Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.5, No.6, 2014

Classification Accuracy Effects of Q-Matrix Validation and Sample Size in DINA and G-DINA Models

Tahsin Oğuz Başokçu [1*]

1. Department of Assessment and Evaluation in Education, Ege University, 35040, İzmir, Turkey

* tahsin.oguz.basokcu@ege.edu.tr

Abstract

This article studies the extend of change in latent classes, relating to students, which were calculated using DINA and Generalized-DINA(G-DINA) Models under different distributions and sub-sample sizes which were calculated using DINA and Generalized-DINA(G-DINA) Models. Main focus of this study is the results of practical application rather than statistical structure of Cognitive Diagnostic Models (CDM). The attribute the individuals master that take the test in CDM are determined categorically. For this reason, both the fit of Q matrix with data and the effect of sample size are searched in modelling the students' category. In the case of low model data fit and inadequate sample size, the findings of this research will be a guide in how the decisions change about which attribute a student master or not. To this end, a mathematic test consisted of 18 multiple choice questions taken by a group of 1000 examinee was employed. Analyses were carried out using 5 different Q-Matrices, for which relations between test items and attributes were determined by experts, and latent classes determined by both DINA and G-DINA models were compared. Comparisons were made with a view to accuracy of values between classes associated with examinees in different sample sizes drawn from the same population and values obtained for population. Thus, for both models, whether they lead to independent results from the samples was tested for sample sizes of 30, 50, 100, 200 and 400 and effects of Q matrix- data fit on analysis results were determined. Results of analysis showed Q-matrix – data fit had significant impact on decisions about students for both models.

Keywords: Cognitive Diagnostic Models, DINA model, G-DINA model, Q Matrix

1. Introduction

Cognitive Diagnostic Models (CDM) have received ever increasing attention after "No Child Left Behind" Act of 2001 in USA. Main objective of this approach is to provide cognitive feedback about students to students, teachers, and families (Embretson, 1991, 1998).

CDM is based mainly on latent class analysis. Latent class analysis is a statistical method determining sub-groups using multivariable categorical data and utilizing interrelations (Cheng 2010). CDM was developed to measure specific knowledge structures or skills the student mastered, and provides information about their cognitive strengths and weakness (Leighton and Gierl, 2007).

CDM is designed to discriminate students according to latent classes based on two parameter attributes. Attributes, represented by latent variables constituting a vector determining expertise, define skill set underlying diagnostics for students. Latent variables denoted as "attributes" here, may be defined as traits, competencies, task, sub-task, cognitive process or skill (Tatsuoka, 1995a).

In a test developed by CDM, instead of total score or subscale scores, which skills, each individual taking the test, mastered and which non-mastered were analyzed. In this regard, results of CDM analysis will not only allow assessment process but also help in determining education needs for each student (de la Torre, 2009a).

CDM is more convenient for cases the test measures more than one interrelated structures. Each item in test is designed to measure those structures or cognitive components. Each item in CDM may measure more than one aspect to be assessed by the test (Rupp & Templin, 2008). CDM determines student's performance on each cognitive item instead of focusing on students' skills levels in latent scale. Probabilities thus obtained may be transformed so as to profile the skills student mastered. In CDM, items are matched with attributes to be measured by Q-Matrix (K. Tatsuoka, 1985).

In Q matrix, each column is a vector of attribute or skill and each row represents an item. Attributes are traits, procedures, method of discovery, strategies, skills and other cognitive components determined by experts of the field (Embretson, 1984). Q matrix will show whether any attributes exist for item by binary coding using 1-0 (Tatsuoka, 1990). This coding approach initially was denoted as "weighting" by Fisher (1973) and if attribute *k* exists for item i, it was coded as 1 and if does not exist it takes 0.

Q matrix with 3 attributes and 5 items are show in Table 1.1. As it was shown in Q matrix it requires the first attribute to correctly answer the 1st item. For the second item, students must have attributes 1 and 2.

Table 1.
Q Matrix Example

items	α_1	α_2	α_3
1	1	0	0
2	1	1	0
3	1	1	1
4	0	1	1
5	0	0	1

Q matrix above indicates α_1 attribute is required for correct answering of item 1 and both α_1 and α_2 attributes are required for correct answering of item 2. 2^k latent classes are defined for k attributes in Q matrix. As a result of analysis based on Q matrix example, respondents shall be placed in 2^3 latent classes.

Latent classes are determined as (000), (100), (010), (001), (110), (101), (011) and (111). Latent classes exactly represent which attributes student mastered and which non-mastered. A student classified under (000) in latent classes defined in the above example possesses no attributes. Similarly under (100) latent class only students having attribute α_1 were assigned whereas under (011) class students who have no α_1 attributes but possess α_2 and α_3 were classified.

This explains diagnostic function of CDM. In this respect, test analyzed with CDM will not only allow assessment process but also help determine each students' educational needs (Cheng & Chang, 2007; Huebner, 2010). De la Torre and Douglas (2004) maintained that specification of Q matrix directly determines prediction of students' skills profile.

