Tabaran Institute of Higher Education
International Journal of Language Testing
Vol. 6, No. 2, October 2016
ISSN 2476-5880

Cognitive Diagnostic Modeling of L2 Reading Comprehension Ability: Providing Feedback on the Reading Performance of Iranian Candidates for the University Entrance Examination

S. Jamal Hemmati[1], Purya Baghaei[*2], Masoumeh Bemani[3]

Received: 2 April 2016 Accepted: 16 September 2016

Abstract

Research has been conducted on linguistic aspects of L2 reading comprehension ability but studies which concern cognitive aspects of this skill are rather scarce. Efforts should be made to have sound understanding of the attributes and subskills needed for successful performance on L2 reading comprehension tests. In this study the reading attributes underlying the reading comprehension section of the Iranian National University Entrance Examination (Konkoor) held in 2011are investigated. The G-DINA model, a general cognitive diagnostic model, is applied and the CDM package in R is used to run the analyses. The findings indicate problematic areas in reading comprehension at a national level. Students who desire to pursue their studies at national universities, the teachers who train the candidates at high schools and the ministry of education can benefit from such feedbacks.

Keywords: *Cognitive Diagnostic Modeling, Reading comprehension, UEE*

1. Introduction

Second language (L2) reading ability, as one of the major language skills has been the subject of many studies and researchers try to shed light on different aspects of this complex skill. Lots of studies have been conducted on linguistic aspects of L2 reading ability but studies which address cognitive aspects of this skill are rather scarce and still lots of efforts are needed in order to have a sound understanding of cognitive attributes involved in successful performance on L2 reading tests.

Here, focusing on the cognitive aspects of reading comprehension assessment, we need to develop our understanding of the cognitive abilities involved in successful performance on reading comprehension test items. When teachers, students, curriculum developers and authorities who are in charge of preparing instructional materials and books, know about the

[1]*English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran*
** Corresponding author, Email: baghaei@mshdiau.ac.ir*
[2]*English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran*
[3]*English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran*

attributes and subskills involved in reading comprehension and they have access to the learners' mastery profiles in detail, it would be possible for them to concentrate more on the reading subskills. In such a case, particularly the teachers are able to help students improve their reading comprehension ability by taking remedial treatments based on individual performance on a reading comprehension test. To this end, in the present study, the cognitive diagnostic models serve as a means to provide better understanding of the attributes involved in reading comprehension. To answer a specific item in a test, test-takers need to use certain underlying attributes. In other words, test-takers should master those attributes to be able to answer each specific item. Attributes are domain-specific skills and knowledge that are needed in order to demonstrate mastery in a cognitive domain (Leighton &Gierl, 2007).

We are not able to determine test-takers' mastery levels of cognitive attributes only based on the row scores they get. To investigate the hypotheses about cognitive attributes involved in a specific language skill andto provide attribute mastery profiles for individual test-takers, psychometricians have developed a method called cognitive diagnostic modeling (CDM). CDMshave the capability to provide detailed diagnostic feedback about the reason why a test taker might succeed or fail on any given test (Ravand&Robitzsch, 2015). Cognitive diagnostic analysis promotes assessment for learning and the learning process as opposed to the assessment of learning outcomes (Jang, 2008). CDMs have been developed by a consensus in the measurement community that the psychology of test performance should be better understood to aid in constructing and scoring tests and interpreting test validity (Lee &Sawaki, 2009a). In this respect CDMs are multi-disciplinary as they integrate cognitive psychology with psychometric modeling **.**

The data used in this study is gathered from the candidates' responses to the reading comprehension section of the University Entrance Examination (Konkoor) for English programs at the Iranian state universities. The majority of candidates are those students who have just graduated from high schools; therefore, this sample well represents average reading ability of Iranian high school graduates. As a result, the findings of this study will give useful feedbacks to the educational system.The results of this study will provide further insight into the specific areas of reading that require more improvement on the part of the examinees.Being aware of high school students' strengths and weaknesses, program administrators will be able to develop a curriculum that meets students' needs.

The outcome will also provide university EFL instructors with valuable diagnostic feedback on test taker's reading ability. This feedback enables teachers and instructors to remedy their students' reading deficiencies. In other words, this technique can serve as a needs analysis device by which the reading comprehension subskills, that learners need to focus on, are determined.

