e-International Journal of Educational Research
Volume: 5 Issue: 1 Winter 2014 pp. 1-19

Comparison of Validity of Classification Made According to DINA Model with Criterion-Referenced and Normative Assessment

Assist.Prof.Dr. Tahsin Oguz Başokçu[1]
Ege University-Turkey
tahsin.oguz.basokcu@ege.edu.tr

Assoc.Prof.Dr. Hülya Kelecioğlu
Hacettepe University- Turkey
hulyaebb@hacettepe.edu.tr

Abstract

The purpose of this research is to examine the validity of classifications by DINA Model as a Cognitive Diagnostic Model and by traditional methods. In order to make comparisons between DINA Model and criterion-referenced and normative models, a measurement tool that belongs to "measurement and assessment in education" class which is appropriate for DINA Model analyses is developed. Properties necessary for answering each item of measurement tool are determined by scholars and the Q matrix which shows item-property relation is prepared considering the compatibility of decisions made by various scholars. Above mentioned measurement tool is applied to 471 undergraduates from the Faculty Education and Arts and Sciences of Ege University. Raw scores of undergraduates are classified according to their success, that is whether they passed the class or failed, through criterion-referenced and normative assessment. And then this classification is compared with other classifications based on DINA Model. The comparison of assessment by normative assessment and classifications by DINA Model shows that the results are different for 50 undergraduates who failed and 28 who passed the class. As a consequence of the study, it is observed that the inconsistency between the normative assessment and DINA Model is 16.5% for the whole group. In assessments done by criterion-referenced, for the students who failed, two methods give the same results. However, for 87 students who are supposed to pass according to criterion-referenced assessment, DINA Model results show that these undergraduates are not qualified to give the right answers of the items. The incompatibility between assessments according to criterion-referenced and classifications according to DINA Model is calculated as 20%.

Key Words: Cognitive Diagnostic Models, DINA Model, Classification validity, Q Matrix

INTRODUCTION

Cognitive diagnostic models (CDMs) are multi-dimensional implicit class analysis models being used to determine both decisions about students and students' incompetence. Recently CDM as a commonly used model considers relationship between individuals' performance and skills being assessed by test item, instead of relationship between performances of individuals being tested and group or test items. In these models students' skills are not tried to be identified on a continuous scale, instead properties that student should have to answer each questions are

[1] This *paper is based* on the *my Ph.D dissertation*

identified and at the end of the analysis it is determined that which properties being identified for the test are acquired by student. In this sense it is possible to examine the validity of classification on a decision stage with assessment instrument developed according to CDM and with test parameters determined by this model. Concurrently it is expected to make well-timedly determinantion of incompetence of student and teaching in the case of whether having a particular property or not by using these models.

In CDM, most commonly used Deterministic Input Noisy And gate (DINA) model is an analysis method based on implicit class analysis and grounded on multi-dimensional item-response theory. This model presents useful and new point of view to classify students according to particular cutoff score or to determine both curriculum' and students' incompetence in teaching period.

DINA Model developed by Haertel (1989) is an implicit class model similar to dual abilty models. For this reason it is closely in relation with item-response theory (IRT) models (Haertel, 1990). However unlike item-response theory DINA model does not assume that students have various sized and continuously distributed abilities; it mostly divides students into well determined, limited number of different implicit classes.

DINA model is based on item – property relation like seen in most CDMs. Better functioning of the model depends on correct identification of properties that is necessary to answer item correctly. It is necessary to prepare Q matrix to show distribution of items correlated to properties in DINA model. Apart from this, there are some other methods in which Markov Chain Monte Carlo estimation (MCMC) is used. Q matrix is 1 – 0 pattern in which necessary properties to answer each questions correctly are determined. In this matrix an item is represented in association to only one property or more than one property. Moreover with developed modifications of DINA such as G-DINA and HO-DINA there is a chance to contribute more than one properties related to one item either by weighted or as hierarchically into the analysis (de la Torre, 2010, de la Torre, & Douglas, 2004).

Parameter estimation in DINA Model

DINA model aims to reveal implicit property underlying the observed ability of respondent. In this sense, model probably grounds the relationship between implicit property and observed property, and it offers to classify two item parameters for each item. These are *s* "slipping" and *g* "guessing" parameters.

$s_j = P\left[Y_{ij} = 0 | \eta_{ij} = 1\right]$ and

$g_j = P\left[Y_{ij} = 1 | \eta_{ij} = 0\right]$,

sj reflects the probability of giving wrong answer of individual having implicit property to j item (incorrect positive possibility), and gj means the case of possibility of giving correct answer of individual not having an implicit property (correct positive possibility). The ηij is a latend response which is defined deterministically through the equation $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$. The lower the sj parameter, the possibility of giving correct answer of indiviuals having required properties increases (de la Torre, 2009).

The *gj* parameter is also called guessing parameter. Maris (1999) alternatively explains *gj* parameter as a successfully usage of mental guessing ability. In this case guessing parameter has different structure from chance parameter in IRT (item response theory). In model the *g* parameter means not only to be answered items correctly by individual not having required property to answer item correctly. At the same time it means that individual answers item correctly by using different properties other than required properties. This shows that different properties not being identified in Q matrix are also sufficient to answer item correctly. The g parameter, proximal to "1", for an item does not mean that item is answered by individuals not having a required property, it is also interpreted that some required properties to answer item correctly can not be determined.
Item-response function in DINA model is as follows:

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

In length K_j^{*} - K denotes the number of attributes measured by the test-, in the case that $1_{K_j^{*}}$ becomes a vector of an individual, the gj expresses the probability of an individual guessing correctly despite not having at least one of the required properties for item j, and 1- sj reflects the possibility of individuals who have all required properties for an item respond the item incorrectly despite not making a slip (de la Torre, Hong, & Deng, 2010).