Several different CDMs were developed in recent years. Reviews by Junker (1999), diBello, Roussos,and Stout (2007), Roussos et al. (2007), and Rupp et al. (2008) provide exhaustive reviews of different CDMs and their statistical qualities. non-compensatory Deterministic inputs noisy and-gate (DINA) and Noisy-input deterministic-and-gate (NIDA) models (e.g. Macready et al., 1977, Haertel, 1989; Junker & Sijtsma, 2001; de la Torre, 2009b)are mostly used. Deterministic inputs noisy or-gate (DINO) and Noisy-input deterministic-or-gate (NIDO) models (e.g. Maris, 1999). Hartz (2002) developed the Fusion model which is also known as reduced re-parametrized unified model (RUM) (Roussos et al., 2007).

Some frameworks such as, the flexible family of general diagnostic models (GDM, van Davier, 2005, 2007), the generalized DINA Model (de la Torre, 2008), as well as the log-linear framework for CDMs (Henson, Templin, & Willse, 2009).were proposed for CDMs. Additionally, some extensions of CDMs for multiple-choice items were proposed in the literature of CDM (e.g. Bolt & Fu, 2004 for the fusion model; de la Torre, 2009b for the DINA model).

1.1. DINA model

DINA model was developed by Haertel (1989). DINA model (deterministic inputs, noisy "and" gate) is one of models developed for cognitive diagnostics (Junker & Sijtsma, 2001). This model is a latent class analysis similar to binary skills model. DINA model is closely related to Item Response Theory (IRT) (Haertel, 1989). Nevertheless DINA model, differing from IRT models, does not assume continuous distribution of different skill sizes of students. Instead, students are dichotomously assigned to small number of latent classes. DINA model classifies respondents into two dimensional classes for each attributes. First class is "Non-Mastery", namely class of respondents lacking specified trait, and the other is "Mastery", namely class of respondents possessing the specified trait. As is understood, DINA model does not define attributes that students possessed as continuous variables but as a categorical variable.

DINA model can be simply defined as follows: Let X_{ij} denotes response of respondent i to item j, and i= 1,…,I and j= 1,…, J. Denote respondent's binary attributes vector as α_i={ α_{ik} }, for k= 1,…,K when respondent's k^{th} entry is 1 it will denote k^{th} attribute possessed and when it is 0, not possessed (de la Torre 2009a).

Item Response function for DINA model is given by:

$$P(\alpha_{ij}^*) = \begin{cases} g_j & \text{if } \alpha_{ij}^* \prec 1_{K_j^*} \\ 1 - s_j & \text{otherwise} \end{cases}$$

Where 1K is a vector of K_j^* g_j is the probability that individuals who lack at least one of the prescribed attributes for item j will guess correctly, and $1-s_j$ is the probability that individuals who have all the required attributed will not slip and get the item wrong (de la Torre 2011).

Main difference of DINA model from other CDMs, is classifying a respondent under Non-Mastery class even if respondent does not posses only one attributes prescribed for an item. In other words, only the respondent who mastered all attributes to answer correctly has correct answering probability near to 1. Function of correct answering probability of an individual who possess all attributes is given by:

$$P\left[Y_{ij} = 1\middle|\eta_{ij}, s_j, g_j\right] = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1-\eta_{ij}}$$

Where P is the probability of a student who possess all prescribed attributes to answer correctly η_{ij} is latent answer, determined by α, and attribute for item i and a vector of qj. Row of item j in Q matrix can be shown as:

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

Main advantage of DINA model over other CDMs, is both application and interpretation processes involved lower level of complexity. Yet, De la Torre and Douglas (2004, 2008) showed DINA model achieved higher model-data fit and facilitated easy adaptation to different strategies with some modifications. Besides, de la Torre maintained that there is a high level similarity between DINA model results and new and more complex CDMs' results (de la Torre, 2008b, 2009b; de la Torre & Liu, 2008; de la Torre, 2013). As well, the studies on dina model applications have proceeded for ex. Differential item functioning in DINA (Feming li, 2008; Hou 2013) and natural network with CDM (Lamb, Annetta, Vallett, & Sadler, 2014; Shu, Henson, & Willse, 2013) etc. All those accounts pinpoint outstanding importance of DINA model among all other CDMs and constituted main starting point in selecting DINA model in this study.

1.2.G- DINA model

G-DINA model is a generalization of the DINA model with more relaxed assumptions. As many cognitive diagnostic models, this model is also based on *JxK* Q matrix. G-DINA model discriminates latent classes into $2^{K_j^*}$ latent groups. Each latent group is reduced to a skill vector represented by$\propto_{lj}^*$. Each latent group has probability of correct answering represented by $P\left(\propto_{lj}^*\right)$ (de la Torre, 2011).

