Exploring Diagnostic Capacity of a High Stakes Reading Comprehension Test: A Pedagogical Demonstration

Hamdollah Ravand [1], Hossein Barati [2], Wahyu Widhiarso [3]

Abstract

Traditional analysis of test results with Item Response Theory (IRT) models, and classical test theory (CTT) provide unidimensional ability values, which allow for ranking of students according to their abilities. Cognitive diagnostic assessment (CDA) is aimed at providing formative diagnostic feedback through a fine-grained reporting of learners' skill mastery. The present study, through application of deterministic input, noisy ''and'' gate [de la Torre & Douglas, 2004 (DINA) model], investigated the degree to which the items of a high stakes reading comprehension test can provide diagnostically useful information. Through expert rating, a Q-matrix including five subskills was developed. Using DINA model, student's skill mastery profiles were produced and issues related to the diagnostic capacity of the reading test were discussed. Meanwhile, this study pedagogically demonstrates the application of the DINA model employing R software.

Key words: *Attribute, Cognitive Diagnostic Model (CDM), DINA, Skill, Q-matrix*

1. Introduction

Cognitive Diagnostic Assessment (CDA) has received much interest in recent years. It has been praised for providing fine-grained diagnostic feedback through reporting of learners' skill mastery profiles by different researchers (i.e. DiBello, Roussos, & Stout, 2006; Embretson, 1991, 1998; Hartz, 2002; Nichols, Chipman, & Brennan, 1995; Tatsuoka, 1983). The CDA approach is intended to promote assessment for learning and the learning process as opposed to assessment of learning outcome through providing information needed to modify instruction and learning in classrooms by teachers (Jang, 2008). "The advantages of these models are most apparent in situations where diagnostic feedback needs to be provided to respondents, and criterion-referenced interpretations of multiple proficiencies or

1 *English Department, University of Isfahan, Iran,* [*ravandhamdollah@gmail.com*](mailto:ravandhamdollah@gmail.com) *(Corresponding author).*

2 *English Department, University of Isfahan, Iran.*

3 *English Department, University of Jena, Germany.*

dispositions are most needed" (Rupp, 2007, p.79). Formative diagnostic information can be used to identify strengths and weaknesses of individual learners in the target area of learning and instruction, such as English language learning and redesign instructional approaches, evaluate instructional resources, and to take appropriate measures to remedy students' weaknesses accordingly. Unlike aggregated test scores based on unidimensional scaling, a diagnostic assessment provides detailed feedback about the current state of knowledge and skills of learners (Jang, 2008) so that they and their teachers can take appropriate actions to remedy their weaknesses in various aspects of second language ability.

CDA has been applied to second language assessment in two ways: (Jang, 2008; Lee & Sawaki, 2009b); (1) inductively to develop a set of diagnostic items or tasks that allow us to infer the knowledge structures and skill processes involved in doing the items or tasks; and (2) a retrofitted approach (or reverse-engineered approach) to extract diagnostic information from existing non diagnostic tests. Until now, most of the CDA analyses in language assessment have focused primarily on retrofitting CDAs to existing tests that were not initially developed for the purposes of cognitive diagnosis (Buck & Tatsuoka, 1998; Buck et al., 1998; Gao, 2007; Jang, 2005; Kasai, 1997; Lee & Sawaki, 2009a; Sheehan, 1997; von Davier, 2005).

The General English Test (GET) is developed and administered by university of Isfahan (UI) as part of the screening process for the PhD programs at this university. The reading comprehension test used in the present study was part of GET. It was designed to assess examinees' understanding of college-level reading texts and administered to about 1,500 PhD applicants to university of Isfahan (UI) yearly, hence a high stakes test. This study is the first attempt at investigating the diagnostic potential of the UI reading test. In order to maximize the instructional and washback values of the UI reading test, it is useful to explore how the CDMs can be used with the reading part of GET test .

## 2. Cognitive Diagnostic Models: An Overview

Rupp and Templin (Rupp & Templin, 2008) define CDMs as "Diagnostic classification models are probabilistic, confirmatory multidimensional latent-variable models with a simple or complex loading structure" (P. 226).

They are probabilistic models in that each CDM expresses a given respondent's performance level in terms of the probability of mastery of each attribute separately, or the probability of his or her having a specific skill-mastery profile or belonging to each latent class(Lee and Sawaki, 2009) . For example, in the present study for k=5 attributes we have 2^5= 32 skill mastery profiles (since each of the k skills may be assigned two levels (mastery/nonmastery), there are 2^k possible skill mastery patterns, where k is the number of attributes) representing a latent class each. As you will see later in Table 3, Respondent 5 is classified into the first latent class with skill profile of (00000) which indicates mastery of none of the skills. The probabilities of belonging to Latent Classes 1, 2, 3, 5, 6, 9, 10, and 13 are about .60,.008,.004,.005,.37, .001, .004 and .003 respectively and 0 for the rest of the latent classes. Therefore she is assigned to the latent class with the highest probability, that is, Latent Class 1. For this respondent CDM provides the probabilities that the respondent possesses skills 1 and 2 as .0013, 0.07, respectively (below we will explain how to compute probabilities for the other three skills).

Cognitive diagnostic models are also confirmatory in nature, like confirmatory factor analysis models, in the sense that latent variables in CDMs are defined a priori by the Q-

matrix. A Q-matrix (Tatsuka, 1985) is the loading structure of a CDM. It is a hypothesis about the required skills for getting each item right (Li, 2011). Another feature of CDMs is their multidimensionality. They are multidimensional latent variable models because they estimate unobservable learner profiles on more than one attribute. Unlike item response theory (IRT) models which assign to respondents a single score on a continuous scale which represents a broadly defined ability, CDMs assign respondents to multidimensional skill profiles by classifying them as masters versus non masters of each skill involved in the test. CDMs are notably different from multidimensional item response theory models and factor analysis in terms of characteristics of latent variables (lee and Sawaki, 2009b). Latent variables in CDMs are discrete or categorical (e.g., masters/non-masters), whereas ability estimates (θ) in multidimensional item response theory models and factor scores in factor analysis are continuous. Moreover, they have complex loading structures because each item can be specified in relation to multiple attributes. For example, in Figure (1) items two and five measure two skills and items four and eight measure three skills hence these four items have got a complex loading structure.

Figure 1. Complex loading structure of CDMs
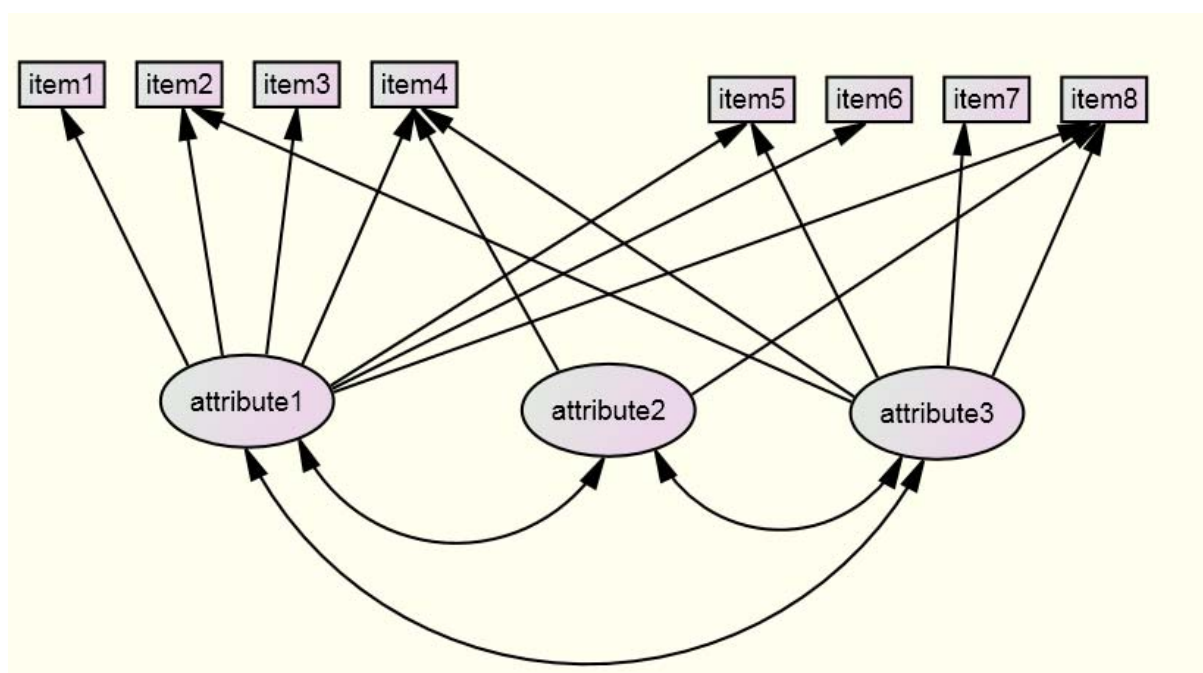
CDMs classify respondents into classes representing specific mastery/nonmastery profiles for the set of attributes specified in the Q-matrix (Henson, Templin, & Willse, 2009; von Davier, 2005). For example, for $K = 5$ skills/attributes, a respondent assigned the vector (skill mastery patterns) $\alpha = (1,1,0,1,0)$, has been deemed a master of the first, second and fourth skills and a non-master of the third and fifth skills. Since respondents' class membership is unknown, these skill mastery patterns into which respondents are assigned are referred to as latent classes. For example, in Table 3 Respondent 1 is classified as belonging to the latent class with the highest probability, that is, Latent Class1.

