

# Comparison of Cognitive Diagnosis Models under Changing Conditions: DINA, RDINA, HODINA and HORDINA

Ömür K. Kalkan<sup>1</sup>, Hülya Kelecioğlu<sup>2</sup> & Tahsin O. Başokçu<sup>3</sup>

<sup>1</sup> Department of Educational Sciences, University of Pamukkale, Denizli, Turkey

<sup>2</sup> Department of Educational Sciences, University of Hacettepe, Ankara, Turkey

<sup>3</sup> Department of Educational Sciences, University of Ege, İzmir, Turkey

Correspondence: Ömür K. Kalkan, Department of Educational Sciences, University of Pamukkale, Denizli, Kınıklı Campus, 20700, Turkey. Tel: 90-258-296-1000. E-mail: kayakalkan@pau.edu.tr

Received: January 25, 2018

Accepted: March 24, 2018

Online Published: May 29, 2018

doi:10.5539/ies.v11n6p119

URL: <https://doi.org/10.5539/ies.v11n6p119>

## Abstract

The application of CDMs to fraction subtraction data revealed problems on the classification of examinees, latent class sizes, and the use of higher-order models. Additionally, selecting the most appropriate model assumes critical importance if there are several appropriate models available for the data. In the present study, DINA–RDINA and HODINA–HORDINA models were compared under changing conditions (i.e., number of attributes,  $g$  and  $s$  item parameter values, and number of items) with simulated and real data. The results show that for conditions where the  $g$ – $s$  parameter values and the number of attributes were low (0.1 and 3, respectively), the reparameterized models generated values that were virtually identical to those obtained using DINA models. However, when the  $g$ – $s$  parameter values and the number of attributes were increased (0.5 and 5, respectively), the parameter estimations obtained from the models, latent class estimates, AIC, and BIC show differences through the values from the models.

**Keywords:** DINA model, reparameterized DINA model, higher-order DINA model, higher-order reparameterized DINA model, latent class estimates

## 1. Introduction

Recently, there has been a growing interest in Cognitive Diagnosis Models (CDM), which are psychometric models used to determine the strengths and weaknesses of the candidates according to a certain set of attributes (Chen & de la Torre, 2013; Chiu, 2008; de la Torre 2008, 2009, 2011; de la Torre & Lee, 2010, 2013; Huebner & Wang, 2011; von Davier, 2005). CDMs are traditional extensions of traditional item response models (Embretson & Reise, 2000; van der Linden & Hambleton, 1997). CDM based on the multidimensional item response theory (MIRT) model proposed by Tatsuoaka (1983) is an extended form of latent class analysis (LCA) that allows the classification of item response patterns and the analysis of more than one latent variable together (von Davier, 2005). CDMs are also known as multiple classification models or multiple classification latent class models (Maris, 1999), restricted latent class models (Haertel, 1989), cognitive diagnosis models (Nichols, Chipman, & Brennan, 1995) or cognitive psychometric models (Rupp, 2007). In the present study, *deterministic inputs, noisy, “and” gate* (DINA; Haertel, 1989; Junker & Sijtsma, 2001) model, higher-order DINA (HODINA; de la Torre & Douglas, 2004) model, reparameterized DINA (RDINA; DeCarlo, 2011) model, and higher-order RDINA (HORDINA; DeCarlo, 2011) model are discussed.

### 1.1 The DINA and HODINA Models

Although many CDMs are proposed in the literature, the DINA model is highly preferred by researchers because of its easy interpretation, good model-data fit, and as one of the simplest multiple classification models (DeCarlo, 2011, 2012; de la Torre, 2009; de la Torre & Douglas, 2004, 2008; de la Torre & Lee, 2010; Henson, Templin, & Willse, 2009; Huebner & Wang, 2011; Rupp & Templin, 2008).

Similar with most CDMs, the DINA model also requires a  $Q$ -matrix (Tatsuoaka, 1983) configuration with  $j \times k$  ( $j$ . column and  $k$ . row) dichotomous values.  $q_{jk}$  specifies whether the  $k$  attribute is required for the correct answer of item  $j$ . In the DINA model the latent response equation  $\eta_{ij}$ , which is exactly specified by  $\alpha_i$ , determines whether examinee  $i$  possesses the attributes required for the item  $j$ .

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{a_{jk}} \quad (1)$$

If an examinee possesses all the required attributes for item  $j$ ,  $\eta_{ij} = 1$ ; and if at least one of the attributes required for the item is missing,  $\eta_{ij} = 0$  (de la Torre, 2009; de la Torre, Hong, & Deng, 2010; de la Torre & Lee, 2010). The parameters required for correct response to item  $j$  are expressed by  $g_j$  (guess) and  $s_j$  (slip) parameters.

The  $g_j$  parameter is the probability that an examinee who does not possess all the required attributes for item  $j$  will respond correctly to the item. The  $s_j$  parameter is the probability that an examinee who possesses all the required attributes for item  $j$  will answer wrongly (de la Torre & Douglas, 2004, 2008; Huebner & Wang, 2011). The parameters  $s_j$  and  $g_j$  are defined as follows.

$$s_j = P(Y_{ij} = 0 | \eta_{ij} = 1) \text{ ve } g_j = P(Y_{ij} = 1 | \eta_{ij} = 0) \quad (2)$$

The item response function for the item  $j$  can be written as

$$P(Y_{ij} = 1 | \alpha) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad (3)$$

The joint likelihood function of the DINA model can be expressed as follows.

$$L(s, g; \alpha) = \prod_{i=1}^N \prod_{j=1}^J [s_j^{1 - y_{ij}} (1 - s_j)^{y_{ij}}]^{\eta_{ij}} [g_j^{y_{ij}} (1 - g_j)^{1 - y_{ij}}]^{1 - \eta_{ij}} \quad (4)$$

de la Torre and Douglas (2004), by adding an IRT model for joint distribution of attributes,

$$P(\alpha | \theta) = \prod_{k=1}^K P(\alpha_k | \theta) \quad (5)$$

have obtained a HODINA model that assumes the dependence of cognitive attributes on one or more latent traits. Examinees with higher-order  $\theta$  at this point are more likely to possess latent attributes in comparison the examinees with lower level  $\theta$  (DeCarlo, 2011; de la Torre & Douglas, 2004). Thus, the HODINA model can be used to classify examinees over specific attributes and estimate their latent trait (Li, 2008).