The original formulation of the G-DINA model based on $P\left(\propto_{lj}^*\right)$can be decomposed into the sum of the effects due the presence of specific attributes and their interactions. Probability formula for G-DINA model is given by:

$$P\left(\alpha_{lj}^*\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\ \alpha_{lk}\alpha_{lk'} \ldots. + \delta_{j12\ldots.K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

δ_{j0} = is the intercept for item j

δ_{jk} = is the main effect due to αk

$\delta_{jkk'}$ = is the interaction effect due to αk and αk

$\delta_{j12\ldots.K_j^*}$ = is the interaction effect due to α1,...,.,$\alpha_{K_j^*}$

Estimation code of G-DINA is an implementation of EM algorithm. In analysis procedure, first $P\left(\propto_{lj}^*\right)$ values with standard errors are calculated, then posterior probabilities of skills are determined and latent classes of students and goodness of fit statistics for item and test are calculated according to those probabilities.

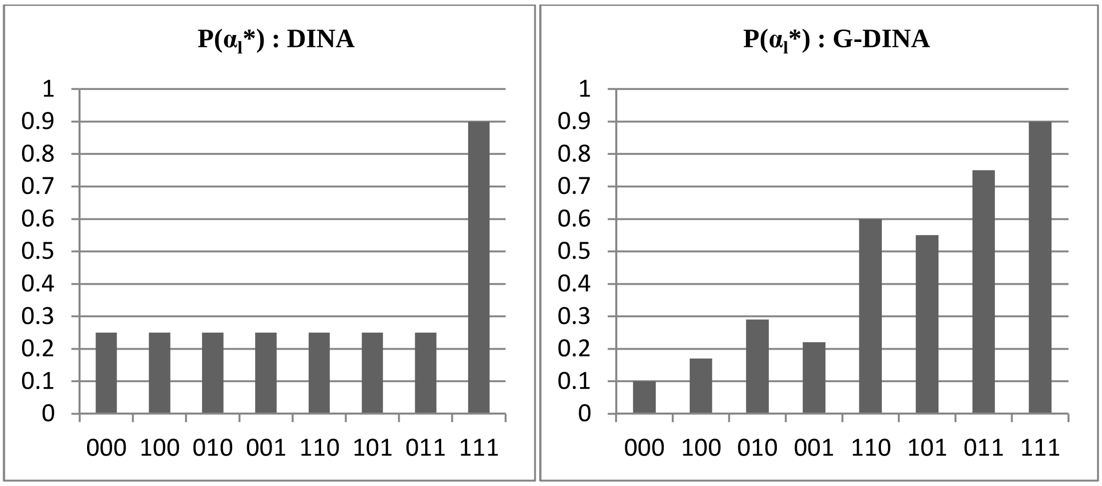

Figure 1

Above Figure 1 shows distribution of success probabilities against levels of attributes mastered for DINA and G-DINA. For three attributes probability of correct answer is maximum only for students who mastered all three attributes and for all other cases probability is at minimum level. In G-DINA model contribution of each attributes to probability of correct answering is different and in case student mastered one or more attributes probability of correct answering depends on weight of the attribute. In G-DINA model probabilities for each $P(\propto_{lj}^{*})$ case which respondent may have are calculated.

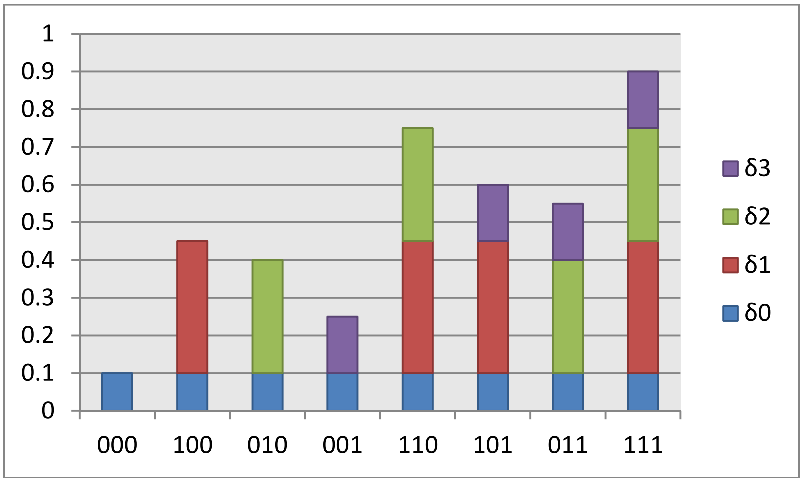

Figure 2

Figure 2 below shows δ values for each items determined by G-DINA model.

As shown in graphics, contributions of attributes to probability that a student gets the item correct are not equal. As an example, in columns 110 and 101, probability that students get item correct if they have first two attributes is higher than probability that students get the item correct if they have attributes first and third(de la Torre, 2011).

