 |
| **19** | 0.4723 | 0.0158 | 0.1506 | 0.0096 | 0.5325 | 0.0141 | 0.1379 | 0.0102 |
| **20** | 0.2377 | 0.0120 | 0.2970 | 0.0134 | 0.2121 | 0.0116 | 0.3392 | 0.0135 |
| **21** | 0.6204 | 0.0130 | 0.0970 | 0.0089 | 0.5931 | 0.0139 | 0.1132 | 0.0093 |
| **22** | 0.3207 | 0.0148 | 0.1884 | 0.0106 | 0.3954 | 0.0139 | 0.1688 | 0.0110 |
| **23** | 0.6208 | 0.0189 | 0.0809 | 0.0066 | 0.6426 | 0.0164 | 0.0733 | 0.0065 |
| **24** | 0.2971 | 0.0195 | 0.3315 | 0.0114 | 0.3192 | 0.0172 | 0.3186 | 0.0117 |
| **25** | 0.5102 | 0.0139 | 0.2732 | 0.0129 | 0.5183 | 0.0134 | 0.2587 | 0.0140 |
| **26** | 0.5544 | 0.0157 | 0.2110 | 0.0107 | 0.5891 | 0.0139 | 0.2007 | 0.0115 |
| **27** | 0.2615 | 0.0125 | 0.3703 | 0.0138 | 0.2741 | 0.0122 | 0.3443 | 0.0152 |
| **28** | 0.6582 | 0.0149 | 0.0862 | 0.0075 | 0.6991 | 0.0129 | 0.0776 | 0.0079 |

*Note. Maximum and minimum guess and slip parameters are underlined.*

Comparison of the DINA model with the non-parametric conjunctive models indicates that all of the Penalized models with different slip weights performed better than the Plain and the Weighted model in terms of both attribute pattern classification and individual attribute classification as presented in the left side of Table 3. Additionally, the Plain model accomplished slightly better classification accuracy than the Weighted model. Furthermore, the Penalized model with $\omega_s = 3$ acquired the highest attribute-wise agreement rate of 0.86 and pattern-wise agreement rate of 0.67. Although all of the three conjunctive Penalized models showed very high classification accuracy according to both agreement indices, the one with $\omega_s = 3$ had the best performance, and therefore this suggest that slips deserve a moderate penalty. These results imply that the conjunctive Penalized model with a moderate slip weight can be considered as an effective alternative of the DINA model in diagnosing latent skills on the ECPE test.

*Table 3. Agreement indices between the non-parametric and the parametric models.*

| **Models compared** | | **Agreement Indices** | | **Models compared** | | **Agreement Indices** | |
|---|---|---|---|---|---|---|---|
| **With DINA**[+] | | **PAR** | **AAR** | **With DINO**[++] | | **PAR** | **AAR** |
| **Plain** | | 0.4442 | 0.7294 | **Plain** | | 0.3395 | 0.6828 |
| **Weighted** | | 0.4158 | 0.7058 | **Weighted** | | 0.3060 | 0.6562 |
| | $\omega_s = 2$ | 0.6410 | 0.8326 | | $\omega_s = 2$ | 0.5647 | 0.8146 |
| **Penalized** | $\omega_s = 3$ | 0.6728 | 0.8617 | **Penalized** | $\omega_s = 3$ | 0.7009 | 0.8865 |
| | $\omega_s = 4$ | 0.6249 | 0.8458 | | $\omega_s = 4$ | 0.7067 | 0.8789 |

[+] *The conjunctive non-parametric models are only compared with the DINA model.*
[++] *The disjunctive non-parametric models are only compared with the DINO model.*
*Note. Three slip weights,* $\omega_s = 2,3,4$ *used for the Penalized Hamming distance method.*
*Note. Maximum agreement indices are underlined.*

As in the comparison of models with the conjunctive condensation rule, comparison of the DINO model with the non-parametric disjunctive models indicates that all of the Penalized models with different slip weights outperformed the Plain and the Weighted model in terms of both agreement indices as presented in the right side of Table 3. Besides, the Plain model performed slightly better than the Weighted model. Moreover, all of the Penalized models achieved very high classification accuracy rate, especially the ones with $\omega_s = 3$ and $\omega_s = 4$ , as reported by the agreement indices though there is an ignorable small difference between them. The disjunctive penalized models with moderate slip weights reached outstanding classification accuracy in that almost 70 percent of attribute pattern and 88 percent of individual attribute classifications of these models were same with the DINO model. Similar to the results of conjunctive models, high classification accuracy of the disjunctive Penalized models with moderate slip weight implied that a moderate level of penalizing the slips yield to an improvement in classification accuracy, in turn, such a non-parametric disjunctive model presumably is a competent alternative of the DINO model in diagnosing the ECPE test.

In addition to the results of agreement indices, further investigation of proportion of subjects possessing each attribute $P(\alpha_k)$ and attribute pattern probabilities could be valuable in support of this analysis and for guiding future studies on the ECPE test. In doing so, diagnosis results of both parametric models and only the conjunctive and disjunctive Penalized models with a moderate slip weight (i.e., $\omega_s = 3$) are examined in detail.