In DINA model conditionally distributed item-response variable Yij depends on both α_{ij} and η_{ij}. This is an extension of possibility function of DINA model. Conditional independence showing independency among subjects can be written as in follows:

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

In DINA model s and g parameters emerge at a level of an item. Each item divides population into two classes, and possibility of giving correct answer to item of the students in same class becomes equal. Students' performance in exam is not a complete indicator of vector of an attribute being predicted in test. For this reason a model based on probability only allows to see possibility of s and g. Case of "slipping" occurs when student responds to subtask or an item incorrectly while he/she has required properties in an item. "Guessing" is a situation in which student can complete the subtasks or respond the item correctly despite he/she does not have required properties. In determination of selected model, thus, it is decided whether s and g parameters substantiate at the level of a subtask or an item.

Implicit classes and skill mastery estimation in DINA Model

In DINA model students fall within two basic classes for each item. First of these two classes is null class, that is, the group formed by students not having any of the required skills, and the other class is full class that is formed by students having all skills. In DINA model one who does not have one of the skills is included in null class. The function below shows that the possibility of being answered an item correctly by an individual having all required skills:

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

The P is the probability of being answered item correctly by student having all required skills. η_{ij} is an implicit respond being determined by α and i quality of subject and vector of qj. The rank corresponds to *j* item of Q matrix can be demonstrated like as follows:

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

Tatsuoka (1982) identifies $\alpha_i = (\alpha_{i1}, \dots, \alpha_{iK})$ as "knowledge states". In this equation, the case $\alpha_{iK} = 0\ or\ 1$ depends upon whether student *i* has *k;* $\eta_i = (\eta_{i1}, \dots, \eta_{ij})$ quality or not. *j* = shows item number. Each element of αi indicates whether the ith student could master the kth skill or not and also help to determine observed score Yij. For specific *k* property there is *2k* possible knowledge model, i.e. implicit class.

DINA model determines *2k* number of implicit class on the basis of *k* number properties for a developed or an implemented test. For example if it is thought that only 3 of the properties are tested in one test, in this case test takers are classified into 8 implicit classes. Possible classes for 3 properties are ranged as "000", "100", "010", "001", "110", "011", "101" ve "111". Individuals who do not have any of the properties are settled in 1st class while individuals who have only first and third properties settle in 7th class.

Determination period of DINA model whether respondent has specific property or not is combination of different processes. Whether student settles in class 0 or in class 1 in terms of property is a probability value. This value may change, but in general use .50 thresholds is used as a base. The probability of having a property for student is less than .50 student gets involved in class 0 while the probability of having a property for student is equal or is more than .50 he/she gets involved in class 1. DINA model does not ground correct respond rate of item representing that property while making estimation α about student. The probability of having a property for student is related to difficulty of item represented that property (de la Torre 2008c).

In this study DINA model parameters are predicted via test developed according to CDM principles. This study having the meaning of real life application aims to suggest applicability of mostly theoretical properties related to CDM. Comparison of application results being done under real conditions contribute to exhibit similarities and differences or processing and inoperative parts of the theory. For this reason, in study the research question, in which level the consistency of classification being done via DINA model and by using normative and criterion-referenced evaluation according to results of "Measurement and evaluation in education" test developed via DINA model, was tried to be answered.

METHOD

The aim of this research is to reveal existent condition in terms of specifying congruity level of a developed test to CDM. For this reason, this research can be thought as a descriptive research. Moreover it is handled as a theoretical research because it gives opportunity to compare different classification techniques related to classification validity.

The population of research consists of 471 students taking measurement and assessment course in Ege University. After being developed, measurement tool was implemented to study group being formed by 471 students from different departments. Descriptive statistics obtained from application of the test are presented in Table 1.

Table 1.
Descriptive statistics of the test

Number of item	50
Respondent	471
Mean	32.12
Variance	59.6
Standart deviation	7.72
Skewness	-.87
Kurtosis	1.20
Alpha reliability	.85
Standard error	2.96
Average p	.64
Average rpb	.50

When descriptive statistics related to applied test are considered, it is observed that distribution is negatively skewed and sharp. Mean of item difficulty indexes shows that test is slightly more than intermediate difficulty. Reliability coefficient of test scores is .85. This coefficient is accepted as high level for achievement tests (Murphy & Davidsofer, 2001).

To determine Q matrix showing property – item relationship for model used in the research, expert opinions are asked. Experts identified 7 properties for 50 items and associate these properties to items. Thus 75 associations are emerged. Hereby 29 items are associated to 1, 17 items are associated to 2 and 4 items are associated to 3 properties. Three of the experts agreed upon 59 associations of 75 associations. It is observed that In 19 of the associations two of the experts reach an agreement. Q-matrix being obtained by relating items to properties is shown in Table 2.

Table 2.
Q-Matrix intra testing instrument

	Properties								Properties						
Item	1	2	3	4	5	6	7	**Item**	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0	**26**	0	0	0	0	0	0	1
2	0	1	0	0	0	0	0	**27**	0	0	1	0	0	0	0
3	1	0	0	0	0	0	0	**28**	0	0	1	1	0	0	0
4	0	0	0	0	0	1	0	**29**	0	0	1	0	0	0	0
5	1	1	0	1	0	0	0	**30**	0	0	1	0	0	0	0
6	1	1	0	1	0	0	0	**31**	0	0	1	0	0	0	0
7	1	1	0	1	0	0	0	**32**	0	0	0	1	0	0	0
8	0	1	0	0	0	0	0	**33**	0	0	1	0	0	0	0
9	1	0	0	0	0	0	0	**34**	0	0	1	1	0	0	0
10	1	1	0	0	0	0	0	**35**	0	0	1	0	0	0	0
11	0	1	0	0	0	0	0	**36**	0	0	1	0	0	0	0
12	1	0	0	0	0	1	0	**37**	0	0	1	1	1	0	0
13	1	0	0	0	0	1	0	**38**	0	0	0	1	0	0	0
14	0	0	0	0	0	1	0	**39**	0	0	0	1	1	0	0
15	0	1	0	0	0	1	0	**40**	0	0	0	1	1	0	0
16	0	1	0	0	1	0	0	**41**	0	0	1	0	0	0	0
17	0	0	0	0	0	0	1	**42**	0	0	0	1	0	0	0
18	0	0	0	0	0	1	0	**43**	0	0	1	1	0	0	0
19	0	0	0	0	0	1	0	**44**	0	0	0	1	1	0	0
20	0	0	0	0	0	1	0	**45**	0	0	0	1	0	0	0
21	0	0	0	0	0	1	1	**46**	0	0	0	1	1	0	0
22	0	0	0	0	0	1	1	**47**	0	0	0	0	1	0	0
23	0	0	0	0	0	0	1	**48**	0	0	0	1	1	0	0
24	0	0	0	0	0	1	1	**49**	0	0	0	1	1	0	0
25	0	0	0	0	0	0	1	**50**	0	0	0	1	0	0	0