Characteristics of CDMs explained so far, are common among all CDMs. CDMs are classified into two classes according to their assumed attribute structure, or the assumed relationships among the attributes. One way might be to distinguish models that assume

conjunctive relationships from those that assume disjunctive relationships (Fu & Li, 2007). Under a conjunctive attribute structure, an item can be successfully answered only if all the required attributes for the item have been successfully mastered and executed. Thus, a conjunctive structure is noncompensatory in the sense that one attribute cannot be completely compensated for by other attributes in terms of item performance; that is all the attributes must function in conjunction with each other (conjunctively). In contrast, in a disjunctive case, successfully executing only one or some of the attributes required for an item is sufficient to answer the item correctly. In other words, the attribute structure is compensatory in that strength in one attribute may compensate for weakness in another, thus mastery of all attributes involved in an item is not necessarily required for a test taker to answer the item correctly (Rupp and Templin, 2008). When attributes involved in a test are disjunctively related, mastery of one or some of the attributes can compensate for non mastery of the other attributes. Put simply the respondent has got a choice to master any, some, or all the attributes in order to provide the correct answer to the items on the test.

2.1. DINA model

DINA is an acronym that stands for the Deterministic Input, Noisy ''And'' gate (de la Torre & Douglas, 2004). Deterministic input of the DINA model indicates whether respondents in latent class c have mastered all measured attributes for item i or not. ''And'' gate refers to conjunctive relationship among the attributes and Noisy refers the stochastic element of the DINA model which points to the fact that response behavior of the respondents is not deterministic (i.e., error-free or nonstochastic) but probabilistic. This stochasticity is due to the noise introduced in the process as a result of slip and guessing parameters—that is, respondents who have mastered all the skills required for an item can slip and miss the item, and those who lack at least one of the skills required can guess and get the item right with typically nonzero probabilities (Rupp, 2007). Put another way, deterministically those who have mastered all the attributes involved in an item must get the item right, but stochastically they are likely to slip and guess.

Through the years different versions of the DINA model have been developed by different researchers. Generalized DINA (G-DINA model) was proposed by de la Torre (de la Torre, 2011), as a generalization of the DINA model. G-DINA can be categorized among general CDMs which unlike specific CDMs do not assume restrictive relationship such as conjunctive or disjunctive among the attributes involved in providing the correct answer to an item. Other examples of general models include general diagnostic model (GDM; von Davier, 2005), and log-linear cognitive diagnostic model (LCDM) (Buck & Tatsuoka, 1998). G-DINA model relaxes the assumption of equal probability of success for all those who lack some or all of the attributes required for an item (de la Torre 2011). The DINA model partitions each of the 2^k latent classes into two latent groups; the group comprised of attribute mastery profile with all the attributes required for item j , $\xi_{ij}=1$ or the group who have not mastered at least one of the attributes required for an item, $\xi_{ij} = 0$. de la Torre argues that this assumption might not hold for the group $\xi_{ij} = 0$ because in this group respondents have got different degrees of deficiency with respect to the attributes required hence their probabilities of getting the item right may be different. For example, in the reading test under study in the present paper the Q-matrix entry for item 27 is [1 0 1 1 0]. According to the DINA model to get this item right one has to have mastered skills one, three, and four. DINA model differentiates between those who have mastered all the three skills and those who have not mastered at least one of the skills. DINA doesn't further differentiate among the respondents with varying degrees of attribute deficiency. In this
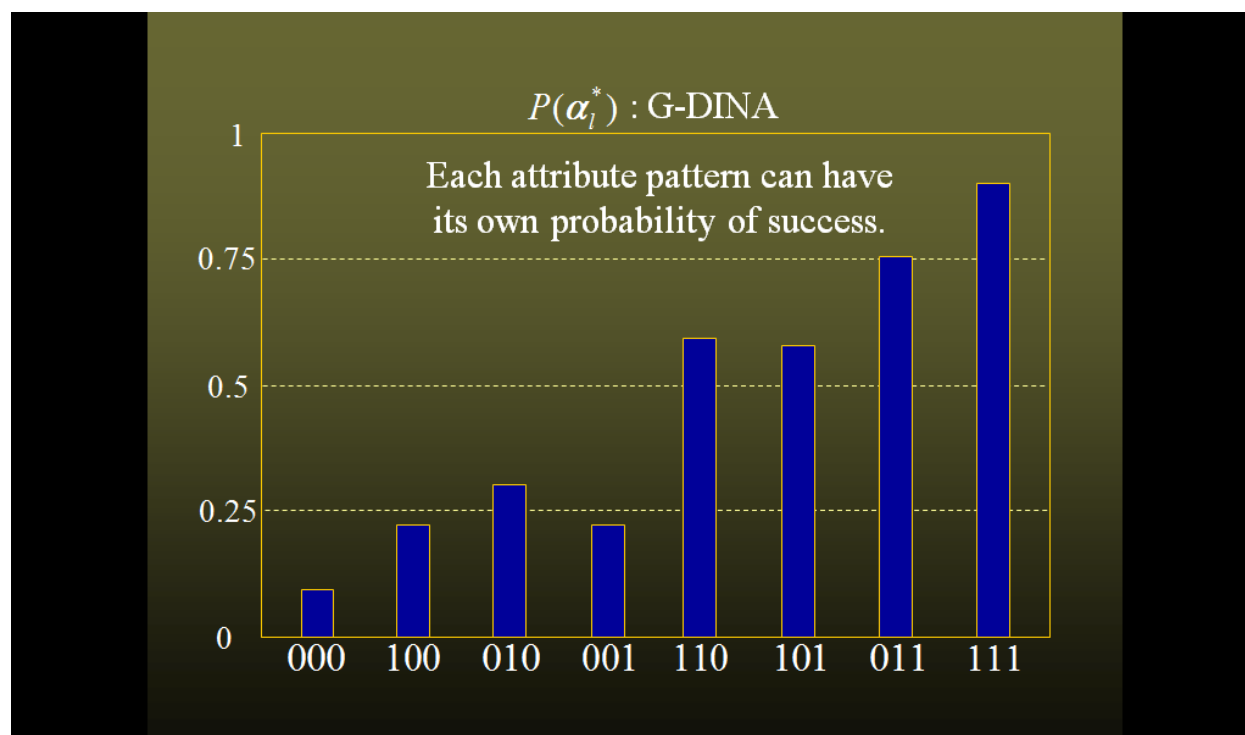
model, the probability of success for a respondent who has mastered only one skill and another respondent who has mastered two skills is the same. In G-DINA, the probability of success for respondents in the second group $\xi ij = 0$ is not the same. A respondent who has mastered two of the skills required has got higher probability of success than a respondent who has mastered only one of the skills. Figures 5 and 6 schematically show how success probabilities vary for the group who have not mastered at least one skill required by an item.

Figure 2. DINA model success probabilities for different skill profiles. Reproduced from de la Torre (2012) with permission
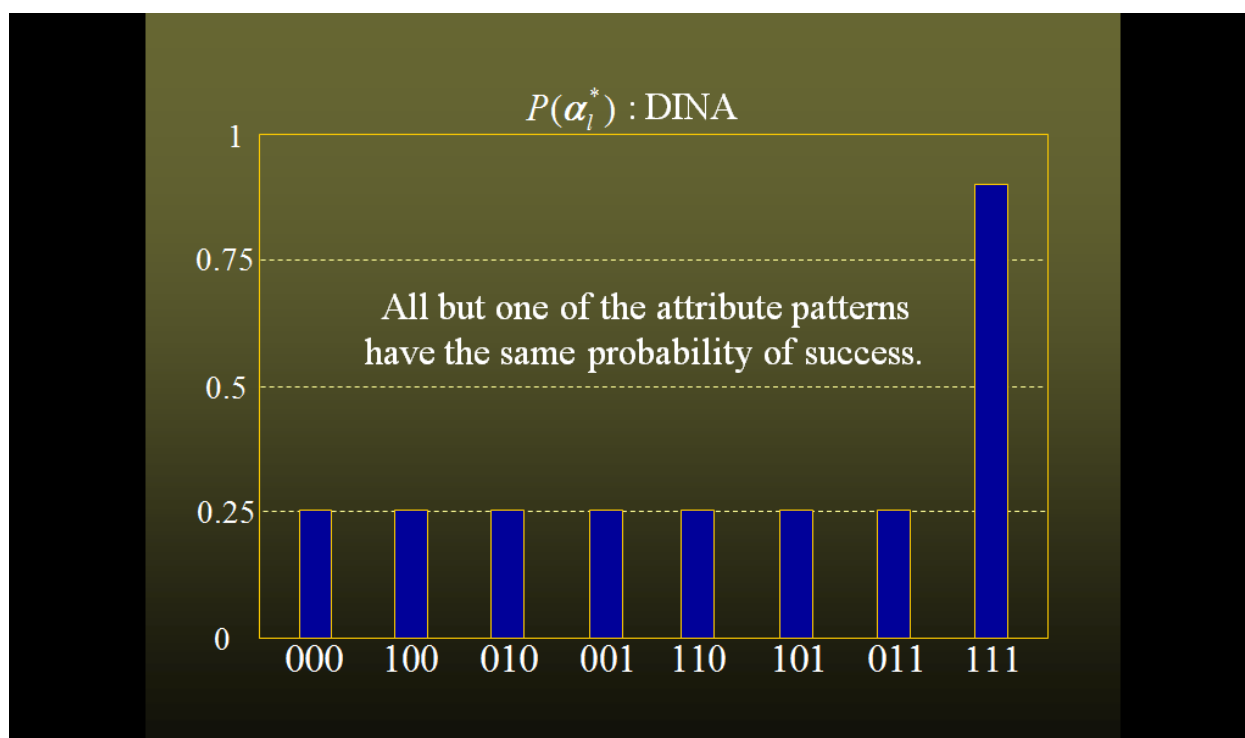
As you can see the probability of success for all latent classes who have not mastered at least one of the skills is the same. But in Figure 5 probabilities of success for those who have not mastered at least one skill, vary depending on how many and which skills they have not mastered.

