A COMPARATIVE ANALYSIS OF THE MODEL OF A QUALITY ASSESSMENT OF MATHEMATICS SUBJECT: AN APPLICATION OF DIFFERENTIAL ITEM FUNCTIONING AND DIFFERENTIAL DISTRACTOR FUNCTIONING

Ruangdech Sirikit[1]

Shotiga Pasiphol[2]

Sirichai Kanjanawasee[3]

Abstract: This study aimed to compare the efficiency of model of quality assessment of mathematics subject when applying 4 models of quality assessment. The data used in this study was secondary data from Trends in International Mathematics and Science Study (TIMSS) of 2007. Samples were composed of students, mathematic teachers who taught student samples and executives of 150 educational institutions where student samples were studying as well as 5,412 students. The study focused on mathematics subject with 14 papers of knowledge evaluation test on mathematics. There were 3 steps of data analysis – 1) investigate differential item functioning and differential distractor functioning with DDFS program 1.0; 2) analyze value added in the model of quality assessment using Value-Added Model with Hierarchical Linear Modeling (HLM) and 2 levels of analysis; and 3) investigate results of quality assessment of mathematics subject in the institutions with 4 different models which were 1) Undetected DIF-DDF & Adjusted; 2) Detect DIF & Adjusted; 3) Detect DIF-DDF & Unadjusted; and 4) Detect DIF-DDF & Adjusted.

The results showed that the model of quality assessment of mathematics which had effect control in the level of students and educational institutions provided similar outcome of quality assessment of mathematics subject. The model 4 had the highest coefficient of determination (R2) at 52.04% followed by Model 1 and 2 which had the coefficient of determination (R2) at 51.96% and 51.86% respectively.

Keywords: Differential Item Functioning, Differential Distractor Functioning

Introduction

Educational quality and standards are something that all parties involved would like to see it happen to the educational management in educational institutions of Thailand as seen from the National Education Act 2542 (1999) (revised edition B.E. 2545 (2002)) in Section 4 defining the “Educational standards” that means specifications of educational characteristics, quality desired, and proficiency required of all educational institutions as to serve as means for equivalency for purposes of enhancement and monitoring, checking, evaluation, and quality assurance in the field of education. The stated meaning is to distinctly confirm the obvious intention of this National Education Act which would like to see all educational institutions throughout Thailand enabling to manage education with quality and in same standards.

Due to the fact that the concept of differential item functioning has been continuously developed, some of measurers have amplified the body of knowledge in order to have the most standard test. They generally give priority to the correct choices more than distractors but some other measurers believe that studying correct choices only may provide incomplete information. If there is additional study on differential item functioning in another dimension – analyzing choices which are distractors will help increase significant information and also make analysis results of differential item functioning more important resulting in the fairest test for testees.

The Value-added Model was the method which helped show the result on the information of educational management – how many institutions were able to create value-added in learning results by comparing actual scores or observed scores to predicted scores from factors of students’ background, community and society contexts, or existing achievements (Sirichai Kanjanawasi, 2007). Using the Value-added Model in the education was to compile statistic techniques from students’ scores to assess the effect size of educational institution or teacher (MaCaffrey, Lockwood, Koretz, & Hamilton, 2003).

The study aimed to extend the concept and method of analyzing data on the quality assessment in educational institutions by applying 2-level analysis of value-added – students and educational institutions as well as analysis techniques of differential item functioning and differential distractor functioning to increase the reliability and fairness of measurement results. Each component included 4 models in which had following details:

Model 1 “Undetected DIF-DDF & Adjusted” was the model of quality assessment of mathematics subject which did not eliminate items of differential functioning and analyzed value-added by controlling factors in the

[1] Ph.D. Candidate in Educational Measurement and Evaluation, Faculty of Education, Chulalongkorn University, Thailand

[2] Associate Professor, Ph.D., Department Educational Research and Psychology, Faculty of Education, Chulalongkorn University, Thailand

[3] Professor, Ph.D., Department Educational Research and Psychology, Faculty of Education, Chulalongkorn University, Thailand

Table 1: Number of Samples to Collect Data for Assessment

Samples	Number of Institutions	Number of Classes	Number of Teachers	Number of Students
Targeted Data	150	150	150	5,579
Actual Data	150	150	150	5,412

level of students and educational institutions (Adjusted).

Model 2 “Detect DIF & Adjusted” was the model of quality assessment of mathematics subject which eliminated items of differential functioning by analyzing both correct choices and distractors and analyzed value-added by controlling factors in the level of students and educational institutions (Adjusted).

Model 3 “Detect DIF-DDF & Unadjusted” was the model of quality assessment of mathematics subject which eliminated items of differential functioning by analyzing both correct choices and distractors but did not analyze value-added (Unadjusted).