*Table 4. Proportion of subjects possessed each skill $P(\alpha_k)$ for the parametric CDM's and the non-parametric Penalized Hamming distance models with $\omega_s = 3$.*

| | Parametric CDM's | | Penalized Hamming Distance Models | |
|---|---|---|---|---|
| **Skill** | **DINA** | **DINO** | **Conjunctive** | **Disjunctive** |
| **1** | 0.50 | 0.45 | 0.40 | 0.32 |
| **2** | 0.61 | 0.57 | 0.73 | 0.71 |
| **3** | 0.63 | 0.54 | 0.71 | 0.56 |

*Note. skill 1: morphosyntactic rules, skill 2: cohesive rules, Skill 3: lexical rules.*

According to diagnosis results of the DINA model on the ECPE test, Table 4 demonstrates that proportion of the subjects acquired each skill was moderate for each skill. That is, 50 percent of the subjects possessed "morphosyntactic rules" ($\alpha_1$), 61 percent of the subjects possessed "cohesive rules" ($\alpha_2$), and 63 percent of the subjects possessed "lexical rules" ($\alpha_3$). Similar to the DINA model, skill class probabilities of the DINO model also revealed that the proportion of the subjects obtained each latent skill was moderate per skill. However, both of the conjunctive and the disjunctive Penalized models with $\omega_s = 3$ produced varying skill class probability. For the conjunctive Penalized model: 40 percent of the subjects possessed “morphosyntactic rules” ($\alpha_1$), 73 percent of the subjects possessed “cohesive rules” ($\alpha_2$), and 71 percent of the subjects possessed “lexical rules” ($\alpha_3$); whereas, for the disjunctive Penalized model these percents are 32, 71, and 56 respectively. As a result, both of the Penalized models suggested that the least possessed skill was “morphosyntactic rules” ($\alpha_1$) and the most possessed skill was “cohesive rules” ($\alpha_2$), conversely, the percent of subject possessed “lexical rules” ($\alpha_3$) was found to be fairly higher for the conjunctive Penalized model compared to the disjunctive counterpart.

The investigation of skill pattern probabilities for the DINA model as represented in Table 5 indicated that 45 percent of the subjects acquired all of the skills, pattern (111), and 31 percent of the subjects not acquired any skill, pattern (000). In addition, percent of the subjects acquired any of the other 6 skill patterns were very low. Similar to the DINA model, the DINO model results revealed that the skill patterns (111) and (000) were possessed by 42 and 37 percent of the subjects, respectively. The other patterns were acquired by a very small portion of the subjects. Furthermore, the total percent of the subjects possessed either all of the skills or not any skill for the DINA and the DINO model were 76 and 79, respectively. These results asserted that both of the parametric CDM’s classified most of the subjects into two extreme skill mastery patterns. It appears that skill pattern variability among the subjects is surprisingly low in regards of analysis of both the DINA and the DINO model. Besides, these findings can be easily grasped by visual examination of skill pattern probabilities as shown in Figure 1.

On the contrary to analysis of the parametric CDM’s, the conjunctive and disjunctive Penalized models generated quite high skill pattern variability among the subjects. Specifically, the two mostly obtained attribute patterns according to the conjunctive Penalized model were

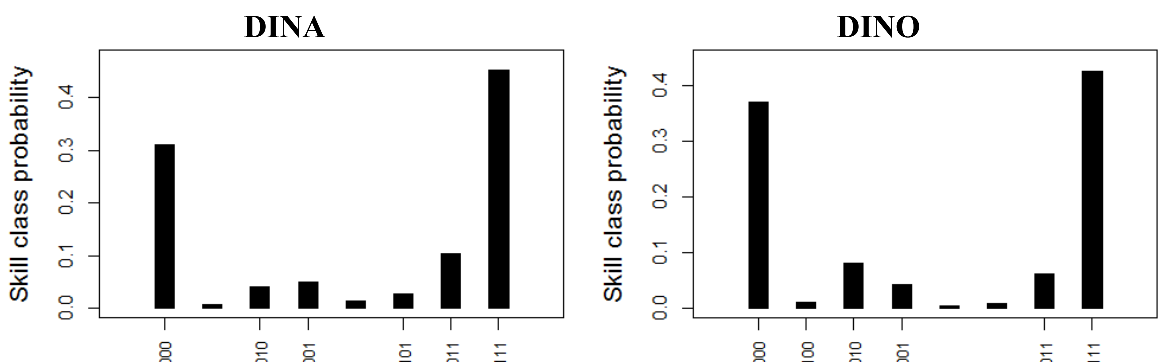
(111) and (011) with 31 and 27 percent; correspondingly, same attribute patterns were obtained by 26 and 18 percent of the subjects regarding the disjunctive Penalized model. Additionally, both of the Penalized models classified a very low proportion of the subjects into (000) pattern, not possessing any of the attribute pattern. The two mostly possessed skill patterns for these Penalized models were same: (111) and (011); whereas, it was (111) and (000) for the parametric models.

