Bulletin of Applied Mathematics and Mathematics Education
Volume 3, Number 1, 2023 pp. 31-44
https://dx.doi.org/10.12928/bamme.v3i1.8964
e-ISSN 2776-1029, p-ISSN 2776-1002

Investigating the 8th-grade Afghan students' mathematics status and skills using the cognitive diagnostic model

Mohammad Nasim Wafa

Ghor Institute of Higher Education, Pozaliche, Firozkoh City, Ghor, Afghanistan

*Corresponding E-mail: nasim_wafa58@yahoo.com

ARTICLE INFO

Article History
Received 1 September 2023
Revised 6 November 2023
Accepted 29 December 2023

Keywords
Cognitive diagnosis models
DINA model
DINO model
TIMMS

How to cite this article:
Wafa, M. N. (2023). Investigating the 8th-grade Afghan students' mathematics status and skills using the cognitive diagnostic model. *Bulletin of Applied Mathematics and Mathematics Education*, *3*(1), 31-44.

ABSTRACT

Cognitive diagnostic models (CDMs) are multidimensional multivariate verification flow models with complex structure. In this research, these models were used to investigate the status of eighth grade high school students in mathematics using the TIMMS questionnaire. The cognitive diagnostic test based on 13 attributes including 32 questions was conducted on a sample of 274 students who were selected based on the multi-stage cluster sampling method among the students of Firuzkoh city. IRT and RESMA models were used to determine the psychometric properties of the questions. Data analysis using DINA and DINO models in cognitive diagnostic modeling of mathematics showed that 13 attributes explain the mathematical performance of eighth-grade students. The result shows that Afghan students have a weak mastery level in most attributes compared to 45 other countries (the countries that were included in the TIMMS questionnaire) also general results show that the examinees perform better in the field of numbers (0.49), while they perform worse in data and chance (0.12). Moreover, there exists some difference in estimating item parameters under the DINA and DINO models, such as Item 3 and Item 27. One possible explanation is that the DINA model is completely compensatory while the DINO model is fully non-compensatory. Similar to the results under the DINA model, the SEs of guessing parameters are lower than those of slipping parameters under the DINO model.

Introduction

Cognitive diagnosis model

Cognitive diagnostic models (CDM) are among several discrete latent variable models that aim to investigate and diagnose subjects' mastery over a group and a set of discretely defined characteristics by providing detailed information about their specific strengths and weaknesses is done (Wafa, 2019; Wang, 2010). Since when this model has been used, it has not been determined definitively that these independent developments emerged from different directions and different dates. First, from the theory of ideas of identity and classification, which can be seen in the mastery model by Macready & Dayton (1977) and in terms of limited latent class models by Haertel (1989) has been found. Secondly, this model of item response theory with early approaches in the multi-component model by Whitely (1980) and that this model from the point of view of mathematical psychology, and here especially in the context of knowledge space theory, and also about the nature of this model e.g. research article.

Based on the multiplicity and application of this model from different approaches (Doignon & Falmagne, 1999), CDMs have many names, for example diagnostic classification models, cognitive psychometric models or structured item response theory models. This model can discuss in all CDM methods that a set of fundamental and latent group skills (i.e. competencies) underlie the tested ability. Furthermore, all CDM methods determine the possession and non-possession of these skills (i.e. the skill classes) in the test population and for the individual students.

In the assessment of cognitive diagnosis, in addition to estimating a person's ability and skills in the structure or structures being measured, a profile is provided for each person that shows his mastery or lack of mastery in a set of basic skills or cognitive attributes that are predetermined to respond to It specifies that the test questions are necessary and necessary. To be more precise, in these models, according to the person's mastery or lack of mastery of the skills and attributes needed to answer the question correctly, the probability of the person's correct answer to the question is estimated. In these models, based on the mastery status of people in each of the basic skills and attributes required by the questions, they can be divided into two groups, dominant and non-dominant, and if necessary, into 3 groups, dominant, non-dominant and indeterminate.

DINA model

Here we introduce the model called Deterministic Input Noisy Output "AND" gate (DINA) this models the first time introduced by (Haertel 1989; Junker & Sijtsma 2001, din function in R package). In literature, if strategies to solve the items are considered, the DINA models can be classified into two categories: the single-strategy DINA models and the multiple-strategy DINA models. A distinguishing feature of the multiple-strategy DINA model is that it incorporates multiple Q-matrices in order to specify the different strategies that suffice to solve the examination problems. Interested readers in the multiple-strategy DINA models can refer to (de la Torre & Douglas, 2008). In this article, we only consider the single strategy DINA.

