Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)
2022, Volume 7, Issue 11, e001877
DOI: https://doi.org/10.47405/mjssh.v7i11.1877

Validating Reading Comprehension Assessment Under the GDINA Model

Renukha Nallasamy[1*], Ahmad Zamri Bin Khairani[2]

[1]Pusat Pengajian Ilmu Pendidikan, Universiti Sains Malaysia (USM), Pulau Pinang, Malaysia.
Email: renukha18@student.usm.my
[2]Pusat Pengajian Ilmu Pendidikan, Universiti Sains Malaysia (USM), Pulau Pinang, Malaysia.
Email: ahmadzamri@usm.my

CORRESPONDING AUTHOR (*):
Author Name
(renukha18@student.usm.my)

KEYWORDS:
Reading comprehension assessment
Cognitive diagnostic model
GDINA model
Q-matrix

CITATION:
Renukha Nallasamy & Ahmad Zamri Khairani. (2022). Validating Reading Comprehension Assessment Under the GDINA Model. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 7*(11), e001877. https://doi.org/10.47405/mjssh.v7i11.1877

ABSTRACT

Cognitive diagnosis models (CDMs) are latent variable models mainly developed to assess students' specific strengths and weaknesses in a set of skills or attributes within a particular domain. In this study, the reading comprehension assessment is diagnostically designed, constructed, and developed from the very first step. The predetermined attributes or sub-skills are explicitly defined in the construction phase as they should align with the instructional goals. Using R package CDM, the Generalized-DINA model (GDINA) was applied to the reading comprehension assessment. A total of 900 Year 4 primary students from the Eastern District of Pulau Pinang national and vernacular schools sat for this assessment. Through the cognitive analysis, the study is expected to provide detailed diagnostic feedback on students' strengths and weaknesses in the underlying skills identified in the reading comprehension assessment. Such detailed information can help teachers in classroom teaching, designing remedial courses, and developing material according to the student's needs.

Contribution/Originality: This study investigated the usage of the Cognitive Diagnostic Models (CDMs) in reading comprehension assessments. CDMs is used to provide fine-grained diagnostic feedback on students' subskills and to provide insights on remedial instructions in specific domains.

1. Introduction

Traditional assessments such as item response theory (IRT) and classical test theory (CTT) typically provide unidimensional ability values, which allow for a ranking of students according to their abilities. These assessments commonly offer information such as single overall scores that indicate students' relative positions on an ability scale. However, these assessments fail to provide diagnostic information on the mastery or non-mastery of skills. Moreover, identifying if a student has mastered specific skills required to solve a question item varies from the objective of traditional measurement models, which posits a student on a continuous scale to gauge relative achievement. Hence,

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

language testing experts have persuasively called for more comprehensive score reporting to improve instruction and learning (Alderson, 2005). Cognitive diagnostic models (CDMs) can compensate for this limitation by delivering fine-grained formative feedback on students' strengths and weaknesses. Furthermore, CDMs promise to provide rich diagnostic information to aid learning and instruction (Rupp & Templin, 2008).

Cognitive diagnosis models (CDMs) are latent variable models developed primarily for cognitive diagnostic assessments to assess student mastery and non-mastery of finer-grained skills. CDMs can provide more targeted information in the form of score profiles that can effectively measure student learning and progress, design better instruction, and possibly intervention to address individual and group needs (de la Torre, 2009; 2011). Hence, in recent years, cognitive diagnostic models have been widely applied in the field of language measurement, which is of great significance to the theoretical research in language testing. Therefore, this study adopted the cognitive diagnostic model, GDINA, to make a cognitive analysis of an English reading comprehension assessment among year 4 primary students. The assessment data will be used to obtain students' cognitive performance and evaluate the assessment's quality.