Purpose of this study is to determine impact of changes in Q matrix, which determines interrelations among measured attributes and items in CDM, on skills estimation power of models. Recent studies on CDMs have focused on new method searches to increase level of representation of items in the test by Q matrix (DeCarlo, 2011; Close et al., 2012; de la Torre, 2008a; de la Torre et al. 2010; Rupp & Templin, 2008; Chen, de la Torre, & Zhang, 2013). This study focuses on finding convincing evidences of conformity between Q matrix and the test. In addition, effects of sample size on skill estimation of models and on changes brought about by misspecification of Q matrix were investigated. To this end, this paper illustrates, using real data, practical results about Q matrix which is keystone for CDM.

2.Method

This paper focuses on practical issues, using real data, on model misspecification of Q matrix and investigates solutions to remedy Q-matrix.

2.1.Data

Data for this study is taken from yearly OKS examination (elementary school student selection and placement examination) taken by primary school 6, 7, 8 grade students. OKS is an achievement test consisting of Mathematics, Science, Turkish Language, Social Sciences and English Language sub-tests. Data of this study are responses of randomly selected 1000 examinees who got grade 8 Mathematic sub-test in OKS 2008 examination. Mathematical test used in the study is in the form of 18 multiple choice items. Sample size of 1000 examinees is preferred to ease comparison of latent classes that examinees are classified.

2.2.Procedure

In this study, 5 different Q matrices with different model data fits were employed to compare latent classes under which students were classified for different sample sizes. To implement analysis first Q matrices are determined. As real data were used in this study help of field expert was obtained for determination of Q Matrix pertaining to Mathematics sub-test items of OKS examination. Field experts first determined attributes measured in the test. Field experts divided the test into “numbers” (α_1), “geometry” (α_2), “probability- statistics” (α_3) and “algebra” (α_4) learning fields. In the next step, experts determined requisite attributes to answer an item correctly. Q matrices determined by field expert is given in Table 2.

Table 2.
5 Q matrices specified by expert opinion

Q1				Q2				Q3				Q4				Q5			
α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4
0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	1	0
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	1	0	1
1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0
1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	1	0	1
1	0	0	1	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	0
1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1
1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0
0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	0	1
0	1	1	1	0	1	1	1	0	1	1	1	1	0	0	0	0	0	1	1
1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	1	0	0	1
0	1	0	0	0	1	0	0	1	1	0	0	0	1	0	0	0	1	1	1
0	1	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	1	1	0
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0
0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	1	1	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0
0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	1	0	0	1	0
1	0	1	0	1	0	0	1	1	0	1	1	0	0	1	0	1	0	1	0

Model data fit indices for five different Q matrices calculated as a result of analysis using DINA and G-DINA models are given by Table 3.

Table 3.
Test-Level Fit Statistics for Q matrices determined by Expert Opinion

	DINA			G-DINA		
	-2LL	AIC	BIC	-2LL	AIC	BIC
Q1	18136,0493	18238,0493	18488,3448	17753,7869	17885,7869	18209,6987
Q2	18597,7407	18699,7407	18950,0362	18238,5150	18370,5150	18694,4268
Q3	18553,8703	18655,8703	18906,1659	18542,1316	18650,1316	18915,1504
Q4	20522,5875	20624,5875	20874,8830	19635,9509	19763,9509	20078,0472
Q5	21011,5629	21113,5629	21363,8584	19784,7675	19948,7675	20351,2034

Q matrices are in descending order from higher fit to lower fit in Table 3. Comparing DINA and G-DINA model results it is observed that fit statistics had similar orders for DINA and G-DINA models. Moreover G-DINA model fit statistics are lower comparing to DINA model. Lower values in fit statistics indicate better model fit.

To investigate effects of model-data misfit on latent classes to which students are classified, samples in different sizes were drawn from data. Sub-samples were generated in 30, 50, 100, 200 and 400 examinee groups randomly selecting from the data of 1000 examinees. Total of 25 sub-samples were used in analyses drawing 5 samples for each of other group sizes. For example, 5 samples of 30 examinees were drawn (30a, 30b, 30c, 30d, 30e). Sub-samples were analyzed separately by using Q1, Q2, Q3, Q4 and Q5 matrices for DINA and G-DINA models.

Means and standard deviations pertaining to specified groups in different sample sizes are given in Table 4. At the first row of Table descriptive statistics for the data of 1000 examinees were provided.

Table 4.
Descriptive statistics for sub-samples drawn for different sample sizes

	N		Mean	Std. Dev.	Skewness	Kurtosis
population	**1000**		**6,92**	**3,60**	**0,82**	**0,57**
Sample	30	a	6,80	3,40	0,98	0,87
		b	6,71	3,60	0,66	1,31
		c	6,47	2,89	0,70	-0,06
		d	6,77	3,79	0,64	0,23
		e	7,03	4,03	0,85	0,60
	50	a	6,64	3,49	0,46	-0,29
		b	6,94	4,00	0,60	-0,16
		c	6,44	3,42	0,35	-0,11
		d	6,84	3,59	0,86	0,47
		e	7,24	3,87	0,63	0,24
	100	a	6,97	3,59	0,68	0,54
		b	6,78	3,75	0,75	0,23
		c	7,32	3,49	0,69	0,48
		d	7,23	3,61	0,83	0,51
		e	6,57	3,41	0,97	1,02
	200	a	6,96	3,45	0,72	0,48
		b	6,90	3,75	0,86	0,49
		c	6,66	3,40	1,05	1,19
		d	6,92	3,63	0,92	0,66
		e	6,81	3,72	0,77	0,44
	400	a	6,72	3,57	0,97	0,94
		b	6,85	3,60	0,89	0,65
		c	6,89	3,64	0,79	0,58
		d	6,76	3,47	0,83	0,73
		e	6,99	3,54	0,85	0,65