Model 4 “Detect DIF-DDF & Adjusted” was the model of quality assessment of mathematics subjects which eliminated items of differential functioning by analyzing both correct choices and distractors and analyze value-added by controlling factors in the level of students and educational institutions (Adjusted).

Aims

To compare the efficiency of model of quality assessment of mathematics subject when applying 4 models of quality assessment

Definitions in the Research

Efficiency of model of quality assessment means coefficient of determination (R^2) from the model of quality assessment which is the multi-level analysis model with factor control in the level of students and educational institutions affecting the learning of students.

Methods

Study Samples

This study was conducted by using secondary data received from Trends in International Mathematics and Science Study (TIMSS) of 2007 which was a project organized to assess mathematics achievement of students in Matayom 2 (Grade 8).

Research team in Thailand which was the Institute for the Promotion of Teaching Science and Technology (IPST) delivered the data on number of educational institutions in Thailand to Statistics Canada for a random selection of samples to collect data for assessment. The samples were composed of students, mathematic teachers who taught student samples and executives of 150 educational institutions where student samples were studying. Once the data had already been collected, it was found that the number of student samples was decreased from targeted number but it was still possible to take the obtained data for result analysis according to the Table 1:

Tools for collecting data were:

1.1) Test – TIMSS 2007 test comprised 430 items of mathematics subject which were plentiful for students. Therefore, in managing students to complete every test within determined time frame of 1 hour and 30 minutes (45 minutes per each subject), it was divided into 14 clusters including multiple-choice item and constructed-response item. Creating test was originated by synthesizing contents and curricula from other participating countries with further details in appendix. Test in each Cluster had appropriate proportion of content and learning behavior according to the assessment of mathematics achievement in the project. In collecting data for assessment, there were 14 papers of test set each of which contained between 28-30 items.

Analysis of Data

This research entitled “Comparative Analysis of the Model of a Quality Assessment of Mathematics Subject: An Application of Differential Item Functioning and Differential Distractor Functioning” had 4 steps of analysis as follows:

Step 1 Scrutinizing information according to study factor

Step 2 Investigating the Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF) as well as estimating students’ competency

Step 3 Analyzing Value-added in the model of quality assessment

Step 4 Investigating results of quality assessment in educational institutions

Step 1 Scrutinizing information according to study factor

I managed to analyze basic statistic value of data with the analysis program SPSS 11.5 for windows as for analyzing fundamental data by means of descriptive statistics i.e. frequency, percentage, mean, standard deviation, skewness and kurtosis, and highest and lowest values.

Step 2 Investigating the Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF)

2.1 Investigating the Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF)

Investigating the Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF) by DDFS 1.0 program (Penfield, 2010), which was applied for analysis by means of Mantel-Haenszel Method together with odds ration indicated the formula as below:

$$\hat{\alpha}_{MH_i} = \frac{\sum_{s=1}^{s} R_{1s}F_{0s}/n_s}{\sum_{s=1}^{s} R_{0s}F_{1s}/n_s} \qquad (1)$$

And Mantel-Haenszel Method together with logarithm of odd ratio indicated the formula as below:

$$\hat{\lambda}_j = \ln(\hat{\alpha}_j) \qquad (2)$$

As the analysis of Mantel- Haenszel Method together with logarithm of odd ratio was to compare the response of Reference group with Focal group, interpreting results could be done as follows:

Interpreting results for the Differential Item Functioning (DIF)

$\ln(\hat{\alpha}_j) = 0$ means No differential item functioning found

$\ln(\hat{\alpha}_j) > 0$ means Reference group had more opportunity to give correct answer than Focal group

$\ln(\hat{\alpha}_j) < 0$ means Focal group had more opportunity

Interpreting results for the Differential Distractor Functioning (DDF)

$\ln(\hat{\alpha}_j) = 0$ means No differential distractor functioning found

$\ln(\hat{\alpha}_j) > 0$ means Reference group had more opportunity to choose answer than Focal group

$\ln(\hat{\alpha}_j) < 0$ means Focal group had more opportunity to choose answer than Reference group

2.2 Eliminating items after the analysis of Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF)

According to the results of Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF) in each paper, I chose to eliminate some items of differential functioning according to terms of models of education by considering the elimination based on 2 criteria: 1) select items of differential functioning with highest effect size, and 2) take into account the structure of test which would still stay the same after having been eliminated as well as the eliminated items should not be more than 20 percent (Clauser (1993), cited in Naraya & Swaminathan, 1994) of those in such test. The deletion of items of different functioning must be aware of the old structure by classifying them by proportion of contents in the test.