2. Literature Review

Diagnostic Cognitive Models are latent class models (Wang, Shu, Shang, & Xu, 2015) that offer a new approach to analyzing test scores. These models can categorise students as masters or non-masters on a set of sub-skills/ attributes and present more fine-grained diagnostic information about the quality of the question items and attributes measured by the question items (DiBello et al., 2007; Lee & Sawaki, 2009a; Rupp et al., 2010). Basically, CDM promotes assessment for learning and the learning process as opposed to assessment of learning outcomes (Jang, 2008), and it is an interdisciplinary diagnostic assessment method. It lies at the interface between cognitive psychology and statistical analysis. The advanced statistical analysis examines the relationship between performance on specific test items and the psychological processes and strategies that underlie responses to those test items.

CDMs are categorized into three categories that are compensatory(disjunctive), non-compensatory (conjunctive), and general models. In a compensatory model such as the Addictive Cognitive Diagnostic Model (ACDM) (de la Torre, 2011) and the Deterministic Inputs, Noisy "Or" Gate (DINO) (Templin & Henson, 2006), mastery of one or some of the attributes for a test item to be correct can compensate for the non-mastery of the other attributes. Nevertheless, these models do not increase the probability of providing the test item with the correct answer because mastery of any subset of attributes is the same as mastering the requisite attributes required for that item. In contrast, non-compensatory/conjunctive models, the Rule Space Model (Tatsuoka, 1983), Attribute Hierarchy model (AHM) (Leighton, Gierl, & Hunka, 2004) and Reparameterized Unified Model (RUM) or Fushion Model (Hartz,2002), lack of mastery of one attribute and cannot be compensated by other attributes (Lee & Sawaki, 2009a; Ravand, 2016; Li & Hunter, 2016). Therefore, to get a question item correctly, a student must successfully have all the required attributes for the specific question item (Lee & Sawaki, 2009a). As for the last category, the general CDMs allow compensatory and non-compensatory relationships within the same test (Ravand & Robitzsch, 2015) as it allows each test item to pick the model that best fits it instead of assigning a single model to all the items. Some of the general CDM models are Generalized Deterministic, Input, Noisy, and Gate (GDINA) (de la

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Torre, 2011), Log-linear CDM (LCDM) (Henson, Templin & Willse, 2009), and General Diagnostic Model (GDM) (Von davier, 2005).

CDMs can apply in two ways : (a) non-diagnostically constructed and (b) diagnostically-constructed. Presently, the application of CDMs studies is more on math and language skills compared to other subjects. The non-diagnostically constructed, commonly referred to as retrofitting of existing non-diagnostic tests, is reverse engineering. The method is frequently used in high-stakes examinations such as MELAB, TOEFL, IELTS, and TMISS. This approach mainly retrofits a set of existing achievement/ proficiency tests to extract and report cognitive sub-skills that are anticipated to be measured through the test. The sub-skills, generally known as attributes, are extracted post hoc through students' think-aloud protocols and expert judgments. In the next step, through the selected CDM, fine-grained information on the student's mastery classification is gathered. In diagnostically-constructed designs, attributes are explicitly defined in the test construction phase. The predetermined attributes must be in line with the instructional goals. Therefore data are analysed with an appropriate CDM, and the scores are reported in a fine-grained diagnostic system.

Four major components of cognitive diagnosis approaches are used to design a new set of test items or to retrofit existing non-diagnostic tests (DiBello et al., 2007). George et al. (2016) simplified the four steps involved in CDA analysis into two stages; qualitative and quantitative. In the qualitative phase, targeted attributes are defined, and Q-matrix is constructed and validated. The quantitative phase, the data is analysed with a selected CDM and provides diagnostic reporting.

In this study, generalized deterministic inputs “and” gate, or GDINA, is adopted to conduct cognitive diagnosis on the diagnostically-constructed reading comprehension assessment. GDINA is a compensatory model, and it is under the general model category, in which each attribute of the question items contributes to the answering of the question items. In other words, it allows the interaction between multiple attributes involved in a test or assessment but still considers the positive or negative effect that the interaction of attributes will have on the possible correct answer to the question item. Therefore by adopting this model, the relationship among reading comprehension attributes will be explored more satisfactorily so that detailed feedback about the student’s cognitive performance in the assessment can be obtained and the quality of the assessment can be evaluated.

