A cognitive analysis of an English reading test through the G-DINA model

Haizhen Zhao¹, Weize Wang² and Yuyuan Huang³

¹School of Foreign Studies, Guilin University of Electronic Technology, Guilin, China
²School of Economic and Management, Guangxi Normal University, Guilin, China
³School of Foreign Studies, Guilin University of Electronic Technology, Guilin, China

Corresponding author and e-mail: Haizhen Zhao, 7629332@qq.com

Abstract. The development of language testing calls for more in-depth diagnostic information of language skills to provide targeted feedback for test developers and test takers as well. Cognitive diagnostic assessment is capable of evaluating language skills (cognitive attributes) involved in a language test and assessing the cognitive features of test takers. Therefore, the introduction of cognitive diagnostic assessment to language testing is of great significance. The study is aimed to make a cognitive analysis of an English reading test for first-year college students through a cognitive diagnostic model named G-DINA. Through the feedback information of cognitive analysis, the study is expected to provide scientific and targeted feedback for students and teachers, helping to improve language testing, teaching and learning.

1. Introduction

At present, the construction and evaluation of language testing at home and abroad are mostly based on classical test theory or item response theory. These assessments usually provides a general evaluation of test scores or language ability but failed to produce diagnostic information which contributes to identifying the cognitive structure and level of subjects. The emergence of a new testing theory, namely cognitive diagnostic assessment, however makes up for the shortcomings of traditional measurement methods and conforms to the need for more precise assessment information. A big advantage of cognitive diagnostic assessment is that it is capable of providing more in-depth and detailed diagnostic information. Particularly, it can provide implicit and potential analysis of the cognitive structure of a subject based on his performance in a test. In this way we can know the strengths and weaknesses of the cognitive development of the subject in the measured fields after analyzing his cognitive structure. Furthermore we can also interpret which aspect of a subject’s cognitive development is better and which aspect needs to be improved [1]. Therefore, in recent years, cognitive diagnosis assessment has been widely used in the field of language measurement, which is of great significance to the theoretical research of language testing, the development of language tests, language teaching, language learning and so on [2].

Therefore the study adopted the cognitive diagnostic model named G-DINA to make a cognitive analysis of an English reading test for first-year college students. And the study tries to collect and interpret the test data from two perspectives. One is to get a detailed feedback information about students’ cognitive performance in the test. The other is to evaluate the quality of the test through data analysis.
2. Cognitive diagnostic model

The realization of cognitive diagnostic assessment is inseparable from specific cognitive diagnostic models (CDM). A cognitive diagnostic model is intended to evaluate the cognitive or skill mastery of subjects from the perspective of cognitive psychology and psychometric model [3]. So far, there are more than 120 cognitive diagnostic models, such as fusion model, attribute hierarchy model (AHM), G-DINA model, etc. These models play an important role in the field of education and psychometric theory [4]. Among the above, Fusion model is commonly used in foreign language listening and reading tests. Li et al compared two non-compensatory models (DINA, RRUM) and two compensatory models (DINO, ACDM) with the saturated model (G-DINA), and the result shows that G-DINA and ACDM are superior to other models [5].

The study adopted G-DINA in view of its advantages. As a general model of DINA, G-DINA was proposed by de la Torre [6]. It is a compensatory model, in which each cognitive attribute of test questions contributes to the answering of test questions. In addition, G-DINA is a saturated model, taking into account the interaction between multiple cognitive attributes involved in a test [7]. In other words, the probability of a right response is not only the sum of probability of mastering the attributes involved. It also considers the positive or negative effect that the interaction of attributes will have on the possible right response to test questions. Therefore, it can be concluded that G-DINA is more consistent with the characteristics of language testing since micro language skills (attributes) are closely related to each other.

3. Methodology

3.1. Subjects and test materials

The subjects of this study are 304 freshmen in the Guilin University of Electronic technology. They attended the final English examination in the first semester and their test scores of reading comprehension are collected. The test material are two reading passages. Each passage consists of 5 multiple choice questions.

3.2. The identification of cognitive attributes and establishment of Q-matrix in G-DINA model

Cognitive attributes refer to a series of skills, knowledge points, thinking process and cognitive strategies that are needed to complete a test task correctly. The determination of cognitive attributes of test materials and the establishment of cognitive Q-matrix are the key steps of cognitive diagnosis. Q-matrix is a two-dimensional matrix table, reflecting the logical relationship between attributes and test items [8]. In other words, the Q-matrix shows the corresponding response of each item to the cognitive attributes involved. If a question item tests an attribute, it is marked as 1; if not, it is marked as 0. Since Q-matrix will directly affect the analysis of examinee's cognitive ability, it is the first step for a test developer to determine the cognitive attributes of each question item in a test.