In DINA Model, concordance of Q-matrix to data and the model is examined by different techniques. First of all DINA model statistics to compare data concordance are calculated during test phase. These statistics provide opportunity to compare different models being used for data. At the same time parameters of DINA model give information about item representation level of properties defined with q-matrix. Information related to model-data concordance and representation level of q-matrix is conveyed in data analysis part.

Model-data fitting

In research concordance of model indicated with q-matrix is examined by using q-matrixes being prepared by experts and q-matrix prepared according to expert concordance. AIC and BIC statistical values acquired by results of analysis being done by using q-matrix prepared by three experts and q-matrix determined in the light of expert concordance are given in Table 3.

Table 2.
AIC and BIC Criterias for Model-Data Comparison

	Expert 1 Q Matrix	Expert 2 Q Matrix	Expert 3 Q Matrix	Expert concordance Q Matrix
AIC	28791.02	28790.10	28796.21	28789.31
BIC	29756.17	29746.25	29769.37	29732.46

As it is seen in Table 4 above, model determined by considering concordance among experts corresponds to data better.

Parameters of DINA Model

Q-matrix prepared within the research and gathered data were tested via DINA model unit software developed for oxedit programme and thus parameters related to measurement tool. The s, g and (1-s) values determining qualities of measurement tool and Q-matrix are given in Table 4.

Table 3.
DINA Model parameters related to testing instrument

Item	*g*	*s*	*1-s*	Item	*g*	*s*	*1-s*
1	.74	.08	.92	**26**	.40	.27	.73
2	.14	.64	.36	**27**	.49	.13	.87
3	.48	.11	.89	**28**	.45	.25	.75
4	.50	.23	.77	**29**	.32	.24	.76
5	.59	.06	.94	**30**	.59	.18	.82
6	.45	.18	.82	**31**	.68	.16	.84
7	.76	.12	.88	**32**	.31	.11	.89
8	.26	.55	.45	**33**	.43	.24	.76
9	.83	.63	.37	**34**	.73	.27	.73
10	.75	.02	.98	**35**	.50	.31	.69
11	.49	.27	.73	**36**	.61	.22	.78
12	.52	.09	.91	**37**	.34	.13	.87
13	.68	.22	.78	**38**	.32	.24	.76
14	.32	.49	.51	**39**	.42	.42	.58
15	.47	.34	.66	**40**	.62	.33	.67
16	.43	.03	.97	**41**	.55	.22	.78
17	.24	.15	.85	**42**	.80	.26	.74
18	.51	.18	.82	**43**	.58	.46	.54
19	.40	.24	.76	**44**	.81	.41	.59
20	.40	.36	.64	**45**	.80	.33	.67
21	.43	.24	.76	**46**	.67	.18	.82
22	.52	.14	.86	**47**	.99	.49	.51
23	.38	.03	.97	**48**	.84	.33	.67
24	.35	.35	.65	**49**	.29	.26	.74
25	.71	.30	.70	**50**	.30	.17	.83
Mean					.52	.25	.75

When the s and the g parameters of DINA model relating to measurement tool are examined, it is seen that g value falls between .14 ile .99. s parameter values relating to measurement tool vary between the range of .02 and .64. While looking at the mean of the parameters, it is seen that mean of g value is .52 and mean of s value is .25. Wenmin (2006) said that low s values and high g values are indicators of the simplicity of test. When s and g parameters are examined, it is seen that test is some more easy than average.

Criterias related to pass-fall decisions

Within the scope of research because validity of decisions related to course success of group to which measurement tool is applied is being examined, according to normative and criterion-referenced assessments about students for analysis criterias are taken differently.

Normative assesment:

Within the context of Ege University in the parts used in normative assessment criteria score is 60 point over 100 point. For this reason in the scope of study passing decision for student 60 point is accepted as criteria.

Criterion-referenced assessment:

In the system of normative assessment of Ege University students' letter grades are determined according to predetermined level criterias. Class level is predicated on mean of raw scores and identified as in Table 5.

Table 4.
Criterion referenced assessment class level

Level	Mean of raw scores
Excellent	70-100
Very good	62,5-70
Good	57,5-62,5
Above average	52,5-57,5
Average	47,5-52,5
Poor	42,5-47,5
Worse	0-42,5

With which criterias students' letter grades are determined in criterion-referenced assessment is given in Table 6.

According to students' letter grades being determined according to criterias given in Table 6, students, except students with letter grade FF and FD, are supposed as successful in that lesson. In research findings related to criterion-referenced assessment are calculated according to explanations given above.

DINA Model:

In literature there are no any studies having aim to make pass-fall decision about individuals who take test by using properties determined by DINA model or having aim to compare these properties with a different technique. For this reason to compare number of properties that students possess with the results of normative and criterion-referenced assessment it is tried to determine a criteria related with DINA model by asking experts' opinions.