2.1. GDINA Model

The G-DINA is a generalization of the DINA model proposed by de la Torre (2011). The DINA is a non-compensatory model that classifies students into two categories for each question item. Those who have mastered the subskills required by the question item and those who have not mastered at least one of the needed subskills. In the concept of the DINA model, “lacking one required attribute (subskill) for an item is considered as lacking all the required attributes for the item” (de la Torre, 2011). Therefore, unlike the DINA model, by ignoring the assumption, G-DINA does not assume an equal probability of success for those who have not mastered any, some, or all the required attributes for a question item. The item response function of the G-DINA model is based on P (Ij*) and is as follows:

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

$$P(\alpha_{Ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\ \alpha_{Ik} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\ \alpha_{Ik}\alpha_{Ik'} \ldots + \delta_{j12..K_j^*} \prod_{k=1}^{K_j^*} \alpha_{Ik}$$

The function above can be decomposed into the sum of the effects due to the presence of specific attributes and all their possible interactions. δ_0 represents the baseline probability (i.e., probability of a correct response when none of the requisite attributes is present), which can be regarded as the guessing parameter; δ_k is the main effect of mastering a single-skill attribute (i.e. α_k); $\delta_{kk'}$ is the first-order interaction effect due to the mastery of both α_k and $\alpha_{k'}$; and $\delta_{12\ldots K_j^*}$ represents the highest-order interaction effect due to the mastery of all the required attributes (de la Torre, 2011). Hence, the characteristics of the G-DINA model can cater to the characteristics of reading comprehension skills and detect the interactive relationship among them.

3. Methodology

3.1. Participants and test material

The participants for this research are 900 Year 4 ESL students from the Eastern District schools of Penang state. The schools are chosen using stratified sampling as there are national schools known as *Sekolah Kebangsaan* and vernacular schools comprising Mandarin-medium and Tamil-medium national-type schools. The students are selected using a simple random sampling method from ten national and six vernacular schools. As for the assessment, the material contains four passages comprising linear and non-linear texts, each with five multiple-choice items. The students are required to complete this assessment in an hour.

3.2. Procedure

3.2.1. Identification of Attributes

A panel of four teachers from three different primary schools was assigned to extract five attributes to be tested in this assessment. These teachers hold a Bachelor of Education majoring in English with 15 to 25 years of teaching from Year 1 to Year 6 with extensive classroom-based test development experience. The attributes were extracted based on Year 4 curriculum specification. Below are the five attributes that were tested in the assessment.

i. Vocabulary
ii. Syntactic Knowledge
iii. Understand Explicit Information
iv. Connecting and Synthesizing
v. Making Inferences

These attributes are measurable skills specified at a fine-grained level to elicit the students' cognitive processes on the item questions to identify their weaknesses and strengths. All the attributes are arranged in sequence of easy to complex using Bloom's Taxonomy and Barrett's Taxonomy to reflect the learning progression of the knowledge and skills in mastering reading comprehension.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

3.2.2. Student's Think-aloud verbal protocols

A pilot test is carried out with 12 volunteer students that are not involved in this research and have no contact with the 900 students, as every piece of information regarding the students and schools is confidential. Think aloud session was carried out on these students. The verbal report provides insight into a student's cognitive processes that are not usually apparent from the written response. This selected group of students is a mix of high ability and moderate performance based on the background information from their subject teachers. The objective is to assess how well the cognitive processes stated by the students match the attributes and hierarchies stated by the panel of teachers.

3.2.3. Q-matrix and item assessment development

The primary and crucial part of cognitive diagnosis is determining the attributes to be tested and establishing the Q-matrix. Q-matrix is a two-dimensional matrix table that reflects the logical relationship between test items and attributes. Basically, Q-matrix shows the corresponding response of the test item to the attributes that are engaged. If a question item tests an attribute, it is marked as '1' if not as '0'. After establishing the Q-matrix, the CDM package in the statistical tool R is used to analyse students' answering data with the G-DINA model.

