Development and Validation of Reading Comprehension Assessments by Using GDINA Model

Renukha Nallasamy^{1*} , Ahmad Zamri Bin Khairani²

¹Pusat Pengajian Ilmu Pendidikan, Universiti Sains Malaysia (USM), Pulau Pinang, Malaysia.

Email: renukha18@student.usm.my

²Pusat Pengajian Ilmu Pendidikan, Universiti Sains Malaysia (USM), Pulau Pinang, Malaysia.

Email: ahmadzamri@usm.my

ABSTRACT

CORRESPONDING

AUTHOR (*):

Renukha Nallasamy
(renukha18@student.usm.my)

KEYWORDS:

Cognitive diagnostic model
Reading comprehension
assessment
GDINA model

CITATION:

Renukha Nallasamy & Ahmad Zamri Khairani. (2022). Development and Validation of Reading Comprehension Assessments by Using GDINA Model. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 7(2), e001278. <https://doi.org/10.47405/mjssh.v7i2.1278>

This research is about developing and validating reading comprehension assessments using the GDINA model. Reading comprehension assessment has stayed relatively unaltered over the years, despite changes recommended by legislators or trends in the education industry. As a result, assessments are still used to evaluate students' performance, whether norm- or criterion-referenced; to inform instruction; to find out if students may acquire access to the right programmes; and even to evaluate the programme, which is still employed. The current reading comprehension assessment in Malaysia does not provide instructors with adequate information to make educated decisions. There was not enough data or information provided by the reading comprehension examinations to support the students individually, as a class, or even as a whole school, despite the teachers' sincere attempts to assist their students in performing and succeeding and becoming strong readers. We need better assessments to help instructors who are desperate for effective tools to help them fulfill individual needs. Reading comprehension is a mental process that occurs when a reader engages with the material. It's a problem for people who have trouble reading. Thus, the discussion is about developing and validating reading comprehension, which can be improvised using the GDINA model.

Contribution/Originality: This study is one of the few studies that investigated the usage of the Cognitive Diagnostic Models (CDMs) in reading comprehension assessments. CDMs can be used in two ways: retrofitting existing non-diagnostic tests or designing a set of question items from the beginning. However, most of the applications of CDMs in language testing are retrofitting on existing non-diagnostic tests.

1. Introduction

It is critical in the education system to determine if specific instructional activities can achieve the desired learning outcomes (William, 2011). Teachers and policymakers are eager to know if the instructional activities may help children improve their test scores

and overcome limitations. Teachers use this method to acquire information on their students' performance.

Students' feedback helps teachers adjust their lesson plans and instruction to meet better their students' needs, such as going back to fundamentals on specific topics, dividing students into groups based on proficiency, and offering remedial sessions. Current educational practice favours summative evaluations over formative evaluations, even in primary school (Eng, Mohamed & Mohamed Ismail, 2016). After the learning process has been completed, an assessment is conducted to offer information and comments on how well the teaching and learning process worked out. Using a set of norms or expectations, rubrics are created for the summative evaluations.

Students can be provided with these rubrics ahead of time to know the expectations for each criterion before they begin working on an exam, assignment, or project. As a result of following the same standards as students use, rubrics can help teachers be more objective when determining a final, summative mark. In addition to providing school accountability, evaluating student progress, and reporting to parents, these summative evaluations also serve various other functions.

The summative exam has a considerable impact on what is taught in the classroom, and the test performance's summary statistics reflect the school's accountability. As a result, teachers focus their lessons on material that will be covered in the final exams. As a result, they place more emphasis on methods that they feel can help students prepare for tests and examinations (Ofsted, 2008). Students' general competence is the primary focus of these exams, which are used to rank and compare students. When it comes to evaluating a student's cognitive abilities, test specifications often just focus on the content requirements (Nicholas, 1994).

In summary, summative assessments are more product-focused and focus on the finished product. After the project is finished, there are no more changes that may be made. In this way, students are evaluated solely based on their content knowledge and not based on their methodological approaches (Briggs, 2007). It's possible, however, for students to take advantage of this opportunity to address their inadequacies if the evaluation's idea is formative because formative assessment focuses on product completion. Students' ability to master 21st-century skills, such as learning how to learn, planning, thinking about one's own thinking, and monitoring and evaluating their own thinking and understanding, can only be achieved through formative assessment, which researchers Boud and Falchikov (2006) and Stiggins (2002) agreed on. On the other hand, formative assessment can collect data while students are actively engaged in the learning process.