| Item | Attribute | A1 | A2 | A3 | A4 | A5 |
|------|-----------|----|----|----|----|----|
| R01  |           | 0  | 0  | 0  | 1  | 1  |
| R02  |           | 0  | 0  | 1  | 1  | 0  |
| R03  |           | 0  | 0  | 1  | 1  | 0  |
| R04  |           | 0  | 0  | 1  | 1  | 0  |
| R05  |           | 0  | 0  | 1  | 0  | 1  |
| R06  |           | 1  | 0  | 0  | 1  | 0  |
| R07  |           | 0  | 0  | 1  | 1  | 0  |
| R08  |           | 1  | 0  | 0  | 0  | 0  |
| R09  |           | 0  | 1  | 0  | 0  | 0  |
| R10  |           | 0  | 0  | 1  | 0  | 1  |
| Frequency |      | 2  | 1  | 6  | 6  | 3  |
Under the proposition framework of the test, the study defined five cognitive attributes in the test by combining relevant research and the teamwork of project members [9]. They are extracting direct information (A1), predicting (A2), inferring information (A3), integrating multiple information (A4) and overall understanding (A5). The model went through modification based on preliminary operation results. Finally, the Q-matrix was established as shown in Table 1 with five cognitive attributes corresponding to ten question items (from R01 to R10).

After determining the Q-matrix, the CDM package in the statistical tool R is used to analyze students’ answering data with the G-DINA model. The results are shown below.

3.3. Analysis of students' performance

First of all, statistics show that the overall RMSEA index of the model is 0.001, and the index of each item is less than 0.05, indicating that the fitting degree is good [10]. Next, we will show the response data from different aspects.

3.3.1. Overall mastery probability of each cognitive attribute. After the diagnosis analysis in G-DINA model, the subjects’ overall mastery probability of five cognitive attributes is shown in table 2.

### Table 2. Overall mastery probability of reading cognitive attributes.

| Attribute number | A1      | A2      | A3      | A4      | A5      |
|------------------|---------|---------|---------|---------|---------|
| Mastery probability | 0.7268  | 0.4397  | 0.6232  | 0.5148  | 0.4749  |

It can be seen from table 2 that students participating in this English test show relatively outstanding performance in extracting direct information (A1) and inferring information (A3), followed by integrating multiple information (A4). However, they show weaker performance in predicting (A2) and overall understanding (A5). The data shows that students' ability to integrate multiple information is low, even lower than the ability of overall understanding of a reading text.

3.3.2. Distribution of potential types. In addition, cognitive diagnosis can also calculate distribution probability of different potential types. As shown in table 3, the five numbers in each potential type represent the cognitive attributes ranked from A1 to A5. And mastering is labeled as “1” while not mastering is labeled as “0” respectively. Thus altogether 32 potential types are formed. Due to individual differences, students' mastery of five cognitive attributes is not the same. Some students only master one of the five attributes like 10000, and some master several attributes at the same time such as 11110, 10011, etc.

### Table 3. Distribution of potential types.

| Potential type | Distribution probability | Potential type | Distribution probability | Potential type | Distribution probability | Potential type | Distribution probability |
|----------------|-------------------------|----------------|-------------------------|----------------|-------------------------|----------------|-------------------------|
| "00000"        | 0.120838                | "10010"        | 0.028691                | "11100"        | 0.195825                | "01011"        | 0.014837                |
| "10000"        | 0.004759                | "10001"        | 0.778126                | "11010"        | 0.000284                | "00111"        | 0.000231                |
| "01000"        | 0.000299                | "01100"        | 0.000494                | "11001"        | 0.3002                  | "11110"        | 0.000816                |
| "00100"        | 0.000203                | "01010"        | 0.015193                | "10110"        | 0.110268                | "11101"        | 0.0423391               |
| "00010"        | 0.000332                | "01001"        | 0.000342                | "10101"        | 0.080887                | "11111"        | 0.018984                |
| "00001"        | 0.049717                | "00110"        | 0.108416                | "10011"        | 0.050896                | "10111"        | 0.005847                |
| "11000"        | 0.016772                | "00101"        | 0.0003348               | "01110"        | 0.026874                | "01111"        | 0.0027282               |
| "00000"        | 0.028441                | "00011"        | 0.0003339               | "01101"        | 0.000318                | "11111"        | 0.0031623               |

It can be seen from table 3 that 10001 is the type with the largest distribution probability, indicating the two reading skills involved in this test are significantly dependent on each other. Next are 11001 and 10110, reflecting that students have a good command of lower level attributes, which conforms to
the general rule of cognitive process. In addition, the distribution probability of several potential types (01000, 00100, 00010, 01100, 01001, 00101, 00011, 00111) is close to zero. These types contain higher level cognitive attributes and the first or second attribute is basically zero, which indicates the lower level cognitive attributes play a fundamental role in mastering higher level cognitive attributes.