Assume a test contain N students and J items which require K attributes. Let Xij denote the manifest response of student i to item j and $\alpha_i = [\alpha_{i1}, \dots \alpha_{iK}]$ denote the student's possessed skills. Let the Q-matrix be a $J \times K$ matrix with the j, k entry if the correct application of attribute k influences the probability of correctly answering the j_{th} item, and equals 0 otherwise. The vector $q_j = [q_{j1}, \dots q_{jK}]$ denotes the q_{th} row in the Q-matrix.

Building the formula below, two core procedures are considered, namely, the deterministic and probabilistic parts (George et al., 2016). In the deterministic process, the ideal responses are determined by the inputs α_i, q_j and the conjunctive rule. Specially, the ideal response for the DINA models is specified as,

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

Based on the above equation, the students who possesses these all required attributes are expected to master the item, that is, $\eta_{ij} = 1.$ For the students who are not expected to master the item, $\eta_{ij} = 0$.

For the probabilistic parts, the probability of lucky guesses or careless slips (Templin, & Henson, 2010), which is quantified as the guessing(g) parameter and the slipping (s) parameter.

10.12928/bamme.v3i1.8964

In the DINA model, $s_j = P\left(X_{ij} = 0 \middle| \eta_{ij} = 1\right),\ j = 1,2, \ldots . J$, which denotes the probability of students who have all attributes slip and answer to an item incorrectly, $g_j = P\left(X_{ij} = 1 \middle| \eta_{ij} = 0\right),\ j = 1,2, \ldots , J$, which denotes the probability of the students who don't have all attributes guess and answer to an item correctly.

The DINA models combined these two parameters and ideal responses calculates the probability of the student i to solve the item j, as

$$P\left[X_{ij} = 1 \middle| \boldsymbol{\alpha_i}, s_j, g_j\right] = (1 - s_j)^{\eta_{ij}} \cdot g_j{}^{(1-\eta_{ij})} = \begin{cases} 1 - s_j & for\ \eta_{ij} = 1 \\ g_j & for\ \eta_{ij} = 0 \end{cases}$$

The joint likelihood function is

$$P[X|\boldsymbol{\alpha_i}, s, g] = \prod_{i=1}^{N}\prod_{j=1}^{J}\left[(1 - s_j)^{x_{ij}} \cdot s_j{}^{1-x_{ij}}\right]^{\eta_{ij}}\left[g_j{}^{x_{ij}}(1 - g_j)^{1-x_{ij}}\right]^{1-\eta_{ij}}$$

We remind you that in this model, the required number of attributes is different depending on the item, that is, the feature of one item will be different from another item. In this DINA model, (Guess and slip) two parameters are needed for each item. In this model, it is assumed that the probability of a correct answer to the question, provided that at least one skill is not mastered, does not depend on the number and type of required skills that the person has not mastered. In fact, this model divides people into two classes in each question. In one class, there are people who have mastered all the skills or attributes measured by the question, and in the other class there are respondents who have not mastered at least one of the skills required by the question. The important feature of this model is that no distinction is made between people who do not have the different attributes or skills required by the question. That is, all respondents who lack at least one of the skills measured by the question are placed in the same class, regardless of which skills or how many skills they lack, when the skills required for a question are of equal importance, the DINA model it is a suitable model. Various studies have examined the DINA model.

DINO model

Compared with the DINA model, the DINO model also separates the latent classes into two groups for each item based on ideal responses. However, for the DINO model, it is supposing that an item can be answered correctly if at least one of the required attributes involved in the item has been mastered. According this assumption, the ideal response for the DINO model is defined as,

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

Given the slipping and guessing parameters

$$s_j = P\left(X_{ij} = 0\middle|_{\eta_{ij} = 1}\right),\ j = 1,2, \ldots . J, \text{and } g_j = P\left(X_{ij} = 1 \middle| \eta_{ij} = 0\right),$$

the joint likelihood is written as,

$$P[X|\boldsymbol{\alpha_i}\, s, g] = \prod_{i=1}^{N}\prod_{j=1}^{j}\left[(1 - s_j)^{x_{ij}} \cdot s_j{}^{1-x_{ij}}\right]^{\eta_{ij}}\left[g_j{}^{x_{ij}}(1 - g_j)^{1-x_{ij}}\right]^{1-\eta_{ij}}.$$

The special attribute of the DINO model is that this model is used to analyze the answers of psychological research. That the probability of a negative answer as a sliding parameter $\eta_{ij} = 1$ and the probability of a positive answer by the group as a guessing parameter that is by the group is $\eta_{ij} = 0$. Based on this, the item response function of the DINO model calculates the probability of endorsing item j for the given group η_ along with the guessing parameter g_j and the sliding parameters_j.