Figure 3. G-DINA model success probabilities for different skill profiles. Reproduced from de la Torre (2012) with permission

2.1.1. Formal representation of the DINA model

The DINA models the probability of getting item j right for respondent i, denoted Xij, as

$$P(Xij = 1 \mid \xi ij) = \left(1-s_j\right)^{\xi_{ij}} g_j^{\,1-\xi_{ij}} \qquad \text{(stochastic part of the model)} \qquad (1)$$

In this equation ξij is zero or one depending on whether or not the ith respondent possesses all the necessary skills for item j.

$$\xi ij = \prod_{k=1}^{K} \alpha_{ik}^{\,q_{ij}}, \qquad \text{(deterministic part of the model)} \qquad (2)$$

where s_j is the probability of a slip (an incorrect response although the respondent possesses all the skills required for that item), and g_j is the probability of a guess (a correct response although the respondent does not possess all the skills required for that item).

$\prod(\cdot)$ indicates that the expression following it is multiplied across all attributes from attribute 1 (α = 1) to attribute A. If an attribute is not measured by an item, then $q_{ia}= 0$, which implies that the value of α_{ik} does not matter. If an attribute is measured, then $q_{ia}= 1$, which implies that it matters whether $\alpha_{ik}= 0$ or $\alpha_{ik} =1$. Because the product term is defined over all attributes, $\xi_{ij} = 1$ occurs only when all product terms are 1, which means that all measured attributes for item i have been mastered by respondents in Latent Class c. For example, consider the 34- item reading test diagnosing five skills in the present study. As you can see in Table 2, Item 5 requires skills 2 and 3. Suppose Respondent 1 possesses all five skills. Then,

$$\xi_{ij} = \prod_{k=1}^{K} \alpha_{ik}{}^{q_{jk}} = 1^0 \times 1^1 \times 1^1 \times 1^0 \times 1^0 = 1,$$

indicating that the respondent has mastered all the skills required. (Note that any number raised to the power of 0 is 1 and 0 raised to the power of any number is 0.) In contrast, suppose respondent 2 possesses skills 1, 3, and 4. Then, for item 5,

$$\xi_{ij} = \prod_{k=1}^{K} \alpha_{ik}{}^{q_{jk}} = 1^0 \times 0^1 \times 1^1 \times 1^0 \times 1^0 = 0$$

indicating that the respondent has not mastered at least one of the skills required. Equation 2 represents the deterministic part of the model. Deterministically, if $\xi_{ij}=1$, respondent should correctly respond to the item and if $\xi_{ij}=0$ respondent should respond to the item incorrectly. However, the DINA model allows for the possibility that respondents who have mastered all measured attributes (i.e., respondents in latent class c who have $\xi_{ij}=1$) "slip" and incorrectly answer the item as well as for the possibility that respondents who have not mastered at least one of the measured attributes (i.e., respondents in latent class c who have $\xi_{ij} = 0$) "guess" and correctly answer the item nevertheless (Rupp, Templin & Henson, 2010). Formally the slipping parameter (s_i) and guessing parameter (g_i) are defined as follows:

$S_i = P(X_{ic}=1 | \xi_{ic}=0)$

which expresses probability of answering item i for respondents in class c correctly ($X_{ic}=1$) when at least one of the attributes required by the item has not been mastered by respondents in latent class c ($\xi_{ic}=0$).

$S_i = P(X_{ic}=0 | \xi_{ic}=1)$

expresses probability of answering item i for respondents in class c incorrectly ($X_{ic}=0$) when all of the required attributes by the item has been mastered by respondents in latent class c ($\xi_{ic}=1$).

According to the stochastic part of the model, Equation (1), for a respondent in latent class c who has mastered all necessary attributes ($\xi_{ic}=1$) the response probability is

$$P(X_{ij} = 1 | \xi) = \left(1 - s_j\right)^1 g_j{}^{1-1} = 1 - s_i$$

And for a for a respondent in latent class c who has not mastered all necessary attributes ($\xi_{ic}=0$) the response probability is $P(X_{ij} = 1 | \xi) = \left(1 - s_j\right)^0 g_j{}^{1-0} = g_j$

Table 1. Response probabilities in the DINA model (Rupp, Templin, & Henson, 2010).

	$X_{ic}=1$	$X_{ic}=0$
	(correct response)	(incorrect response)
$\xi_{ic}= 1$ (mastery of all measured attributes)	$1\text{-}s_i$	s_i
$\xi_{ic}= 0$ (nonmastery of all measured attributes)	g_i	$1\text{-}g_i$

Thus in DINA model the probability of answering an item correctly is a function of two different error probabilities depending on which of the two groups distinguished by the model, the respondent examinee belongs to; the guessing probability (gj) which represents the probability of answering an item correctly when at least one the attribute required has not been mastered, and the slipping probability (sj) which represents the probability of failing to answer the item correctly when all the attributes required have been mastered.

3. Purpose of the study

The primary focus of the present paper is to investigate the diagnostic capacity of a high stakes reading comprehension test which was administered to PhD program applicants at university of Isfahan (UI), Iran. Specifically we address the following research question in this study:

Q. To what extent the items of UI reading comprehension test can provide diagnostically useful information?
        To answer this question we specifically focus on the slipping and guessing parameters generated by R for each item and Item Discrimination Index (IDI) which we compute manually. The secondary purpose of the present paper addresses the problem that has contributed to the underutilization of CDMs. de la Torre (2009) noted that CDMs have remained underutilized because of two major limitations. First, as compared to traditional IRT models, CDMs are relatively novel and in some cases, more complex. Consequently, many researchers lack familiarity with these models and their properties. Second, unlike traditional IRT models, which can be analyzed using commercially available software such as BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 1996), accessible computer programs for CDMs are not readily available. As a result, implementations of these models have been hampered.

To address this issue, we demonstrate application of DINA model using R freeware. The DINA model was chosen for two reasons; (1) statistically, it is one of the least complex CDMs (Rupp, 2007), as compared with more complex models and as a result of its simplicity of estimation and interpretation DINA model has enjoyed much attention in the recent CDM literature (Huebner, 2010) and, 2) to demonstrate the application of a CDM we had to choose a software that is freely available. We decide to choose R for this purpose

because it can be freely downloaded from the internet and has got a wide range of uses and new packages by experts all over the world are increasingly developed that can carry a wide variety of statistical analyses. Currently the only CDM package available for R does only DINA and DINO models.

3.1. Data

Data analyzed in the present study were from 1500 PhD applicants who took a large scale reading comprehension test at IU in 2009. Previously admission to PhD programs at any of the Iranian universities required the applicants to undergo a strict three-stage screening process wherein the applicants had to take a General English test first. The reading part of the GET contained of 34 multiple choice items and some other cloze and short answer items. For the purpose of the present study polytomously scored items (cloze and short answer) were ignored and only the dichotomously scored multiple choice items were used. Participants were 56% male and 44% females and all Iranian nationals. In terms of age they ranged from 24 to 50.