Using this strategy, teachers and students alike can see how far they've come as instructors. If the teacher is unsure whether or not a lesson or activity should be utilised again or modified, they might poll or monitor the students. The goal of formative assessment is to discover potential areas for development to meet students' educational requirements better. Students' progress toward mastery and the efficiency of instruction can be measured with these tests, which are rarely evaluated. "Rethinking" and then "redelivering" that material has allowed teachers to ensure that their students are on track. To ensure that students are prepared for an exam, it's wise to use assessments like these to gauge their current level of understanding before putting them under pressure to perform well.

2. Reading Comprehension Assessment

Reading comprehension assessment has stayed relatively unaltered over the years, despite changes recommended by legislators or trends in the education industry. As a result, assessments are still used to evaluate students' performance, whether norm- or criterion-referenced; to inform instruction; to find out if students may acquire access to the right programmes; and even to evaluate the programme, which is still employed.

The current reading comprehension assessment in Malaysia does not provide instructors with adequate information to make educated decisions. There was not enough data or information provided by the reading comprehension examinations to support the students individually, as a class, or even as an entire school, despite the teachers' sincere attempts to assist their students in performing and succeeding and becoming strong readers. We need better assessments to help instructors who are desperate for effective tools to help them fulfill individual needs, as [Pearson and Hamm \(2005\)](#) correctly put it. Reading comprehension is a mental process that occurs when a reader engages with the material. It's a problem for people who have trouble reading.

Classroom assessment is a critical ongoing process that allows teachers to understand what a learner can and cannot do in the classroom ([Brown, 2004](#); [Popham, 1999](#)). Students benefit from the information gained via assessments, according to [Masters \(2014\)](#). For the most part, the purpose of assessment is not to make comparisons. Instead, the goal is to identify students' strengths and flaws and present them with constructive comments. In Malaysia, Ministry of Education (MOE) introduces School-Based Assessment System (SBA, or better known as PBS in the Malay language) which usher in a new assessment system through a holistic approach. Whereby, this new SBA approachable to access students equally based on their cognitive, affective, and psychomotor domains as stipulated in the National Education Philosophy ([Curriculum Development Division, 2011](#)).

As English is taught as a second language (ESL) in all primary and secondary schools in Malaysia, the Standard- based English Language Curriculum (SBELC, or also known as *Dokumen Standard Kurikulum dan Pentaksiran*) ([Ministry of Education, 2018](#)) is conducted to gather accurate and detailed information about students' performance and to use the data effectively to improve teaching and learning. It's actually involves of gathering information of students knowledge, learning, and skills to inform teachers' instructional practices. The feedback from SBA assessments should focus on helping students to improve their learning. It also specifies that primary school students should be able to read independently by the end of Year 6 of primary school.

In an ESL classroom, primary school students learn how to improve their reading comprehension. The English language proficiency of upper primary children is evaluated in accordance with the Primary School Achievement Test framework (also known as *Ujian Penilaian Sekolah Rendah*). As well as students' understanding of what they read, this test measures their knowledge of vocabulary, grammar, sentence structure, and note expansion. The only way to know how well a student did on a test is to provide a grade, and that's it. The Malaysian Ministry of Education also uses the findings obtained to evaluate the effectiveness of ESL teaching ([Mohd Sofi Ali, 2003](#)). As a result, the ability of a student to comprehend what they have read cannot be evaluated in any meaningful way.

Students learning English as a second or foreign language need to be proficient readers. Readers who have a strong command of the English language have a leg up in the classroom (Puteri et al., 2017). An analogy is drawn between the reading comprehension process and reading performance, defined as the interaction with a text involving a wide range of cognitive skills and processes. According to numerous surveys and research, children with varying reading abilities arrive at the same class. Children have always entered school with a wide range of literacy experiences and abilities, and instructors have struggled for years to satisfy the requirements of all of their learners, according to a case study by Ankrum and Bean (2008).