Although we have discussed both DINA and DINO models, and regarding the duality of these two models with the major difference in how to calculate the latent response variable, you can see Liu, Xu, and Ying (2012).

Duality of the DINA model and the DINO model

The DINA and DINO models are two popular cognitive diagnosis models (CDMs) for educational assessment and represent different views on how the mastery of cognitive skills and the probability of a correct item response are related. Recently, Liu, Xu, and Ying (2012) demonstrated that the DINO model and the DINA model share a "dual" relation and which of the two models is fitted to a given data set is essentially irrelevant because the results are identical.

As J. Liu et al. (2012) discovered and proved, the DINA model and the DINO model are technically identical under certain transformations of (a) the examinees' attribute profiles, (b) their observed item scores, and (c) the model parameters. This means that one model can be expressed in terms of the other and both models can be fitted by the same software. (As an aside, note that the characterization of the special relation between the DINA model and the DINO model as "dual" deviates from the well-defined meaning of this term in operations research; for details, consult Papadimitriou & Steiglitz, 1998.)

Item response theory (IRT) model

In all cognitive diagnosis models based on IRT (CDMI), the probability of correct answer to the question is defined as a function of a set of discrete attributes measured by the question. The IRT model has been widely used in large-scale assessments to measure the ability of students participating in group tests. The main idea of IRT is the item response function (IRF), which IRT examines to determine the probability of a given answer as a function of the actual ability of the student participating in the test.

The simplest IRT model for binary responses ($Y = 0$ if the question or item has been answered incorrectly and $Y = 1$ if it has been answered correctly) is the one-parameter logistic (1PL) model with an item difficulty parameter for each item, most commonly known as the Rasch (Rasch, 1961) model. Under the Rasch model, the probability that person j with latent ability θj gives a correct response ($Yij = 1$) to item i with difficulty βi is

In addition, these models require a Q matrix that specifies the attributes required for the questions. Q matrix is a two-symbol $n \times k$, where k is the number of attributes or skills to be measured and n is the number of test questions. For a specific element of the Q matrix in the n_{th} row and k_{th} column, the number 1 indicates that the question measures k attribute or skill, and the number zero indicates that the question does not measure the desired attribute. In other words, in this matrix, the number 1 indicates that to give the correct answer to the nth question (Y=1 for coreect item Y=0 is for incorrect item), the k-th attribute or skill is needed, and the number zero indicates that the k-th attribute is not required to give the correct answer to the nth question.

$$P(Y_{ij} = 1|\theta_j = \frac{exp(\theta_j - \beta_i)}{1 + exp(\theta_j - \beta_i)}$$

Such a model with two item parameters, item difficulty and item differentiation, is called a two-parameter logistic model (2PLm) (Birnbaum, 1968) and is defined as such.

$$P(Y_{ij} = 1|\theta_j = \frac{exp\{\alpha_i(\theta_j - \beta_i)\}}{1 + \{\alpha_i(\theta_j - \beta_i)\}}$$

Here, it is assumed that the answers related to an item are independent of the answers to any

 10.12928/bamme.v3i1.8964

other item, subject to the individual's ability, which is called local independence. The joint probability of a response vectory_j, given the latent ability θ_j, can be expressed as follows.

$$P(y_j|\theta_j) = \prod_{i=1}^{I} p_{ij}^{y_{ij}} \{1 - p_{ij}\}^{1-y_{ij}}, \text{ where } p_{ij} = P(y_{ij} = 1|\theta_j)$$