3.3.3. The probability distribution of the correct answer to each test item and mastery of corresponding cognitive attributes. Through the diagnostic analysis of G-DINA model, we can analyze the probabilistic relationships between the correct answer to each test item and the mastery of cognitive attributes involved. In other words, the function and contribution of certain cognitive skills can be figured out through data analysis. The results are shown in Table 4. Attribute 1 attribute 2 in the table refers to the specific and corresponding reading cognitive skills required for each test item as identified in table 1. Attribute 1 × 2 represents the probability of correct answer to each item under the interaction of two attributes involved in a certain test item. Guessing refers to the probability of correct answer to each item when not a single cognitive skill is mastered.

Table 4. Probability of accuracy to each test item.

| test item | guessing | attribute 1 | attribute 2 | attribute 1 × 2 |
|-----------|----------|-------------|-------------|-----------------|
| Item1     | 0.0001   | 0.0001      | 0.1190      | 0.1310          |
| Item2     | 0.4550   | 0.0001      | 0.0890      | 0.0617          |
| Item3     | 0.4279   | 0.5411      | 0.0791      | 0.0590          |
| Item4     | 0.0783   | 0.5518      | 0.0473      | 0.0581          |
| Item5     | 0.3901   | 0.0011      | 0.1173      | 0.1111          |
| Item6     | 0.6502   | 0.0345      | 0.1301      | 0.0269          |
| Item7     | 0.0001   | 0.8635      | 0.0806      | 0.0393          |
| Item8     | 0.0666   | 0.4165      | —           | —               |
| Item9     | 0.0001   | 0.0018      | 0.1003      | 0.1290          |
| Item10    | 0.0001   | 0.0470      | 0.2103      | 0.0455          |

From the guessing probability in table 4, it is easy to find out any possible problems with each test item and sort out the reason behind it. As shown in the table, the guessing probability of item 2, item 3 and item 6 is larger than 0.4000. That means students can easily choose the right answers without mastering any cognitive skills required for those test items. The question for item 3 is that “According to the passage, which of the following may NOT be an occasion for public communication?” The four choices for students to choose are as follows. A). Expressing opinion on a topic in class. B). Trying to sell a new product to customers. C). Talk about the holiday plan with friends. D). Discussing a project with a group of colleagues. Students may easily choose the right answer (C) based on their common sense about public communication even if they didn’t understand the whole reading text. Because most college students know for sure that talking about the holiday with friends is a private issue instead of a public communication. Therefore, the problem for this test item is identified and figured out. In the same way, the author can also sort out those well-designed items base on the probability in table 4. For example, for items 1, 8 and 9 students have a larger probability of choosing the right answers if they mastered one or both cognitive attributes involved. That means that the questions for these 3 items are well designed and can effectively test student’s reading cognitive ability. Next item 10 is taken as an example to interpret the data in a detailed way.
As shown in figure 1, there are two cognitive attributes (A3 and A5) involved in item 10. If not a single attribute is mastered by students (A00), the probability of correct answer is close to zero. While the probability of success reached 0.99 if attribute A5 (A01) is mastered. The probability of success reached 0.95 if both attributes (A11) are mastered by students. The data indicates that attribute A5 plays a bigger role in helping students choose the right answer to test item 10.

(4) Mastery probability of individual test takers

Cognitive diagnostic assessment can also interpret the cognitive features of individual test takers. Next 3 test takers numbered 10, 20, and 50 are taken as an example.
improve the cognitive attribute A3. The student numbered 20 performed well in almost all the cognitive contributes except for A4.

4. Findings and implications
As seen from the above analysis, the use of cognitive diagnostic model is based on the identification of certain cognitive skills (reading skills in this test), which is of great importance for providing diagnostic information. At the same time, the information of students’ mastery of certain cognitive skills will make teachers and students focus more on the acquisition of cognitive abilities rather than simply test scores. This is exactly what an indirect and vertical language test lacks [11]. The following are the research findings and implications based on cognitive diagnostic theory and test statistics above. The findings will be presented from two perspectives. First, the study will summarize students’ reading cognitive ability according to their test performance. Secondly, the study will examine the quality of the test and evaluate whether the test effectively examined each cognitive ability concerned in the test.

4.1. Students’ test performance
It can be seen from the above statistics that students perform much better in distracting direct information and inferring information than the other 3 reading skills. This proved that for students lower level reading skills (extracting direct information and inferring information) are easy to master compared with high level skills, which need more global understanding of reading materials [12]. According to the College English Teaching Guide (2017), the higher goal for reading comprehension ability requires that students are expected to understand implied meaning and the logical structure of an article, understand different reading materials in a comprehensive way. From the analysis of the test, students perform better in understanding implied meaning and relating multiple information. However there is still room for improvement in reading cognitive skills like global understanding of a reading text.