### 1.2 The RDINA and HORDINA Models

DeCarlo (2011) obtained the following RDINA model by reparameterized the DINA model as a latent class logistic regression model.

$$\text{logit } p(Y_{ij} = 1 | \eta_{ij}) = f_j + d_j \eta_{ij} \quad (6)$$

In this equation, the items are used to determine the attribute sets. The  $f_j$  parameter gives the logodds of the false alarm, which is the probability of correctly responding to examinee  $j$  that does not have the required attributes. In addition, DeCarlo (2011) obtained the HORDINA model by including the latent continuous variable  $\theta$  into the model for situations where the probability an examinee to possess the attributes is determined by the examinee's latent trait  $\theta$ .

Despite their potential benefits, there are some limitations to the practice of CDMs in education. These limitations include the complexities of the CDM and the choice of the wrong model (de la Torre & Douglas, 2004; de la Torre et al., 2010). Additionally, the application of CDM to fraction subtraction data of Tatsuoka (1990) reveals some problems on the classification of examinees, latent class estimations, and using higher-order models (DeCarlo, 2011). Majority of these problems involve misclassification of  $Q$ -matrix specifications and other specifications of the model, and examinees who get all of the items incorrect are classified as possessing most of the skills by some CDMs. Although the use of higher-order models has been shown to provide a limited ameliorated classification problem, it cannot be precisely specified that this may be due to the misspecification of the  $Q$  matrix (DeCarlo, 2011).

Furthermore, the selection of the most appropriate model among the many appropriate models available in the

CDM literature is unclear. Consequently, if there are a few models that are fitted to the data, the most appropriate model selection becomes a critical issue (de la Torre & Lee, 2013). Simulation studies have shown that while wrong model selection leads to poor attribute classification, correct model selection leads to better estimations (e.g., higher correlation and lower root mean square error [RMSE]) and increase in the correct classification rates of attributes (de la Torre & Douglas, 2004). Therefore, the appropriate model selection, item parameter estimates, and correct classification of attributes in the education decisions to be taken based on the findings obtained from the CDM models have great importance (de la Torre et al., 2010). Another discussion topic that is increasingly important in CDM model literature is information criteria. The general approach in these discussions is that the tests modeled with the same  $Q$  matrices are evaluated under the same model fit statistics (de la Torre & Douglas, 2008).

In the context of these limitations expressed in relation to CDM, there is a need to investigate how the model parameters are affected by the characteristic variations underlying the attribute distribution, and that it is necessary to regularly write and analyze new items in real tests (DeCarlo, 2011; de la Torre et al., 2010; de la Torre & Lee, 2010). Furthermore, it is stated that the effect of variables, such as the number of attributes, ideal test length, slip and guess parameter level, and the structure of attributes, needs to be examined in the classification of examinees (de la Torre et al., 2010; Huebner & Wang, 2011). Furthermore, DeCarlo (2011, 2012) stated that it is necessary to investigate different types of  $Q$  matrix, a large number of attributes, and the relationship between other higher-order factors. de la Torre (2011) stated that it would be useful to conduct studies that examine information criteria (e.g., Akaike's information criterion [AIC]) for multiple comparisons at the item and test levels and fit measures.

For this purpose, in the scope of the research, DINA–RDINA and HODINA–HORDINA models were compared using simulated and real data under changing conditions, such as the number of attributes,  $g$  and  $s$  parameter values, and the number of items to provide more reliable and valid inferences to evaluate the parameters and help to select the most appropriate model. The comparison of the models is aimed at providing a better understanding of the extent to which some of the theoretical features of the model are realized in practice, what the missing aspects are, and how the parameters are affected by the characteristic changes. Although DINA and RDINA models have the same likelihood functions, it is desired to investigate whether models can provide the same estimates in practice.

## 2. Method

### 2.1 Simulation Study

The simulation study, in the case of changes to the number of attributes (3, 4, 5),  $g$ - $s$  parameter values (0.1-0.5), and item numbers (20, 30), aims to investigate whether the  $g$  and  $s$  parameter values, the AIC and BIC, and latent class estimates obtained from DINA-RDINA and HODINA-HORDINA models differ or not. For this purpose, 5 different values were determined for each of the  $g$  and  $s$  parameter values from 0.1 to 0.5, with 0.1 increment. Then, 25 different  $g$ - $s$  combinations meeting these values were obtained. Thus,  $3 \times 25 \times 2 = 150$  conditions (3 attribute number  $\times$  25  $g$ - $s$  values  $\times$  2 item sets) were tested. The sample size was set to 2000 to obtain accurate parameter estimates from the models (de la Torre et al., 2010). Simulation data were obtained with Ox (Doornik, 2002) based on DINA model and it was assumed that the attributes were independent. Additionally, DINA and HODINA models'  $g$ - $s$  parameter and latent class estimates, AIC and BIC information criteria were obtained with Ox. LatentGold (Vermunt & Magidson, 2005) was used for RDINA and HORDINA models analysis. However, latent class estimates are not obtained directly at the output of the LatentGold. Latent class estimates have been obtained using a macro written in Excel, which uses a posterior classification matrix obtained from the LatentGold output. If the posterior mean of an examinee ( $\alpha_{ik}$ ) is equal to or greater than 0.5, it assumed that examinee  $i$  possesses  $k$  attribute; if  $\alpha_{ik}$  is lower than 0.5, then it is assumed that examinee  $i$  has not possessed the attribute  $k$  (DeCarlo 2011; de la Torre & Douglas, 2004; de la Torre et al., 2010). Thus, the percentage of examinee classification assigned to latent classes by the models for the same data set was compared. To evaluate the model-data fit, the two most frequently used information criteria in the statistical literature (Hagenaars & McCutcheon, 2002), AIC proposed by Akaike (1973) and BIC proposed by Schwarz (1978), were preferred. DeCarlo (2011) stated that the RDINA model provides virtually identical values to the item parameter values obtained using the DINA model (de la Torre & Douglas, 2004). Therefore, it is expected that the item parameter ( $g$ ,  $s$ ) estimation differences obtained from DINA-RDINA and HODINA-HORDINA models will be zero. Therefore, the differences of the parameter estimates obtained from the models are shown graphically.