2.3 Estimating students' competency

To estimate competency of students from the assessment of their mathematics proficiency by Trends in International Mathematics and Science Study (TIMSS) 2007, I utilized MULTILOG program to analyze competency of testees and conducted the analysis for 3 times – firstly, to analyze the testees' competency before eliminating items of differential functioning; secondly, to analyze the testees' competency after deleting items of differential functioning and consider correct choices only; and thirdly, to analyze the testees' competency after eliminating items of differential functioning and consider their competency happening to both correct choices and distractors.

Step 3 Analyzing Value-added in the model of quality assessment

In analyzing value-added in the model of quality assessment, I applied Hierarchical Linear Modeling (HLM) with 2 levels of analysis for value-added in 4 models of quality assessment by means of HLM program.

Step 4 Investigating results of quality assessment in educational institutions

Competency in describing variance of variables with predictor variable or coefficient of determination (R^2) in each model had the following equation:

Variance of residual value reduced when with predictor variable

Variance of residual value reduced when without predictor variable

Explained variance of variables with predictor variable in each level of both students and educational institutions were able to be calculated from product between explained variance according to hypothetical model and all variances in such level received from fully unconditional model as follows:

Explained variance from the level of i = $\sigma_i^2(\exp lained) \times \sigma_i^2$

All explained variances by both levels of model were equal to the summation of explained variable from the level 1 and level 2.

Results

Results on comparing efficiency of model of quality *assessment of mathematics subject*

Results on comparing efficiency of 4 models of quality assessment of mathematics subject were concluded as follows:

The analysis results on comparing efficiency from 4 models of quality assessment showed that variance of result scores of mathematics evaluation between educational institutions in Model 3, the model which eliminated items of differential functioning with both correct choices and distractors but did not control the effect of predictor variable in the level of students and educational institutions. The variance between educational institutions (level 2) was at 0.29852 with statistical significance or as intraclass correlation was at 0.4475 (44.75%) meaning that there was variance of result scores of mathematics evaluation between educational institutions at 44.75% and when comparing to Model 4, which had the same characteristics as Model 3 but could control the effect of predictor variable in the level of students and educational institutions, it was indicated that the variance of result scores of mathematics evaluation was decreased and variance between educational institutions (level 2) was at 0.16232 with statistical significance or as intraclass correlation was at 0.3199 (31.99%) meaning there was variance of result scores of mathematics evaluation between educational institutions at 31.99% with coefficient of determination (R^2) in level 1 at 6.41%. The level 1 was able to explain variance of dependent variables at 6.41% of intraclass correlation in the level 1 (68.01%) while coefficient of determination (R^2) in the level 2 was at 45.63% meaning that the level 2 could explain variance of dependent variables at 45.63% of intraclass correlation in the level 2 (31.99%). The overall coefficient determination of Model 4 was at 0.5203 (52.03%) meaning that Model 4 (both levels) was able to explain variance at 52.03%.

When comparing models which had effect control in the level of students and educational institutions between Model 1, which did not eliminate any item of differential functioning; Model 2, which eliminated items of differential functioning with correct choices; and Model 4, which eliminated items of differential functioning with both correct choices and distractors, it was found that the variance between educational institutions in Model 1 was at 0.17316 or as intraclass correlation was at 0.3405 (34.05%). Model 2 had the variance between educational institutions at 0.16014 or as intraclass correlation at 0.3163 (31.63%) while Model 4 had the variance between educational institutions at 0.16232 or as intraclass correlation at 0.3199 (31.99%). Furthermore, when comparing coefficient determination between models that had effect control in the level of students and educational institutions, it was indicated that there was coefficient determination (R^2) at 0.51962 (51.96%) in Model 1, at 0.5186 (51.86%) in Model 2, and 0.5203 (52.03%) in Model 4 respectively as shown in Table 2 as follows:

Table 2: Summary of variance of explained dependent variables in Model 1 to 4

variance components	Model 1	Model 2	Model 3	Model 4
Level 1 – Variance within educational institutions(R)	0.33539	0.34614	0.36862	0.34501
Level 2 – Variance between educational institutions (U0)	0.17316	0.16014	0.29852	0.16232
Variance level of students' competency from mathematics proficiency evaluation at				
within educational institution (Level 1)	0.6595 (65.95%)	0.6837 (68.37%)	0.5525 (55.25%)	0.6801 (68.01%)
between educational institution (Level 2)	0.3405 (34.05%)	0.3163 (31.63%	0.4475 (44.75%)	0.3199 (31.99%)
Proportion of all variance of explained dependent variables (R^2)				
Level 1	0.07040 (7.04%)	0.06484 (6.48%)	NA	0.06405 (6.41%)
Level 2	0.44922 (44.92%)	0.45376 (45.38%)	NA	0.45625 (45.63%)
Both Levels	0.51962 (51.96%)	0.51860 (51.86%)	NA	0.52030 (52.03%)