TIMMS 2011

Trends in International Mathematics and Science Study (TIMSS), which is an international assessment of mathematics and science in the fourth and eighth grades, and this assessment is conducted in 63 countries. This program started in 1995 and is implemented every four years. That is, in the years 1999, 2003, 2007, 2011, 2015, 2019 and 2023, countries such as Singapore, South Korea, China, the United States of America and other countries were evaluated, and in 2011, Singapore took the first place. In 2011, national student representative samples from 63 countries and 14 benchmark entities (such as states) participated in TIMSS. Countries and benchmarking participants in the Criterion can participate in the Fourth Grade, Eighth Grade or both: Fifty-two countries and seven benchmark assessments, with fourth- five countries and fourteen benchmarks participating in the Eighth-grade assessment. Several of the countries were fourth and eighth-grade students were expected to find the TIMSS assessments too difficult, administered the fourth and eighth-grade assessments to their sixth and ninth-grade students.

This assessment is done every four years, which started in 1995. It is an international assessment of mathematics and science in the fourth and eighth grades. This assessment is shared by 63 countries and 14 benchmarking bodies (countries' regional jurisdictions, such as states) every four years. Participating countries can participate in 4th grade, 8th grade or both. In the recent TIMMS 2011 assessment, Singapore had the highest correct assessment among 63 countries. But unfortunately, Afghanistan is not included in this evaluation. We tried to prepare this questionnaire with the presence of researchers and professors of Afghan schools and removed some questions according to the eighth-grade curriculum so that the eighth grade level of Afghan schools is standard and equal. Fifty-two countries and seven benchmarks participate in the eighth-grade assessment, with countries from fourth through fifth and fourteen benchmarks. In several countries, 4th and 8th grade students were expected to find the TIMSS assessment very difficult.

In the released TIMSS 2011 math test, there are five areas in mathematics (number and operations, algebra, geometry, measurement and data analysis, and probability) that in total, more than 600,000 students participated in this assessment. TIMSS 2011 is the continuation of the series of international assessments in mathematics and science conducted by the International Association for the Assessment of Educational Progress (IAAEP). For more information see the TIMMS 2011 book.

Item fit

RMSEA is one of the CDMs in the R package as an item fit statistic, which stands for root mean square error of approximation (George ,et all 2016), which indicates how An item with a good model. The item-fit RMSEA for item-j- compares the model-predicted item response probabilities $P(Xj = 1|\alpha l)$ with the predicted absolute number of correct responses $N'(Xj = 1|\alpha l)$ in each skill class αl

$$RMSEA_j = \sqrt{\sum_{l=1}^{L} p(\boldsymbol{\alpha}_l)\left[P(X_j = 1|\boldsymbol{\alpha}_l) - \frac{N'(X_j = 1|\boldsymbol{\alpha}_l)}{N'(X_j|\boldsymbol{\alpha}_l)}\right]^2}$$

Here $p(\boldsymbol{\alpha l})$ frequency of students classified in skill class α_l and $N(Xj|\boldsymbol{\alpha l})$ the predicted total number of responses (i.e., correct and incorrect ones) to item j given by students in skill class $\boldsymbol{\alpha_i}$. Kunina-Habenicht et al. (2009) the evaluate below the rule items recommend RMSEA values. If the RMSEA values of the items are below 0.10, the fit indices of the items with the values of $RMSEAj > 0.10$ indicate a poor fit, and the items below 0.05 indicate a good fit.

Item discrimination index (IDI)

In DINO and DINA models the additional constraint $gj < 1 - sj$ ensures that the probability of mastering an item in possession of all required skills without slipping is higher than the probability of guessing an item while lacking at least one required skill. The extent to which this limitation is not considered in the estimation process may be checked with the item discrimination index $\mathrm{IDI_j} = 1 - \mathrm{s_j} - \mathrm{g_j}$ (Lee, de la Torre, and Park 2012, function summary.din), where negative values IDI indicates a violation. From this limitation, the IDI may also be considered as a diagnostic index that reports for each case how the difference between students who have all skills (i.e. response probability $1 - \mathrm{s_j}$) and students who do not have at least one skill. (i.e. guessing with probability g_j) is distinguished). Thus, IDIs approaching 0 indicate low discrimination. While those approaching 1 are indicative of good differentiation or "identification" of the item.