### 2.2 Real Data

The real data set are obtained from the English grammar test applied to 565 examinees at Ege University, School of

Foreign Languages. The test was developed according to the CDM. The  $Q$  matrix consisted of 5 attributes and 24 items that were determined by 3 experts.

### 3. Results

#### 3.1 Simulation Study Results

##### 3.1.1 Item Parameters

Findings are given below when the number of attributes is 3, the number of items is 20, and the item parameters  $g$  and  $s$  are changed in 0.1 and 0.5 range.  $g$  and  $s$  parameter values, for example  $g = 0.1, s = 0.1$ , were abbreviated and represented as  $g1s1$ . The  $g$  parameter estimated differences obtained from the DINA and RDINA models are given in Figure 1, where they are shown pointwise on the axis for 20 items in each data set.

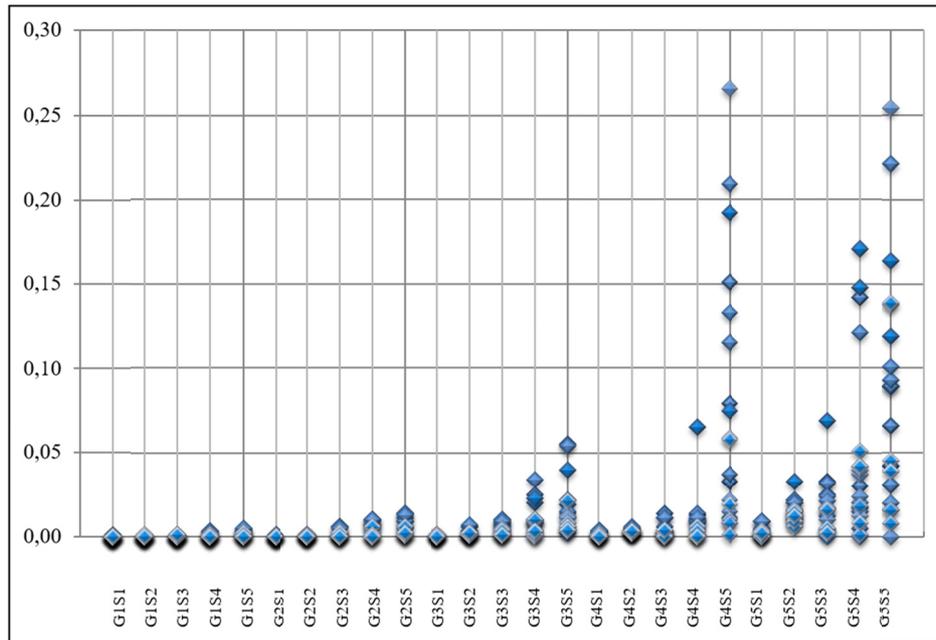


Figure 1. DINA and RDINA Models  $g$  parameter differences

The  $g$  parameter estimates of the two models are virtually identical when the  $g$  and  $s$  parameters are low. When the sum of the  $g$  and  $s$  parameters is 0.4 or less ( $g1s1, g1s2, g1s3, g2s1, g2s2, g3s1$ ), the average of the differences between the  $g$  parameter estimates varies from 0.0001 to 0.0007. If the sum of the  $g$ - $s$  parameters exceeds 0.8 ( $g4s5, g5s4, g5s5$ ), the average of the differences between the  $g$  parameter estimates varies between 0.0472-0.0845. For the 25 data sets, the average of the  $g$  parameter estimation differences obtained from both models is 0.0123.

Similarly, when the  $g$  and  $s$  parameters in HODINA and HORDINA models are low, the two models yielded virtually identical  $g$  parameter estimates (not shown here). For example, the average of the estimates of  $g$  parameter differences when the  $g$  and  $s$  parameters are  $g1s1, g1s2, g2s1$ , and  $g5s1$  vary between 0.0006-0.001. When the sum of the  $g$  and  $s$  parameters exceeds 0.8 ( $g4s5, g5s4, g5s5$ ), the average of the differences between the  $g$  parameter estimates varied from 0.0355 to 0.0528. For the 25 data sets, the average of the  $g$  parameter estimation differences obtained from both models is 0.0099. The differences of  $s$  parameters obtained from DINA and RDINA models are given in Figure 2.