As shown in Table 4, means and standard deviations for sub-samples take on values close to population distribution of 1000 examinees. Means varied in the range of 6,44 to 7,32 in the sub-samples whereas standard deviations varied between 2,89 and 4,00. This indicates sub-sample distributions are drawn from the same population.

Analysis results of latent classes pertaining to students determined by each sub-sample were compared with analysis results of latent classes of students determined by population. For example, Students' latent classes are determined by analysis performed using Q_1 matrix for the first sample of 30 students. Then latent classes pertaining to the same students using Q_1 matrix in the distribution of 1000 examinees are determined. In this case, two latent classes for each student in the group of 30 examinees were obtained using DINA model for Q_1 matrix. The same procedure was repeated for five different Q matrices and for each sub-sample set both using DINA and G-DINA models.

In this study, for sample selection and descriptive statistics SPSS software package was used whereas estimation of DINA and G-DINA model parameters were made using codes running under OX EDIT software.

3.Results

Firstly this study investigated extends of impacts of different Q matrices on decisions made about students. To this end, changes in 5 different Q matrices due to DINA and G-DINA classifications pertaining to data of 1000 examinees were calculated. To this end, classification accuracy of results of highest fit Q_1 matrix with other matrices were determined. Result of analysis is given in Table 5.

Table 5.
Q matrices interrelations in DINA and G-DINA models

		Q2	Q3	Q4	Q5
DINA	Latent Class	16	16	14	8
	Classification accuracy	85,4	84,6	54,4	13,4
G-DINA	Latent Class	16	16	16	13
	Classification accuracy	90,9	87	81,2	13,6

Number of latent classes, DINA and G-DINA models assigned students to, are shown in latent class row in Table 5. As there are four attributes in the test for both models, there are 24 = 16 latent classes. An examination of Table reveals that in DINA model Q_2 and Q_3 matrices placed students in 16 classes but for Q_4 and Q_5 matrices the same students are assigned to less number of classes. For G-DINA model less variability was observed. Determination of probabilities for less number of latent classes can be interpreted as diminished sensitivity on discriminating individual differences. Classification accuracy rows of Table show the concordance between classes assigned by models using Q1 matrix and other matrices. As it is readily seen changes in matrices have significant impacts on decisions about students.

3.1.Findings regarding Different Sample sizes:

Correlations between calculated fit statistics pertaining to results of all analyses implemented by this study and classification accuracy percentages are investigated. Results were given in Table 6.

Table 6.
Fit statistics and classification accuracy percentage correlation.

N	Fit İndex	DINA	G-DINA
30	-2LL	$-,855^{**}$	$-,427^{*}$
	AIC	$-,855^{**}$	$-,657^{**}$
	BIC	$-,855^{**}$	$-,716^{**}$
50	-2LL	$-,869^{**}$	$-,522^{**}$
	AIC	$-,869^{**}$	$-,566^{**}$
	BIC	$-,869^{**}$	$-,598^{**}$
100	-2LL	$-,667^{**}$	$-,779^{**}$
	AIC	$-,667^{**}$	$-,819^{**}$
	BIC	$-,667^{**}$	$-,777^{**}$
200	-2LL	$-,659^{**}$	$-,744^{**}$
	AIC	$-,659^{**}$	$-,746^{**}$
	BIC	$-,659^{**}$	$-,760^{**}$
400	-2LL	$-,755^{**}$	$-,756^{**}$
	AIC	$-,755^{**}$	$-,770^{**}$
	BIC	$-,755^{**}$	$-,779^{**}$

For five different Q matrices and five different samples drawn randomly for each sample size, individual analyses were implemented. For each analysis between data and test, 2 log likelihood (-2LL), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were calculated. There is high negative relation between fit statistics and classification accuracy percentage as shown in Table 6. This indicates that as model-data fit deteriorates, classifications accuracy decreases.

Table 7 shows results of DINA model analyses for different sample sizes.

Table 7.
Classification accuracy percentage of latent classes determined by DINA model for different Q matrices and sample sizes.