According to Santhi (2011), all students have different abilities and progress at different rates. For the purpose of improving student performance, teachers in Malaysia are shifting their attention from evaluation of learning to assessment for learning, as reported by Charanjit et al. (2017). Many students have many intelligences, and the results of one exam cannot reflect all of these intelligences. Because students are typically only given a composite grade, the assessment results unquestionably do not show the students' genuine ability. As a part of the Primary School Achievement Test in Malaysia, students receive grades based on their proficiency in English. Because just a tiny portion of the test is dedicated to reading comprehension, these composite grades cannot accurately assess students' reading abilities.

As Lin et al. (2016) point out, grades alone provide only a limited amount of information on students' levels of reading ability. Even if two students get the same grade on a test, that doesn't mean they have the same degree of reading proficiency. In general, grades don't give enough information about a student's skills and limitations. Even, the Standard-based English Language Curriculum (SBELC) document incorporates a mapping of the English Language Content and learning standards as well as provides a descriptor for performance standards to teachers; nevertheless, the descriptors lack information and do not guide the instructors to assess students' capabilities correctly (Ministry of Education Malaysia, 2017).

Consequently, educators continue to fumble around in the dark, and the problem of students not knowing their strengths and limitations while answering reading comprehension questions endures. A second attempt was made by the Ministry of Education Malaysia (2013), this time with the goal of categorising schools according to their performance on a performance scale (Band 1 to Band 7). Only the overall performance in the many disciplines taught in the schools determines the school performing bands, not a student's reading comprehension level.

Teachers aren't going to benefit from this endeavour. To boost academic performance, it only categorically identifies the school as a whole. As a solution, the researcher plans to reexamine reading comprehension assessments and fill the void by creating an accurate reading evaluation system for Malaysian primary schools. Ahlam et al. (2017) stated that teachers need to keep track of students' progress to justify their instruction and keep it in sync with their development as they advance through their education. Additionally, the researcher plans to design a credible reading assessment system that will help ESL teachers overcome their difficulties in creating reading comprehension assessments.

Reading professionals conducted the majority of the qualitative research on taxonomies of reading comprehension skills. In the past, reading specialists have used a series of passages and questions designed to test different levels of comprehension to assess

reading comprehension skills or abilities. According to [Davis \(1968\)](#) and [Munby \(1978\)](#), there are eight different types of reading "skills," while [Alderson and Lukmani \(1989\)](#) derived their eight-skill taxonomy from Bloom's Taxonomy of Educational Objectives in the Cognitive Domain ([Bloom, 1956](#)), and [Heaton \(1991\)](#) summarised his fourteen skills taxonomy from the classifications as mentioned earlier.

However, these qualitatively mapped taxonomies of reading comprehension skills need to be used with caution. Their beginnings are more likely to be derived from a theorist's armchair than an actual study as a first step. As a second disadvantage, they are frequently vague or nonexistent, making them appear more distinct than they are. Thirdly, it is difficult to get all experts to agree on what skills each test item operationalizes. To sum up, the following taxonomies are essentially lists of talents with no well-defined relationship between them.

Diverse statistical approaches were used to do quantitative research on taxonomies of reading comprehension skillsets. Although studies using factor analysis and multiple regression (such as those conducted by [Pierce Katzir Wolf and Noam \(2010\)](#) and by [Spearitt \(1972\)](#) and [Williams, Eaves, and Cox \(2001\)](#)) were able to identify reading comprehension abilities in their subject matter empirically, those studies did not go into detail about the connections between the various skills. There may be a relationship between reading comprehension skills and a static construct in the minds of the subjects, but studies using multitrait-multimethod correlational analysis (Marsh, Butler) and structural equation modeling ([Song, 2008](#); [van Steensel, Oostdam & van Gelderen, 2013](#)) may not reveal the relationship between reading comprehension skills and a dynamic construct in the process of solving a specific i.

3. Approaches in Diagnostic Assessments

In educational assessment, the primary goal is to help students learn rather than evaluate them. As a result, educators place a high value on requiring students to submit formative and diagnostic assessment data as part of the process of improving classroom instruction and student learning. According to [Keeley and Tobey \(2011\)](#), diagnostic assessments can inform instructional decisions by providing information about students' views, reasoning methods, and educational requirements.