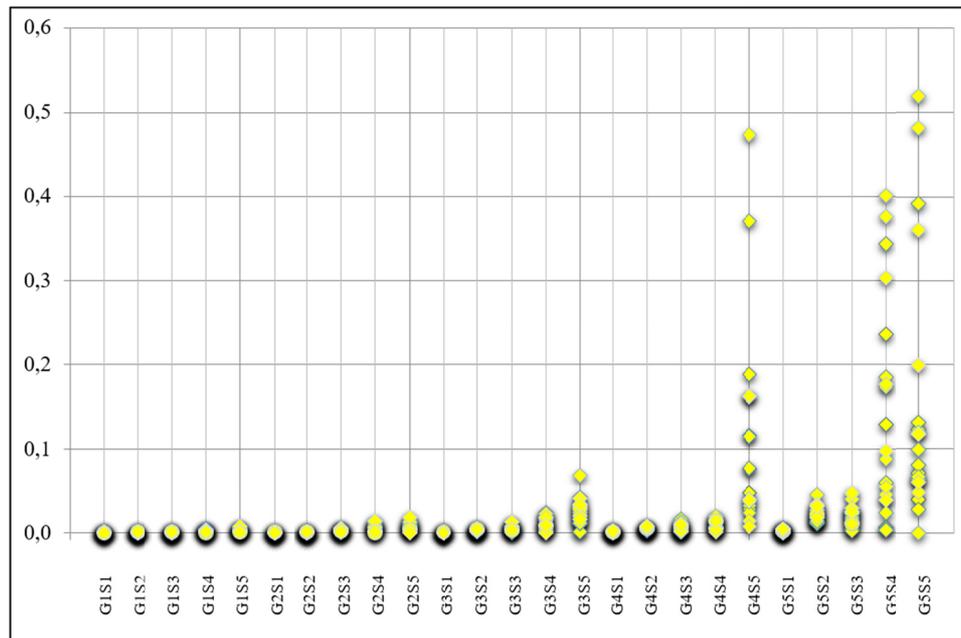


Figure 2. DINA and RDINA Models  $s$  parameter differences

Figure 2 shows that the item parameter estimates of the two models are getting closer to the condition that the  $g$  and  $s$  parameters are low. The average of the difference between the  $s$  parameter estimates when the sum of  $g$  and  $s$  is equal 0.4 or less varies in the range of 0.0002-0.0007. For the 25 data sets, the average of the differences of  $s$  parameter estimates obtained from both models is 0.0216. HODINA and HORDINA models provided virtually identical  $s$  parameter estimates (not shown here) when the  $g$  and  $s$  parameters are low ( $g1s1$ ,  $g2s1$ ). When the sum of  $g$  and  $s$  is equal to 0.8 and higher, the average of the differences between the  $s$  parameter estimates varied from 0.0226 to 0.0517. For the 25 data sets, the average of the differences of the  $s$  parameter estimates obtained from both models is 0.0131. Consequently, when the  $g$  and  $s$  values are low, the models yielded very close parameter estimates; when the values of  $g$  and  $s$  increase, the parameter values obtained from the models also differentiated.

### 3.1.2 Latent Class Estimates

Latent class estimates obtained from DINA-RDINA models are given in Figure 3.

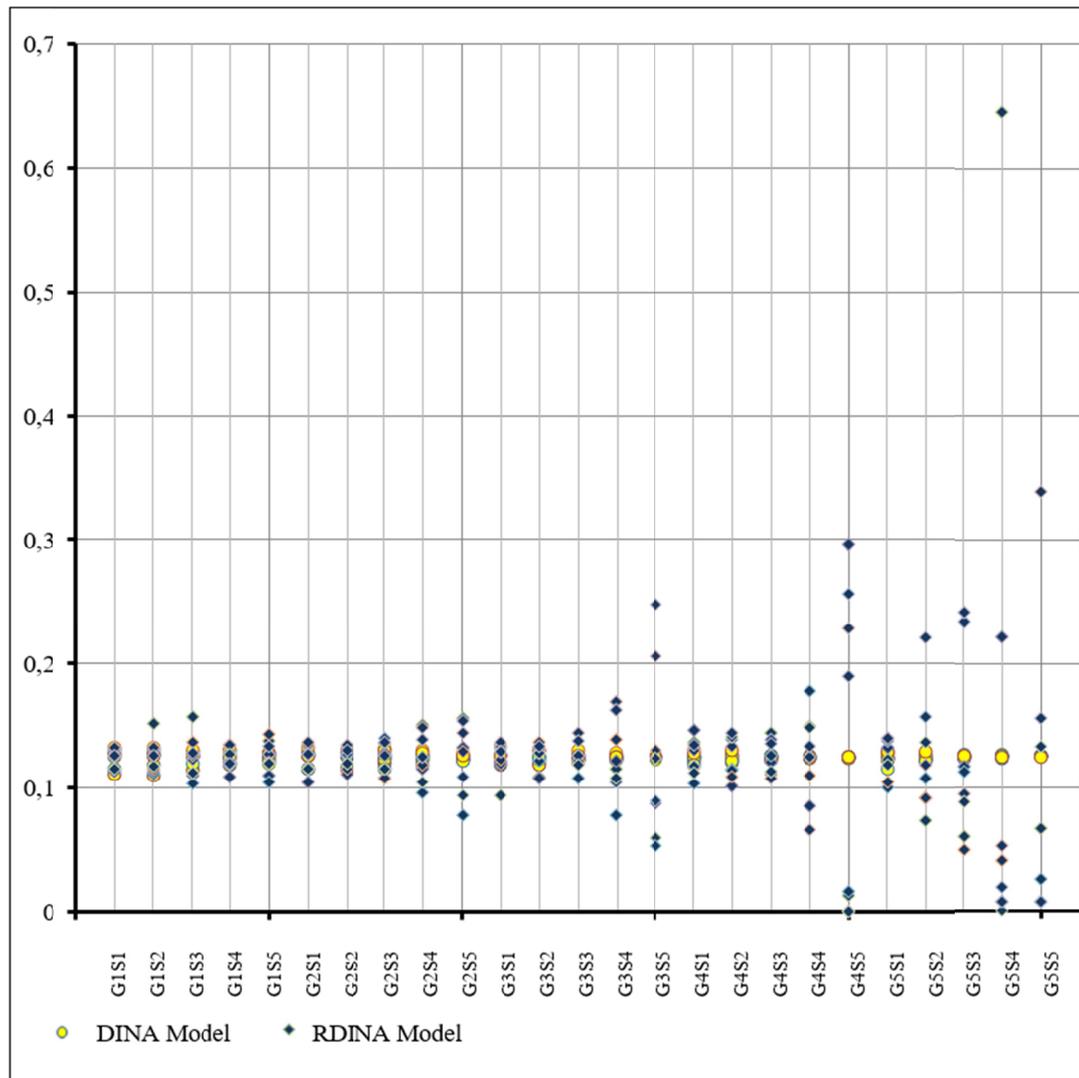


Figure 3. DINA and RDINA Models latent class estimates

Figure 3 shows 8 latent class estimates obtained from two models in terms of 25 data sets and 3 attributes pointwise on the axis. The latent class estimates of both models change in the range of 0.1-0.15 when the sum of the  $g$ - $s$  parameters is equal to 0.5 and lower. However, the differences are observed in the latent class estimates when the sum of the  $g$ - $s$  parameters is 0.8 and higher. Latent class estimates obtained from the HODINA-HORDINA models vary by 0.1-0.15 when the sum of the  $g$ - $s$  parameters is 0.6 or lower in total (not shown here). However, when the sum of the  $g$ - $s$  parameters is 0.8 and higher, the differences are observed between the latent class estimates obtained from higher-order models.

