Consistency and Ability of Students Using DINA and DINO Models

Mohammad Nasim Wafa, Ziauddin Zia and Fazlahmad Frozan

***Abstract* — Cognitive diagnosis models (CDMs) are restricted latent class models that can be used to analyze response data from educational or psychological tests. The Deterministic Input Noisy Output "AND" gate (DINA) model and the Deterministic Input Noisy Output "OR" gate (DINO) model there are two popular cognitive diagnosis models (CDMs) for educational and evaluation assessment. The aim of this study is to compare the mentioned models and girls and boys with the modeling of cognitive diagnosis. The aims of this study was to identify differences in performance of Afghan boys' and girls' students in the basic mathematical attributes and cognitive skills of the eighth grade in TIMMS (2011). As well as rankings among the countries that participated in TIMMS. Two commonly used CDMs were employed to fit the response data, including these two models (DINA and DINO). With the assistance of CDMs, we could obtain not only the item parameters, but also the skill profile for each student. Results show that the examinees do best in number domain while do worst in data and chance.**

***Keywords* — CDMs; DINA; DINO; Q-Matrix; TIMSS 2011.**

I. Introduction

Cognitive diagnostic models (CDMs) are those that study the performance of an experimental test of an individual or student's overall ability with their skills [1], which may be based on each they dominate, conditionally disintegrate, examine. On this direction, a detailed description or characteristics of her characteristics, strengths, and weaknesses in the field of testing ability. A set of possible attribute indicators for a given test shows the intellectual skills classes that can be assigned to them. Psychological measurement focuses on obtain a good understanding of people's latent traits, abilities, personality, or intelligence level. Formally, the primary goal is to infer their unobservable latent traits numerically through the observable responses to a test. Therefore far, a list of statistical tools has been introduced to adjust the measurement error to make a powerful inference. Two widely used test theories: classical test theory [2] and item response theory [3].

Classical test theory (CTT) often assumes that a person's score observed on a test is the sum of a true score (error-free score) and an error score. It also is regarded as a true score theory. However, CTT mainly focuses on the examinees' characteristics which cannot be separated. Moreover, the definition of reliability relies on the form of the test.

The Deterministic Input Noisy Output "AND" gate (DINA) model [4] and the Deterministic Input Noisy Output "OR" gate (DINO) model [5] are two popular cognitive diagnosis models (CDMs). CDMs for educational assessment decompose an examinee's ability in a domain into binary cognitive skills called *attributes*, each of which an examinee may or may not have mastered. Distinct profiles of attributes define different proficiency classes. From the observed item scores, maximum likelihood estimates of the model parameters are obtained that are then used to assign examinees to the different proficiency classes. Software for fitting the DINA model and the DINO model using marginal maximum likelihood estimation via the Expectation Maximization (EM) algorithm (MMLE-EM) is available through the package CDM implemented in R [6].

Nowadays, detailed information on ability or skills, not an overall ability, provokes interest and Cognitive Diagnosis Models (CDMs), as the most important part, has also gained growing interest in test development and in the measurement of human performance [3]. As statistical and psychometric models, CDMs has gradually become a common evaluation tool [4].

Published on July 6, 2023.
M. N. Wafa, Department of Mathematics, Ghor Institute of Higher Education, Afghanistan.
(corresponding e-mail: nasim_wafa58@yahoo.com)
Z. Zia, Department of Mathematics, Ghor Institute of Higher Education, Afghanistan
(e-mail: ziauddin1364@gmail.com)
F.Frozan, Department of physics, Ghor institute of Higher Education, Afghanistan.
(e-mail: fazlahmadfrozan81@gmail.com)

Therefore far, various forms of CDMs have been introduced in the measurement history. Generally, these models cover a variety of situations (such as types of structure, responses, and dimensions) which are of interest to researchers in psychometric and cognitive science.