3.2. Q-matrix construction

Lee and Sawaki (2009b) specify the following four steps for doing cognitive diagnostic analysis using CDMs: (1) identifying a set of attributes involved in a test; (2) constructing Q-matrix based on the attributes required for providing the correct answer to each item in the test ; (3) estimating the skill mastery profiles for individual respondents based on their performance on the test using the CDM; and (4) reporting mastery/nonmastery of the skills to respondents and other score users. To define attributes involved in a test various sources such as test specifications, theories of content domain, item content analysis, think-aloud protocol analysis of respondents' test taking process can be sought (Embretson, 1991; Leighton & Gierl, 2007; Leighton, Gierl, & Hunka, 2004). Since the present study is a case of retrofitting CDA to an existing non diagnostic test (extracting diagnostic information from an existing non diagnostic test), a detailed cognitive model of task performance was not available therefore we got two experienced university instructors to brainstorm on the possible attributes measured by the test. They specified a set of five attributes underlying the reading test, e.g., vocabulary, syntax, extracting explicit information, connecting and synthesizing, and making inferences. Then two other university instructors with over five years of teaching reading comprehension experience were asked to independently specify the attributes measured by each of the 34 reading comprehension items.

As in all CDMs the present study utilizes a Q matrix (Tatsuoka, 1985), a $I \times K$ matrix, to indicates the attributes required to get each item right. The elements of the matrix, q_{ik}, are valued 1 if the ith item requires the kth skill and 0 if not. Table 2 is part of the Q-matrix constructed for the present study. It describes the attributes involved in answering Items 1 to 5 of the reading comprehends test analyzed in this study. As you can see Item 1 requires Skill 4, Items 2, 3, and 4 require Skill 3, and Item 5 requires Skills 2 and 3.

Table 2. Part of the *Q*-matirix used in the present study

	Skill 1	Skill 2	Skill 3	Skill 4	Skill 5
Item1	0	0	0	1	0
Item2	0	0	1	0	0
Item3	0	0	1	0	0
Item4	0	0	1	0	0
Item5	0	1	1	0	0

3.3. Analysis

Using the DINA model, the two item parameters, the guessing and slipping parameters, with their standard errors were estimated for 34 reading comprehension items as shown in Table 3.

Table 3. Guessing and slipping parameters of the DINA model

	guess s	e.guess	slip s	e.slip
item1	0.548	0.032	0.116	0.01
item2	0.541	0.02	0.235	0.016
item3	0.59	0.02	0.055	0.009
item4	0.665	0.019	0.109	0.012
item5	0.421	0.019	0.261	0.019
item6	0.142	0.013	0.733	0.016
item7	0.223	0.016	0.622	0.019
item8	0.465	0.02	0.191	0.015
item9	0.368	0.019	0.551	0.019
item10	0.349	0.019	0.402	0.018
item11	0.142	0.013	0.819	0.015
item12	0.321	0.03	0.412	0.016
item13	0.615	0.019	0.17	0.014
item14	0.424	0.02	0.279	0.018
item15	0.735	0.017	0.009	0.004
item16	0.587	0.019	0.034	0.007
item17	0.383	0.019	0.333	0.019
item18	0.203	0.015	0.663	0.02
item19	0.366	0.02	0.261	0.017
item20	0.562	0.02	0.088	0.011
item21	0.417	0.02	0.273	0.017
item22	0.323	0.019	0.315	0.018
item23	0.321	0.019	0.277	0.017
item24	0.141	0.022	0.569	0.017
item25	0.298	0.018	0.275	0.018
item26	0.145	0.015	0.462	0.019
item27	0.306	0.018	0.642	0.019

item28	0.156	0.014	0.74	0.017
item29	0.127	0.013	0.723	0.016
item30	0.325	0.03	0.306	0.015
item31	0.309	0.019	0.401	0.019
item32	0.321	0.019	0.279	0.018
item33	0.148	0.014	0.692	0.018
item34	0.259	0.018	0.584	0.019
mean	0.36		0.38	

The average values of the guessing and slipping parameters are .36 and .38. The mean guessing parameter indicates that for the students who have not mastered all the required skills for an item, there is still, on average, a 36 percent chance that they will choose the correct response and the average slipping parameter indicates that for the students who have mastered all the skills required for an item, there is still, on average, a 38 percent chance that they will choose the incorrect response. The most informative items on a test are the ones whose slipping and guessing probabilities are low (Rupp, et al., 2010). Generally speaking, small guessing and slipping parameters indicate a good fit between the diagnostic assessment design, the response data, and the postulated DINA model. However there is no hard and fast rule as to what constitutes "small". The results in Table 3 show that some items like 1, 2, 3, 4, 6, 7,… have higher guessing and/or slipping parameters (>.5) compared to others. Slipping and guessing parameters are also schematically represented in Figure 4.

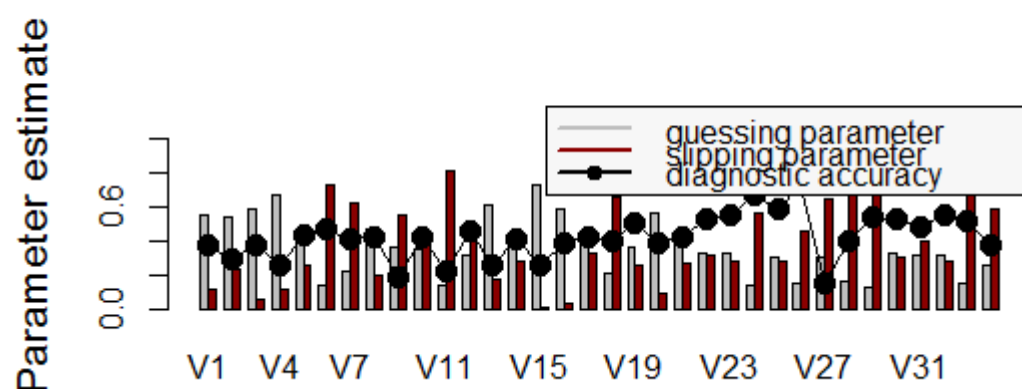**Figure 4.** DINA model parameters

Table 4 shows probabilities that respondent one to five belong to any of the $32(2^5)$ latent classes.

Table 4. Posterior skill distribution of the respondents

Latent Class	Skill Profile	Respondent1	Respondent2	Respondent3	Respondent4	Respondent5
1	00000	0.615	0.59	0.608	0.534	0.602
2	10000	0	0.001	0	0	0.008
3	01000	0.002	0.021	0.007	0.003	0.004
4	00100	0	0	0	0.1	0
5	00010	0.001	0.004	0.004	0.001	0.005
6	00001	0.38	0.365	0.375	0.33	0.372
7	11000	0	0	0	0	0
8	10100	0	0	0	0.002	0
9	10010	0	0	0	0	0.001
10	10001	0	0	0	0	0.004
11	01100	0	0	0	0.002	0
12	01010	0	0.001	0	0	0
13	01001	0.002	0.018	0.005	0.003	0.003
14	00110	0	0	0	0.003	0
15	00101	0	0	0	0.012	0
16	00011	0	0	0.001	0	0
17	11100	0	0	0	0	0
18	11010	0	0	0	0	0
19	11001	0	0	0	0	0
20	10110	0	0	0	0	0
21	10101	0	0	0	0	0
22	10011	0	0	0	0	0
23	01110	0	0	0	0	0
24	01101	0	0	0	0	0
25	01011	0	0	0	0	0
26	00111	0	0	0	0.006	0
27	11110	0	0	0	0	0
28	11101	0	0	0	0	0
29	11011	0	0	0	0	0
30	10111	0	0	0	0	0
31	01111	0	0	0	0	0
32	11111	0	0	0	0.001	0

In Table 4 values for each respondent represent the posterior probability that he belongs to latent class *c* with the given skill profile. For example, for Respondent 1 the skill profile with highest posterior probability is α1=[00000] which classifies him in the first latent class. It means that respondent1 has a 61% chance of actually belonging to latent class one. In fact it means that there is a 61% chance that he has not mastered any of the five skills required by the reading test. For the same respondent the attribute profile with the second highest posterior probability is α6=[00001]. It means that respondent1 has a 38% chance of

actually belonging to latent class six hence having mastered only the fifth skill required by the reading test.

Table 5 represents the estimated occurrence probabilities of the skill classes and the expected frequency of the attribute classes given the model.

Table 5. Class probabilities and expected frequencies

	Class probability	Class exp. frequency
00000	0.224	346.838
10000	0.009	13.859
01000	0.004	5.599
00100	0.007	10.32
00010	0.006	8.971
00001	0.162	251.59
11000	0.001	2.029
10100	0.01	15.907
10010	0.004	6.445
10001	0.005	7.346
01100	0	0.271
01010	0.001	1.118
01001	0.004	5.523
00110	0.008	12.004
00101	0.001	1.431
00011	0.005	8.418
11100	0.002	3.595
11010	0	0.09
11001	0.001	1.602
10110	0.003	3.989
10101	0.002	3.18
10011	0.016	24.455
01110	0.001	1.678
01101	0	0.026
01011	0	0.499
00111	0.028	44.089
11110	0.002	3.577
11101	0.003	5.141
11011	0.006	9.836
10111	0.06	92.471
01111	0.003	5.265
11111	0.421	652.841

Table 5 provides, for each Latent Class c with the skill profile given, the estimated number of respondents in that class and its conversion into a proportion. As we can read from the first column of the table approximately 22% of the respondents in the present study belong to the first latent class with a skill profile of $\alpha_1=[00000]$ hence is expected to have not mastered any of the five attributes. Furthermore latent class 32 with attribute profile of $\alpha_{32}=[11111]$ has class probability of about .42, indicating that approximately 42% of our population is expected to have mastered all five skills. The third column provides the expected count of respondents with each class. This count is found by multiplying the value from the second column, class probability, by the sample size. For example, for our data we see that latent class 1has got class probability of about .22 if we multiply this by the size of the population for the present study which is 1550 we get 346.838 which means that approximately 346.838 respondents are expected to belong to the first class.