Diagnoses provide information about students' past knowledge and skill levels concerning the content of the course. An important aspect of instructional decision-making is the diagnosis ([Ketterlin-Geller & Yovanoff, 2009](#)). Students who may be in danger of failing are identified, and useful instructional material is provided to develop remedial education programmes or extra interventions through the use of this platform. With this information, the teachers may adjust their lessons to prepare students for the real world. Because of this, students' educational needs will be met through various teaching methods ([Fuchs, Fuchs, Hosp & Hamlett, 2003](#)). Due to the difficulty of giving, translating, and applying diagnostic tests, teachers tend to avoid using them as a guide for making instructional decisions. There is also a general lack of understanding of the types of assessments that can be utilised for diagnosis, according to [Ketterlin-Geller and Yovanoff \(2009\)](#).

Response analyses and cognitive diagnostic tests are two of the most commonly utilised methods for providing diagnostic information. In response analysis, students' responses to the instructional relevant items are analysed based on their errors or their skills. On

the other hand, error analysis concentrates on students' deficiencies and helps teachers categorize students' blunders. Students' mastery of individual subskills can be assessed through a skills analysis. Correct or incorrect responses are scored using this simple dichotomous procedure. However, for determining the underlying cognitive processes of students, [Ketterlin-Geller and Yovanoff \(2009\)](#) noted that error analysis and skill analysis methodologies have only limited advantages.

In order to rectify students' current misconceptions, it is possible that the results of analyses of students' replies to problems might be used to adjust education. However, these analyses may only provide limited information regarding students' persistent and systematic thinking errors. It is possible to test individual student-level cognitive processes using cognitive theory and statistical modeling of response patterns, on the other hand, in a cognitive diagnostic assessment. Educators can devise more effective educational plans for students who need extra help using this data. This table compares diagnostic assessment methodologies and describes each approach's relative strengths and weaknesses to make educational judgments. Education professionals can distinguish between these assessment methods and choose the best one for their needs.

In assessing students, teachers are primarily interested in finding out what they can do and how much they already know about a certain topic ([Pellegrino, 2005](#)). Because assessing students' knowledge and educational outcomes is not as straightforward as weighing or measuring, evaluations cannot directly relate a student's thinking. It's challenging to observe the characteristics that are being measured because they are based on mental representations and processes. As a result, an assessment is a tool for observing students' behaviour and generating data to develop fair conclusions about students' knowledge. Because of this, deciding what to assess and the best method for doing so isn't as straightforward as it might appear.

At the moment, the emphasis is on evaluating how students learn cognitively. In order to improve educational measurement, the findings of cognitive science and psychometrics were required. Educational assessment and cognitive science can diagnose children's learning and advise teachers to enhance their teaching and learning processes. Researchers [Nicholas, Chipman, Brennan \(1995\)](#), [Choi, Lee, and Park \(2015\)](#) have supported this. Assessments based on the results of cognitive science research are called cognitive diagnostic assessments (CDAs), and they use psychometric models to analyse the cognitive process underlying students' test results ([DiBello, Roussos, & Stout, 2007](#); [Hartz, 2002](#); [Nicholas, 1994](#); [Embretson, 1991](#)).

The goal of a cognitive diagnostic assessment (CDA) is to accurately determine a student's cognitive strengths and limitations by measuring specific knowledge structures and cognitive skills ([Leighton & Gierl, 2007](#)). Teachers can identify students' misunderstandings and irrational strategies based on information about their cognitive strengths and shortcomings ([Nicholas, 1994](#)). Instead of utilising a highly reliable scale to score students for the sake of selection, classification, or summative evaluation, as is the case with curriculum-based assessments, teachers create these exams primarily to analyse the statistical features of the items.

Even in CDAs, [Leighton and Gierl \(2007\)](#) emphasise that the content domain of test development is founded on empirical and theoretical evidence. Psychological studies involving knowledge structure and skills or traits related to a domain's desired learning outcomes guide item generation in CDAs in this specific domain. Standardised large-scale

testing is distinct from CDAs, according to [Leighton \(2004\)](#). Test developers use table specifications ([Millman & Green, 1989](#)) to define knowledge and skill needs. Specification tables like this one show instructors' use of knowledge and abilities as markers of proficiency in the curriculum specifications and standards.

Discrimination, difficulty, and dimensionality are all met by the items in the item pool that have been picked. It is possible that the information and skills listed in the table specification do not accurately reflect what students really know and can do on the test. There is no assurance that students' response patterns suggest that cognitive processes indicative of competence in the specific area described in the test specification table can be provided to stakeholders ([Nicholas, 1994](#)).