In most parts of Afghanistan, most eighth grade math, fourth grade math, and basic math students with basic math skills have significant disadvantages, and a significant proportion of students fail to learn higher and low level subjects [7]. The dropout rate in this course and the poor results in national and international exams are due to this weakness, and because in the eighth grade, students also learn different subjects of mathematics in addition to mathematics in school, which causes weakness in learning mathematics and also Wafa found old curriculum is taught in Afghanistan. Therefore, this study pays more attention to constructing the hierarchy of this course when developing educational programs and following the cognitive pattern in the learning process, as well as planning to ensure that it is reciprocal. Prerequisite knowledge for curriculum concepts and preplanning skills can have a significant impact on the quality of teaching and learning mathematics in Afghanistan's middle schools.

A. Research Objectives

In this article, to investigate why it is more difficult for Afghan students to learn mathematics, it isnecessary to conduct research focusing on assessing and identifying the \strengths and weaknesses of Afghan students, especially eighth grade students, in characteristics and skills. It is clear that Afghanistan has experienced forty years of war and no research has been done on students' strengths or weaknesses in mathematics.

The main purpose of this study is to provide information about the strengths and weaknesses of students to provide valid educational information that the teacher can use effectively. Sedat and Arican, using CDMs for Mathematics in grade 8. Specifically, this research mainly focuses on the better cognitive diagnosis model that can be used to model students' Mathematical abilities. At the same time, under this model, an evaluation of Afghan secondary school students' level of mathematical ability would be presented. Hopefully, it would also promote teacher enhancement of instructional procedures to match student development in the future.

B. Research Questions

a. In which skills is the best performance of Afghan eighth grade math students in the TIMMS test using DINA and DINO models?
b. What are the Afghanistan eight grade students' weaknesses and strengths based on DINA and DINO models?

II. Preliminaries

A. DINA model

Deterministic input, noisy-or-gate model, known as DINO, is a compensatory CDM [5] because it assumes that lack of one measured attribute can be compensated by another attribute. More specifically, mastery of at least one attribute compensates deficit of all the other measured attributes. Similarly, to the DINA model, the slipping and the guessing parameters are estimated at the item level. The DINO model works with a disjunctive condensation rule in which the presence of at least one measured attribute guaranties a high probability of endorsing an item [8]. DINO model estimates the probability of a correct response for item i in latent class c as follows:

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

Here is student i examinee, with skill profile $\alpha_{ik} = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK}]$ to item j and Assume a test containing *i* students and *J* items which require *K* attributes. Let π_{ij} be the dichotomous response of student i to item j and $\alpha_{ik} = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK}]$ be the corresponding latent profile. Let the *Q*-matrix be a $J \times K$ matrix with the j, entry where J and K are the number of attributes and here j is element of row and k is element of column. If the correct application of attribute *k* influences the probability of correctly responding the jth item, and matches 0 otherwise. The vector j_{ik}=[j_{i1},… j_{ik}] means the qth row in the *Q*-matrix. Given π_{ij} , the probability of a true response $P(X_{ij} = 1|\pi_{ij})$ is defined by the DINA for the (j_{-th}) item as:

$$P\left[X_{ij} = 1 \middle| \boldsymbol{\alpha}_i, s_j, g_j\right] = (1 - s_j)^{\pi_{ij}} \cdot g_j^{(1-\pi_{ij})} = \begin{cases} 1 - s_j & for\ \pi_{ij} = 1 \\ g_j & for \pi_{ij} = 0 \end{cases} \tag{2}$$

Here we know that the s_j slip parameter is the probability of the wrong answer to the j_{th} that the person i_h has obtained all the features measured by the j_{th} item.