Method

In this research, a total of 274 Afghan students in 8 schools participated in the study of the TIMSS 2011 questionnaire. In each classroom, eight different class of Afghanistan mathematics tests were assigned randomly to students. Based on this model, the models were designed by 5 mathematicians with a master's degree and Bachler degree mathematics and mathematical education with 10, 8, 8, 7 and 5 years of teaching and writing experience. The researcher worked with these five people to develop a cognitive model as well as diagnostic test questions. These five teachers prepared a clean questionnaire based on the eighth grade in accordance with the Afghan Curriculum, they choose 32 questionnaires from 88 TIMMS 2011 questionnaire based on the proposed model, thirteen traits were introduced as the underlying traits for high school eighth-grade mathematics. Table 2 shows a list of eight traits and related skills. In order to compare between the subgroups, four of these schools were selected in the rural group and four in the urban schools. Schools that are geographically in faraway areas from Firuzkoh city and also schools that were located in the middle of Firozkoh city.

This test, which is taken from the TIMMS test, has 32 questions and each question has four options, one of which is correct. And when the answer was wrong, it was coded with 0 in Q-matrix and when the answer was correct, it was coded with 1. This number is 274 people, of which 53% are men (that is, 145 people) and 129 are women, of which the corresponding percentage is 47.00%, and the average age of the examinees is about 17-18 years.

 10.12928/bamme.v3i1.8964

Results and discussion

Table 1 shows the rate of correct answers for each item based on the observed answers of all subjects. According to the results of the table the highest frequency of correct response belongs to Item 4 at 0.489 and the lowest at Item 21 which is (0.088).

Table 1. Percentage of correct responses probability

Item	Correct rate	Item	Correct rate
1	0.215	17	0.285
2	0.208	18	0.296
3	0.222	19	0.157
4	0.489	20	0.118
5	0.401	21	0.088
6	0.237	22	0.198
7	0.332	23	0.120
8	0.369	24	0.147
9	0.300	25	0.248
10	0.204	26	0.263
11	0.338	27	0.285
12	0.336	28	0.193
13	0.201	29	0.255
14	0.201	30	0.369
15	0.160	31	0.212
16	0.300	32	0.120

Table 2 shows the percentage of correct responses probability based on attributes Based on the results in the table, the highest probability of mastering is 0.348 at attribute A_N2 and A_N3,and the lowest one is at A_D1 its (0.130). So, it may conclude that the examinees are good at Number domain while do worst at Data and Chance.

Table 2. Percentage of correct responses probability based on attributes

Attribute	Probability	Attribute	Probability
A_{N1}	0.241	A_{A4}	0.234
A_{N2}	0.348	A_{G1}	0.177
A_{N3}	0.348	A_{G2}	0.251
A_{N4}	0.269	A_{G3}	0.200
A_{A1}	0.223	A_{NG4}	0.263
A_{A2}	0.191	A_{D1}	0.130
A_{A3}	0.170		

Table 3. Comparative Fit for Different DINA and DINO model

Models	loglike	AIC	BIC	RMSEA mean
DINA	-4257.720	25025	54881	0.229
DINO	-4266.705	25043	54899	0.191

Table 4. Estimated item parameters and standard errors under the DINA model

Item	Guess est.	Guess SE	Slip est.	Slip SE
01	0.162	0.022	0.418	0.127
02	0.091	0.014	0	0
03	0.17	0.024	0.27	0.081
04	0.364	0.032	0.176	0.071
05	0.19	0.022	0.027	0.010
06	0.084	0.019	0.271	0.045
07	0.203	0.033	0.254	0.044
08	0.263	0.04	0.296	0.053
09	0.274	0.029	0.58	0.126
10	0.183	0.035	0.731	0.062
11	0.002	0.0002	5.05E-11	1.69E-11
12	0.333	0.034	0.663	0.089
13	0.174	0.022	0.476	0.137
14	0.197	0.025	0.753	0.119
15	5.49E-12	1.73E-12	0.502	0.066
16	0.221	0.037	0.355	0.057
16	0.241	0.027	0.261	0.098
18	0.188	0.033	0.255	0.043
19	0.081	0.016	0.056	0.019
20	0.029	0.013	0.436	0.063
21	0.027	0.012	0.658	0.07
22	0.185	0.024	0.637	0.154
23	0.072	0.017	0.373	0.095
24	0.074	0.014	0	0
25	0.203	0.027	0.55	0.113
26	0.248	0.041	0.713	0.058
27	1.87E-08	4.02E-09	0.2	0.034
28	0.169	0.025	0.636	0.111
29	0.142	0.021	0.274	0.075
30	0.186	0.03	0.091	0.019
31	9.10E-16	2.58E-16	0.158	0.028
32	3.19E-110	9.51E-111	0	0
Mean	0.154		0.346	

In this particular case, the AIC and BIC are used to compare the two models and aid in model selection. Table 3 contains a summary of the number of estimated item parameters, AIC, and BIC for the gdina model, for the dina Model and dino model As can be seen from the table, the AIC is very similar between the two different models, although the dina model has a slightly smaller value.