Figure 5. Pattern Occurrence Probabilities
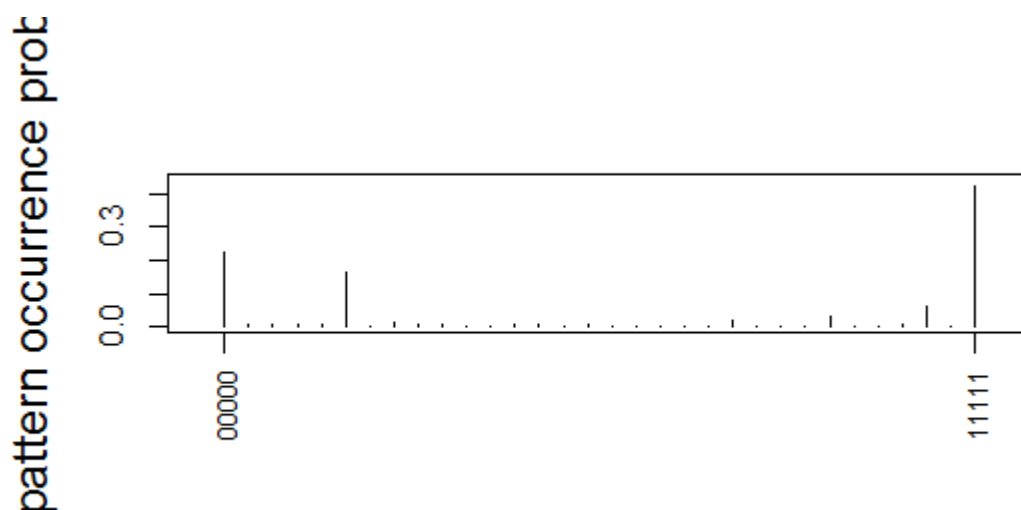
As Figure 5 also shows the two skill patterns with the highest probabilities are skill profile1, $\alpha_1=[00000]$ and skill profile 32, $\alpha_{32}=[11111]$. It indicates that in the present study about 42% of the respondents were classified into latent class32 as masters of all skills and about 22% were classified into latent class1 as masters of non of the five skills. We note ,as tables 4 and 5 show, a large number of respondents are classified as having a "flat" profiles i.e. masters of either none (00000) or all (11111) attributes rather than a particular combination of attributes. This command also gives us skill probabilities in Table 6. The table shows that approximately 54% of the respondents have mastered Skill 1 and 71% have mastered skill5.

Table 6. Skill probabilities

Skill1	Skill2	Skill3	Skill4	Skill5
0.546	0.451	0.552	0.565	0.719

In line with other CDM studies, as we can see in the above plot, in the present study also most of the respondents have been diagnosed as having flat skill profiles (00000) or (11111). The plot shows that skill profile $\alpha32=[11111]$ has the highest probability and the skill profile $\alpha1=[00000]$ has got the second highest and skill profile $\alpha6=[00001]$ the third highest probability.

Current versing of CDM package in R produces diagnostic accuracy for each item along with AIC and BIC fit indices. However, Item Discrimination Indices (IDI) are recommended and preferred to diagnostic accuracy indices, therefore, we computed and used IDIs in this study. Diagnostic accuracy indices are not easy to interpret and are going to be replaced by IDI in the future versions of CDM (Robiztch, September, 2012, personal communication).

In classical test theory (CTT) discrimination is a measure of the relationship between item score and total test score. Unlike many assessments analyzed within CTT or other unidimensional IRT models CDMs measure multiple attributes hence the concept of discrimination can be expressed as "how well does an item differentiates between respondents who have mastered more attributes and respondents who have mastered fewer attributes" (Rupp, et. al., 2010). As a matter of fact IDI is a quick and simple measure of the diagnostic value of an item. There are two different methods for computing IDI in CDM, one global IDI and another method for describing attribute specific item discrimination. DINA model contains only item level parameters whose values don not depend on the number of attributes that each item is measuring. Unlike models like NIDA which compute separate slipping and guessing parameter for each attribute involved in an item, DINA model estimates one slipping and guessing parameter per item regardless of the number of attributes involved in the item. Rupp, et. al. argue that in DINA model maximal attribute discrimination occurs when comparing a respondent who has mastered all attributes to a respondents who has not mastered any measured attributes. Consequently, the values of attribute specific item discrimination indices will be identical for each item and only global IDIs are reported. Since the new version of CDM is not still out we manually computed DINA model IDIs with the following equation;

$$d_{iDINA}= (1-s_i)-g_i \qquad\qquad (3)$$

The last column of Table 7 shows global item discrimination indices. As we can see the lower the slipping and guessing values are the higher the IDIs and the better the item is. The pattern of IDIs show that most of values fall at low end of the distribution but some like 23, and 25 are in better condition than the other items. There is no hard and fast rule as to a cut off for IDIs but you can compare them to decide which items to include in a test, especially if you have the luxury of having more items than necessary (de la Torre, personal communication, September 2012).

Table 7. item discrimination indices

Column1	guess s	slip s	IDI
item1	0.548	0.116	0.336
item2	0.541	0.235	0.224
item3	0.59	0.055	0.355
item4	0.665	0.109	0.226
item5	0.421	0.261	0.318
item6	0.142	0.733	0.125
item7	0.223	0.622	0.155
item8	0.465	0.191	0.344
item9	0.368	0.551	0.081
item10	0.349	0.402	0.249
item11	0.142	0.819	0.039
item12	0.321	0.412	0.267
item13	0.615	0.17	0.215
item14	0.424	0.279	0.297
item15	0.735	0.009	0.256
item16	0.587	0.034	0.379
item17	0.383	0.333	0.284
item18	0.203	0.663	0.134
item19	0.366	0.261	0.373
item20	0.562	0.088	0.35
item21	0.417	0.273	0.31
item22	0.323	0.315	0.362
item23	0.321	0.277	0.402
item24	0.141	0.569	0.29
item25	0.298	0.275	0.427
item26	0.145	0.462	0.393
item27	0.306	0.642	0.052
item28	0.156	0.74	0.104
item29	0.127	0.723	0.15
item30	0.325	0.306	0.369
item31	0.309	0.401	0.29
item32	0.321	0.279	0.4
item33	0.148	0.692	0.16
item34	0.259	0.584	0.157

It is also possible to manually compute the probability that each respondent has mastered each individual skill. Heubner, (2010) argues that "since the latent classes are mutually exclusive and exhaustive, we may simply add the probabilities of the latent classes associated with each skill" (p. 5). For example, for Respondent 5, the probability that he/she has mastered Skill 1 is computed by adding the probabilities of him belonging to the latent classes associated with skill 1. We can read these probabilities from Table 3, from the skill profile column where the values for the first skill are nonzero. As we read down the table we see that Skill 1 is associated with the following Latent Classes: 2, 7, 8, 9, 10, 17, 18, 19, 20, 21, 22, 27, 28, 29, 30, and 32. Now we add up the probabilities of Respondent 5 belonging to each one of these latent classes; P(skill1)= p([10000])+p([11000])+p([10100])+.......+p([11111])

Only three latent classes have nonzero probabilities. Therefore, the probability that Respondent 5 has mastered Skill 1 is: .008+.001+.004= .013

As we see the probability that he has mastered skill one is much less than 1%. Similarly we can calculate probabilities of skill2. As we see in table 3 skill 2 is associated with the following latent classes:3,7,11,12,13,17,18,19,23,24,25,27,28,29,31, and 32. Now we add up probabilities of respondent 5 belonging to each one of these latent classes; P(skill1)= p([01000])+p([11000])+p([01100])+.......+p([11111])

Because only two latent classes have non zero probabilities, to save space we ignored zero probabilities. Therefore probability that respondent five has mastered skill 2 is .004+.003= .007

It is no surprise that the probability that he has mastered the skills is so low because he belongs to the first latent class whose members with skill profile of α1=[00000] are masters of non of the five skills.

4. Discussion

The present study explored the diagnostic capacity of items on the UI reading comprehension test. Rupp et al. (2010) note that diagnostically informative items are those with low slipping and guessing parameters. Another index that shows the diagnostic value of an item is IDIs. Technically speaking IDI is a difference in the probability for a correct response between respondents who have mastered "more" measured attributes for an item and those who have mastered "fewer" measured attributes (Rupp et al., 2010).