In the 1980s, cognitive psychology and psychometrics were combined to create items based on cognitive reasoning ([Butterfield, Nielsen, Tangen & Richardson, 1985](#); [Bejar & Yocom, 1986](#)). This idea of CDA gained traction because of [Embretson's \(1991\)](#), where his well-developed cognitive theory can give a framework for the production of items measuring knowledge structures and procedural skills. According to [Dibello, Roussos, and Stout \(2007\)](#) and [Nicholas, Chipman, and Brennan \(1995\)](#), CDA can identify students' cognitive impairments in a certain learning domain.

Educators and researchers agree that CDAs can assist overcome some of the drawbacks of the traditional assessment method ([Cui, Leighton & Zheng, 2006](#)). Students must master various cognitive abilities and traits to successfully complete and pass each CDA item test, which is a fundamental assumption ([Gierl et al., 2000](#)). For teachers and students alike, the CDA provides a snapshot of how well students are grasping key fine-grained qualities, as reported even by CDA's own [Huebner \(2010\)](#).

The diverse cognitive processes and knowledge required to solve test questions in a particular area of interest can be measured with a well-defined CDA. In addition, CDA, which is coordinated with instruction, can extract information from specific response patterns to provide comprehensive information on the underlying cognition of students. As a result, teachers can use this method to identify students' cognitive strengths and weaknesses in a particular area of study and use this information to help them improve their performance in that area.

CDA focuses on the qualitative and multifaceted character of knowledge structure and its relevance to student performance evaluation, as [Nicholas \(1994\)](#) stated. Combining cognition modelling (cognitive model of learning) and psychometric modelling (cognitive diagnosis) is the primary goal of CDA (CDM). Researchers and educators can study cognitive processes and knowledge structures using CDAs, which combine cognitive theory with statistical modelling of response patterns.

4. GDINA Model in Reading Comprehension Assessment

The G-DINA model is a generalization of the DINA model with a more relaxed assumption to address the strong conjunctive assumption asserted in the DINA model. It also serves as a general framework for deriving other CDM formulations, estimating some commonly used CDMs, and testing the adequacy of reduced models in place of the saturated model. Like all other CDMs, the G-DINA model also requires a $J \times K$, Q-matrix ([Tatsuoka, 1983](#)), in which element in row j and column k in the Q-matrix, q_{jk} , is equal to 1 if k th attribute is required in answering item j correctly, or else it is equal to zero.

The G-DINA classifies the latent classes into $2^{K_j^*}$ latent groups, where $K_j^* = \sum_{k=1}^K q_{jk}$ signifies the number of prerequisite attributes needed for item j . Each latent group is reduced to an attribute vector represented by α_{lj}^* whose elements are required attributes for item j . In this research, it would be adequate to consider the reduced attribute vector $\alpha_{lj}^* = (\alpha_{lj1}, \dots, \alpha_{ljK_j^*})'$ in place of the whole attribute vector $\alpha_{lj} = (\alpha_{lj1}, \dots, \alpha_{ljK})'$. Therefore, each latent group has the probability of correctly answering the item represented by $P(\alpha_{lj}^*)$ (de la Torre, 2011).

Non-compensatory models were found to be more frequently employed in CDA research of reading comprehension. According to the studies that have supported the use of non-compensatory models, linkages between the basic cognitive skills of reading are non-compensatory. In contrast, studies on compensatory models hold that reading's fundamental skills have both compensatory and non-compensatory linkages (Jang, 2009; Li, 2011; Li et al., 2015). While this is true, Li et al. (2015) pointed out that when the relationship between cognitive skills is not fully understood, it is safe to employ saturated (more complicated) CDM, which considers many interpersonal skills. GDINA (de la Torre, 2011) is a saturated CDM that has been used in Ravand's (2016) and Chen and Chen's (2016) experiments to improve reading comprehension. Because this model "is sensitive to the integrative nature of reading comprehension skills and capacity to discern an interacting link among them," this feature coincides with that of reading comprehension (Chen & Chen, 2016).