 10.12928/bamme.v3i1.8964

Thus, the AIC provides weak evidence that the full model should be used. However, the BIC, which has a same penalty for additional parameters, is smaller for the dina model suggesting that the dina and dino model should be used. Therefore, the AIC and BIC seem to suggest that the proposed dina of the model is feasible.

Table 4 shows the estimators and SEs for item parameters under the DINA model. According to Table 1, Item 32 has the lowest guessing parameter, while Item 4 has highest one. In terms of guessing parameter, Item 2 arrived at its highest value and Item 14 has the lowest value.

Actually, the numeric value of items parameters could assess the goodness of fit between the diagnostic assessment design, the response data, and the postulated DINA model. Generally speaking, the smaller the guessing and slipping parameters, the better the model fit (Ravand, Barati, & Widhiarso. 2012). For the specific items, those items with lower slipping and guessing parameters are more informative (Rupp et al., 2010), which is consistent with the item discrimination index proposed by de la Torre (2008). In this point, Item 11 and Item 32 are more informative than other items under DINA model.

We recall that the average parameters of guessing is 0.153 and slip is 0.346. The most informative items in the test are those that are less likely to slip and guess (Rupp et al., 2010). Overall, the guess parameters and small slip indicate a good fit between the diagnostic assessment design, the response data, and the hypothesized DINA model. The average guessing parameter shows that for students who do not master all the skills required for an item, there is still an average of 15.37% chance of selection. The correct answer and mean slip parameter show that for students who master all the skills required for an item, there is still an average of 34.61% chance of choosing the wrong answer.

The SEs of all guessing parameters were below 0.05, while the SEs of some slipping parameters are bigger than 0.10, including Item 1, Item 9, Item 13, Item 14, Item 22, Item 25 and Item 29. For these items, there is one thing in common: two or more than two attributes are measured by them. In context of DINA model, the increment of required attributed would bring burden to estimate slipping parameters, especially under small sample size. So, the results on guessing parameters are more reliable than that on slipping parameters.

Table 5 shows the estimators and SEs for guessing and slipping parameters under DINO model. In terms the estimated item parameters, Item 32 has lowest value of guessing parameter as well as slipping parameters. It indicated that this item is most discriminative among all items. Moreover, there exists some difference in estimating item parameters under the DINA and DINO model, such as Item 3 and Item 27. One possible explanation is that the DINA model is completely compensatory while the DINO model is fully non-compensatory. Similar to the results under DINA model, the SEs of guessing parameters are lower than that of slipping parameters under DINO model.

Table 5. Estimated item parameters and standard errors under the DINO model

Item	Guess est.	Guess SE	Slip est.	Slip SE
01	0.118	0.021	0.574	0.0662
02	0.011	0.003	0.326	0.0543
03	0.087	0.021	0.579	0.0563
04	0.374	0.034	4.29E-16	1.21E-16
05	0.292	0.031	0.109	0.0265
06	0.124	0.02	0.172	0.0727
07	0.232	0.027	0.151	0.0671
08	0.284	0.03	0.197	0.0868
09	0.264	0.034	0.611	0.0653
10	0.178	0.025	0.663	0.1171
11	0.212	0.027	0.276	0.055
12	0.264	0.031	0.442	0.074
13	0.148	0.027	0.721	0.056
14	0.036	0.013	0.549	0.055
15	0.013	0.002	0.302	0.076
16	0.231	0.028	0.395	0.102
17	0.246	0.036	0.649	0.051
18	0.272	0.032	0.410	0.086
19	0.081	0.021	0.492	0.065
20	0.031	0.009	0.437	0.106
21	0.016	0.005	0.001	0.0002
22	0.182	0.032	0.779	0.044
23	0.029	0.011	0.453	0.063
24	0.075	0.018	0.624	0.067
25	0.219	0.029	0.665	0.08
26	0.23	0.028	0.537	0.121
27	0.17	0.023	0	0
28	0.14	0.025	0.590	0.071
29	0.133	0.025	0.499	0.064
30	0.221	0.027	0	0
31	0.029	0.006	0	0
32	1.08E-145	4.70E-146	0.385	0.057
Mean	0.119		0.346	

As shown in Table 4 and Table 5, the highest DINA - IDI belongs to item 27 and the lowest IDI belongs to item 26 and the highest DINO- IDI belongs to item 21 and lowest IDI belongs to item 22.