4. Data Analysis

4.1. Model Fit

Data are analysed using the R-package "GDINA," version 4.0.3 (Ma & de la Torre, 2020). The R package employs marginal maximum likelihood estimation using the EM algorithm for fitting the models (George et al., 2016). Therefore, the fit of a model can be ascertained by examining absolute fit and relative fit statistics. Absolute fit checks whether the model fits the data, and relative fit compares the model with other rivalry models.

Table 1 illustrates the relative fit indices for GDINA and three rival models based on the Log-likelihood ratio test, the conventional Akaike Information Criterion (Akaike, 1974), and the Bayesian Information Criterion (BIC) (Schwarz, 1978). The model that reports the minor information criteria is considered the best model to fit the data and the preferable one. A thorough examination of fit indices shows that the values were higher for the ACDM, RRUM, and DINA models than for the GDINA model, which reveals that G-DINA had a better model fit. However, the GDINA model did not show a low value for BIC, which can be explained by the sensitivity of BIC to highly parameterized models (Li et al., 2015).

Table 1: GDINA relative fit indices

Model	Loglikehood	Deviance	AIC	BIC
GDINA	-8971.25	17942.49	18189.49	18480.59
DINA	-9257.67	17963.54	18657.34	18398.34
ACDM	-8982.77	17725.95	18295.54	18609.64
RRUM	-9233.60	18609.21	18609.21	18509.24

The significance level of the Z-score can be adjusted by means of the Bonferroni correction. For α=0.01, 0.05, and 0.1, critical Z-scores are 4.17, 3.78, and 3.61 respectively. Table 2 indicates that the G-DINA model fits the data well, as the maximum Z-score values of the three statistics are below the critical values, and the adjusted *p*-value > 0.05. Based

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

on the analysis, the GDINA model provides a good fit to data based on proportion correct values Max Z = 0.7315 < 4.17, and the adjusted *p*-value is more than 0.05. As for the transformed correlation and Log-odds ratio, the values of Max Z is below the critical values; Max Z = 2.63 < 3.78 and Max Z = 2.26 < 3.61 respectively for each statistics. This result reveals no Q-matrix misspecification, poorly written item questions, or too small a sample size in this research.

Table 2: GDINA absolute fit indices

	Mean [stats]	**Max [stats]**	**Max [z.stats]**	**p-value**	**adj.p-value**
Proportion correct	0.0012	0.0056	0.3431	0.7315	1
Transformed correlation	0.0197	0.0880	2.6342	0.0841	1
Log odds ratio	0.0959	0.4058	2.2582	0.0239	1

Note: p-value and adj.*p*-value are associated with max[z.stats]. adj. p-values are based on the Bonferroni method.

4.2. Attribute Prevalence

The overall students' diagnostic test performance is analysed to examine the population's probability of mastery for each attribute. Table 3 below shows the comprehensive students' probabilities of the five listed attributes obtained by analysing the students' response data using this R package. Level 1(**0**) indicates non-mastery of the attributes and Level 1(**1**) mastery of the attributes.

Table 3: Attribute Prevalence

Attribute	**Level 1(0)**	**Level 1(1)**
Vocabulary	0.229	0.771
Syntactic Knowledge	0.246	0.754
Understand Explicit Information	0.328	0.672
Connecting and Synthesizing	0.471	0.529
Making Inferences	0.531	0.469

As evident in Table 3, vocabulary and syntactic knowledge are the easiest attributes to be mastered, with 77% and 75% of the students having mastered these attributes, respectively. Making inferences, mastered by only 46% of the students, was the most challenging attribute to dominate. Understand explicit information and connecting and synthesizing come after syntactic knowledge, with 67% and 58% mastery levels.