### 3.1.3 AIC and BIC

For the 25 data sets, AIC and BIC were obtained from the models. According to the AIC results, in all data sets (number of attribute 3, number of items 20), the RDINA model provided better fit values compared with the DINA model. In higher-order models, the HODINA model provided better fit values compared with the HORDINA model in 21 of the 25 conditions. The results of BIC show that considering all the conditions (number of attribute 3, number of items 20), RDINA and HODINA models provide better fit values compared with the DINA and HORDINA models, respectively. Additionally, while the number of attributes in the simulation studies is 3, the DINA model and the reparameterized models provide fairly close AIC and BIC, but these values are differentiated when the number of attributes is increased.

### 3.2 Real Data Results

#### 3.2.1 Item Parameters

The  $g$  and  $s$  parameters and latent class estimates and the AIC and BIC obtained from the models in the real data set are given below. The  $g$  parameter estimates obtained from the DINA and RDINA models for the 24 items are given in Figure 4.

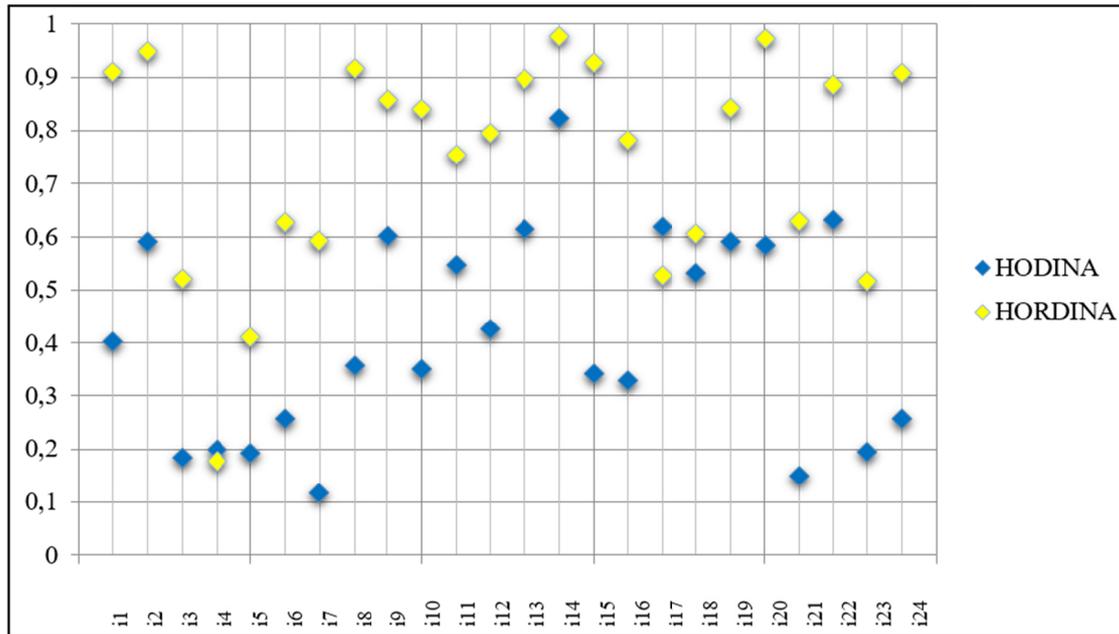


Figure 4. DINA and RDINA models  $g$  parameter estimates

Figure 4 shows that the differences in the  $g$  parameter estimation obtained from both models are 0.008 and 0.011 for items 18, and 22, respectively. There are differences between the  $g$  parameter values obtained for the other items. The average of the  $g$  parameter estimation differences for 24 items is 0.19. The  $g$  parameter estimation differences obtained from the HODINA and HORDINA models are 0.024 and 0.073 for items 4 and 18, respectively (not shown here). There are significant differences between the item parameters obtained for other items. The average of the  $g$  parameter estimation differences obtained with two models from 24 items is 0.34. The  $s$  parameter estimates obtained from the DINA and RDINA models are given in Figure 5.

Figure 5 shows that the  $s$  parameter estimation difference obtained from both models is 0.023 for item 4, 0.02 for item 14, 0.003 for item 17, and 0.019 for item 20. There are significant differences in the  $s$  parameters obtained for other items. For all items, the average of the differences of the  $s$  parameter estimates obtained from the two models is 0.12. The differences in the  $s$  parameter estimates obtained from the HODINA and HORDINA models are 0.003 and 0.025 for items 4 and 17, respectively.