On the other hand we know that the g_j (guess parameter), the probability of the correct answer for the item j_{th}, when the person i_{th} does not master all the properties measured by the item j_{th}.

$$s_j = P\left(X_{ij} = 0 \middle| \pi_{ij} = 1\right), \qquad j = 1,2, \dots . J$$

Here also we show the guessing parameter,g_j:

$$g_j = P\left(X_{ij} = 1 \middle| \pi_{ij} = 0\right), \qquad j = 1,2, \dots , J$$

B. DINO Model

DINO model estimates the probability of a correct response for item i in latent class c as follows:

$$\pi_{ic} = P\left(X_{ic} = 1 \middle| \pi_{ij}\right) = (1 - s_j)^{\pi_{ij}} g_j{}^{1-\pi_{ij}}$$

$wher$
$$s_j = P(X_{ij} = 0 | \pi_{ij} = 1)$$
$$g_j = P(X_{ij} = 1 | \pi_{ij} = 0)$$

C. The Duality of the DINA Model and the DINO Model

As mentioned before, the DINA and DINO model, they are two specific models in the conative diagnosis models (CDMs) to assess educational affairs, which look differently at the way the mastery of cognitive skills as well as correct item response are associated. Recently [9], have shown that DINA and DINO could be explained according to one another and which of the two models fit the dataset is not fundamentally relevant since they have similar results. The DINA-DINO duality as an instant practical outcome could fit the same software. If the models tend to be similar under some linear transformations, they are supposed to share similar theoretical properties. A methodological advance related to one model mechanically applies to the other one. Therefore, instead of proving them individually, they can be proved as a single [9]. Ko˙han et al. also mentioned that the two models are the same under the following transformations: (1) the profiles of exam takers' attributes, (2) their scores of the observed items and (3) the parameters of the models. The real proving of the duality of the DINA model and the DINO model introduced here is the correction of the main proof by Liu, J., Xu, G. & Ying, Z. (2012) in link with Asymptotic Classification Theory of Cognitive Diagnosis (ACTCD).

D. Q-matrix

The analysis of most CDMs is based on an item attribute incidence matrix called a *Q*-matrix [15]. The diagnostic power of CDMs relies on the construction of a *Q*-matrix with attributes that is theoretically appropriate and empirically supported [8]. Studies on the *Q*-matrix can be normally categorized as exploratory approaches intend to discover the *Q*-matrix from the data when whole *Q*-matrix is unknown. Confirmatory approaches aim to purify a certain *Q*-matrix in which some elements of the *Q*-matrix are assumed to be known. Although an entirely exploratory approach obtains no information about the number of attributes in advance, an approach given the number of attributes is still regarded as exploratory here as long as it estimates the whole *Q*-matrix [15]. After defining, determining and identify the *Q*-matrix for measuring the test, the next step is to construct the *Q*-matrix.

In this study, to form a Q matrix, after translating the questions into Persian language, the was question coding 1 or 0, a copy of the eighth grade TIMMS-2011 math questions with features, and protocol was coded, which included 32 questions with 13 attributes. Different sections and provided to 5 math teachers. Master's and bachelor's degrees, which had 15 years of teaching and training experience, were 12 years, 12 years, and 11 years, respectively [3], [10]. They are asked to create a Q-matrix separately and independently. To encode in a two-dimensional matrix in which columns include those skills, each question measures the properties of the question rows by specifying 1 or 0.

Table II shows the original Q matrix for this example. For the 32 items in this study and assessment, the skill requirements table (Q-matrix) for each item forms the Q matrix. AA_{N1}, A_{N2} , A_{N3} and A_{N4} are the properties of the Number domain. A_{A1}, A_{A2}, A_{A3} and A_{A4} are the relation to algebra. A_{G1}, A_{G2}, A_{G3} and A_{G4} are the characteristics of the relation of geometry. And A_{D1} are the properties of the *Data and Chance* domain.

