Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 191 (2015) 121 – 125

Procedia
Social and Behavioral Sciences

WCES 2014

A study of Quality Assessment of Science Instructional Management in Thailand: An analysis of Differential Item Functioning and Test Functioning in Mixed format Tests

Panat Chanpleng [a], Nuttaporn Lawthong [b]*, Sungwon Ngudgratoke[c]

[a] *Ph.D. Student in the Department of Educational Research and Psychology, Chulalongkorn University, Bangkok 10330, Thailand*
[b] *Assistant Professor in the Department of Educational Research and Psychology, Chulalongkorn University, Bangkok 10330, Thailand*
[c] *Lecturer in the School of Educational Studies, Sukhothai Thammathirat Open University, Nonthaburi 11120,Thailand*

Abstract

The international education assessment is important for educational improvement when it is used appropriately by participating countries to identify weaknesses of educational systems. However, its usefulness is dependent on validity of the test itself. One aspect of validity issues worth being investigated is the fairness which assesses whether or not assessment items/test is fair to all subgroups of examinees. Differential item functioning (DIF) and test functioning (DTF) methods are usually used to assess fairness. Test items that are not fair will be flagged as DIF. Similarly, tests that are not fair will be flagged as DTF. The objects of this research were to apply differential item functioning and test functioning methods to analyze the extent of differential item functioning and test functioning in science assessment data. The data that were explored in this research was the secondary data from the Programme for International Student Assessment (PISA) in 2009 and was analyzed using Differential item functioning analysis system (DIFAS) version 5.0. It was found that mixed format tests used by PISA favored some groups of examinees over other groups which indicate the degree of unfairness across groups of students with different backgrounds. The recommendation for the assessment of student performance is that DIF items be removed before the score reporting is calculated and that the value added model be used to remove factors that are unfair to different groups of test takers.

Selection and peer-review under responsibility of the Organizing Committee of WCES 2014

Keywords: differential item functioning, differential test functioning, mixed format tests;

* Nuttaporn Lawthong. Tel.: +6-684-134-5933; fax: +6-622-182-801.
E-mail address: Nuttaporn.l@chula.ac.th

Selection and peer-review under responsibility of the Organizing Committee of WCES 2014
doi:10.1016/j.sbspro.2015.04.348

1. Introduction

The Programme for International Student Assessment (PISA) was established by the cooperation and expertise of OECD and non-OECD countries to create a common set of tests with capacity to indicate the quality of education in countries participating. The goal of PISA identify a quality index for member countries to determine the future rather than to determine only the quality of students at any particular level. The PISA assessment take place every three years in a literacy perspective which emphasizes on the knowledge and skills of students learning and practiced at school(OECD, 2012). According to the assessment results, Thai students participating in the program in 2009 were found to have scientific literacy scores below the OECD average. When the aforementioned assessment results are compared to current Thailand's target on scientific literacy, which focuses on cognitive thinking and creativity, we find a vital need to bring the assessment results under consideration for greater instructional management in scientific subject(OECD, 2012). Nevertheless, scientific literacy assessment is based on test scores alone. It is also necessary to consider the context of the fundamental cultures of students in each country. Furthermore, test fairness should be given top priority. Test quality is important for educational assessment and evaluation. Tests need to be composed of reliability and validity meeting set standards. Consideration should also be given to potential bias during test administration. Differential Item Functioning (DIF) is one statistical method for investigating item bias and is very useful for educational testing. DIF occurs when examinees from different groups (e.g., gender, demographic, socioeconomic status, wealth) with the same abilities have a probability of giving a specific item response differently(Zumbo, 1999; Millsap & Everson, 1993; Dodeen & Johnson, 2003; Kamata & Vaughn, 2004). Furthermore, when any item or test set is detected that have differently test function, the item is removed or revised and put back into the original test set. The goal is so the test set is equally fair for every group of test examinees. However, if it is found after differential test function of the entire test set that the test set that it is not suitable to be used, a method suitable for differential function the entire test set is therefore sought, and it will be useful as information for the judging of instrument effectiveness, in priority ranking or in selecting. For these reasons, the researcher is interested in studying DIF with dichotomously and polytomously scored items and the DTF from the scientific literacy test of the Programme for International Student Assessment (PISA) to present the analysis results on bias due to gender and wealth which indicate the degree of unfairness across groups of students with different backgrounds. The findings can be used in conjunction with scientific subject's administration in the context of Thailand.

2. Literature review

The review of documents and researches related to differential item functioning and differential test functioning can be summarized as follows:

2.1. Differential Item Functioning (DIF) Analysis

DIF detection is the study of test content validity and fairness (AERA, APA, NCME, 1999). A test set that is used to assess any particular field of ability must have content validity for students with differences and diversity. The group of examinees representing the main group in the population is called the reference group and the other group representing the sub-group in the population is called the focal group which is the group of interest for studying DIF. In each testing, the people taking the test in the sub-groups might have different characteristics (Swaminathan & Rogers, 1990; Gallagher & Kaufman, 2005). Furthermore, the variety of DIF detection have been developed for dichotomously and polytomously scored items (Millsap & Everson, 1993; Penfield & Algina, 2006). Some techniques are based on IRT (e.g., SIBTEST, Lord's χ^2 test), others do not use IRT, such as Mantel-Haenszel method.

2.2. Differential Test Functioning (DTF) Analysis

DTF detection can provide valuable data for the content validity process in a number of ways as follows: 1) All information related to DIF is provided by combining all test items in the test set; 2) DTF provides a potential impact index of classification under the differential item functioning results; and 3) Measuring the area where the DTF