Developing a reading comprehension test for cognitive diagnostic assessment: A RUM analysis

Fatemeh Ranjbaran*, Seyyed Mohammed Alavi

Department of Foreign Languages and Literature, University of Tehran, Iran

ARTICLE INFO

Keywords:
Attributes
CDA
Formative assessment
CDM
Fusion model
Q-matrix
Reading comprehension test
RUM
Second language

ABSTRACT

A critical issue in cognitive diagnostic assessment (CDA) lies in the dearth of research in developing diagnostic tests for cognitive diagnostic purposes. Most research thus far has been mainly carried out on large-scale tests, e.g., Test of English as a Foreign Language (TOEFL), Michigan English Language Assessment Battery (MELAB), International English Language Testing System (IELTS), etc. In particular, CDA of formative language assessment that aims to inform instruction and to discover strengths and weaknesses of students to guide instruction has not been conducted in a foreign (i.e., second) language-learning context. This study explored how developing a reading comprehension test based on a cognitive framework could be used for such diagnostic purposes. To achieve this, initially, a list of 9 reading attributes was prepared by experts based on the literature, and then the targeted attributes were used to construct a 20-item reading comprehension test. Second, a tentative "Q-matrix" that specified the relationships between test items and the target attributes required by each item was developed. Third, the test was administered to seven language-testing experts who were asked to identify which of the 9 attributes were required by each item of the test. Fourth, on the basis of the overall agreement of the experts' judgments concerning the choices of attributes, review of the literature and results of student think-aloud protocols, the tentative Q-Matrix was refined and used for statistical analyses. Finally, the test was administered to 1986 students of a General English Language Course at the University of Tehran, Iran. To examine the CDA of the test, the Reparameterized Unified Model (RUM) (also known as the Fusion Model), a type of cognitive diagnostic measurement model (CDM), was used for further refining the Q-Matrix for future data analyses and, most importantly, for diagnosing the participants' strengths and weaknesses. Data analysis results confirmed that the nine proposed reading attributes are involved in the reading comprehension test items. Such diagnostic information could be helpful for teachers and practitioners to prepare instructional materials that target specific weaknesses and inform them of the more problematic areas that need to be emphasized in class in order to plan for better L2 reading instruction. Further, such information could inform individualized student instruction and produce improved diagnostic tests for future use.

1. Introduction

Over recent decades, testing and assessment in general have been employed to obtain overall, average or individual scores of achievement for examining student's acquired knowledge. However, assessment that can function as formative to inform instruction has currently become the focus of attention for diagnostic purposes and is intended to assess strengths and weaknesses of students to guide instruction. From another stance, teacher-made tests and quizzes have recently gained attention because of their formative assessment role in aiding students' learning (Huang & Wu, 2013). To make the process of learning easier and more effective, teachers are expected to be competent in test construction and learning diagnostics in class. As these tests are expected to

detect student errors, incomplete understandings, and misconceptions during the learning process, the use of diagnostic tests to improve student's conceptual understanding has become highly valued and recognized in many fields (Hartman, 2001). Hence, researchers and practitioners are increasingly focusing on the integration of cognitive psychology, educational pedagogy, and educational measurement for the improvement of learning and instruction (Leighton, Gierl, & Hunka, 2004; Mislevy, 2006; Tatsuoka, 1995; Snow & Lohman, 1989).

Reading is a fundamental skill for gaining knowledge in all academic fields. Therefore, it is necessary to determine the various components of reading ability for a better understanding of reading and to find language learner problems when teaching second language (L2) reading. If problematic areas of reading proficiency are diagnosed

during the instructional term, sufficient and timely feedback can be provided to students in order to improve learning and eliminate weaknesses during the learning process.

Criticism occurs because the main goal of educational tests is usually to provide quantitative assessment of a student's general overall, often unidimensional ability and proficiency as compared to other students in a normative group. This type of norm-referenced testing has been used extensively for ranking and selecting students for various educational decisions. In addition to merely providing general summarizing and usually unidimensional information about students' skills and their ability to perform on a test, these assessments are invariably incapable of providing necessary detailed information about students' strengths and weaknesses that could possibly help them in improving their skills or that might also assist the teacher in instructional planning. Recently, scholars have suggested that cognitive diagnostic assessment has a key role in improving the informational value of assessment (Alderson, 2010; de la Torre, 2009; Jang, 2005; Leighton & Gierl, 2007; Rupp, Templin, & Henson, 2012). In his commentary on "Cognitive Diagnosis and Q-Matrices in Language Assessment," Alderson (2010) shows his disappointment of there being very few truly diagnostic tests in existence. In fact, nearly all studies carried out thus far have been on existing large-scale assessments and proficiency tests, not those developed for low-stakes formative assessment. He argues that far more studies focus on developing and researching high-stakes proficiency tests aimed at placement, achievement, or aptitude than are specifically constructed for cognitive diagnosis in the form of classroom-based formative assessments. The most desirable cognitive diagnostic assessment is the one that is diagnostically designed, constructed, and scored from the initial phase. In such an approach, cognitive attributes are explicitly defined to be targeted in the test construction phase. These predetermined attributes should be in line with the instructional goals. When the attributes are set, the data are to be analyzed with an appropriate CDM. Afterwards, the scores are to be reported in a fine-grained diagnostic system. While fine-grained cognitive diagnostic assessment is intended to inform instructional settings in this way, diagnostically constructed designs have hardly been discussed in the literature. A few tests, however, have been designed in order to fulfill the needs of diagnostic analysis (e.g. DIALANG by Alderson, 2005; Alderson & Huhta, 2005; DELNA (www.delna.auckland.ac.nz/uoa); and DELTA by Urmston, Raquel, & Tsang, 2013), while, none have yet provided individualized score reports to enhance learning and teaching at the classroom level. This study responds to the call for cognitive diagnostic assessment using a specially constructed diagnostic test, one that will attempt to provide detailed information about students' strengths and weaknesses in L2 reading comprehension, and perhaps in reading comprehension in general.