ISSN: 2736-5484

TABLE I: Q-MATRIX FOR EACH CONTENT DOMAIN

	attribute												
$Item$	A_{N1}	A_{N2}	A_{N3}	A_{N4}	A_{A1}	A_{A2}	A_{A3}	A_{A4}	A_{G1}	A_{G2}	A_{G3}	A_{G4}	A_{D1}
$Item$ 1	1	1	1	0	0	0	0	0	0	0	0	0	0
$Item$ 2	0	0	1	1	0	0	0	0	0	0	0	0	1
$Item$ 3	1	0	0	0	1	0	0	1	0	0	0	0	0
$Item$ 4	0	1	1	0	0	0	0	0	0	0	0	0	0
$Item$ 5	0	1	1	0	0	0	0	0	0	0	0	0	0
$Item$ 6	0	0	0	1	1	0	0	0	0	0	0	0	0
$Item$ 7	0	0	0	1	0	0	0	0	0	0	0	0	0
$Item$ 8	0	0	0	1	0	0	0	0	0	0	0	0	0
$Item$ 9	0	1	1	1	0	0	0	0	0	0	0	0	0
$Item$ 10	0	0	0	1	0	0	0	1	0	0	0	0	0
$Item$ 11	0	0	0	0	0	0	0	0	0	1	0	1	0
$Item$ 12	0	0	0	0	0	0	0	0	0	1	0	1	0
$Item$ 13	0	0	0	0	0	0	0	0	0	0	1	1	0
$Item$ 14	0	0	0	0	0	0	0	0	1	1	1	1	0
$Item$ 15	0	0	0	0	1	0	0	0	0	0	0	0	0
$Item$ 16	0	0	0	0	0	1	0	0	0	0	0	0	0
$Item$ 17	0	0	0	0	1	1	0	1	0	0	0	0	0
$Item$ 18	0	0	0	0	0	0	1	0	0	0	0	0	0
$Item$ 19	0	0	0	0	0	1	1	0	0	0	0	0	0
$Item$ 20	0	0	0	0	0	1	0	0	0	0	0	0	0
$Item$ 21	0	0	0	0	0	0	1	0	0	0	0	0	0
$Item$ 22	0	0	0	0	1	1	1	1	0	0	0	0	0
$Item$ 23	0	0	0	0	0	1	1	0	0	0	0	0	0
$Item$ 24	0	0	0	0	0	0	0	0	1	1	1	0	0
$Item$ 25	0	0	0	0	0	0	0	0	1	0	1	0	0
$Item$ 26	0	0	0	0	0	0	0	0	0	0	1	0	0
$Item$ 27	0	0	0	0	0	0	0	0	0	0	1	0	0
$Item$ 28	0	0	0	0	0	0	0	0	0	0	0	0	1
$Item$ 29	0	0	0	0	1	1	1	1	0	0	0	0	0
$Item$ 30	0	0	0	0	0	0	0	1	0	0	0	0	0
$Item$ 31	0	0	0	0	0	0	0	1	0	0	0	0	0
$Item$ 32	0	0	0	0	0	0	0	1	0	0	0	0	1

E. Item fit

Root mean square error of approximation (item-fit RMSEA); [9]and [5] is a common criterion to quantify the goodness of an item in the model. The expression of RMSEA is given as follows,

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

Where $N(X_j|\alpha_l)$ is the number of responses to item j given by students in latent class α_l. If RMSEA is below 0.05, it often means a good fit. Conversely, if RMSEA is above 0.1, a worse item fit presents [11]. If the value is larger than 0.05 but lower than 0.1, it often is regarded as a moderate item fit.

F. Item Discrimination Index (IDI)

The Item Discrimination Index (IDI), $IDI_j = 1 - s_j - g_j$ [16] is also a frequently used criterion in the DINA and DINO models. Especially, in DINA model, IDI describes the discrimination of each item when measuring the differences between the students having all required attributes and the students not having at least one. The IDI is closed to 1 means that an acceptable discrimination of the item, or better diagnositicity. However, if the IDI is close to 0, it means a low discrimination.