4.2. Test assessment and implications
The test will be assessed from the following two aspects. Firstly, overall quality of the test is good. The test is aimed to examine students' comprehensive reading ability. The test results show that the skills involved in this reading test are indeed hard to be separated and interact with each other, which also reflects the compensatory and saturation characteristics of G-DINA model, and the appropriateness of using this model in the reading test. Secondly, the diagnostic data in this study verified the quality of the test. The overall mastery probability of reading cognitive attributes in table 2 indicates that students have higher ability of sentence processing like extracting direct information and understanding implied information while have lower ability in integrating multiple information and global understanding of a reading text. This result proves that the five cognitive attributes involved are well identified and the Q-matrix of the test is reasonably established. However, the guessing probability of certain test items is too large, indicating that the setting of multiple question is not reasonable and the reading skills required for these items failed to be tested as a result. The mastery probability of some items is small when both attributes involved are mastered by students while the mastery probability of the same items is rather larger when only one of the needed attributes are mastered. This is possibly because that other reading skills are needed in answering these items or because the given choices for these items are not well designed.

In summary, the overall quality of the test is good. It effectively checked the reading skills intended for the test. However, there is still room for improvement such as the design of given choices and the construction of cognitive attributes.

5. Conclusions
In this study, a cognitive diagnostic model is used to analyze the performance of first-year college students in a large-scale English reading test, aiming to find more detailed feedback information. From
a macro perspective, the study presents students' performance in various reading skills as well as the distribution of students' cognitive potential types, which provides a reference for improving the quality of English reading tests and students' reading skills. From a micro perspective, analyzing the strengths and weaknesses of individual students is helpful for English teachers to recognize the cognitive features of their students and then provide targeted guidance. In addition, the study also provides reference data for construct identification and test development. In a word, the development of English testing cannot be achieved without the proper use of test analysis and its feedback information.

Acknowledgement

This work is supported by the Teaching Reform Project of Guangxi Higher Education (Grant No. 2020JGZ117), the Teaching Reform Project of Guangxi Higher Education (Grant No.2015JGA222), the Teaching Reform Project of Guangxi Higher Education (Grant No. 2019JGA137), the Teaching Reform Project of Guilin University of Electronic Technology (Grant No. JGB201816) and the Basic Scientific Research Ability Improvement Project for Young and Middle-aged Teachers of Guangxi’s University (Grant No. 2019KY0243).

References

[1] Luo Zhaosheng 2019 Fundamentals of Cognitive Diagnostic Assessment (Beijing: Beijing Normal University Publishing Group).
[2] Chen Huilin and Chen Jinsong 2013 A Study of the Construct Validity of TEM-4 Cloze Test: A Process Approach. J. Modern Foreign Languages vol 29 pp 71-77.
[3] Lee and Sawaki Y 2009 Application of Three Cognitive Diagnosis Models to ESL Reading and Listening assessments J. Language Assessment Quarterly vol 6 pp 239-263.
[4] Chen Huilin and Zhao Guanfang 2013 Application of Diagnostic Approach: A new stage of Language Testing Research J Foreign Language Testing and Teaching pp 1-10.
[5] Cai Yan, Ding Shuliang and Xu Dongbo 2011 A Cognitive analysis of English Reading Comprehension J Journal of Psychological Science.
[6] De la Torre J 2011 The Generalized DINA Model Framework vol 76 pp 179-199.
[7] Chen Huiyan, Chen Jinsong 2016 Exploring Reading Comprehension Skill Relationships through the G-DINA model. J. Educational Psychology vol 36 pp 1049-1064.
[8] Buck G, Tatsuoka K and Kostin I 1997 The Subskills of Reading: Rule-space Analysis of a Multiple-choice Test of Second Language Reading Comprehension J language Learning. vol 47 pp 423-466.
[9] Chen Xiaokou and Li Shaoshan 2006 On Establishing Construct Validity of Language Testing J Journal of PLA university of Foreign Languages. pp 45-49.
[10] Yi Y S 2017 Probing the Relative Importance of Different Attributes in L2 Reading and Listening Comprehension Items: An Application of Cognitive Diagnostic Models.
[11] Shohamy E 1982 Predicting Speaking Proficiency from Cloze Tests: Theoretical and Practical Considerations for Tests Substitution vol 3 pp 161-171.
[12] Alderson J C and Lukmani 1989 Cognition and Reading: Cognitive Levels as Embodied in Test Questions. J. Reading in a Foreign Language. vol 5 pp 253-270.