	DINA					
N	30 M of %*	50	100	200	400	Toplam
Q_1	82,0	86,8	91,4	94,1	94,5	89,76
Q_2	75,3	86,4	85,6	90,6	92,5	86,08
Q_3	74,0	81,6	83,6	89,9	91,5	84,12
Q_4	31,3	52,4	67,4	86,6	86,5	64,84
Q_5	29,3	25,6	55,0	60,7	65,5	47,22

- Average of 5 samples for each sample size

5 Different Q matrices used in analyses are denoted as Q_1, Q_2, Q_3, Q_4 and Q_5. Symbols 30, 50, 100, 200 and 400 denote classification accuracy averages of five different groups determined for each sample size. For example analysis with Q_1 in sample sizes of 30 examinees averagely resulted in 82% accuracy in 5 samples for the decisions made about the same students. In other words, the same 30 students are placed, in a rate of 82%, in the same classes as DINA model evaluated them both in the group of 1000 students and in sample size of 30 students.

As Table 7 indicates, differences among classifications get higher as Q matrix fit for each sample size deteriorates. Although classification accuracy rises as sample size increases, when Q_5 matrix, which gave lowest data fit, was used, differences between classifications increased to significant levels.

Table 8.
Classification accuracy percentage of latent classes determined by G- DINA model for different Q matrices and sample sizes.

	G-DINA					
N	30 M of %*	50	100	200	400	Toplam
Q_1	81,3	85,6	91,8	92,9	96,6	89,6
Q_2	74,0	82,4	88,8	90,5	94,3	86,0
Q_3	79,3	79,2	83,8	87,0	90,0	83,9
Q_4	78,0	76,8	82,8	84,7	92,0	82,9
Q_5	26,0	68,4	52,4	53,7	90,3	58,2

- Average of 5 samples for each sample size

Tables 8 shows classification accuracy of decisions made about students between sub-groups and data of 1000 students by results of analyses for G-DINA model with 5 different Q matrices. An examination of Table reveals that G-DINA model gave highest classification accuracy with Q_1 matrix. On the other hand, poorest classifications accuracy in G-DINA model, was obtained using Q_5 matrix. Besides, for G-DINA model ,accuracy percentage increases as n becomes larger.

3.2.DINA and G-DINA models comparison results:

Following results are obtained from comparison of classification accuracy rates of models for different sample sizes.

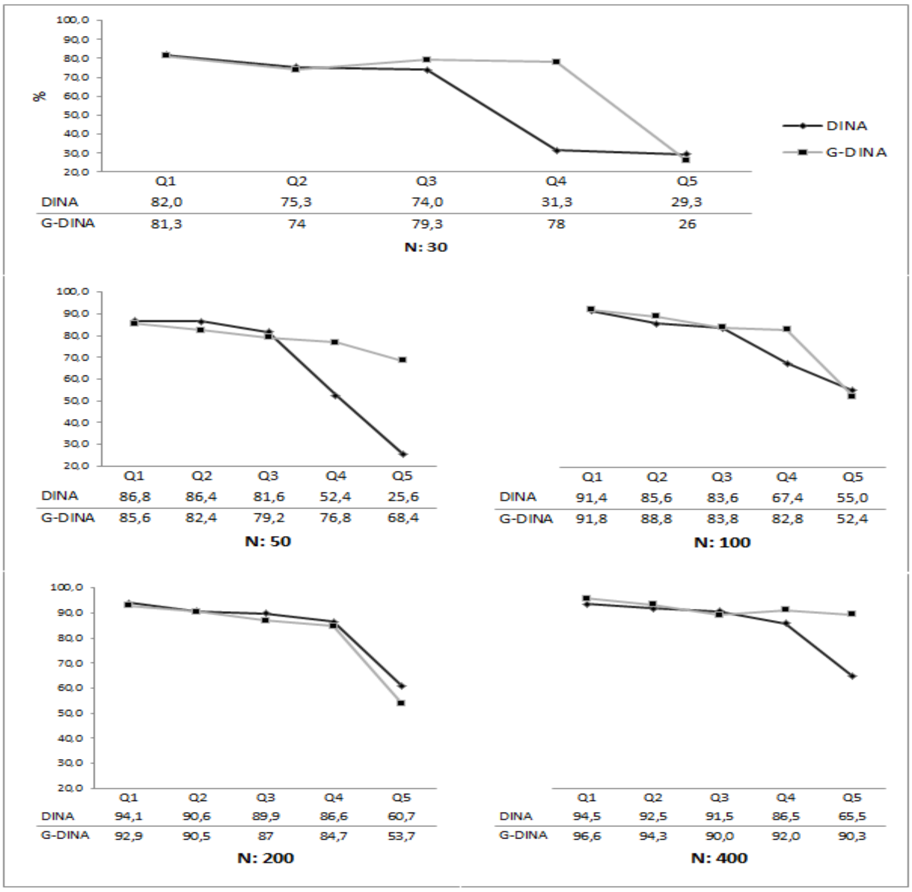

Figure 3 illustrates classification accuracy for students determined by DINA and G-DINA models for various sample sizes. First graphics in Figure 1, is based on results of analysis for 5 different Q matrices using both models for n=30. First points in the graphics are average of classification accuracy percentages obtained by Q_1 matrix pertaining to 5 different samples of 30 examinees. As readily observed in graphics, classification accuracy by DINA model significantly deteriorates with Q_4 matrix. For the same matrix, G-DINA models still give high level classifications accuracy. For both models it can be said that there is discrepancy in decisions made about students using Q_5 matrix. For n=50, DINA and G-DINA models achieved consistent decisions about students using Q_1, Q_2 and Q_3 matrices whereas for DINA model calculations made with Q_4 and Q_5 matrices classification accuracy significantly deteriorated. Although higher accuracy for all matrices was observed in G-DINA model, misclassification increases as model-data fit deteriorates.