III. METHOD

This study focuses mainly on the TIMSS 2011 questionnaire in Afghanistan. It was the same questionnaire used to compare math performance between Korean and Turkish students as well as the research has been done with American and Turkish students in Taiwan. There have been eight booklets and

ISSN: 2736-5484

every student was asked to respond to one of them, and each student was asked to answer only one in eight booklets. The present study selected only booklets of 1, 2, 25 & 7.

The experiment consisted of 32 items, including 15 multiple-choice and 17 constructed response items. And when the answer is correct, it was coded by 1, and 0 otherwise. There are 274 examinees, 52.92% of them are males (i.e., 145), and 129 are female, of which the corresponding percentage is 47.08%, where

TABLE II: THE RATE OF CORRECT RESPONSES FOR EACH ITEM

Items	Correct rate	Items	Correct rate
1	0.215	17	0.284
2	0.208	18	0.295
3	0.223	19	0.157
4	0.49	20	0.117
5	0.401	21	0.09
6	0.237	22	0.197
7	0.332	23	0.12
8	0.369	24	0.15
9	0.295	25	0.248
10	0.204	26	0.263
11	0.336	27	0.285
12	0.336	28	0.193
13	0.201	29	0.255
14	0.201	30	0.367
15	0.16	31	0.212
16	0.292	32	0.12

The average age of examinees is around 17 years old. Table II presents the rates of correct responses for each item based on all examinees' observed responses. As claimed by the results in the table, the highest frequency of correct response befits into Item 4 at (0.49) and the lowest at Item 21, which is (0.09).

A. Q-Matrix

After collecting the response data, CDMs analysis may require the construction of a Q-matrix to obtain the relationships between the items and attributes. The four content domains tested in TIMSS 2011 were Number (30%), Algebra (30%), Geometry (20%), and Data and Chance (20%). The attributes measured in each domain were specified based on the Common Core State Standards of Mathematics (CCSSM; Common Core Standards Initiative, 2010). In total, 13 attributes were measured in the test. Table III presents the detailed attributes, description, editing feature, and specifications for each content domain reported by TIMSS researchers [10], which were identified by four experts in mathematical education [14].

The procedure of constructing Q-matrix is mainly based on the empirical Q-matrix validation method [15]. The detailed Q-matrix is shown in Table III. The attributes in the Q matrix are created independently by considering the steps required to solve each case. For example, in the section of geometry at Item 17, students were given a picture of a rectangular garden that had a $(x + 4) - M$ width and an xM height (we see in Fig. III1).

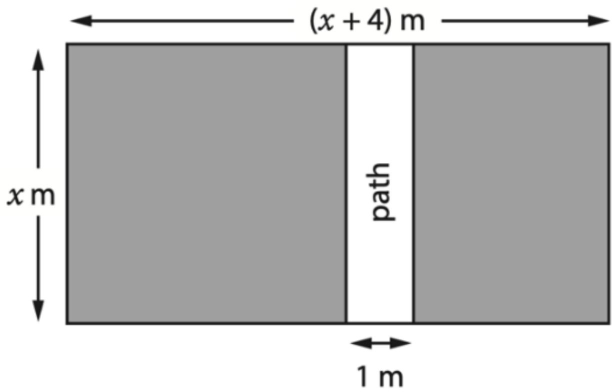

Fig. 1. A sample of TIMMs questionnaire.

B. Estimation

Two CDMs were employed to fit the response data, including the compensatory model, DINA model, and the noncompensatory one, DINO model. Both of these selected models have two parameters only per item, so they are more concise compared with those structured CDMs. Under both models, Expectation-Maximization (EM) algorithm performed the estimation of item Parameters by using Marginal Maximum Likelihood (MML) estimation [8]. Believed that the Standard Error (SE) estimate is premeditated by the empirical cross-product approach for the item parameter [15] said that 'CDM' package implemented the detailed computation procedures in R software environment.