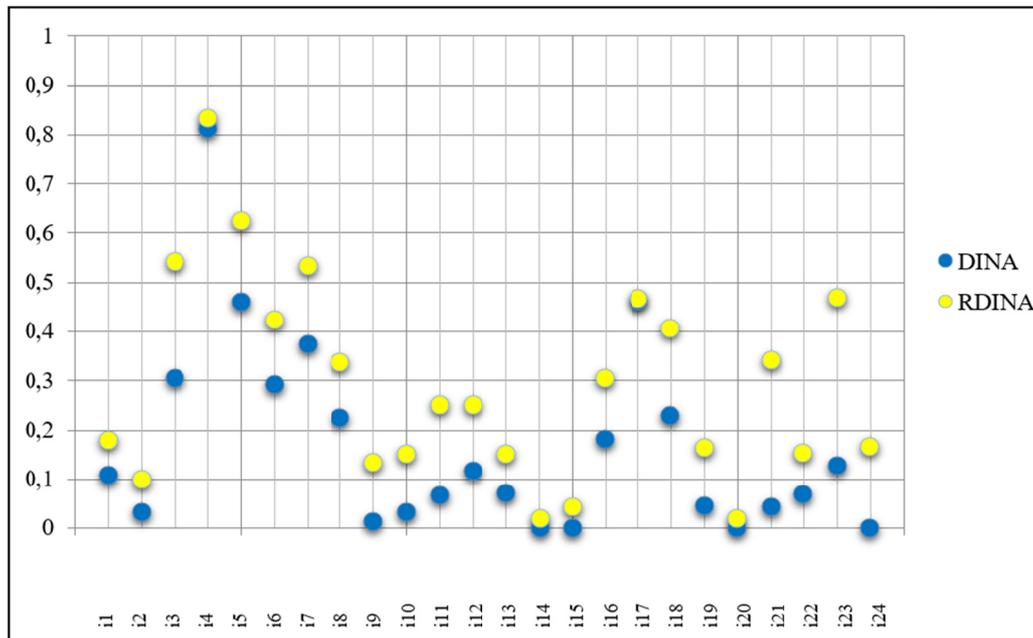


Figure 5. DINA and RDINA models  $s$  parameter estimates

There are also differences between the  $s$  parameters obtained for other items. The average of the differences of the  $s$  parameter estimates obtained from the two models for all items is 0.25.

### 3.2.2 Latent Class Estimates

Latent class estimates of DINA-RDINA models are given in Figure 6, which shows the 32 latent class estimates obtained from the two models with respect to the 5 attributes in the  $Q$  matrix of the real data set. Figure 6 shows that the size of 22 latent classes by the RDINA model is estimated as zero. The difference between the latent class estimates obtained by the two models is changing between 2-58%.

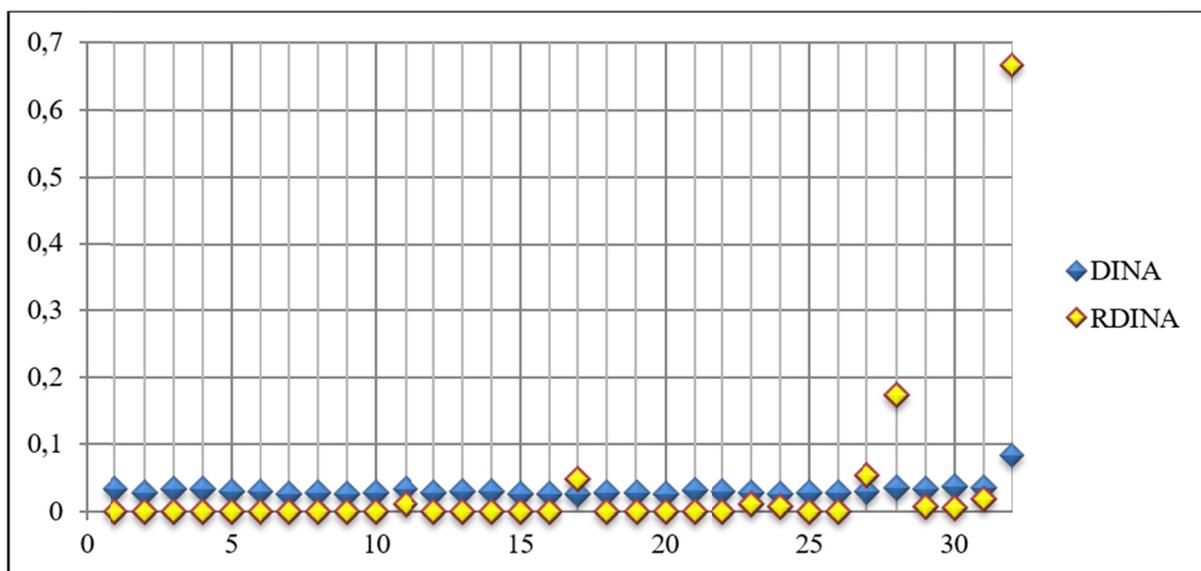


Figure 6. DINA-RDINA models latent class estimates

The HORDINA model obtained 16 estimates of the latent class as zero (not shown here). Differences between the latent class estimates obtained by the HODINA and HORDINA models range between 0.01 and 34%.

### 3.2.3 AIC and BIC

The AIC results show that the RDINA (15,343) and HODINA models (15,220) provide better fit values compared with the DINA (15,631) and HORDINA models (15,251), respectively. The BIC results also show that the RDINA (15,573) and HODINA models (15,454) provide better fit values compared with the DINA (15,973) and the HORDINA models (15,502), respectively.

## 4. Discussion

### 4.1 *g* and *s* Parameter Estimates

While the joint distribution of skills is based on a multinomial distribution in the DINA model, the HODINA model is based on a higher-order latent proficiency. The DINA model estimates are based on the mode (maximum), whereas the HODINA model estimates are based on the mean (expected value). While the HODINA model uses Markov Chain Monte Carlo (MCMC) algorithm for item parameter estimates, the DINA model uses the Expectation-Maximization (EM) algorithm. Parameter estimates for MCMC and standard errors are obtained by calculating posterior means and standard deviations. Although it is expected that algorithms may give different results due to these important differences, DINA and HODINA models provide considerably similar estimates, which indicate that EM and MCMC algorithms can be used to obtain accurate parameter estimates (de Torre & Douglas, 2004; de la Torre, 2009).

In all analyses performed using the simulated data, the item parameter estimates obtained from the models provided virtually identical values when the number of attributes and the parameter values of *g-s* were low. The findings obtained with respect to the low *g-s* parameter values were consistent with the findings of de la Torre et al. (2010) and Huebner and Wang (2011). However, in conditions (*g4s5*, *g5s4*, *g5s5*) where the number of attributes and item parameter values (particularly sum of *g* and *s* parameters is equal or higher 0.8) were increased, *g* and *s* parameter estimates obtained from the models showed differences. Furthermore, in real data analysis, the differences between the item parameter estimates obtained from the DINA-RDINA and HODINA-HORDINA models were also remarkably high. Similarly, de la Torre and Lee (2010) stated that although the invariant property of the DINA model parameters is provided in the simulated data, there were inconsistencies in the item parameter estimates in the real data; therefore, this condition was not fully provided. However, this should not be seen as a reason to downplay the practical usefulness of the DINA model (de la Torre & Lee, 2010).