Table 5.
"t" Threshold value of letter grades

Level	FF	FD	DD	DC	CC	CB	BB	BA	AA
Excellent	<34	34-38,99	39-43,99	44-48,99	49-53,99	54-58,99	59-63,99	64-68,99	≥69
Very good	<36	36-40,99	41-45,99	46-50,99	51-55,99	56-60,99	61-65,99	66-70,99	≥71
Good	<38	38-42,99	43-47,99	48-52,99	53-57,99	58-62,99	63-67,99	68-72,99	≥73
Above average	<40	40-44,99	45-49,99	50-54,99	55-59,99	60-64,99	65-69,99	70-74,99	≥75
Average	<42	42-46,99	47-51,99	52-56,99	57-61,99	62-66,99	67-71,99	72-76,99	≥77
Poor	<44	44-48,99	49-53,99	54-58,99	59-63,99	64-68,99	69-73,99	74-78,99	≥79
Worse	<46	46-50,99	51-55,99	56-60,99	61-65,99	66-70,99	71-75,99	76-80,99	≥81

For this purpose opinion related to this subject of three experts instructing "measurement and assessment in education" lesson in two different universities is asked. Experts that examine the measurement tool developed for this research and Q matrix being determined according to items in measurement tool achieve a consensus on to define possessing any four of seven properties identified by DINA model for being successful as a threshold.

In accordance with experts' opinion in parallel with threshold 60% identified for absolute criteria it is decided to accept possessing at least four of seven properties determined for measurement tool as success criteria in lesson for DINA model.

FINDINGS AND COMMENTS

After application of measurement tool the comparison of pass-fall decisions being made about students according to absolute and relative criterias with implicit classes students belong to being determined by DINA model is examined.

In Table 7 below, the distribution of students being decided for pass or fall based on absolute criteria related to testing and measurement in education is seen.

Table 6.
Pass-fall ratio per absolute criteria

Decision	Frequency	Percentage
Fall	138	29.2
Pass	333	70.6
Total	471	100

As is seen in Table 7, 333 persons in a group are assumed to be succeeding in lesson while 138 persons are assumed to be failed. Success rate of the group is at the level of 70%. Distribution of students in study group according to possessed properties being determined by DINA model is given in Table 8.

e-International Journal of Educational Research
Volume: 5 Issue: 1 Winter 2014 pp. 1-19

Table 7.
Absolute criteria and ratio of students being made decision for pass or fall

		DINA Model		
		Fall	**Pass**	**Total**
Absolute criteria	**Fall**	88(%18.7)	50 (%10.6)	138(%29.3)
	Pass	28 (%5.9)	305 (%64.8)	333(%70.7)
	Total	116(%24.6)	355 (%75.4)	471 (%100)

When the entire group is taken into account, 50 of 138 students being decided for fall amongst total of 471 students and 28 of 333 students being decided for pass are classified as inconsistent according to having the number of properties being determined via DINA model. In this sense it is observed that 16.5% of classifications being done by absolute criteria for 471 students are inconsistent in terms of DINA model classifications. The correlation between number of properties that students have and raw scores that students have taken from the test is shown in figure 1.

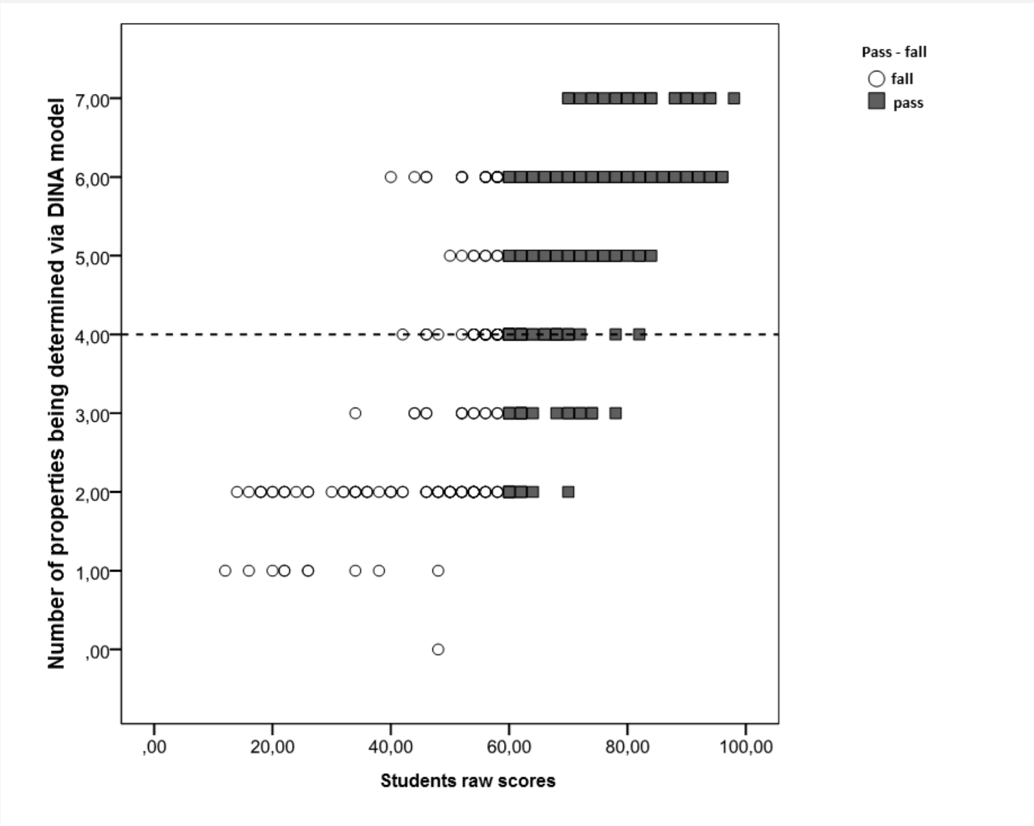

Figure 1.
The correlation between number of properties of students and raw scores in terms of pass – fall decision via absolute criteria

The slot in Figure 1 is shown minimum property number, 4, that students should have to accept succeed in the lesson. Individual expressed by square in graph are succeed in lesson while individuals expressed by circle are fail in the lesson. In this sense squares under the slot and circles above the slot represent inconsistent

e-Uluslararası Eğitim Araştırmaları Dergisi
Cilt: 5 Sayı: 1 Kış 2014 ss. 1-19

classifications between two methods. In view of error margin of two models, 17 students assumed as a fail in the lesson in spite of having six properties and 11 students assumed as succeeded in the lesson despite having two properties give an idea about significant level of inconsistency ratio between two methods. The ratio of pass-fall decision mistakes for normative assessment is at the level of 6%.