2. Method

2.1. Participants

The sample included 10000 candidates. The candidates were high school graduates who desired to pursue their studies for a bachelor's degree in English in Iranian state universities. 28% of the candidates were male and 72 % of them were female. The test is held annually all over the

country and the participants in the test are from different cities of Iran; therefore, we have a good representative sample of Iranian student population.

The test-takers had studied English as a foreign language at school. As a fixed part of their curriculum, Iranian high school students receive at least 3 hours of English instruction each week. However, most of the schools hold extra classes to make the candidates prepared for the examination, particularly in the last year of high school before the Konkoor. What is more, some of the students attend English classes in private institutes since their childhood.

The participants were chosen randomly among thousands of candidates who had taken part in the exam so they well represented the population. The sample size is large enough to account for the generalizability of the results of the study.

2.2. Instrument

The test analyzed in this study was the reading comprehension section of the University Entrance Examination (Konkoor) specific for English studies held in 2011. This Iranian nation-wide test is held by the National Educational Measurement Organization (Sazman e Sanjesh). Through this high-stakes examination, candidates, based on their performance on the test, are admitted to different state universities for bachelor programs in English studies. It is a four-option multiple-choice test and, is held annually.

The test consisted of 70 items and the candidates should have answered the questions in 105 minutes. The test included grammar, vocabulary, sentence structure, language functions, cloze test, and reading comprehension. Each section consisted of 10, 15, 5, 10, 10, and 20 items, respectively. The 20 reading comprehension items were selected for this study. This section included 3 texts. The texts contained 428, 503, and 466 words. The construct validity of the reading comprehension section of the test was confirmed in another research conducted by Hemmati and Baghaei (2016).

2.3. Procedure

Data for this study was obtained from the National Education Measurement Organization (Sazman e Sanjesh). For implementation of CDMs, we need to have test-takers' responses at the item level, not their total scores. It other words, we need to know whether they have answered each item correctly or not. Therefore, data at the item level were obtained.

To define attributes involved in a test, different sources might be used, including theories of content domain, test specifications, content analysis of the test items and think-aloud protocol analysis (Leighton, Gierl, &Hunka, 2004; Leighton &Gierl, 2007). In this study we used the results of a non-diagnostic test, a test which was not developed initially for diagnostic purposes, to extract diagnostic information about test takers' cognitive reading abilities (retrofitting case).

To determinethe attributes underlying the reading comprehension test items, in the initial step we referred to previous literature on language ability models. Other useful sources for identifying the attributes are the previous research conducted in the field of reading comprehension. Therefore, we referred to numerous studies in which the reading comprehension attributes were investigated (Baghaei&Ravand, 2015; Buck, Tatsuoka, &Kostin,1997; Clark, 2013; Jang, 2005; Kim, 2014; Lee &Sawaki, 2009a; Ravand, 2015; Ravand, Barati, &Widhiarso 2013; Sheehan, 1997; VanderVeen, Huff, Gierl, McNamara, Louwerse, &Graesser, 2007; Zheng & De Jong, 2011). At this stage we had a pool of reading comprehension subskills among which we had to

choose the attributes required for answering the test items. To this end, a team of experts, including the researchers, and two experienced English instructors determined the attributes measured by the test items based on the content analysis of each item. The process resulted in a list of 5 attributes which should be mastered by the candidates to successfully answer the test items. The attributes are: making inferences, extracting explicit information, identifying word meaning from the context, identify pronominal references, and evaluating response options.

As in all CDMs, in the present study, too, we used a Q matrix; an I × K matrix which indicates the attributes required to get each item right. The elements of the matrix, Ks, are valued 1 if the Ith item requires the kth attribute and 0, if not. To construct the Q-matrix the team of experts selected among the 5 attributes defined in the previous step and assigned them to the items. Then they checked the Q-matrices together and came to an agreement about the final Q-matrix which consisted of 5 columns representing the 5 attributes, and 20 rows, representing the items.

3. Analyses and Results

In the present study, we used a mixed research method, consisting of two main phases, the qualitative phase and the quantitative phase. In the qualitative phase the attributes required for answering the reading items are determined. In the quantitative phase the Q-matrix and the data were analyzed using the G-DINA cognitive diagnostic model.

In this study the generalized deterministic input, noisy and gate (G-DINA) model proposed by de la Torre (2011) is used to analyze the data. The G-DINA model, as a general CDM, allows the researcher to hypothesize varying relationships among the attributes across the items. In order to run the analysis, we used the "CDM" package version 4.5-0 (Robitzsch, Kiefer, George, &Uenlue, 2015) from the R freeware (R core team, 2015) version 3.2.2. R is a free open source tool which is utilized to conduct a wide range of statistical analyses and graphics.