The IDIs, as shown in Table 7, are well below .50. IDIs generally vary between 0 and 1; the higher the IDI the better the diagnostic value of the item is. As we can see from Equation 3 above IDI is a function of guessing and slipping parameters for each item. The lower the guessing and slipping parameters are the higher the IDI will be. And the reason for low IDIs in the present study is high guessing and slipping parameters. As we can see from Table 2 guessing and slipping parameters are relatively high. Although there is no consensus on how low the guessing parameters should be to be considered low, as a rule of thumb we can consider those above the midpoint ($>.5$) as high. For example, for item four, for those who have not mastered all or any of the required skills by the items, still there is about 65% chance of getting the item right which sounds unreasonable. But in the case of Item 6 for those who have mastered all the subskills required by the item there is about 73% chance of getting the item wrong which is also unreasonably high.

Two possible explanations for the high guessing parameters can be presented. It may be that the required skills are compensatory instead of conjunctive for those items and students (which violates the model assumption about the nature of the skills), so the students do not necessarily need to have all the required skills to answer these items correctly. An important consideration before analyzing the data is determining the attribute structure or the relationships among attributes involved in a test.

If we assume a conjunctive noncompensatory relationship among the subskills, models like Latent Class Analysis (LCA) (Yamamoto, 1989), Rule Space Model (RSM) (Tatsuoka, 1983), Deterministic Inputs, Noisy- And gate (DINA) model (de la Torre & Douglas, 2004; Sijtsma & Junker, 2006) or Non Compensatory Reparameterized Unified Model (NC-RUM) which is also known as the Fusion Model (DiBello, Stout, & Roussos, 1995; Hartz, 2002) can be used. But if we assume a disjunctive compensatory relationships among the attributes models like Deterministic Noisy Or-gate (DINO) model (Templin & Henson, 2006) or Compensatory Reparameterized Unified Model (RUM) (Hartz, 2002; Templin, 2006) can be used. For a good review of CDMs, their features and the available software refer to Lee and Sawaki (Lee & Sawaki, 2009a) and Rupp and Templin (2008). "Non compensatory models have been preferred for cognitive diagnostic analysis, as they can generate more fine grained diagnostic information" (Li, 2011,p.40).

There is no clear cut answer as to whether noncompensatory models are superior to compensatory models with reading tests. Lee and Sawaki (2009a) compared respondent classification results across LCA, NC-RUM, and compensatory General Diagnostic Model (GDM) (von Davier, 2005) for TOEFL iBT reading and listening data. They found highly similar results across the three models. Another study by Jang (Jang, 2005) showed that the attribute structure of the TOFEL iBT is a mixture of compensatory and non compensatory relationships. Li (2011) suggests that it may seem reasonable to hypothesize that "the relationships among subskills depend on the difficulty level of the subskills needed for solving a particular item" (p.40). Thus this relationship may change across items within the same test. Henson, Templin, and Willse (2008) introduced Log linear Cognitive Diagnostic Model (LCDM) which assumes no restrictive relationship such as conjunctive or disjunctive among the attributes involved in providing the correct answer to an item. Henson, Templin, and Willse (2008) found that the DINA model does not seem reasonable for all items. In addition, the LCDM provided some insight as to what model could be more appropriate for some items. Thus with LCDM the choice between a compensatory and non compensatory model is not an issue of concern and subskills can have varying relationships from item to item. To test whether choice of the model accounts for the high slipping and guessing parameters, instead of the DINA model one can fit its compensatory analog DINO to the data and then compare the general fit indices of the two models such as Akaike Information Criterion (AIC) and Bayesian Information Criterion(BIC) (Huebner, 2010). If the fit indices of the DINO model are smaller we can conclude that the attribute structure of the test is disjunctive compensatory rather than conjunctive noncompensatory.

Another possible reason for high slipping and guessing parameters is the misspecification of the Q-matirix (Rupp & Templin, 2008). The reason for the high slipping parameters could be incompleteness of the Q-matrix—some unspecified skills or strategies could be required to respond correctly. Rupp and Templin in a study aimed at investigating the effects of Q-matrix misspecifications on parameter estimates and misclassification rates for the DINA model found that incorrectly omitting the attributes in the Q-matrix the slipping parameter for a misspecified item was overestimated most strongly since the item turned out to be much easier making it more likely that slipping must have occurred. In contrast, the

incorrect addition of attributes in the Q-matrix for a particular item resulted in a strong overestimation of the guessing parameter for the misspecified item because the item turned out to be much more difficult making it more likely that guessing must have occurred.

Another significant finding of the present study is the prevalence of "flat" skill mastery profiles, namely, "nonmaster of all skills" $\alpha_1=[00000]$, and "master of all skills" $\alpha_{32}=[11111]$, which is in line with all other CDM studies carried out by other researchers (e.g., Lee & Sawaki, 2009b; Li, 2011) . Approximately 42% of the respondents were classified as masters of none of the skills, and 62% were classified as masters of all skills. Two explanations have been put forward for this finding; 1) it could be due to high positive correlations among the attributes (Rupp, et al., 2010), and 2) unidimensionality of the measure used, where a master of one skill tends to be a master of another skill, or vice versa (Lee & Sawaki, 2009b).

Since the secondary purpose of the present paper was to demonstrate the application of CDM with R freeware, our choice of the model was limited to the DINA model which is the only R package available for cognitive diagnostic modeling. Future studies may compare results obtained from DINA which assumes conjunctive noncompensatory relationships among the attributes involved in a test with LCDM (Henson, Templin, & Willse, 2008) which assumes no restrictive relationship such as conjunctive or disjunctive among the attributes involved in providing the correct answer to an item.

Another limitation of the present study was related to attribute definition and Q-matrix construction. To define attributes in the present study we solely relied on content area experts' judgment. Future studies may combine expert judgment with think aloud protocol analysis of respondents' test taking processes. We could have also submitted the Q-matrix to empirical analysis based on Fusion model. Fusion model provides item difficulty and discrimination parameter estimates which are helpful in identifying potential Q-matrix misspecification.

References

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing, 15*(2), 119-157. doi: 10.1177/026553229801500201

Buck, G., VanEssen, T., Tatsuoka, K., Kostin, I., Lutz, D., & Phelps, M. (1998). Development, selection and validation of a set of cognitive and linguistic attributes for the SAT I Verbal: Sentence completion section. Princeton, NJ:: Educational Testing Service.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115-130. doi: 10.3102/1076998607309474

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333-353. doi: 10.1007/bf02295640

de la Torre, J. (2011). The Generalized DINA model framework. *Psychometrika, 76*(2), 179-199. doi: 10.1007/s11336-011-9207-7

de la Torre, J. (2012, July). *Recent developments in the G-DINA model framework.* Paper presnetd at the V European Congress of Methodology, Santiago de Compostela, Spain.

DiBello, L. V., Roussos, L. A., & Stout, W. (2006). A review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. Volume 26, pp. 979-1030): Elsevier.

DiBello, L. V., Stout, W., & Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. Nichols, S. Chipman & R. Brennan (Eds.), *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*(3), 495-515. doi: 10.1007/bf02294487

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. [Empirical Study]. *Psychological Methods, 3*(3), 380-396. doi: 10.1037/1082-989x.3.3.380

Gao, L. (2007). *Cognitive psychometric modeling of the MELAB reading items.* Dissertation Dept. of Educational Psychology University of Alberta.

Hartz, S. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practice.* Unpublished doctoral thesis, University of Illinois at Urbana-Champain.

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191-210. doi: 10.1007/s11336-008-9089-5

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation, 15*, 1.

Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG-TOEFL.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign, Urbana, IL.

Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle, Y. R. Chung & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment*. Ames, IA: Iowa State University.

Kasai, M. (1997). *Application of the rule space model to the reading comprehension section of the test of Englishas a foreign language (TOEFL).* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign, Urbana, IL.

Lee, Y.-W., & Sawaki, Y. (2009a). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*(3), 239-263. doi: 10.1080/15434300903079562

Lee, Y.-W., & Sawaki, Y. (2009b). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly, 6*(3), 172-189. doi: 10.1080/15434300902985108

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education : Theory and applications*. Cambridge, MA: Cambridge University Press.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement, 41*(3), 205-237.

Li, H. (2011). A Cognitive Diagnostic analysis of the MELAB reading test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 9*, 17–46.

Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively diagnostic assessment.* Hillsdale, NJ: Lawrence Erlbaum.

Rupp, A. A. (2007). The answer is in the question: A guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing, 7*(2), 95-125.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: The Guilford Press.

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives, 6*(4), 219-262. doi: 10.1080/15366360802490866

Sheehan, K. M. (1997). A tree-based approach to proficiency scaling and diagnostic assessment. *Journal of Educational Measurement, 34*(4), 333-352.

Sijtsma, K., & Junker, B. W. (2006). Item response theory: past performance, present developments, and future expectations. *Behaviormetrika, 33*, 75-102.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*(4), 345-354.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational and Behavioral Statistics, 10* (1), 55-73. doi: 10.3102/10769986010001055

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. [Mathematical Model]. *Psychological Methods, 11*(3), 287-305. doi: 10.1037/1082-989x.11.3.287

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

Yamamoto, K. (1989). HYBRID model of IRT and latent class model *ETS research report series (RR 89-41)*. Princeton, NJ: Educational Testing Service.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items [Computer software]. Chicago: Scientific Software International.