During the examination of pass-fall decision according to criterion-referenced assessment firstly the results of criterion-referenced assessment for the entire group are taken into consideration. First of all while mean of the group is being determined by considering average value-added criteria (point 15) for criterion-referenced assessment method in Ege University, scores of students in line 347 and 302 are excluded to mean. New raw score value for group is found as 64.44 so the group is included in “very good” class level in the criterion-referenced assessment method of Ege University. As said in method part of research, in classes level of which is defined as “very good” minimum passmark is determined as 41 (DD grade). In Table 9, frequencies of pass-fall decisions given by criterion-referenced assessment method considering these criterias for the entire group is seen.

Table 8.
Relative criteria pass-fall ratio

Decision	Frequency	Percentage
Fall	29	6.2
Pass	442	93.8
Total	471	100

As is seen in Table 9, while relative criteria are used, passed decision is given for 94% of the group.

When passed decision given according to relative criteria are examined, 87 students are assumed as suceeded in lesson despite they have less than half of the properties being determined as necessary for answering items correctly. Consistency of pass-fall decisions given by using DINA model and relative criteria is shown in Table 10.

Table 9.
Ratio of students given pass-fall decision as per relative criteria and DINA Model

		DINA Model		
		Pass	**Fall**	**Total**
Relative criteria	**Fall**	29(6.2%)	0(0%)	29 (6.2%)
	Pass	87(18.5%)	355(75.4%)	442(93.8%)
	Total	116(24.6%)	355 (75.4%)	471 (100%)

As in Table 10, while between two methods any inconsistency for students about who pass-fall decision is given is not seen, it is observed that concurrence between DINA model and relative criteria for students about who pass-fall decision is given is corrupted at the rate of 19%.

In Figure 2, the correlation between number of properties that students have and their raw scores in test is shown considering pass-fall decision given by using relative criteria.

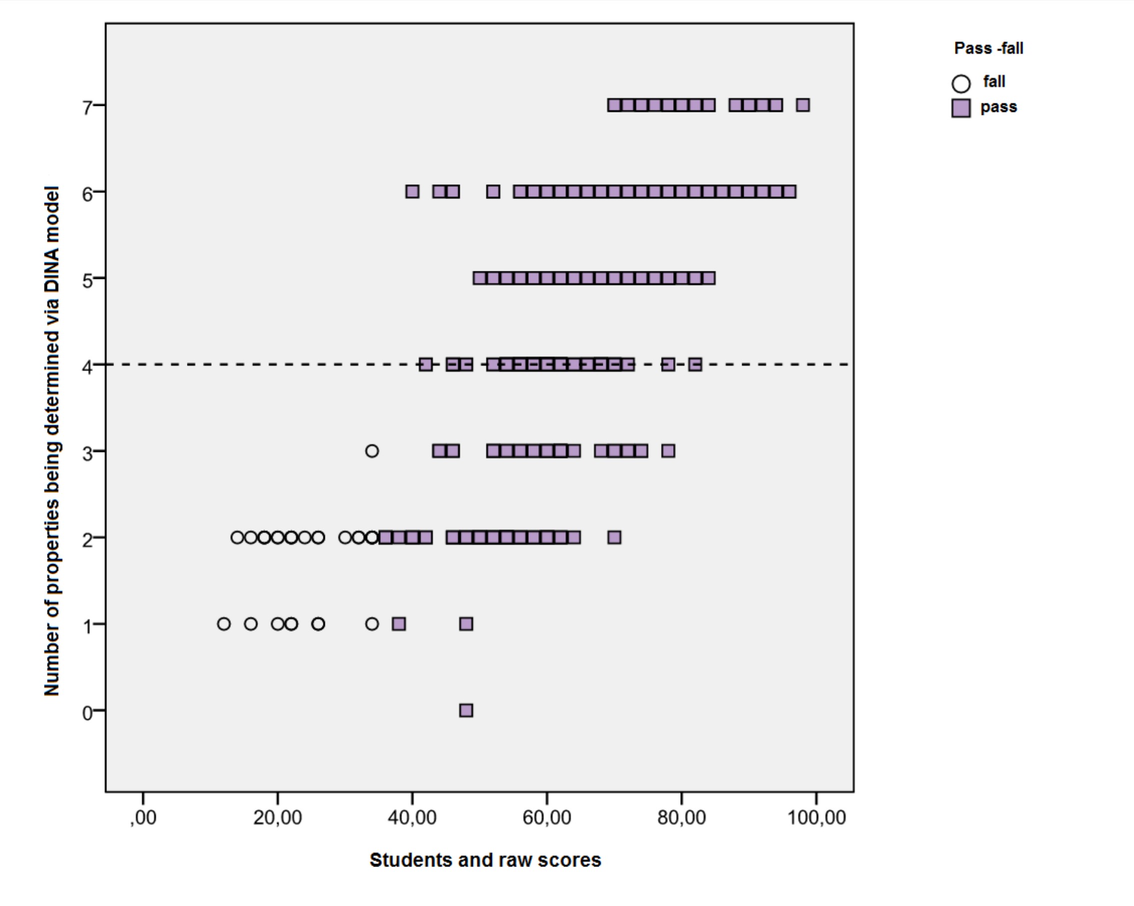

Figure 2.
Number of properties and raw score correlation according to pass-fall decision by using norm-regerenced

In Figure 2, according to DINA model classifications square symbols under the slot showing four properties that are minimum value necessary for students to be assumed as a succeeded in lesson are indicated inconsistent classifications. As is seen in the figure, when relative criteria is used, about individuals having none of the properties or having one or two properties passed decision is given. These 57 students show critical inconsistency between two methods. The ratio of inconsistency between two methods is calculated as 12%.