A reading comprehension test may theoretically use the GDINA paradigm. It is determined that the G-DINA model has been successfully used in reading comprehension tests, even though there are only a few sources to look into. When Chen and Chen (2016) applied the G-DINA model and analysed the test response data of 1,029 British secondary school students' performance on a reading test (i.e., the Program for International Student Assessment), they found that the five reading subskills (i.e., identifying explicit information, generalising main idea, interpretation, and explanation, getting inference, and evaluation and comment) specified and defined by six content experts were all related. According to the findings, the G-DINA model is well-suited to assessing reading subskills and even suggests that it can be used to evaluate examinations that require a hierarchy of abilities. The G-DINA model is well-suited for modeling reading comprehension subskills, according to research findings.

However, G-DINA has been used in only a few educational projects. In order to obtain more granular diagnostic information about students, this model might be used for the current study to examine their performance on the reading comprehension examination. The goodness of fit of many more frugal forms of the G-DINA model will be assessed further to find the best-fitting model.

A latent variable model created for cognitive diagnostic examinations, cognitive diagnosis models (CDMs), is used to measure students' proficiency in a collection of finer-grained skills. In the form of score profiles, CDMs can provide more specific information that can be used to accurately measure student learning and progress, create better education, and possibly intervene to address individual and group needs (de la Torre, 2009, 2011). The fact that reading comprehension is a contentious skill makes using CDMs in tests of understanding a difficult issue for CDM researchers.

It's general knowledge that reading is a two-step process: vocabulary acquisition and cognitive comprehension. A standard description of the latter entails parsing sentences, interpreting sentences in context, constructing a context structure, and finally connecting this new understanding to what one already knows (Gough, Juel, & Griffith, 1992). A high correlation between vocabulary abilities and academic reading performance has been established. It may be assumed that vocabulary is not all that reading examinations are designed for under more abstract comprehension skills. That's why this study concentrated on the understanding components of the study.

Even though many specialists in reading (Davis, 1968; Heaton, 1991; Hughes, 1989; Munby, 1978) considered reading comprehension to be multi-divisible, other experts maintained that reading comprehension was unitary or holistic. As far as we know, there is no evidence to support the existence of discrete independent abilities and, instead, reading is an all-encompassing aptitude. It was shown that judges of reading comprehension were often unable to agree on assigning certain skills and techniques for specific test items (Alderson, 1990a, 1990b). Divisibility was suggested by Weir and Porter (1994) to be dependent on how proficient readers were, suggesting that while reading was one-dimensional for proficient readers, it wasn't for less proficient readers. Selecting appropriate CDMs and Q-matrices that take into account the qualities of reading comprehension skills would help cognitive diagnostic on comprehension tests be more applicable and accurate. As a result, the G-DINA model, a saturated CDM, was used in this work to perform cognitive diagnosis on reading comprehension tests. The G-DINA model may aid in better understanding the interrelationships between reading comprehension skills and other traits, allowing researchers to make more informed decisions about how to teach reading comprehension skills and how to solve reading problems in general.

5. Conclusion

Cognitive Diagnostic Models give diagnostic information by providing examinees' mastery profiles on a set of preset skills. Using CDMs, educators can obtain more accurate information in the form of a score profile. Hence, enabling them to effectively measure students' learning and progress and designing better instructional to cater to the students' needs, and likely conducting some intervention programs to address an individual or group needs (de la Torre, 2011).

Acknowledgement

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

Funding

This study received no funding.

Conflict of Interests

The authors declare no conflict of interest in this study.