4.3. Latent Classes Profiles and Posterior Probabilities

Furthermore, students can be categorised into latent classes, representing specific mastery/non-mastery profiles for the set of sub-skills specified in the Q-matrix (von Davier, 2005; Huebner, 2010). As previously described, the GDINA model defines 2^k possible latent classes for each cognitive domain. Since there are five attributes in the current research, so K= 5 defined attributes, 32 latent classes were expected. Students who mastered all the skills belong to the '11111' latent class, and non-masters of all skills are categorised in the '00000' latent class. For instance, if the latent class is 11000, it indicates the student has mastered the first two sub-skills stated in the final Q-matrix. Table 4 shows 32 possible latent class and their probabilities.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Table 4: Latent Classes and Posterior Probabilities

	Latent Class	Probability		Latent Class	Probability
1	00000	0.0004	17	11100	0.0910
2	10000	0.0699	18	11010	0.0310
3	01000	0.0005	19	11001	0.0100
4	00100	0.0029	20	10110	0.0082
5	00010	0.0010	21	10101	0.0039
6	00001	0.0004	22	10011	0.0018
7	11000	0.0631	23	01110	0.0033
8	10100	0.0022	24	01101	0.0016
9	10010	0.0014	25	01011	0.0003
10	10001	0.0007	26	00111	0.0009
11	01100	0.0080	27	11110	0.1435
12	01010	0.0072	28	11101	0.0390
13	01001	0.0068	29	11011	0.0151
14	00110	0.0044	30	10111	0.0021
15	00101	0.0033	31	01111	0.0002
16	**00011**	**0.0000**	32	**11111**	**0.4759**

Based on Table 4, latent class 11111 is dominant with the highest-class probability (0.4759), which means about 428 students were estimated to have mastered all the attributes. Another prevalent class latent is ' 11110', to which 14% of the students have mastered attribute one to attribute four (vocabulary, syntax, extracting explicit information, and connecting and synthesizing). The latent class '00011' shows the lowest posterior probability, 0.0000, which indicates the likelihood no students are categorised in this latent class. Students can't master the complex attributes (connecting and synthesizing, making inferences) without mastering the essential three primary attributes (vocabulary, syntactic knowledge, and understand explicit information). Based on the posterior probabilities of each latent class, as shown in Table 4, students tend to perform well on the diagnostic test as they master the main three attributes or the first two primary attributes; vocabulary and syntactic knowledge. This can be seen from the probabilities value of the latent classes 11100, 11010, 11001, 11110, and 11101 compared to other probabilities latent classes.

4.4. Individual Performance

An individual student mastery pattern can be obtained using this R package software. Table 5 demonstrates the attribute mastery pattern of two individual test-takers that score the same marks (i.e., 70). Both students have the same score but different attribute mastery patterns. The student with the Stud_id 12 has successfully mastered the three primary and fourth attributes; connecting and synthesizing. But a student with the Stud_id 112 scores the same mark as Stud_Id 12 has a different mastery profile. This student has mastered the first three attributes and the last attribute (making inferences) but fails to master the fourth attribute, as shown in Table 5. These results suggest that students with the same total score do not necessarily have the same cognitive skill profiles, mainly when a test measures various cognitive attributes.

Therefore, these diagnostic performance reports help teachers identify each student's strengths and weaknesses. Furthermore, each report's outcome will guide the teachers to improve their instructional design and guide their students' learning by improving the skills they are weak.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Table 5: Attribute mastery patterns for individual students.

Student Id	Exam Score(%)	Attribute Mastery Pattern
Stud_id 12	70%	11110
Stud_id 112	70%	11101

4.5. Items Parameters Estimates

The guessing and slipping parameters can be estimated with their standard error for 20 reading comprehension item questions in this diagnostic test.

Slipping parameter or incorrect answering an item question by a student who has mastered all the attributes occurs when one of the attributes is deleted from a Q-matrix incorrectly. In contrast, guessing parameter or answering an item question correctly by a student who has not mastered all its attributes would be seen if attributes are added to a Q-matrix incorrectly. Rupp and Templin (2008) pointed out that the lack of misspecification of attributes in Q-matrix can result in guessing and slipping parameters. For this reason, researchers and educators must carefully consider all the attributes during the Q-matrix construction phases, as adding or deleting attributes may result in the misclassification of the students.