To have a better understanding of the candidates' performance on the reading comprehension test, the main descriptive statistics of the test scores are provided in Table 1.The reliability of the test, which was calculated through Cronbach's Alpha, is 0.86.

Table1.*Descriptive Statistics*

N	Minimum	Maximum	Mean	Std. Error	Std. Deviation	Variance
10000	.00	18.00	2.4917	.03387	3.38749	11.475

The output of the analysis illustrates the latent class probabilities. The respondents' class membership is unknown; therefore the skill mastery patterns into which respondents are assigned are referred to as latent classes. The output which represents the latent class probabilities and frequencies is illustrated in Table 2. The first column of Table 2 shows the numbers allocated to each of the possible latent classes. As it was already mentioned, the number of latent classes varies as a function of the number of required attributes. In the present study the test-takers are classified into 2^5=32 different latent classes.

The attribute profiles related to each latent class are illustrated in the second column of Table 2. The attribute profiles show the mastery status of each attribute for individual test-takers. Mastery of each attribute is illustrated by 1 and non-mastery is illustrated by 0. In the present study we have five attributes therefore each mastery profile contains five digits. Imagine an examinee who has mastered attributes two and four, but he has not mastered attributes one, three, and five. The attribute mastery profile for such a candidate is (01010).

The probabilities of each latent class are in the third column of the table and its frequency is shown in the fourth column. Table 2 indicates that the first latent class with the attribute profile of (00000) has the highest class probability which is about 57%. It means that approximately 57% of the candidates of the Konkoor are classified in this latent class. As the fourth column shows it is expected that 5768 candidates belong to the first latent class, the members of which, have not mastered any of the attributes.

The second latent class with the highest probability is the one with the attribute profile of (11111), latent class 32 with the probability of approximately 10%. The candidates who belong to this latent class are expected to have mastered all of the attributes. Therefore in the presentstudy the attribute profiles of (00000) and (11111) have the highest probability in comparison to other thirty classes with an expected total frequency of approximately 6746 candidates. These two types of mastery profiles are called *flat profiles.* Latent class (00000) represents the mastery profiles of non-masters of all attributes and (11111) represents the profiles of masters of all the 5 attributes.

Table2.*Class Probabilities*

Latent Class	Skill Pattern	Probabilities	Class Expected Frequency
1	00000	0.5768	5768
2	10000	0.0129	129
3	01000	0	0
4	11000	0.0232	232
5	00100	0	0
6	10100	0.0062	62
7	01100	0	0
8	11100	0.0135	135
9	00010	0.0126	126

ISSN 2476-5880

10	10010	0	0
11	01010	0	0
12	11010	0	0
13	00110	0.0335	335
14	10110	0.0357	357
15	01110	0	0
16	11110	0.0443	443
17	00001	0.0039	39
18	10001	0	0
19	01001	0.0075	75
20	11001	0.0005	5
21	00101	0	0
22	10101	0	0
23	01101	0	0
24	11101	0.0001	1
25	00011	0.0291	291
26	10011	0	0
27	01011	0.0323	323
28	11011	0	0
29	00111	0.0257	257
30	10111	0	0
31	01111	0.0343	343
32	11111	0.1078	1078

The fact that flat profiles have the highest probability and the most frequency may be interpreted as evidence for unidimensionality of reading comprehension skill because based on the results of this study most of the test-takers belong to one of these two profiles and they have not/have mastered the reading comprehension as a whole, unit skill. In other words, a test-taker who has mastered the reading comprehension skill, actually has mastered all of the pre-assumed subskills which form a unit skill; and in fact the subskills cannot be separated from each other.

4. Discussion

As an important step of cognitive diagnostic approach, we specified the attributes required for successful performance on the reading comprehension items of University EntranceExamination. We had a list of reading comprehension attributes which was the result of review of literature and content analysis of the test items, conducted by a team of experts. The attributes were assigned to the items of the test at the Q-matrix specification stage. By specifying the Q-matrix, we actually determined the relationship among the test items and the required attributes which are involved in successful performance on the test. In the current study, five attributes were defined as the required subskills which the candidates needed to master for correct responding to the reading comprehension items of University Entrance Examination. These attributes were: *making inferences, extracting explicit information, identifying word meaning from context, identifying pronominal references*, and *evaluating response options*. Identifying these subskills of reading comprehension draws a framework to focus on.Konkoor candidates who try to improve their level of reading comprehension ability as a key to achieving better results on the test can benefit from our findings. When the candidates know about the subskills of reading comprehension which they need for answering the Konkoor reading comprehension items, they have higher chances of success on the test. The instructors who try to make their students prepared for the Konkoor, also can benefit from the results of this study since they know which attributes should be mastered by the candidates, therefore the teachers invest more time on these subskills.