Appendix

DINA application with R

R is a freeware software that can be found at: http://www.r-project.org/. There are two windows that you will mainly use in R. The main window is called the **console** and it is where you can both type commands and see the results of executing these commands (in other words, see the output of your analysis). Rather than writing commands directly into he console you can also write them in a separate window (known as the **editor** window). Working with this window has the advantage that you can save collections of commands as a file that you can reuse at another point in time. As you open R automatically, the console window also opens. To open the editor window first open R then go to **File.** If you want to modify already saved R commands click on **open script** and then go to the path where you have previously saved the editor but if you want to write new commands click on **new script**.

R comes with some base functions ready for you to use. However to get the most out of it we need to install packages that enable us to do particular things. For example,, to do cognitive diagnostic modeling we need to install CDM package. We can install packages in two ways: through the menus or using a command. If you know the package that you want to install then the simplest way is to execute this command:

Install.packages("package.name")

For example, to install CDM package we replace "package. name" with "CDM", therefore we execute:

Install.packages("CDM")   (1)

Note that the name of the package must be enclosed in speech marks. To execute a command we put the cursor on it and press ctrl+R.

Alternatively we can install a package through the menu system. To do so go the R window select **Packages→install package(s)** a window will open that first asks you to select the country where CRAN(Comprehensive **R** Archive Network) is located. For security reasons identical versions (mirrors) of CRAN are stored in different locations across the globe. As a resident of Iran I would likely access CRAN in Iran, whereas if you are in a different country you can get access to the copy of the CRAN in your own country(or one nearby). Having selected the CRAN nearest to you from the llis and clicked on **OK** , a new dialog box will open that lists all of the available packages. Click on the one or one that you want(you can select several by holding down the Ctrl key as you click) and then click on **OK.** This will have the same effect as using the *Install.packages(" ")* command.

Once a package is installed you need to load it for R to know that you are using it. You need to install the package only once but you need to load it each time you start a new session of **R.**

To load a package we simply execute this general command:

library(package.name)

For example, to load CDM package we simply replace *package.name* with CDM:

library(CDM)   (2)

Alternatively we can load packages through the menu system. You can load packages by **selecting Packages→Load package** which opens a dialog box with all of the available packages that you could load.

Every time you want to do anything(e.g., load or save a file) from **R** you have to establish a working directory in which you want to store your data files, any scripts associated with the analysis or your workspace. we can set the working directory, either through the menus or using a command. We can use the **setwd()** command to specify the working directory:

setwd("location of your file")

To specify the location of your file you can simply go to the drive and folder where your file is to be stored click in the address bar in the window copy and paste the file path into the brackets. For example, in this case we have saved my data file and Q-matrix file in drive H and folder *DINA* therefor I go to the folder *DINA* click in the address bar as shown in the figure below, the file path gets blue. I copy the path and past it as follows:

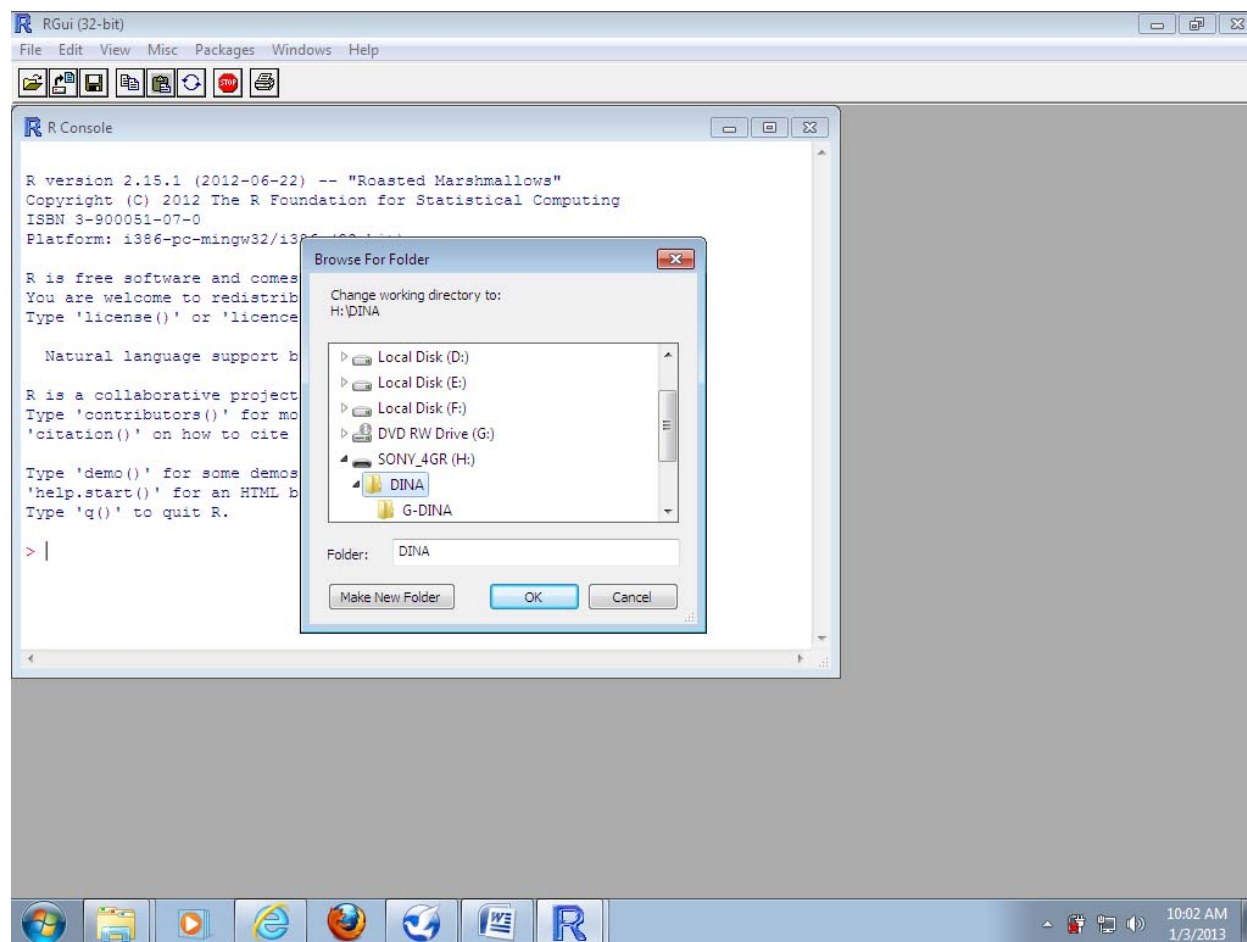setwd("H:/DINA")   (3)