There are two versions of determining the quality of these parameters. According to de la Torre (2009), high-quality items is defined as $0 < g+ s < 0.2$, medium quality $0.2 < g+ s < 0.4$, low quality $0.4 < g+ s < 0.6$ and problematic items $g+ s > 0.6$. However, Ravand, Barati, and Widhiarso (2013) suggested high-quality item is $g+ s < 0.6$, and low-quality item is $g+ s > 0.6$. For this researcher, the terms defined by de la Torre(2009) will be applied.

The average slipping parameter for this diagnostic test is 0.0973 (Table 6). In other words, there are chances that 9.7 % of students answer the item questions incorrectly, even though they have mastered all the required attributes. As for guessing parameters, the average is 0.2008; there is a chance that 20 % of students will guess the answers even though they did not master all the attributes needed for the item questions. Based on Table 6. five high-quality item questions are questions 5,10, 12, 17, and 18, question no. 15 falls under the category of low-quality item questions. Question no. 8 is considered a problematic item as the added value of guessing and slipping parameters is 0.9610, more than 0.6. The remaining thirteen item questions are under the category of medium-quality items.

Table 6: Guessing and slipping parameters of the reading comprehension assessment.

Item No.	Guessing (g)	Slipping (s)	$g+ s$	1- $g+ s$
Item 1	0.1889	0.0637	0.2526	0.7474
Item 2	0.2900	0.0115	0.3015	0.6985
Item 3	0.1263	0.1796	0.3059	0.6941
Item 4	0.0001	0.3256	0.3257	0.6743
Item 5	0.0001	0.0258	0.0259	0.9741
Item 6	0.2730	0.0001	0.2731	0.7269
Item 7	0.2307	0.0533	0.2840	0.7160
Item 8	0.5006	0.4604	0.9610	0.0390
Item 9	0.1753	0.2181	0.3934	0.6066
Item 10	0.0001	0.0949	0.0950	0.9050
Item 11	0.2345	0.0086	0.3381	0.6619

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Item 12	0.1552	0.0001	0.1553	0.8447
Item 13	0.3712	0.0218	0.3930	0.6070
Item 14	0.1226	0.1981	0.3207	0.6793
Item 15	0.5833	0.0001	0.5834	0.4166
Item 16	0.0928	0.1431	0.2359	0.7641
Item 17	0.0001	0.0467	0.0468	0.9532
Item 18	0.1508	0.0001	0.1509	0.8491
Item 19	0.2135	0.0652	0.2787	0.7213
Item 20	0.3066	0.0372	0.3438	0.6562
Mean	**0.2008**	**0.0973**	**0.2981**	0.6968

4.6. Discussion

In recent years, cognitive diagnosis assessment has been widely used in the field of language measurement, which offer great implication to the theoretical research of language testing, the development of language tests, teaching, and learning. The realization of cognitive diagnostic assessment is inseparable from specific cognitive diagnostic models (CDM). A cognitive diagnostic model aims to evaluate the cognitive or skill mastery of subjects from the cognitive psychology perspective and psychometric model is incorporate to provide fine-grained information about student's performance on various tests. The information obtained from CDMs benefits stakeholders, educators, curriculum developers, and students who do not have access to this type of information through traditional scoring systems. Despite the potential usefulness CDMs in advancing educational measurement, there are some limitations and drawbacks of CDM.

Majority of CDMs statistical package such as R-package, Arpeggio, Mplus and Ox code demands large scale of participants to provide the desirable outputs (Aryadoust, 2011). These demands are impossible to cater when assessment is carried out in small scale. Furthermore, each school or learning institutions conduct their very own assessment according to their own syllabus, time and their students' needs especially in higher learning institutions. This is will be quite impossible and hindrance for educators in learning institution to adapt this CDM into their assessment. Hence the usage of CDMs only can be implemented on high-stakes examinations such as MUET and *Sijil Peperiksaan Malaysia* or commonly known as SPM.

Defining the attribute that needed to be tested is very tedious and time- consuming task. Basically, an educator needs to identify the attributes that able to diagnose and provide information about their students' cognitive strengths and weaknesses. Moreover, an educator needs to set an assessment that follows the syllabus requirements and able to diagnose and provide information on students' cognitive strengths and weaknesses. Therefore the procedure of extracting and defining attributes, constructing Q-matrix and developing item questions consume a lot of time and research. As misspecification of Q-matrix could lead to various misclassification of the students.