References

- Ahlam, A. S. H., Harvinjitt, Kaur a/p D. S., Ibrahim, M. S., & Zazali, M. (2017). Teachers' perception towards the use of classroom-based assessment in English reading. *International Journal of Education and Research*, 5(11), 153-168.
- Alderson, J. C. (1990a). Testing reading comprehension skills (part one). *Reading in a Foreign Language*, 6, 425-438.
- Alderson, J. C. (1990b). Testing reading comprehension skills (part two). *Reading in a Foreign Language*, 7, 465-503.
- Alderson, J.C. & Lukmani, Y. (1989). Cognition and reading: cognitive levels as emodied in test questions. *Reading in a Foreign Language*, 5 (2), 253-70.
- Ankrum J. W., & Bean, R. M. (2008). Differentiating reading instruction: What and how. *Reading Horizons*, 48(2), 133-146.
- Bejar, I. I., & Yocom, P. (1986). A generative approach to the development of hidden-figure items. *ETS Research Report Series*, 1986(1), i-38.
- Bloom, B. (1956). Taxonomy of educational objectives, handbook I: The cognitive domain. New York, NY: David McKay.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & evaluation in higher education*, 31(4), 399-413.
- Briggs, D. (2007). Assessing what students know, how they know it, or both?. *Measurement: Interdisciplinary Research and Perspectives*, 5(1), 62-65.
- Brown, H. D. (2004). Language assessment: Principles and classroom practices. New York: Pearson Education.
- Caldwell, J. S. (2014). *Reading Assessment: A Primer for Teachers in the Common Core Era (3rd ed.)*. New York: The Guilford Press.
- Butterfield, E.C., Nielsen, D., Tangen, K.L., & Richardson, M.B. (1985). Theoretically based psychometric measures of inductive reasoning. In S.E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics*. New York: Academic.
- Charanjit Kaur Swaran Singh, Othman Lebar, Napisah Kepol, Rafiah Abdul Rahman & Kurotol Aini Muhammad Mukhtar (2017). An observation of classroom assessment practices among lecturers in selected Malaysian higher learning institutions. *Malaysian Journal of Learning and Instruction*, 14(1), 23-61.
- Chen, H., & Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly*, 13(3), 218-230.
- Choi, K. M., Lee, Y. S., & Park, Y. S. (2015). What CDM can tell about What students have learned: An Analysis Of TIMSS Eight Grade Mathematics. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(6), 1563-1577.
- Cui, Y., Leighton, J. P., & Zheng, Y. G. (2006). *Simulation studies for evaluating the performance of the two classification methods in the AHM*. San Francisco, CA: NCME.
- Curriculum Development Division. (2011). *Standard Document of Primary School Curriculum: Basic Core Module of English Language for national Schools*. Putrajaya.
- Davis, F. B. (1968). Research in comprehension in reading. *Reading Research Quaterly*, 3, 499-545.
- 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 model framework. *Psychometrika*, 76(2), 179-199.
- DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychmetric models. In C. R. (Eds), *Handbook of statistics, vol. 26: psychometric* (pp. 979-1030). Amsterdam, Netherlands: Elsevier.

- Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 495-516.
- Eng, L. S., Mohamed, A. R., & Ismail, S. A. M. M. . (2016). Systematic Tracking of Malaysian Primary School Students' ESL Reading Comprehension Performance to Facilitate Instructional Processes. *International Journal of Instruction*, 9(1), 149-162.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Hamlett, C. L. (2003). The potential for diagnostic analysis within curriculum-based measurement. *Assessment for Effective Intervention*, 28(3-4), 13-22.
- Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2000). An NCME Instructional Module on Exploring the Logic of Tatsuoka's Rule-Space Model for Test Development and Analysis. *Educational Measurement: Issues and Practice*, 19(3), 34-44. doi:10.1111/j.1745-3992.2000.tb00036.x
- Gough, P., Juel, C., & Griffith, P. (1992). Reading, speaking and the orthographic cipher. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 35-48). Hillsdale, NJ: L. Erlbaum.
- Hartz, S. (2002). *A Bayesian framework for the Unified Model for Assessing Cognitive Abilities: Blending Theory with Practicality* (Unpublished doctoral dissertation). University of Illinois at Urbana-Campaign, IL.
- Heaton, J. B. (1991). *Writing English language test*. New York, NY: Longman
- 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.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fushio model application to LanguEdge assessment. *Language Testing*, 26(1), 031-073.
- Keeley, P., & Tobey, C. R. (2011). *Mathematics Formative Assessment, Volume 1: 75 Practical Strategies for Linking Assessment, Instruction, and Learning*. Corwin Press.
- Ketterlin-Geller, L., & Yovanoff, P. . (2009). Cognitive diagnostic assessment in mathematics to support instructional decision making. *Practical Assessment, Research, & Evaluation*, 14(16), 1-11.
- Leighton, J. P. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, 4, 6-15. doi:10.1111/j.1745-3992.2004.tb00164.x.
- Leighton, J. P., & Gierl, M. J. (2007). Defining and evaluating models of cognition used in educationl measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, 26, 3-16.
- Li, H. (2011). *Evaluating language group differences in the subskills of reading using a cognitive diagnostic modeling and differential skill functioning approach*. University State College, Penn State, PA: Unpublised Doctorate.
- Li, H., Hunter, C. V., & Lei, P. W. (2015). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*.
- Lin, S. E., Abdul, R. M., & Shaik, A. M. M. I. (2016). Systematic tracking system of Malaysian primary school students' ESL reading comprehension performance to facilitate instructional processes. *International Journal of Instruction*, 9(1), 149-162. <https://doi.org/10.12973/iji.2016.9112a>
- Masters, G. N. (2014). *Assessment: Getting to the essence*, ACER. Retrieved October 7, 2016, from https://www.acer.org/files/uploads/Assessment_Getting_to_the_essence.pdf

- Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed), *Educational measurement (3rd ed)* (pp. 335-366). New York: Council of Education/Macmillan.
- Ministry of Education Malaysia (2013). Malaysia Education Blueprint 2013-2025. Putrajaya: Kementerian Pendidikan Malaysia
- Ministry of Education Malaysia. (2017) *Dokumen Standard Kurikulum Dan Pentaksiran Bahasa Inggeris SK Tahun 1*. Putrajaya: Malaysia
- Ministry of Education. (2018). *Dokumen Standard Kurikulum Dan Pentaksiran Bahasa Inggeris SK Tahun 4*. Putrajaya: Malaysia.
- Mohd Sofi Ali (2003). English language teaching in primary schools: Policy and implementation concerns. Retrieved from <http://www2.moe.gov.my/~ipba/EJournal/Mohdsofi.pdf/>
- Munby, J. L. (1978). *Communication syllabus design*. Cambridge, UK: Cambridge University Press.
- Nicholas, P. D., Chipman, S. F. and Brennan, R. L. (eds). (1995). *Cognitive Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Association, Publishers.
- Nichols, P. D. (1994). A framework for developing cognitive diagnostic assessment. *Review of Educational Research*, 64, 575-603.
- Ofsted (Office for Standards in Education) (2008). Mathematics: Understanding the Score. London: Ofsted. Retrieved on October 10, 2021 from [http://nmpartnership.co.uk/pdf/Mathematics_understanding_the_score_-_primary\[1\].pdf](http://nmpartnership.co.uk/pdf/Mathematics_understanding_the_score_-_primary[1].pdf)
- Pearson, P. D., & Hamm, D. N. (2005). The Assessment of Reading Comprehension: A Review of Practices-Past, Present, and Future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13-69). Lawrence Erlbaum Associates, Publishers.
- Pellegrino, J. W. (2005). The challenge of knowing what students know. *Measurement: Interdisciplinary research and perspectives*, 1(1).
- Pierce, M., Katzir, T., Wolf, M., & Noam, G. (2010). Examining the construct of reading among dysfluent urban children: A factor analysis approach. *Journal of Literacy Research*, 42, 124-158. doi:<http://dx.doi.org/10.1080/10862961003796140>
- Popham, W. J. (1999). *Classroom assessment: What teachers need to know (2nd ed.)*. Boston: Allyn and Bacon.
- Puteri, R. M. A. R., Faridah, Y., & Zuraida, D. (2017). Facilitating reading comprehension among ESL learners using graphic organizers. *Malaysian Journal of ELT Research*, 13(1), 30-42.
- 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
- Santhi, G. (2011). Problems of teaching English in mixed ability classrooms at polytechnic level in India.
- Song, M. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25, 435-464. doi:<http://dx.doi.org/10.1177/0265532208094272>.
- Spearitt, D. (1972). Identification of subskills of reading comprehension by maximum likelihood factor analysis. *Reading Research Quarterly*, 8, 92-111.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Tatsuoka, K. (1983). Rule Space; An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.

- van Steensel, R., Oostdam, R., & van Gelderen, A. (2013). Assessing reading comprehension in adolescent low achievers: Subskills identification and task specificity. *Language Testing*, 30, 3–21. doi:<http://dx.doi.org/10.1177/0265532212440950>
- Weir, C. J., & Porter, D. (1994). The multi-divisible or unitary nature of reading: The language tester between Scylla and Charybdis. *Reading in a Foreign Language*, 10, 1–19.
- William, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3-14.
- Williams, T. O., Eaves, R. C., & Cox, C. (2001). Factor structure of the woodcock reading mastery tests-revised, forms G and H. *Assessment for Effective Intervention*, 26, 35–42. doi:<http://dx.doi.org/10.1177/073724770102600303>