Due to make interpretation about accrual level of distortion ratio letter grades of students in relative system is viwed, and the correlation between letter grades and DINA model tacit classes is examined.

Table 11 shows frequency of students' letter grades for all classes and number of properties determined by DINA model that these students have.

Table 10.
Students' relative criteria letter grades and number of properties to DINA Model

Number of property	Letter grades									Total
	FF	FD	DD	DC	CC	CB	BB	BA	AA	
0	0	0	1	0	0	0	0	0	0	1
1	5	3	1	1	0	0	0	0	0	10
2	11	10	10	10	17	13	2	1	0	74
3	1	0	3	3	10	7	6	1	0	30
4	0	0	0	5	20	23	8	3	1	60
5	0	0	1	1	11	25	25	18	0	81
6	0	0	0	5	16	30	64	38	36	190
7	0	0	0	0	0	0	5	9	11	25
Total	17	13	16	25	74	98	110	70	48	**471**

As Table 11 is examined, some extreme observations are drawn attention. It is observed that one of the students getting DD letter grade has five properties in total. Concurrently it is seen that students who has only four properties gets AA letter grade. Dark cells in Table 11 show students having 3 properties or less than 3 properties. When Table is examined in this way, low observation rate in lower-left cell and top right cell is regarded as an indicator of claasification validity. As the property of students about who fall decision is given is examined, any inconsitency is seen in classifications. In case of students about who passed decision is given, it is seen that 10 students have higher letter grades in spite of having less property. 47 students get CC and CB, 29 students gets DD and DC and thus they are assumed as succeeded in lesson. As is seen, the inconsistency among achievement level of students as well as pass-fall decisions being given about students in criterion-referenced assessment is observed. CC-CB letter grade interval corresponds to 60-73 score interval in absolute system. In this case, 13 individuals having only two of the seven properties necessary to answer items correctly are assumed as succeeded in 66-73 score interval.

In research when comparison of consistency of number of properties determined by DINA model that students have with pass-fall decisions given via absolute and relative criteria is done, firstly by using absolute and relative criteria pass-fall decisions given about students and achievement percentage of departments are calculated. It is examined that to what extent these percentages correspond to total number of possessed properties determined by DINA model on a department level.

Achievement percentages of classes according to both relative and absolute criteria, and mean of properties possessed by students in departments determined by DINA model are given in Table 12 below.

Table 11.
Achievement percentages of departments via absolute and relative criteria, and mean of possessed properties

Department	Absolute	Relative	Mean of possessed properties
Biology	40.32	90.32	3.12
Computer and teaching technologies	84.93	94.52	4.85
Geography	73.33	100,00	4.73
Sociology	81.25	93.75	5.56
Physics	64.58	93.75	4.83
Chemistry	63,64	87.88	4.30
Mathematics	87,36	100,00	5,59
History	71,05	92,11	4,71
Philosophy	79,41	100,00	4,85
Art history	47,06	79,41	4,05
Mean	69,29	93,17	4,66

As is seen in Table 12, achievement ratios of students in lesson vary according to absolute and relative criteria in each department. For example, biology department has the lowest achievement level vis-à-vis other departments. Concordantly biology department has seven necessary properties to answer test items correctly in the ratio of 3.12. In view of criterion-referenced assessment, biology department is more successful than art history department despite possessed properties of art history department is more than biology department. According to type of assessment when the correlation among means of possessed properties of classes is examined, between DINA model mean of property with normative assessment spearman rho rank differences correlation coefficient is calculated as .94 and with criterion-referenced assessment it is calculated as .79. According to these results it is said that possession level of property determined by DINA model is more consistent with normative assessment.

These results are similar to Nartgün's study findings in that Nartgün (2007) examined the consistency among assessments done with relative and absolute criteria to represent students' achievement levels. In Nartgün's research it has been seen that normative assessment results is more representative for students' achievement level in comparison with criterion-referenced assessment.

CONCLUSION

In case of being used absolute and relative criteria when achievement level of departments in lesson and means of properties determined by DINA model that they possess are examined, it is observed that classifications being made via normative assessment are more consistent with DINA model classification. When achievement level of departments is examined, it is seen that biology department students succeed at the level of 90% in criterion-referenced assessment and its mean of possessed property is calculated as 3.12. The achievement level of same department is 40% according to normative assessment. Mean of possessed property of chemistry department students is 4.30 while achievement level of this class is 87% according to criterion-referenced assessment. It can be said that DINA model results is good at reflecting real condition related with possession of students to determined properties, that is, determining tacit properties that students posses (de la Torre; 2008a, 2008c, 2009b; de la Torre & Douglas, 2008; Cheng, 2010, Huebner, Wang, & Lee, 2009; Wenmin, 2006). In this sense it is seen that classifications being made by using absolute criteria give more appropriate classification validity.