Considering results of sample size n=100, for DINA model, significant discrepancies were observed in decisions about students with Q_4 and Q_5 matrices. On the other hand, higher accuracy results were obtained for first four matrices with G-DINA model but it deteriorates significantly when Q_5 was used. As for averages of sample size n=200, both DINA and G-DINA models give similar results. Lastly for the sample size n=400, classification accuracy deteriorated in DINA model when Q_5 matrix was used, yet, classification accuracy for all matrices in G-DINA was above 90%. All in all, it can be said that Q matrix changes had less impact on G-DINA than DINA model. However, the impacts of sample sizes on both models are similar.

4.Conclusion

Considering the findings of this study, significant impact of model-data fit on decisions about students in CDM was observed. Analyses summarized in Table 5 indicate that two students classified under “1111” class, namely determined as having all attributes using Q_1 for the population, were assigned to “1010” class when Q_2 matrix was used. As matrix misfit increases, misclassification increases significantly as well. Decision made regarding any student changes as data-fit of Q matrix deteriorates.

Deterioration in model-data fit lead to changes in estimated profiles for the same students with the same Q matrix when different sample sizes were analyzed. Results of analysis showed that a student who was classified as having all four attributes (1111) in the group of 1000 students, was classified as having no attributes (0000) when evaluated in the sub-group of 400 students. This revealed importance of Q matrix in practical applications. In DINA model results, classes determined by analysis for population using Q_4 and Q_5 matrices were observed to differentiate approximately 50% in sub-samples. This differentiation was also observed in large samples of 400 students. In other words, this rate indicates that model decisions were changed for half of the group. It can be said that when a Q matrix with lower data-fit was used, models cannot give sample- independent results.

When relations between DINA and G-DINA models are considered, generally classifications by G-DINA model will be affected less by changes in Q matrix. This also conforms to the fact that fit-indices of G-DINA is lower than DINA model's as shown in this study. At this stage, it can be assumed as an indication that data and items studied here better fits G-DINA model.

Determination of Q matrix- item relations by exclusively expert view may lead to thoughts that decisions made by CDM models on students are open to dispute. In this study, fit statistics showing data-fit of Q matrix was found to have high correlation with population and sub-sample classification intersections. In this sense, statistical evidences are needed to determine Q matrix specification. Researchers and practitioners should take the studies on this issue (DeCarlo, 2011, 2012; Close et al. , 2012; de la Torre, 2008a; de la Torre at al. 2010; Rupp & Templin, 2008) into consideration during construction of their models to ensure reliability of their outcomes. Those studies regarding structure of Q matrix embody milestones for “diagnostic analysis”.

Besides, the researchers’ obtaining new findings about the model fit and sample size with the researches based on the real application data in the field of CDM and its appropriate usage in the educational areas will increase the usage of these approaches. Meanwhile, the other approaches’ comparison that developed in CDM will be useful for practitioners.

References

Bolt, D. & Fu, J. (2004). A Polytomous Extension of the Fusion Model and Its Bayesian Parameter Estimation. Paper presented at the Annual meeting of the National Council on Measurement (NCME) in Education, San Diego, CA, April.

Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and Absolute Fit Evaluation in Cognitive Diagnosis Modeling. Journal of Educational Measurement, 50(2), 123-140.

Cheng Y.& Chang H.(2007). The modified maximum global discrimination index method for cognitive diagnostic computerized adaptive testing. Presented at the CAT and Cognitive Structure Paper Session, June 7.

Cheng, Y. (2010). Improving cognitive diagnostic computerized adaptive testing by balancing attribute coverage: the modified maximum global discrimination index method. Educational and Psychological Measurement, Vol. 70 (6), p. 902-913

Close, C.N., (2012). An exploratory technique for finding the Q-matrix for the DINA model in cognitive diagnostic assessment: Combining theory with data. Unpublished doctoral dissertation, University of Minnesota.

de la Torre, J. & Douglas, J. (2008). Model Evaluation and Multiple Strategies in Cognitive Diagnosis: An Analysis of Fraction Subtraction Data. Psychometrika. Vol,73, N3, p. 595-624.