Finally in Table 6 shows, the Overall Mean score percentage of TIMMS (2011) in this study was compared with 45 countries and 14 states. The highest score (80.59) in this list belonged to

 10.12928/bamme.v3i1.8964

Singapore while Afghanistan lied in the last part of the list with a score of (24.34).

Table 6. Overall mean score of TIMMS

Number	Countries	Overall Mean of TIMMS	Number	Country	Overall Mean of TIMMS
1	Singapore	80.593	31	Dubai, UAE	47.375
2	Korea, Rep. of	77.750	32	Sweden	46.937
3	Chinese Taipei	77.625	33	Romania	46.468
4	Hong Kong SAR	75.593	34	Lebanon	46.406
5	Japan	66.625	35	United Arab Emirates	44.968
6	Russian Federation	65.656	36	New Zealand	44.312
7	Massachusetts, USA	63.781	37	Alabama, USA	44.031
8	North Carolina, USA	62.937	38	Georgia	42.906
9	Minnesota, USA	62.875	39	Malaysia	42.406
10	Cuebec, Canada	59.531	40	Norway	42.093
11	Indiana, USA	58.125	41	Turkey	41.250
12	Israel	57.093	42	Abu Dhabi, UAE	41.218
13	United States	55.125	43	Macedonia	40.750
14	Connecticut, USA	54.843	44	Tunisia	38.562
15	Florida, USA	54.218	45	Thailand	37.250
16	Colorado, USA	53.937	46	Qatar	37.125
17	Hungary	53.750	47	Palestinian Nat'l Auth.	35.562
18	Italy	52.937	48	Jordan	35.406
19	Ontario, Canada	52.750	49	Iran, Islamic Rep. of	34.343
20	Alberta, Canada	52.750	50	Bahrain	34.343
21	Lithuania	52.656	51	Saudi Arabia	34.031
22	Kazakhstan	52.312	52	Chile	32.437
23	Finland	52.062	53	Botswana	32.400
24	Slovenia	51.468	54	Austria	32.310
25	Armenia	51.281	55	Oman	31.125
26	California, USA	51.125	56	Indonesia	30.906
27	Australia	50.531	57	Syrian Arab Republic	30.156
28	England	49.718	58	Morocco	29.843
29	Czech Republic	49.541	59	Ghana	25.718
30	Ukraine	49.406	60	Afghanistan (Ghor)	24.343

Conclusion

In this study, we mainly focus on assessing the questionnaires of TIMSS 2011 in Afghanistan. Two commonly used CDMs were employed, including DINA and DINO models. With the aid of CDMs, not only item parameters, but also skill profile for each student could be estimated. According to the result, Item 32 is more informative than other items under both models. On the whole, the examinees are good at Number domain while do worst at Data and Chance. In table 7 and 8 item 32 has lowest value of guessing parameter as well as slipping parameters. It indicated that this item is most discriminative among all items. Moreover, there exists some difference in estimating item parameters under the DINA and DINO model, such as Item 3 and Item 27. One possible explanation is that the DINA model is completely compensatory while the DINO model is fully noncompensatory. Similar to the results under DINA model, the SEs of guessing parameters are lower than that of slipping parameters under DINO model.The result shows that Afghan students have a weak mastery level in most attributes compared to 45 other countries (the countries that were included in the TIMMS questionnaire) also general results show that the examinees perform better in the field of numbers (0.49), while they perform worse in data and chance (0.12). Moreover, there exists some difference in estimating item parameters under the DINA and DINO model, such as Item 3 and Item 27. One possible explanation is that the DINA model is completely compensatory while the DINO model is fully noncompensatory. Similar to the results under DINA model, the SEs of guessing parameters are lower than that of slipping parameters under DINO model.

References

Chiu, C. Y., & Köhn, H. F. (2015). Consistency of cluster analysis for cognitive diagnosis: The DINO model and the DINA model revisited. *Applied Psychological Measurement, 39*(6), 465-479.