### 4.2 Latent Class Estimates

The attribute numbers in simulation data are 3, 4, and 5, and 5 in the real data set. In the simulated data, 50 data sets are analyzed for each attribute. The results show that the latent class estimates obtained from the DINA-RDINA and HODINA-HORDINA models are virtually identical when the sum of the *g-s* parameter values is equal 0.5 and lower, while significant differences are found when they were 0.8 and higher. There are also significant differences in latent class estimates obtained from the real data set. Similar classification problems are also expressed by DeCarlo (2011). In addition, de la Torre et al. (2010) and Huebner and Wang (2011) are also stated that the accuracy of the classification has increased with low level *g* and *s* parameters. Furthermore, many latent classes are estimated to be zero by the reparameterized models when the number of attributes and *g-s* parameter values increased. Consequently, for all conditions in the simulated data, the DINA models provided more consistent latent class estimates in comparison to the reparameterized models. In addition, many latent classes are estimated to be zero by the reparameterized models in the analysis results carried out with the real data set.

### 4.3 AIC and BIC

AIC and BIC fit statistics were used to evaluate the model data fit. de la Torre and Douglas (2004) stated that the choice of appropriate model becomes a critical issue when the number of attributes increased. While the number of attributes in the simulation studies is 3, the DINA model and the reparameterized models provide fairly close AIC and BIC, but the values are differentiated when the number of attributes is increased (despite very low *g* and *s* parameter values).

When all the given conditions are considered for simulated data, the AIC results show that the RDINA model provided better fit values compared with the DINA model in 147 out of 150 conditions. The BIC results show that the RDINA model in all 150 conditions provided better fit values compared with the DINA model. The AIC and BIC in the real data analysis show that the RDINA model also provided better fit values compared with the DINA model. The results of AIC in higher-order models show that the HODINA model provided better fit values in 125 out of 150 conditions than the HORDINA model. The BIC results show that the HODINA model provided better fit values in 149 of the 150 conditions than the HORDINA model. The results of AIC and BIC in the real data analysis show that the HODINA model provided better fit values than the HORDINA model. Furthermore, when

the models are handled according to their structure, the AIC and BIC values for all simulated conditions of the HODINA model provided better fit values compared with the DINA model. de la Torre and Douglas (2004) also stated that the higher-order model provides better fit values than the basic model. In reparameterized models, the results of BIC show that the RDINA model provided better fit values than the HODINA model in all conditions. The results of AIC show that in 144 of the 150 conditions, the RDINA model provided better fit values than the HODINA model. Similar results have been reported by DeCarlo (2011). When all conditions of the research are considered, the BIC provides more consistent results compared with the AIC. Although a large number of studies reporting the superiority of one another in both criteria are found in literature, the findings of the present research are consistent with those results reported by Jedidi et al. (1997), Kuha (2004), Li et al. (2009), McQuarrie and Tsai (1998), Nylund et al. (2007), Tofighi and Enders (2007), Yang (2006).

The results of AIC and BIC determined the most parsimonious model. Nevertheless, the decisions to be made on the  $Q$ -matrix should not be given based only on the information criteria, but the fit statistics, validity studies, and other evidence are also needed (DeCarlo, 2011). The higher  $g$  and  $s$  parameters obtained from CDMs can be considered as empirical evidence of the misspecification of the  $Q$  matrix (Rupp & Templin, 2008). In addition, the  $Q$  matrix structure has a potential to influence item parameter estimates and misclassification (de la Torre et al., 2010). de la Torre and Douglas (2004) stated that the appropriate model choice provides greater consistency in the correct classification rates of attributes and thus provides better estimates (e.g., higher correlation and lower RMSE). Consequently, attention should be given to the selection of the appropriate model and the correct determination of the  $Q$  matrix in the item parameter and latent class estimates obtained from the CDMs.

The findings of the present research revealed that the model-data fit is highly related with the item and latent class estimates. In addition, the decrease of the model-data fit leads to the differentiation of all estimates independent of the preferred mode. This emphasizes the importance of a priori analysis (design of  $Q$  matrix, correct specification of item and attribute relation) for CDM studies (de la Torre et al., 2010; Huebner & Wang, 2011; DeCarlo, 2011, 2012). The most important variables of model-data fit are  $Q$  matrix validity and item quality. In this case, the comparison studies for the models should be done considering the  $Q$  matrix compatibility. The work to be carried out for the situations provided by these conditions will give more realistic results for model comparisons. However, it is difficult to achieve a perfect match between the  $Q$  matrix properties and the latent class structure (de la Torre et al., 2010).

#### 4.4 Item and Attribute Numbers

The number of items used in simulation data studies is 20 and 30. There are 75 different conditions for each item set. The increase in the number of items leads to an increase in the AIC and BIC values. This result is expected because of the increased number of estimated parameters. The increase in the number of items in terms of AIC causes a difference of 3 in 75 cases where the DINA-RINA models are compared and 4 cases in the HODINA-HORDINA model comparisons. The increase in the number of items in terms of BIC does not lead to any difference in the 75 conditions compared with the values obtained from the DINA-RINA models, whereas in the HODINA-HORDINA model comparisons, the result is different only in one condition. Consequently, there was no evidence of a significant effect of the increase in the number of items covered in the study on the conditions where the models were compared to the AIC and BIC.

Similarly, when the number of items is fixed (e.g., 20 and 30), the AIC and BIC show no significant finding of the effect of the increase in attribute numbers (3, 4, 5) on the comparison of models. However, the increase in the number of attributes has led to an increase in the  $g$  and  $s$  parameter differences (DINA-RDINA, HODINA-HORDINA) obtained from the models and differentiation in latent class estimates. Similar results have been reported by Chiu (2008) and de la Torre et al. (2010).