The only aim of this research does not develop criteria to determine the validity of classifications. Especially DINA model and in general CDM (individualistic complete learning model) are assumed as newish methods and studies related to DINA model has become widespread since after 2005. In this respect it can not be said that there is enough scientific research and evidence to use the results of analysis being made by using DINA model to give unrecoverable decisions about students. However, it is sayable that DINA model is useful and sufficiently reliable to determine partial learning and to show superior and weak qualifications of students. In USA with "No Child Left Behind Act" approach usage of CDM models is supported on the purpose of identifying learning deficiencies of primary and secondary school students after taking a test and of giving more detailed feedback to family, teachers and authorities (McGlohen, 2004; Fagan, 2002; Huebner, 2010; Cheng & Chang, 2007). According to PISA 2000, defined as "PISA-shock" in Germany, due to the drop of secondary school students' success the search for educational reform has begun in country and considering cognitive models new performance standarts have been searched (Waldow, 2009; Krüger, 2003). As is seen BTM has been used in a widespread manner on the purpose of getting more and accurate information about students than standart tests.

In this study, the consistency between assessments made by using absolute criteria and tacit classes determined by DINA model is examined. The results of analysis show that the consistency level pass-fall decisions about students determined by absolute criteria and number of possessed properties determined by DINA model is 84.5%. It is determined that in assessment made by absolute criteria the inconsistency ratio especially for students about who fall decision is given is higher. As per absolute criteria it is necessary for a student to answer at least 30 questions correctly in a test with 50 items. However with DINA model decision can be given whether student has determined properties or not by using a lot fewer items via property-item interaction. It can be said that this is the one of the reasons that causes inconsistency between two methods and affects validity of decision phase. The results of study are evaluated in terms of consistency of two methods as well as number of properties that students possess. An important point in this study is students about who fall decision is given in normative assessment in spite of having most of the determined properties.

In case of giving passed decision about students according to relative criteria it is seen that group is substantially successful in the lesson. Students about who fall decision is given have 3 or less properties. This situation shows that unsuccessful students in lesson do not possess enough properties necessary to answer questions correctly. When condition of students about who pass decision is given is examined, contradictory results with DINA model classification can be observed. It is determined by DINA model that 87 of 442 students about who pass decision is given via relative criteria possess 3 or less properties. In this case consistency between DINA model classifications and pass decisions given via relative criteria is calculated at the level of 80%. In conclusion, there are no students assumed as unsuccessful in spite of having necessary properties in criterion-referenced assessment although one out of every five students is assumed as a successful despite of not having necessary properties.

REFERENCES

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*, 70 (6), 902-913

de la Torre, J. (2008a). *The generalized DINA model framework.* Unpublished manuscript. State University of New Jersey.

de la Torre, J. (2008b). *The generalized DINA model*. Uluslar arası Psychometric Society toplantısında sunulmuştur, Temmuz, Durham, NH.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively-based multiple-choice options. *Applied Psychological Measurement*, 33, 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, ss. 115–130

de la Torre, J. (2010). *The Generalized DINA Model Framework*, National Science Foundation Grant tarafından desteklenen araştırma raporu.

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*, Bahar 2010, Sayı 47, No:2, s: 227-249

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

de la Torre, J. .& Douglas, J. (2008). Model Evaluation and Multiple Strategies in Cognitive Diagnosis: An Analysis of Fraction Subtraction Data. *Psychometrika*. V73, N3, s. 595-624.

Fagan, T. (2002). *Implications of the No Child Left Behind Act for States in the Southwestern Region*, Southwest Texas: Educational Development Laboratory, Austin.

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

Haertel, E.H. (1990). Continuous and discrete latent structure models of item response data.

Psychometrika 55, 477–494.

Huebner, A. (2010). An Overview of Recent Developments in Congnitive Diagnostic Computer Adaptive Assessments, *Practical Assessment, Research & Evaluation*, Vol15, Sayı 3. Ocak

Huebner, A., Wang, B. & Lee S. (2009). *Practical Issues Concerning the Application of the DINA Model to CAT Data,* D. J. Weiss (Ed.), Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.

Krüger, S. (2003). *The PISA-shock and its Consequences: The Future of Libraries for Childeren in Germany*, World Library and Information Congress: 69th IFLA General Conference and Council, 1-9 ağustos 2003, Berlin

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

Mcglohen, M.K. (2004). *The Application of a Cognitive Diagnosis Model via an Analysis of a Large-Scale Assessment and a Computerized Adaptive Testing Administration.* Yayımlanmamış Doktora Tezi, The University of Texas Austin

Murphy, K. R. & Davidshofer, C. O. (2001). *Psychological testing*. (5th ed.). New Jersey: Prentice-Hall,Inc

Nartgün, Z. (2007). Aynı Puan üzerinden Yapılan Mutlak ve Bağıl Değerlendirme Uygulamalarının Notlarda Farklılık Oluşturup Oluşturmadığına İlişkin Bir İnceleme. *Ege Eğitim Dergisi* 2007(8), sayı: 1 ss: 19,40

Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement* 20, 345–354.

Waldow, F. (2009). What PISA Did and Did Not Do: Germany after the ‘PISA-shock’, *European Educational Research Journal*, Sayı: 8 No: 3 ss: 476-483

Wenmin, Z. (2006). *Detecting Differential Item Functioning Using the DINA Model.* The University of North Carolina at Greensboro. Unpublished doctorate thesis. Greensboro

e-International Journal of Educational Research
Volume: 5 Issue: 1 Winter 2014 pp. 1-19

Bağıl ve Mutlak Değerlendirme ile DINA Modele Göre Yapılan Sınıflamaların Geçerliğinin Karşılaştırılması

Yrd. Doç.Dr. Tahsin Oguz Başokçu[2]
Ege Üniversitesi-Türkiye
tahsin.oguz.basokcu@ege.edu.tr

Doç.Dr. Hülya Kelecioğlu
Hacettepe Üniversitesi-Türkiye
hulyaebb@hacettepe.edu.tr

Genişletilmiş Özet

Problem: Bu çalışmada, Bilişsel Tanı Modellerinden(BTM) biri olan DINA model ve geleneksel yöntemlerle yapılan sınıflamaların geçerliği incelenmiştir. DINA model, ikili yetenek modellerine benzeyen bir örtük sınıf modelidir. Bu model çoğu BTM'de görüldüğü gibi madde özellik ilişkisini temel alır. Modelin iyi işleyebilmesi, bir maddenin doğru cevaplanması için gerekli olan özelliklerin doğru belirlenmesine bağlıdır. DINA modelde özelliklerle ilişkilendirilen maddelerin dağılımını gösteren bir Q matrisi hazırlamak gerekmektedir. Q matris her bir maddenin doğru cevaplanması için gerekli olan özelliklerin belirlendiği 1-0 örüntüsüdür. Bu matriste bir madde tek bir özellikle ilişkilendirilebildiği gibi birden çok özelliklede temsil edilebilir.