According to the analysis results in Table 2, if we considered ability to explain variance of students' competency from mathematics proficiency evaluation, Models 1, 2 and 4 were model which was able to explain variance quite similarly and all 3 models eradicated resource of errors from the characteristics effect of students and educational institutions in order to be on the same basis. Therefore, it can be said that all 3 models provided accurate and fair assessment results but if we considered coefficient determination (R^2), it would be found that although all 3 models were similar but Model 4 had the highest coefficient determination (R^2) at 52.03% provided that this model not only was able to get away with the resource of errors from the characteristics effect of students and educational institutions but it could also take out errors resource from effect of items of differential functioning with both correct choices and distractors. Therefore, it can be said that Model 4 was the most efficient model tending to provide the most accurate and fairest results of assessment.

Discussion

The results came from study on different models of quality assessment of mathematics subject used in basic educational institutions with two study issues – test quality and value-added analysis model. 4 models of quality assessment of mathematics subject were concluded as below:

With the clear comparison on model of quality assessment of mathematics subject by indicating that models with effect control in the level of students and educational institutions would provide similar results of quality assessment considered from coefficient determination (R^2) in each model, it was revealed that Model 1, which had no analysis of differential item functioning and differential distractor functioning, was the model without control of error caused by bias on a certain group of testees with coefficient determination at 51.96% while Model 2, which had more procedures of analysis than Model 1 – differential items functioning, provided very coefficient determination to Model 1 at 51.86%. And Model 4, which had more procedures of analysis than Models 1 and 2 – differential distractor functioning, showed that despite similar coefficient determination to these two Models but to it had the highest coefficient determination at 52.03% which might explain that if items of differential functioning with correct choices and distractors were eliminated, it would mean there was a control of error resource caused by biased test for a certain group of testees resulting in increasing coefficient determination in the model of quality assessment. However, results on rating quality level and ranking quality of educational management in educational institutions from 3 models were all concordant and, therefore, it can be shown that the quality assessment by Models 1, 2 and 4 provided concordant results of quality assessment and either of them could be applied for assessment depending on readiness component of educational institutions.

The quality assessment by effective analysis of value-added must take into account all details of variables in the level of students and educational institutions as well as application of differential items functioning and differential distractor functioning will also increase reliability of obtained results of assessment.

References

Thai Language

IPST (2009). *Trends in International Mathematics Study.* Bangkok: Institute for the Promotion of Teaching Science and Technology.

Ittirith Phongpiyaratana. (2008). *An Item Analysis and An Investigation of Differential Item Functioning: A Multilevel Analysis.* Doctoral Thesis, Department of Educational Research and Psychology, Graduate School, Chulalongkorn University.

Prakrittiya Tuksino. (2009). A *Quality Assessment of Science Instructional Management in Basic Education Schools: An Application of Differential Item Functioning and Value-Added Model.* Doctoral Thesis, Department of Educational Research and Psychology, Graduate School, Chulalongkorn University.

Rakchanok Yeesunesri. *(2000). An Analysis of Differential Functioning of Items and Tests based on DFIT Procedures in English and Mathematics for University Entrance Examination.* Master Thesis, Department of Educational Research and Psychology, Graduate School, Chulalongkorn University.

Sirichai Kanjanawasi. (2007). *Multi-Analysis.* 4th Edition. Bangkok: Chulalongkorn University Press.

Sirichai Kanjanawasi. (2005). *Classical Test Theory.* 5th Edition. Bangkok: Chulalongkorn University Press.

Sirichai Kanjanawasi. (2007). *Modern Test Theories.* 3rd Edition. Bangkok: Chulalongkorn University Press.

English Language

Camilli G. and Shepard L. A. (1994). *Method for identifying biased test items.* California: SAGE.

Cheong, Y. F. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. *International Journal of Testing* 6(1): 57-79.

Green B. F., Crone C. R., and Folk V. G. (1989). A Method for studying differential distractor functioning. *Journal of Educational Measurement* 26(2): 147-160.

Penfield R. D. (2008). An Odds Ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement.* 45(3): 247–269.

Penfield, R. D. (2007). *DIFAS 4.0 Differential item functioning analysis system: User's manual.* Unpublished manuscript.

Penfield R. D. (2010). *DDFS 1.0, Differential Distractor Functioning Software User's Manual.* Unpublished manuscript.

TIMSS. (2007). *User guide for the International database.* Publisher: TIMSS & PIRLS International Study Center. Lynch School of Education, Boston College.