### 5. Conclusion

Within the scope of the research, the DINA-RDINA and HODINA-HORDINA models were compared under varying conditions, such as the number of attributes in the  $Q$  matrix,  $g$  and  $s$  parameter values, and the number of items. In all analyses performed with the simulated data, the conditions where the number of attributes and  $g$ - $s$  parameter values is low, the item parameter and latent class estimates obtained from the models are virtually identical. However, the  $g$ - $s$  parameter and latent class estimates obtained from the models are differentiated by increasing the number of attributes and item parameter values. In addition, many latent classes are estimated to be zero by the reparameterized model under conditions of increasing attribute number and item parameter values. In all conditions of the simulated data and the real data set, the DINA models provided more consistent latent class estimation values compared with the reparameterized models. The AIC and BIC obtained from the simulated and the real data show that the RDINA and HODINA models provided better fit values compared with the DINA and

HORDINA models, respectively. When all conditions in the research are considered, the BIC provides more consistent results compared with the AIC. In cases where the models are compared based on the AIC and BIC, no significant effect is observed of the increase in item and attribute numbers. However, the increase in the number of attributes leads to an increase in the  $g$  and  $s$  parameter differences and differentiation in latent class sizes.

Consequently, for the lower number of attributes and  $g$ - $s$  parameter values for the simulated data in all conditions, the reparameterized models (RDINA, HORDINA) proposed by DeCarlo (2011), and DINA models (DINA, HODINA) yielded virtually identical  $g$  and  $s$  parameter estimates, latent class estimates, AIC and BIC; whereas, with the increase in the number of attributes and item parameter values, all parameters are differentiated. In the real data set, all values obtained are different.

### Acknowledgments

We would like to express our great appreciation to Hacettepe Technopolis Technology Transfer Center for their substantial contributions to the whole project.

### Compliance with ethical standards

### Funding

This work was supported by Hacettepe University Scientific Research Projects Coordination Unit under code SDK-2015-5395.

### Conflicts of Interest

The authors declare that they have no conflict of interest.

### References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In V. Petrov, & F. Csáki (Eds.), *Proc. 2nd International Symposium on Information Theory* (pp. 267-281).
- Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement, 37*(6), 419-437. <https://doi.org/10.1177/0146621613479818>
- Chiu, C. (2008). *Cluster analysis for cognitive diagnosis: theory and applications* (Doctoral dissertation). Available from ProQuest Dissertation & Theses: Full Text (NR3337778).
- de la Torre, J. (2008). An empirically based method of Q-Matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343-362. <https://doi.org/10.1111/j.1745-3984.2008.00069.x>
- de la Torre, J. (2009). DINA model and parameter estimation: a didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115-130. <https://doi.org/10.3102/1076998607309474>
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 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*(3), 333-353. <https://doi.org/10.1007/BF02295640>
- de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika, 73*(4), 595-624. <https://doi.org/10.1007/s11336-008-9063-2>
- de la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement, 47*(1), 115-127. <https://doi.org/10.1111/j.1745-3984.2009.00102.x>
- 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*(2), 227-249. <https://doi.org/10.1111/j.1745-3984.2010.00110.x>
- de la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item - level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement, 50*(4), 355-373. <https://doi.org/10.1111/jedm.12022>
- 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. <https://doi.org/10.1177/0146621610377081>
- Decarlo, L. T. (2012). Recognizing uncertainty in the Q-Matrix via a Bayesian extension of the DINA model.

- Applied Psychological Measurement*, 36(6), 447-468. <https://doi.org/10.1177/0146621612449069>
- Doornik, J. A. (2002). *Object-oriented matrix programming using Ox, (Version 3.1)*. London: Timberlake Consultants Press.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 301-321. <https://doi.org/10.1111/j.1745-3984.1989.tb00336.x>
- Hagenaars, J., & McCutcheon, A. (2002). *Applied latent class analysis models*. New York: Cambridge University Press.
- 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. <https://doi.org/10.1007/s11336-008-9089-5>
- Huang, H. Y., & Wang, W. C. (2014). The random-effect DINA model. *Journal of Educational Measurement*, 51(1), 75-97. <https://doi.org/10.1111/jedm.12035>
- Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, 71(2), 407-419. <https://doi.org/10.1177/0013164410388832>
- Jedidi, K., Jagpal, H., & Desarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39-59. <https://doi.org/10.1287/mksc.16.1.39>
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272. <http://dx.doi.org/10.1177/01466210122032064>
- Kuha, J. (2004). AIC and BIC comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2), 188-229. <https://doi.org/10.1177/0049124103262065>
- Li, F. (2008). *A modified higher-order DINA model for detecting differential item functioning and differential attribute functioning* (Doctoral dissertation). Available from ProQuest Dissertation & Theses: Full Text (NR200812).
- Li, F., Cohen, A. S., Kim, S. H., & Cho, S. J. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement*, 33(5), 353-373. <https://doi.org/10.1177/0146621608326422>
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2), 187-212. <https://doi.org/10.1007/BF02294535>
- Mcquarrie, A. D., & Tsai, C. L. (1998). *Regression and time series model selection*. Singapore: World Scientific.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum.
- Nylund, K. L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- Rupp, A. A. (2007). The answer is in the question: a guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing*, 7(2), 95-125. <https://doi.org/10.1080/15305050701193454>
- Rupp, A. A., & Templin, J. (2008). The effects of Q-Matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68(1), 78-96. <https://doi.org/10.1177/0013164407301545>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464. <https://doi.org/10.1214/aos/1176344136>
- 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>
- Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge*

- acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.
- Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement, 32*(7), 559-574. <https://doi.org/10.1177/0146621607300286>
- Tofghi, D., & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock, & C. T. Greenwich (Eds.), *Mixture models in latent variable research* (pp. 317-341).
- Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York, NY: Springer Science+Business Media.
- Vermunt, J. K., & Magidson, J. (2005). *Technical guide for Latent GOLD 4.0: basic and advanced*. Belmont Massachusetts: Statistical Innovations Inc.
- Von davier, M. (2005). *A general diagnostic model applied to language testing data (ETS Research Report RR-05-16)*. Princeton, NJ: Educational Testing Service.
- Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*, 1090-1104. <https://doi.org/10.1016/j.csda.2004.11.004>

### Note

Note 1. This article is an extension of corresponding author's doctoral thesis and a part of this study was presented at the LXXXI international meeting of the Psychometric Society (IMPS, 2016) held on 12-15 July 2016 in Asheville, NC, USA.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).