Note that we have to change all the backslashes(\) in the file path into slashes(/).
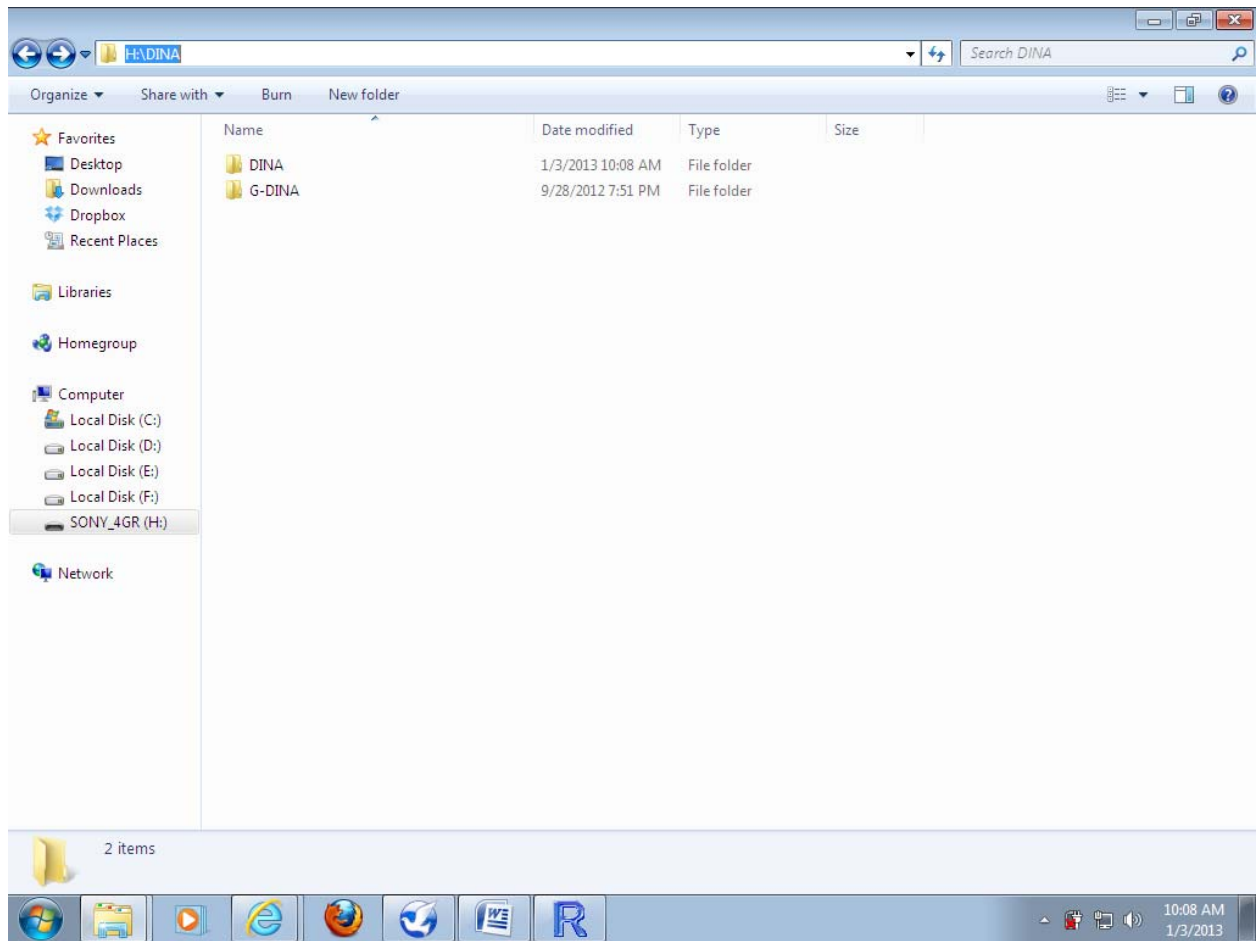
Alternatively we can click anywhere in the R console. In the **File** select **Change dir…**
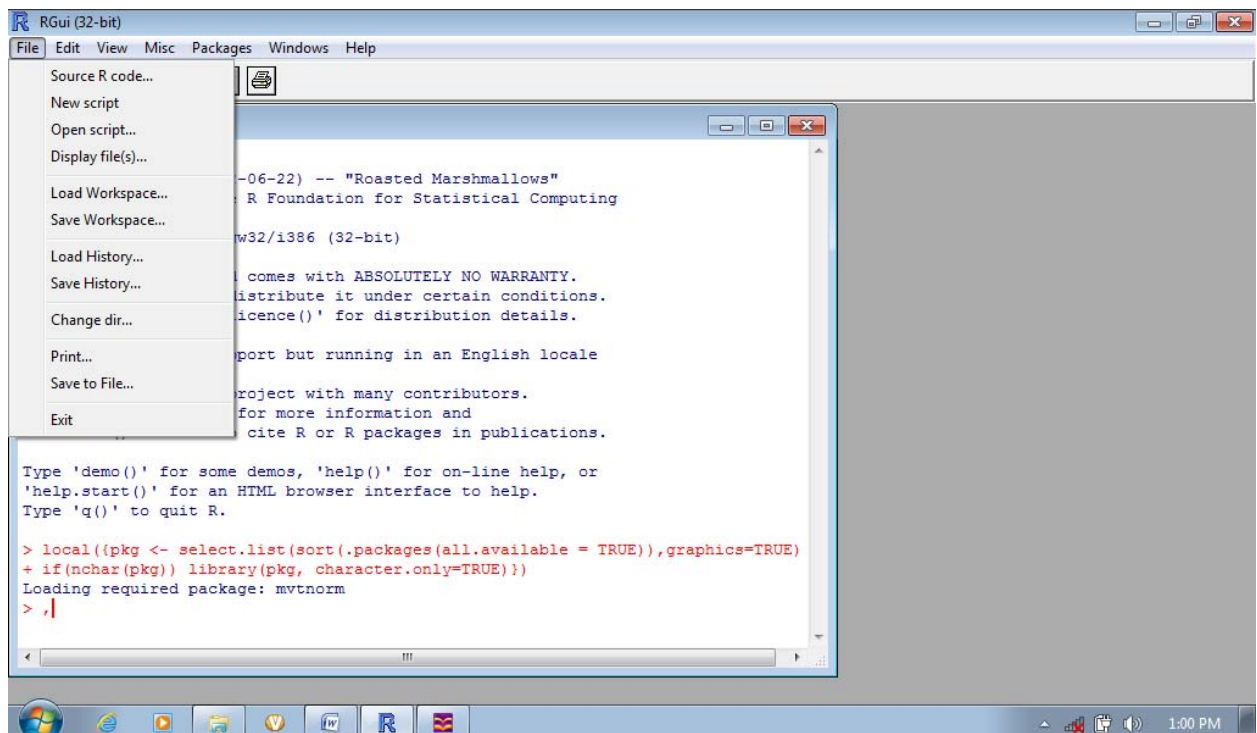
A dialog box will open that asks you to select the path where you want to save your files or where your data and Q-matrix files are stored as follows:

Select the folder and click on **OK**.

The **foreign** package can be used to import directly data files into **R.** The two most commonly used R-friendly formats are tab-delimited test(*.txt* in Excel and *.dat* in SPSS) where values are separated with a tab space and comma separated values (*.csv*) where the values or data are separated with commas.

If we have saved the data as CSV file, then we can import these data to a data frame using the *read.csv* function. The general form of the function is:

Dataframe.name<-read.csv("filename.extension",header=TRUE)

To do cognitive diagnostic modeling we must have two separate files: a data file where the respondents' data are stored and a Q-matrix file where the attributes or subskills measured by each item are specified.

Let's imagine we have stored our data and Q-matrix file in *dat* and *csv* files called *Reading.dat/Q-matrix.csv* respectively in the directory set before. To load these data into a dataframe we could execute the following commands:

setwd("H:/DINA")   (3)

mydata<-read.delim("Reading.dat",header=TRUE) (4)

QDINA<-read.csv("Q-matrix.csv",header=FALSE) (5)

The first command sets the directory to the *D:/DINA* path where our data file (*Reading.dat)* and Q-matrix file (*Q-matrix.csv)*. with these command we tell R where to read the data file and Q-matrix file from and where to save the new data frames.

Prior to explaining the second command lets's diverge a little to present some general points about how commands are produced and interpreted in R. Commands in R are generally made up of two parts: objects and functions. These are separated by '<-' which you can think of as meaning ' is created from'. As such the general form of a command is

Object<-function

Which means 'object is created from function'. An object is anything created in R. Functions are the things that you do in R to create your objects.

The second command will create an object called **mydata** (it could be any name) by reading the data file called 'Reading.dat' which is stored in the directory specified.

The third command will create another object called **QDINA** (it could be any name) by reading the file called 'Q-matrix.dat' which is stored in the directory specified.

The *header = TRUE* in the commands tells R that the data files has variable names in the first row of the file but if you have saved the file without variable names then you should use *header = FALSE*

For the purpose of the present study we used cognitive diagnostic modeling package developed by Robitzsch et al. (2012). The main function of CDM is:

din(data,q.matrix,rule="DINA")

With the objects earlier created(**mydata** and **QDINA**) the function becomes:

dina<-din(data=mydata,Q-matrix=QDINA,rule="DINA") (6)

This command will create an object called **dina** (it could be any name) from the data called *mydata* and Q-matrix called *QDINA* according to DINA rule

To execute the command we put the cursor on the command line and then press Ctrl+R keys. You have to wait for a while till the computations are performed depending of how fast your PC is. As you can see in R console window, after some iterations the model converges.

The guessing and slipping parameters and their standard errors can be estimated by execution of the following command.

dina$coef   (7)

Dollar sign ($) in R means that the operation specified after $ should be carried out within the object named before $. This command tells R to compute the coefficients from the object made in command 6. Table 2 shows the results

Executing the command

dina$posterior

returns a matrix given the posterior skill distribution for all respondents. Table 3 shows the probabilities that each respondent belongs to any of the $32(2^5)$ latent classes. For space considerations the table has been transposed; the rows represent latent classes and the columns represent the respondents. We have also rounded the numbers to two decimal points by executing the following command:

round(dina$posterior[, 1:32],2)

this command tells R to round the output from the command *dina$posterior* to 2 decimal points in a way that rows remain intact(note that in the brackets we don't have anything before comma therefore the command is not applied to the rows) and columns 1 to 32 are rounded.

Executing the following command:

dina$attribute.patt

returns the estimated occurrence probabilities of the skill classes and the expected frequency of the attribute classes given the model, as shown in Table 4.

Executing the following command:

print(din)

returns highest skill pattern probabilities as shown in Table 5.

We can also get a visual inspection of the outputs by executing the following command:

plot(din)

Each time that we press Enter key we get one of the following plots shown above;

There is one more output which can be requested from R by executing the command;

summary(dina)

it returns a table with the diagnostic accuracy of each item along with AIC and BIC fit indices. diagnostic accuracy indices are not easy to interpret and are going to be replaced with Item Discrimination Indices (IDI). Robiztch (personal communication, September, 2012) recommends ignoring the diagnostic accuracy indices and report IDIs instead. Therefore, we manually computed IDIs using Equation 3 above and presented them in Table 7.