The results of present study also indicated that, the two flat skill mastery profiles, “non-master of all skills” and “master of all skills”, had the most frequency. This finding is in line with other CDM studies (e.g., Lee &Sawaki, 2009b; Li, 2011; Ravand et al., 2013; Ravand, 2015). Lee and Sawaki (2009b), consider high frequency of flat skill mastery profiles as evidence for unidimensionality of the measure used, where a master of one skill tends to be a master of another skill, or vice versa.

The main aim of the current study was to investigate the reading comprehension ability of Iranian high-school graduate students’ population, a representative sample of which takes part in the Konkoor every year. The results of the study showed that, more than half of the candidates (57%) have not mastered any of the required attributes. We can say that our students are not well-equipped with the essential subskills which make them prepared for answering the reading questions of a high-stakes English test. Poor performance of the candidates on the reading test and considerable number of candidates who have not mastered any of the reading comprehension subskills is an alarming feedback which should be taken into account by the Iranian educational system. The teachers, curriculum developers, syllabus designers and the learners need to take remedial actions to solve the problem.

The teachers should invest more time and energy on the components and subskills of reading comprehension to improve learners' awareness of the existence of these subskills. The teachersalso need to use exercises which helplearners develop their knowledge of reading comprehension subskills. The experts in the educational system, who are in charge of designing instructional materials for English courses, also should pave the way for the teachers to concentrate more on the reading comprehension subskills by providing them with appropriate exercises in the school books.

References:

Baghaei, P., &Ravand, H. (2015).A cognitive processing model of reading comprehension in English as a foreign language using the linear logistic test model.*Learning and Individual Differences, 43,* 100-105.

Buck, G., Tatsuoka, K., &Kostin, I. (1997). The subskills of reading: Rule-space analysis of a multiple choice test of second language reading comprehension. *Language Learning, 47*(3), 423–466.

Clark, A. K. (2013). *Validation of a cognitive diagnostic model across multiple forms of a reading comprehension assessment* (Doctoral dissertation).Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/15071/Clark_ku_0099D_13062_DATA_1.pdf?sequence=1.

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

Hemmati, S. J., &Baghaei, P. (2016).Investigating the construct validity of the general English section of the Iranian University Entrance Examination. Paper presented at the first national conference on English, literature, and translation studies, Quchan, Iran.

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, Urbana-Champaign, IL.

Jang, E. E. (2008). A framework for cognitive diagnostic assessment.*Towards adaptive CALL: Natural language processing for diagnostic language assessment*, 117-131.

Kim, A. Y. A. (2014). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*.

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*, 172–189.

Leighton, J. P., &Gierl, M. J. (2007).*Cognitive diagnostic assessment for education: Theory and applications.* New York: 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.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL https://www.R-project.org/.

Ravand, H. (2015). Application of a cognitive diagnostic model to a high-stakes reading comprehension test.*Journal of Psychoeducational Assessment*, 1– 18.doi: 10.1177/0734282915623053

Ravand, H., &Robitzsch, A. (2015).Cognitive diagnostic modeling using R. *Practical Assessment Research and Evaluation*, 20, 1-12.

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

Robitzsch, A., Kiefer, T., George, A. C. &Uenlue, A. (2015). CDM: Cognitive Diagnosis Modeling. R package version 4.5-0.http://CRAN.R-project.org/package=CDM

Sheehan, K. M. (1997). A tree-based approach to proficiency scaling and diagnostic assessment.*Journal of Educational Measurement, 34*(4), 333-352.Svetina et al., 2011.

VanderVeen, A., Huff, K., Gierl, M., McNamara, D. S., Louwerse, M., &Graesser, A. (2007). Developing and validating instructionally relevant reading competency profiles measured by the critical reading section of the SAT Reasoning Test™. In D. S. McNamara (Ed.), *Reading Comprehension Strategies,* (pp.137-171). Mahwah, NJ: Erlbaum.

Zheng, Y., & De Jong, J. H. A. L. (2011).*Establishing Construct and Concurrent Validity of Pearson Test of English*. London, UK: Academic Pearson.