2. Literature review

2.1. L2 reading ability

In CDA, the different components of a specific domain (in this case, L2 reading) are referred to as attributes. Attributes are the divided components of a general cognitive ability, which can be defined as "procedures, skills, or knowledge a student must possess in order to successfully complete the target task" (Birenbaum, Kelly, & Tatsuoka, 1993, p.443). Therefore, L2 reading attributes are composed of different types of language knowledge, and reading strategies, which are required in comprehending texts (Birenbaum et al., 1993; Templin, 2004). CDA has been introduced as a new method in educational measurement that can provide fine-grained diagnostic information about test-takers' degree of mastery of domain sub-skills (Lee & Sawaki, 2009). Sub-skills are defined as domain-specific knowledge and skills that are required to indicate mastery in a specific cognitive domain (Leighton & Gierl, 2007). Taking reading skill as a cognitive domain, it is necessary to have knowledge of vocabulary, grammar, and making

inferences in order to fully comprehend a text. These are considered the sub-skills of the reading domain, which are called attributes as well, used interchangeably throughout this paper. The most distinct characteristic of this approach is that it is the point where cognitive psychology and psychometric modeling meet within a single framework, therefore it aims to assess the test-takers' knowledge and underlying cognitive processing sub-skills (DiBello, Roussos, & Stout, 2006; Leighton & Gierl, 2007).

In the assessment of reading comprehension in a second or foreign language, the many underlying cognitive attributes required for reading ability mastery have made it a complex process. Reading ability is a fundamental tool for gaining knowledge and improving learning in everyday academic settings and everyday life in general. Therefore, it comes as no surprise that the nature of reading ability has been the focus of research in applied linguistics, education and psychology for quite some time (Cohen & Upton, 2006). Regardless of the extensive research on reading ability, there is still some debate as to how second language reading ability is defined and how its performance should be evaluated and reported. It seems that teachers, students, and practitioners have not been given diagnostic feedback tools that could be used for improvements in reading ability, specifically for classroom-based profile score reporting. These are issues that mostly need consideration in the context of L2 reading assessment in the Iranian context.

At times, there is such emphasis on reading strategies that other important elements of reading competence such as language knowledge, including pragmatic knowledge and grammatical knowledge, have been have been given less attention. One aspect of second language reading ability is the use of language to understand written text. Therefore, the knowledge of language components and strategic reading competence should both be considered in order to master the written text. While the difficulty of defining the construct of reading ability is clear, other problems have been seen with regards to how L2 reading performance has been analyzed and reported. L2 reading test scores are often reported using a general test score without any detailed information (Goodman & Hambleton, 2004). When an exam provides only one total score, it can serve the test's immediate summative purpose; however, it cannot be easily used to improve reading performance (Stiggins, Alter, & Chappius, 2004). Only providing a total score does not provide information regarding each student's specific strengths and weaknesses (Sheehan & Mislevy, 1990). On the other hand, a detailed score report of each individual, including their level on each reading attribute can be used to both improve individual student reading ability and guide teacher instruction (Snow & Lohman, 1989).

2.2. Frameworks for developing cognitive diagnostic tests

There are two measurement driven approaches that have been widely used for diagnostic test development. One is Embretson's Cognitive Design System (CDS) (Embretson & Gorin, 2001) and the other is Mislevy's Evidence-Centered Design (ECD) (Mislevy, 1994; Mislevy, Steinberg, & Almond, 2002). These two approaches focus on the use of cognition in the process of item and test development, considering the issues of construct definition while item writing, and concluding with validation procedures (Leighton & Gierl, 2007).

CDS and ECD may differ in their emphasis on the various parts of assessment design and their details, but both share the three principles of the assessment triangle. The assessment triangle includes three related elements, which are cognition (theories of learning), observation (test data) and interpretation (the probabilistic model that relates a student's multidimensional latent cognitive learning state to his/her test response pattern) (Pellegrino, Chudowsky, & Glaser, 2001). This NRC panel of researchers state that cognition is related to a cognitive model about how students represent knowledge and how they develop competence in a certain subject (p.44). A cognitive model provides a description of what should be assessed, but it is different from cognition