IV. Discussion

Firstly, in the questionnaire, a descriptive analysis of the DINA model was performed to show the probability level of each attribute. Results were shown that the highest probability of mastery belonged to attribute A_{N_2} and A_{N_3} at (0.348) while A_{D_1} at (0.13) was the lowest probability attribute. After that, the second descriptive analysis was done for clarifying the item probability level in the questionnaire. The mastery's highest probability belonged to item four (0.49) and the lowest probability belonged to item- 21 (0.09).

Secondly, the determination of every item's guessing and slipping parameters was the reason to use the DINA model for the analysis to be calculated. Item (31) with (2.78E_ 08) and item#4 with the value of (0.36) were the lowest and highest guessing coefficients of the evaluation results. In addition, item 2 with

A value of (0) and item (22) with a value of (0.671) were the lowest and highest coefficient slipping parameters. DINO model was also used to determine every item guess and slip parameters by using the same procedure like DINA model. Item (32) with (1.08E-145) and item#4 with the value of (0.375) were the lowest and highest guessing coefficients of the evaluation results. In addition, items (27, 30 & 31) with a value of (0) and item (22) with a value of (0.779) were the lowest and highest coefficient slipping parameters; by these numbers and coefficients, it would be understood that the probability of students to respond the question mistakenly have the skills needed to response the question.

Besides, the levels of IDI under the DINA and DINO models were also shown by the R software analyses. It was found out from the analyses results that the lowest in item (11) and item (21) was the value of (0.929 and 0.984).

Finally, the results of skill probability showed that the high skill probability belonged to the A_{G2} while the lowest belong to the A_{A3}. This result showed the appropriateness of students mastering the skill and Algebra content was regarded the hardest based on this result.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

Reference

[1] Ghosh S, Roy MK, Biswas SC. Determination of the best fit probability distribution for monthly rainfall data in Bangladesh. *American Journal of Mathematics and Statistics*. 2016; 6(4): 170-174.

[2] Wafa MN, Hussani SA, Pazhman J. Evaluation of Students' Mathematical Ability in Afghanistan's Schools Using Cognitive Diagnosis Models. *Eurasia Journal of Mathematics, Science and Technology Education*. 2020; 16(6): 57-64.

[3] Wafa MN, Assessing school students' mathematic ability using DINA and DINO models. *International Journal of Mathematics Trends and Technology.* 2019; 65(12): 153-65.

[4] Rupp AA, Gushta M, Mislevy RJ, Shaffer DW. Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment*. 2010; 8(4): 1-12.

[5] Ravand H, Barati H, Widhiarso W. Exploring diagnostic capacity of a high stakes reading comprehension test: A pedagogical demonstration. *Iranian Journal of Language Testing.* 2013; 3(1):11-37.

[6] George AC, Robitzsch A, Kiefer T, Groß J, Ünlü A. The R package CDM for cognitive diagnosis models. *Journal of Statistical Software.* 2016; 74(1):1-24.

[7] Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*. 1973; 60(2): 255-65.

[8] Choi KM, Lee YS, Park YS. What CDM can tell about what students have learned: An analysis of TIMSS eighth grade mathematics. *Eurasia Journal of Mathematics, Science and Technology Education.* 2015; 11(6):1563-77.

[9] De Boeck P, Wilson M, editors. Explanatory item response models: A generalized linear and nonlinear approach. New York: Springer; 2004.

[10] de la Torre J, Chiu CY. A general method of empirical Q-matrix validation. *Psychometrika*. 2016; 81: 253-73.

[11] Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician.* 2007; 61(1):79-90.

[12] Groß J, George AC. On permissible attribute classes in noncompensatory cognitive diagnosis models. *Methodology*. 2014.

[13] Henson RA, Templin JL, Willse JT. Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*. 2009; 74(1):191-210.

[14] Junker BW, Sijtsma K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement.* 2001; 25(3): 258-72.

[15] Köhn HF, Chiu CY. A proof of the duality of the DINA model and the DINO model. *Journal of Classification*. 2016; 33(1): 171-84.