Choi, K. M., Lee, Y. S., & Park, Y. S. (2015). What CDM Can Tell About What Students Have Learned: An Analysis of TIMSS Eighth Grade Mathematics. *Eurasia Journal of Mathematics, Science & Technology Education, 11(*6).

de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika, 81*(2), 253–273.

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.

George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software, 74*(2), 1–24.

Groß, J., & George, A. C. (2014). On permissible attribute classes in noncompensatory cognitive diagnosis models. Methodology.

Halimi, N. (2013). Mathematics Education in Secondary School in Afghanistan: Teachers' View and Practices on Teaching Problem Solving.

Huang, N. E. (2014). Hilbert-Huang transform and its applications (Vol. 16). World Scientific.

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement, 71*(2), 407–419.

Köhn, H. F., & Chiu, C. Y. (2016). A proof of the duality of the DINA model and the DINO model. *Journal of Classification, 33*(2), 171-184.

Lee, Y.-S., de la Torre, J., & Park, Y. S. (2012). Relationships between cognitive diagnosis, CTT, and IRT indices: An empirical investigation. *Asia Pacific Education Review, 13*(2), 333–345.

Liu, J. Y., Chen, C. H., Lin, C. H., Tsai, H. F., Chen, C. H., & Kamogawa, M. (2011). Ionospheric

 10.12928/bamme.v3i1.8964

disturbances triggered by the 11 March 2011 M9. 0 Tohoku earthquake. *Journal of Geophysical Research: Space Physics, 116*(A6).
Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement, 36*(7), 548–564.
Mansory, A. M. (2010). Do Children Learn in Afghan Schools? *Assessment of Math and Language Achievements of Students at the End of Grades 3 and 6 in SCA Supported Schools*. Swedish Committee for Afghanistan (SCA).
Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 International Results in Science*. ERIC.
Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics* (pp. 139-171). Chestnut Hill, MA: TIMSS & PIRLS International Study Center.
Paek, I., & Cai, L. (2014). A comparison of item parameter standard error estimation procedures for unidimensional and multidimensional item response theory modeling. *Educational and Psychological Measurement, 74*(1), 58-76.
Ravand, H., Barati, H., & Widhiarso, W. (2012). Exploring diagnostic capacity of a high stakes reading comprehension test: A pedagogical demonstration. *Iranian Journal of Language Testing, 3*(1), 12-37.
Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2014). *CDM: Cognitive diagnosis modeling. R package version*, 3.
Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment, 8*(4).
Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher, 41*(8), 294-308.
Sedat, Ş. E. N., & Arican, M. (2015). A diagnostic comparison of Turkish and Korean students' mathematics performances on the TIMSS 2011 assessment. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6*(2).
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.
Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
Terzi, R., & Sen, S. (2019). A nondiagnostic assessment for diagnostic purposes: Q-matrix validation and item-based model fit evaluation for the TIMSS 2011 assessment. *SAGE Open, 9*(1), 2158244019832684.
Toker, T., & Green, K. (2012). An Application of Cognitive Diagnostic Assessment on TIMMS-2007 8th Grade Mathematics Items. Online Submission.
Wafa, M. N., Hussani, S. A. M., & Pazhman, J. (2020). Evaluation of Students' Mathematical Ability in Afghanistan's Schools Using Cognitive Diagnosis Models. Eurasia Journal of Mathematics, *Science and Technology Education, 16*(6).
Wafa, M. N. (2019). Assessing school students' mathematic ability using DINA and DINO models. *International Journal of Mathematics Trends and Technology (IJMTT), 65*(12), 153-165
Wafa, M. N., Zia, Z., & Frozan, F. (2023). Consistency and Ability of Students Using DINA and DINO Models. *European Journal of Mathematics and Statistics, 4*(4), 7-13.
Wafa, M. N. (2020). 基于认知诊断模型框架评价阿富汗学生的数学能力 (Master's thesis,

东北师范大学).

Hussaini, S. A. M., & Wafa, M. N. (2021). Wettability of Femtosecond Laser Direct Fabrication of Micro and Nanostructures on Metal Surface. *Turkish Online Journal of Qualitative Inquiry, 12*(9).

Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology, 71*(2), 262-286.

 10.12928/bamme.v3i1.8964