DINA model cevaplayıcının gözlenen yeteneğinin altına yatan örtük özelliği ortaya çıkartmayı amaçlamaktadır. Bu anlamda model, örtük özellikle gözlenen özellik arasındaki ilişkiyi olasılıkla temellendirmekte ve her madde için iki madde parametresinin sınıflanmasını sağlamaktadır. Bunlar s "kaydırma" (slip) ve g "tahmin" (guess) parametreleridir. Bu parametreler yardımıyla model öğrencileri her madde için temel iki sınıfta değerlendirir. Bu sınıflardan ilki yokluk sınıfı (null class) yani beklenen hiçbir beceriye sahip olmayan öğrencilerin oluşturduğu grup ve diğeri de tam sınıf (full class) yani bütün becerilere sahip olan öğrencilerin sınıfıdır. Model cevaplayıcıların maddelere verdiği yanıtlardan yola çıkarak her bir cevaplayıcının Q matris tarafından tanımlanan özelliklere sahip olup olmadıklarını belirler.

Araştırma karar verme amacı taşıyan testlerde geleneksel yöntemlerle yapılan sınıflamalar ile DINA modelle belirlenen sınıflamaların, tutarlığını incelemektir. Testi alan bireyler hakkında bağıl ya da mutlak değerlendirme ölçütlerine göre verilen geçti kaldı kararlarının öğrencilerin ölçülen özelliğin alt yetenek düzeylerine sahip olup olmamaları bakımından ne derece uyum sağladığı çalışma kapsamında incelenmiştir.

Yöntem: Araştırma, geliştirilen bir testin BTM'ye uygunluk düzeyini belirleme bakımından var olan bir durumu ortaya çıkartmak amacı taşımaktadır. Bu nedenle bu araştırma betimsel bir araştırma olarak değerlendirilebilir. Bunu yanında

[2] Bu makale Doktora tezinden üretilmiştir.

sınıflama geçerliğine ilişkin farklı teknikleri karşılaştırma olanağı da verdiğinden kuramsal bir araştırma olarak da ele alınabilir.

İki yöntem arasında karşılaştırılmaların yapılabilmesi için DINA modelle analiz yapılmaya uygun "eğitimde ölçme ve değerlendirme" dersine ait bir ölçme aracı geliştirilmiştir. Ölçme aracındaki her bir maddeyi doğru cevaplamak için gerekli olan özellikler uzmanlar tarafından belirlenmiş ve madde özellik ilişkisini gösteren Q matrisi uzman uyumu gözetilerek hazırlanmıştır.

Araştırmanın evrenini Ege Üniversitesinde Ölçme Değerlendirme dersi alan 471 öğrenci oluşturmaktadır. Ölçme aracı geliştirildikten sonra 10 farklı bölüm öğrencilerinin oluşturduğu 471 kişilik çalışma grubuna uygulanmıştır. Uygulama sonucunda öğrencilerin ham puanları mutlak ölçüt ve bağıl ölçüt kullanılarak geçti kaldı kararlarına göre sınıflandırılmıştır. Bu sınıflandırma ile DINA modele dayalı sınıflandırmalar karşılaştırılmıştır.

Bulgular ve Sonuç: Mutlak ölçüt kullanılarak yapılan değerlendirme ve DINA model sınıflandırmalarının karşılaştırması sonucunda, öğrenciler hakkında mutlak ölçütle belirlenen geçti kaldı kararlarıyla DINA modelle belirlenen öğrencilerin sahip olduğu özellik sayısı arasında %84.5 oranında bir uyum olduğu gözlenmiştir. Çalışma sonuçları iki yöntemin uyumu kadar, hakkında geçti ya da kaldı kararı verilen öğrencilerin sahip oldukları özellik sayıları bakımından da değerlendirilmelidir. Bu noktada araştırmada dikkat çekilmek istenen önemli bir noktada belirlenen özelliklerden çoğuna sahip olmasına rağmen mutlak değerlendirme içinde hakkında kaldı kararı verilen öğrencilerdir.

Bağıl ölçüt kullanılarak yapılan değerlendirmelerde hakkında kaldı kararı verilen öğrenciler için iki yöntem aynı sonuçları vermiştir. Bununla birlikte DINA model sınıflamalarıyla bağıl ölçütle verilen geçti kararları arasındaki uyum yaklaşık %80 düzeyinde hesaplanmıştır. Sonuç olarak bağıl değerlendirmede gerekli özelliklere sahip olduğu halde dersten başarısız sayılan öğrenci bulunmamakla birlikte hakkında geçti kararı verilen her beş öğrenciden biri gerekli özelliklere sahip olmadığı halde dersten başarılı sayılmıştır.

Mutlak ve bağıl ölçüt kullanıldığı durumda bölümlerin dersten başarı yüzdeleri ile DINA modelle belirlenen sahip oldukları özellik ortalamaları incelendiğinde, mutlak değerlendirme sonucunda yapılan sınıflamaların DINA model sınıflamalarıyla daha uyumlu olduğu gözlenmiştir. Bu anlamda mutlak ölçüt kullanılarak yapılan sınıflamaların gerçek duruma daha uygun bir sınıflama geçerliği verdiği görülmektedir.

Anahtar Kelimeler: Bilişsel Tanı Modelleri, DINA Model, Sınıflama geçerliği, Q Matriks