Even though, cognitive diagnostic models were introduced decade ago into education system, DCMs are used to analyse large-scale of non-diagnostic tests such as MELAB, IELTS, PIRLS and university entrance examinations. The research on CDMS is only conducted in micro scale by certain university research team from language testing department. Hence there is lack of exposure especially towards educators from different level of education system. Whereby every educators needs to know and learn how to apply a CDM software program. Most importantly this software programs demand high

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

level of statistical and psychometric knowledge which is not an easy task (Javidanmehr & Anani Sarab, 2017).

5. Conclusion

In this study, the GDINA model is used to examine the performance of year four students in an English reading comprehension assessment to find more comprehensive feedback information. CDMs can offer rich diagnostic information about the strengths and weaknesses of the student's cognitive skills (Lee, de la Torre, & Park, 2012). The acquired information from profile scores can be utilized in tackling individual and group needs through remedial instruction and improving instruction to enhance learning and advancement (Lee & Sawaki, 2009b). To sum up, test analysis with a CDM will grant an assessment process and assist in resolving each student's educational needs (Huebner, 2010).

Ethics Approval and Consent to Participate

The researchers used the research ethics provided by the Educational Research Development and Culture Unit of the Malaysia Ministry of Education. All procedures performed in this study involving human participants were conducted in accordance with the ethical standards of the Ministry of Education.

Acknowledgement

Part of this article was extracted from a doctoral thesis that will be submitted to the University of Science Malaysia, Penang.

Funding

This study received no funding.

Conflict of Interests

The authors reported no conflicts of interest for this work and declare that there is no potential conflict of interest with respect to the research, authorship, or publication of this article.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London: Continuum.

Aryadoust, V. (2011). Application of the fusion model to while-listening performance tests. *Shiken: JALT Testing & Evaluation SIG Newsletter, 15*(2), 2-9.

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

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

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. (Eds.), *Handbook of statistics* (Vol. 26, pp. 979-1030). Amsterdam, Netherlands: Elsevier.

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

Hartz, S. M. (2002). A Bayesian framework for the unified model for assessing cognitive abilities : Blending theory with practicality. Doctoral dissertation: University of Illinois at Urbana-Champaign.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191-210.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research, and Evaluation, 15*(1), 3. doi:10.7275/7fdd-6897

Jang, E. E. (2008). A Review of cognitive diagnostic assessment for education: Theory and application. *International Journal of Testing, 8*(3), 290-295.

Javidanmehr, Z., & Anani Sarab, M. R. (2017). Diagnostic Assessment: Issues and Considerations. *International Journal of Language Testing, 7*(2), 73-98.

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

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 Quaterly, 6*(3), 172-189. doi:10.1080/15434300902985108

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*, 205-237.

Li, H., & Hunter, C. V. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing, 33*(3), 391-409.

Li, H., Hunter, C. V., & Lei,P. W. (2015). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing, 33*, 391–409.

Ma, W., Minchen, N., & de la Torre, J. (2020). Choosing between CDM and Unidimensional IRT: The Proportion Reasoning Test Case. *Measurement: Interdisciplinary Research and Perspectives, 80*(2), 87-96. doi:10.1109/15366367.2019.1697122

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

Ravand, H. (2016). Application of a cognitive diagnostic model to a high stakes reading comprehension test. *Journal of Psychoeducational Assessment, 34*, 782-799. doi:10.1177/0734282915623053

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

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

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.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Schwarz. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464. doi:10.1214/aos/1176344136

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

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorder using cognitive diagnosis models. *Psychological Methods, 11*(3), 287-305.

von Davier, M. (2005). *A general diagnostic model applied to language testing data (RR-05-16).* Princeton, NJ: Educational Testing Service.

Wang, C., Shu, Z., & Xu, G. (2015). Assessing item-level fit for the DINA model. *Applied Psychological measurement, 39*(7), 525-538. doi:10.1177/0146621615583050

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).