*Table 5. Skill pattern probabilities for the parametric CDM's and the non-parametric Penalized Hamming distance models with* $\omega_s = 3$.

| | Parametric CDM's | | Penalized Hamming Distance Models | |
|---|---|---|---|---|
| **Skill Pattern** | **DINA** | **DINO** | **Conjunctive** | **disjunctive** |
| **000** | **0.31** | **0.37** | 0.11 | 0.15 |
| **100** | 0.01 | 0.01 | 0.02 | 0.01 |
| **010** | 0.04 | 0.08 | 0.12 | 0.25 |
| **001** | 0.05 | 0.04 | 0.11 | 0.09 |
| **110** | 0.01 | 0.00 | 0.04 | 0.02 |
| **101** | 0.03 | 0.01 | 0.03 | 0.03 |
| **011** | 0.10 | 0.06 | **0.27** | **0.18** |
| **111** | **0.45** | **0.42** | **0.31** | **0.26** |

*Note. Two of the mostly possessed skill patterns for each model are bolded.*

*Figure 1. Skill pattern probabilities for the DINA and the DINO models.*




Tetrachoric correlations for the DINA and the DINO model as presented in Table 6 revealed that correlations between any pair of the attributes were adequately high, ranging between 0.88 and 0.96. These correlations implied that presumably there are substantial interactions between these attributes or collapsing two attributes to one attribute is required. Moreover, recognize that these parametric models classified most of the subjects into two extreme attribute mastery patterns —which are mastering all of the three attributes and non-mastering any of them— and this caused the very low variability in attribute pattern distribution. Such a classification manner could be due to the fact that (i) an attribute is a prerequisite of possessing another attribute (ii) there are some meaningful interaction terms (iii) there are misspecifications in the Q-matrix of the ECPE test and some other possible situations.

*Table 6. Lower triangular Tetrachoric correlations for the parametric CDM's.*

| | **DINA** | | | | **DINO** | | |
|---|---|---|---|---|---|---|---|
| | **Skill 1** | **Skill 2** | **Skill 3** | | **Skill 1** | **Skill 2** | **Skill 3** |
| **Skill 1** | 1 | | | **Skill 1** | 1 | | |
| **Skill 2** | 0.8813 | 1 | | **Skill 2** | 0.9137 | 1 | |
| **Skill 3** | 0.9036 | 0.9020 | 1 | **Skill 3** | 0.9587 | 0.9065 | 1 |

*Note. skill 1: morphosyntactic rules, skill 2: cohesive rules, Skill 3: lexical rules.*

For simplicity, considering the first two cases, presence of considerable interactions and a higher order relation between the attributes, the DINA and the DINO model that were used in this study were not able to uncover them due to their simple structure. Instead of supporting the idea of diagnosing the ECPE test by using the DINA and the DINO model and then comparing the classification accuracy between the parametric and the non-parametric CDM's afterward, this analysis suggests conducting another parametric CDM that could generate more accurate classifications. Therefore, the ECPE test could be more accurately diagnosed by conducting a parametric CDM that is more sophisticated in accounting (i) interactions between the attributes such as G-DINA and LCDM (Log Linear Cognitive Diagnosis Model) or (ii) higher order relations between attributes such as HO-DINA (Higher Order DINA) and HO-DINO (Higher Order DINO). By doing so, it appears feasible to obtain more meaningful attribute classifications for a parametric CDM. This could lead to a more meaningful comparison of classification accuracy of parametric versus non-parametric diagnostic models even though the agreement indices regarding classification accuracy of the parametric and non-parametric models were found remarkably high.

**References**


Bradshaw, L., & Templin, L. (2014). Combining item response theory and diagnostic classification models: A psychometric model for scaling ability and diagnosing misconceptions. Psychometrika, 79, 403-425.

Chiu, C. Y., & Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, *30*(2), 225-250.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement,35*(1), 8-26. doi:10.1177/0146621610377081

Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, N.J: Lawrence Erlbaum Associates, Publishers (p.291).

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191-210.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation, 15*(3)

R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Robitzsch, A., Kiefer, T., George, A.C., & Ünlü, A. (2014). *CDM: Cognitive Diagnosis Modeling*. R package version 2.7-7. http://CRAN.R-project.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: Interdisciplinary Research and Perspectives, 6*(4), 219-262. doi:10.1080/15366360802490866

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.

Sijtsma, K., & Junker, B. W. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement,25*(3), 258-272. doi:10.1177/01466210122032064

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. *Diagnostic Monitoring of Skill and Knowledge Acquisition*, pages 453–488.

Tatsuoka, K. K. (2009). *Cognitive assessment: an introduction to the rule space method*. CRC Press.

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

Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice, 32*(2), 37-50.

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika, 79*(2), 317-339.

Zheng, Y., Chiu, C.Y. (2015). *NPCD: Nonparametric Methods for Cognitive Diagnosis. R package version 1.0-9.* http://CRAN.R-project.org/package=NPCD