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, p. 343–362.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively-based multiple-choice options. Applied Psychological Measurement, 33(3), 163–183.

de la Torre, J. (2009b). DINA Model and Parameter Estimation: A Didactic. Journal of Educational and Behavioral Statistics March, Vol. 34, No. 1, p. 115–130.

de la Torre, J. (2011). The generalized DINA model framework. Psychometrika. Vol 76, 2, p. 179-199.

de la Torre, J. (2013). Application of the DINA Model Framework to Enhance Assessment and Learning. In M. M. C. Mok (Ed.), Self-directed Learning Oriented Assessments in the Asia-Pacific (Vol. 18, pp. 87-103): Springer Netherlands.

de la Torre, J. .& Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. Psychometrika. V69, 3, p. 333-353.

de la Torre, J. .& Lee, Y.S. (2010). A note on Invariance of the DINA Model Parameters, Journal of Educational Measurement, Spring 2010, Vol 47, No:1, p. 115-127.

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, 2010, Vol. 47, No:2, p. 227-249.

de la Torre, J., & Liu, Y. (2008). A cognitive diagnosis model for continuous response. Presented at the National Council on Measurement in Education, March, New York, NY.

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, Vol 35, p. 8-26.

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. Applied Psychological Measurement, Vol 36, p. 447-468.

diBello, L., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. V. Rao & S. Sinharay (Eds.), Handbook of Statistics (Vol. 26, Psychometrics) (pp. 979-1027). Amsterdam, Netherlands: Elsevier.

Embretson, S. (1984). A general latent trait model for response processes. Psychometrika, 49, 175–186.

Embretson, S. (1991). A multidimensional latenttrait model for measuring learning and change. Psychometrika, Vol 37, p. 359–374.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. Psychological Methods, Vol 3, p. 380-396.

Fischer, G.H. (1973). The linear logistic model as an instrument in educational research. Acta Psychologica Vol 37, p. 359–374.

Haertel, E.H. (1984). An application of latent class models to assessment data. Applied Psychological Measurement 8, 333–346.

Haertel, E.H. (1989). Using restricted latent class models to map the skill structure of achievement items. Journal of Educational Measurement, Vol 26, p. 333-352.

Haertel, E.H. (1990). Continuous and discrete latent structure models of item response data. Psychometrika Vol 55, p. 477–494.

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

Hou, L. (2013). Dıfferentıal Item Functıonıng Assessment In Cognıtıve Dıagnostıc Modelıng: Applyıng Wald Test To Investıgate Dıf In The Generalızed Dına Model Framework, Doctor of Philosophy, The University of Dalaware, Unpublished doctoral dissertation.

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

Junker, B. W. (1999). Some statistical models and computational methods that may be useful for cognitively-relevant assessment. Prepared for the Committee on the F oundations of Assessment, National Research Council, November 30, 1999 http://www.stat.cmu.edu/~brian/nrc/cfa/

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. Applied Psychological Measurement, Vol 25(3), p. 258-272.

Lamb, R. L., Annetta, L., Vallett, D. B., & Sadler, T. D. (2014). Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames. Computers & Education, 70(0), 92-104.

Leighton, J. P.& Gierl M. J. (2007). Why Cognitive Diagnostic Assessment? Leighton, J. P. Gierl M. J. (Eds). Cognitive Diagnostic Assessment for Education. New York: Cambridge University Press

Li, F. (2008). A Modified Higher-Order DINA model for Detecting Differential Item Functioning and Differential Attribute Functioning. Doctor of Philosophy, The University of Georgia, Unpublished doctoral dissertation.

Macready, G. B, & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics, 2, p. 99-120.

Maris, E. (1999). Estimating multiple classification latent class models. Psychometrika, Vol 64, p. 187-212.

Roussos, L., diBello, L. V., Stout, W., Hartz, S., Henson, R. A., & Templin, J. H. (2007). The fusion model skills diagnosis system. In J. P. Leighton, & Gierl, M. J. (Ed.), Cognitively diagnostic assessment for education: Theory and practice (pp. 275-318). Thousand Oaks, CA: SAGE.

Rupp, A. A., & Templin, J. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. Measurement: Interdisciplinary Research and Perspectives, 6, 219-262.

Shu, Z., Henson, R., & Willse, J. (2013). Using Neural Network Analysis to Define Methods of DINA Model Estimation for Small Sample Sizes. Journal of Classification, 30(2), 173-194.

Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. Journal of Educational Statistics, Vol 12, p. 55–73.

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & Safto, M. (Eds.). Monitoring skills and knowledge acquisition (pp.453-488). Hillsdale, NJ; Erlbaum

Tatsuoka, K. (1995a). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P.D. Nichols, S. F.

Tatsuoka, K. (1995b). Cognitive assessment An Introduction to the Rule Space Method, Routledge, New York: Taylor & Francis Group.

von Davier, M. (2005). A general diagnostic model applied to language testing data (ETS Research Report No. RR-05-16). Princeton, NJ: Educational Testing Service

von Davier, M. (2007). Hierarchical general diagnostic models. (ETS Research Report, RR-07-19). ETS: Princeton, NJ.