Journal of Positive School Psychology
2022, Vol. 6, No. 3, 8767–8773 http://journalppw.com

A Study of Item Bias In Mathematics Subject For Primary School Level Through International Test

Ruangdech Sirikit*[1], Panwasn Mahalawalert[2], Naowarat Chimpree[3]

[1,2]*Educational and Psychological Test Bureau, Srinakharinwirot University, Bangkok, Thailand*
[3]*Saint Gabriel's College, Bangkok, Thailand*

Abstract

The objective was to examine differential item functioning and differential distractor functioning in mathematics, primary school level through charactertistic classification analysis in gender and school size. Secondary data from the project studying on the trends in International Mathematics and Science study 2011 or TIMSS 2011 was used. The sample groups were students, mathematics teachers (who teach sample students) and school executives. The sample students were 4,448 students in grade 4 from 168 schools. Only mathematics subject results from mathematics evaluation (14 tests/primary level) were studied. Data analysis has two steps – 1) data management from studied factors and 2) analysis on differential item functioning and differential distractor functioning.

The research result showed that;

For gender aspect, there were 24 items on differential item functioning and differential distractor functioning from the total of 356 items. The highest topics were number, geometry and data, respectively. For school aspect, there were 54 items on differential item functioning and differential distractor functioning from the total of 356 items. The highest topics were number, geometry and data, respectively..

I. INTRODUCTION

The widely-used testing tool is multiple-choice item. It is used for many important items such as educational achievement evaluation in schools or university entrance examination. Multiple-choice item is used because it can be used for any items with a lot of applicants within limited time. Also, it can examine item takers' latent traits. Multiple-choice item is used for various situations. Hence, bias control is crucial for quality control.

The study on item bias began since 1960 though designing to find item development solutions to match differences in society including culture, ski mcolor, social status, gender, age etc. (Angoff, 1993). There have been a lot of claims against fairness for ages. These claims lead to sues on related organizations such as NAACP to stop or cancel item (Rudner, Getson and Knight cited in Hambleton, Swaminathan and Rogers, 1991). At present, differential item functioning measures are improved. New measures are devised and applied. Each measure has different conditions such as two-value item, multi-value item etc. In Thailand, differential item functioning does not occur by social differences. However, it is caused by other factors such as gender, age or experience etc.

Further studies on differential distractor functioning (DDF) are considered as concept extensions through examining from correct answers to distractors (Bank, 2009). The study advantages are dimensional increase in varied items and fairness promotion. Distractor efficiency for multiple choice item must be determined. According to theoretically significance, those who choose distractors are considered to choose the wrong answers leading to the perception of ineffectiveness (Sirichai

Kanchanawasee, 2005) – coorelated with Dorans, Schmitt and Bleistein (referenced in Penfield, 2008). The studies state item functioning analysis as the complete item must consist of differential item functioning analysis and differential distractor functioning analysis.

Considering the Basic Education Core Curriculum B.E. 2551 on Mathematics, it is found that mathematics promotes human's cognitive development and encourage students to gain creative thinking, cognitive and decision-making reasons, systematic and logical thinking processes. So, students can apply them to their daily lives when planning, decision-making and predicting are required (Ministry of Education, 2008). Besides, mathematics is a tool and a basis for further applications on other fields such as science, technology and social studies etc. So, mathematics is important for human survival and improves quality of living to match social change.

Hence, the researcher is interested in studying on mathematics through gender and school size classification using differential item functioning and differential distractor functioning. The result is another important dimension for students' learning development in mathematics for further better learning.

II. OBJECTIVE

To study differential item functioning and differential distractor functioning in mathematics, primary school level through charactertistic classification analysis in gender and school size.

Research Scope

1. Secondary data from database from the 2011 project studying on mathematics education management tendency or TIMSS, conducted by the cooperation between the International Association for the Evaluation of Educational Achievement (IEA) and the Institute for the Promotion of Teaching Science and Technology (IPST) was used with the objective to evaluate educational achievement in mathematics and science of grade 4 students in terms of subject contents and learning behaviors.

2. The item used was achievement item in the trends in International Mathematics and Science study 2011 or TIMSS 2011. This research examined only mathematics consisting of 14 multiple-choice items and constructed-response items.

III. DEFINITION

Differential Item Functioning (DIF) refers to the item characteristics which enable the students from different group with the same level of competency but different chance to give correct answer. but different chance to achieve item answering. This research used the analysis of differential item functioning investigated through bivariate student characteristics – gender and school size – was used.

Differential Distractor Functioning (DDF) refers to the condition when the members in reference group and focal group with same level of competency but different chance to choose distractors than reference group. Bivariate student characteristics – gender and school size – were examined.

Educational Achievement in Mathematics refers to mathematical item in the trends in International Mathematics and Science study 2011 or TIMSS 2011.

IV. METHODS

The methods of research were divided to 3 Section 1: Research Data, Section 2: Data Collection and Section 3: Data Analysis. The details were shown as follows;

Section 1 Research Data

The research used secondary data from the study on the trends in International Mathematics and Science study 2011 or TIMSS 2011 – a project on educational achievement in mathematics by studying grade 4 students as shown in the following details;

1) Sample Groups

The sample groups included students, mathematics teachers (who teach sample students) and school executives from the sample students' schools. The details were as in Table 1 and 2.

Table 1 The Sample Group Number on Data Collection For Evaluation

Sample Group	No. of Schools	No. of Teachers	No. of Students
Primary 4	168	168	4,448

Table 2 Number of Students and Percentage of Basic Information of the Sample Groups

Students' Basic Information	Number	Percentage
1. Gender		
1.1 Female	2,211	49.70
1.2 Male	2,237	50.30
2. School Size		
2.1 Small (Less than 600)	2,014	45.30
2.2 Large (600 and above)	2,434	54.70

Section 2 Data Collection

***2) Data Collection Tools** included;*

2.1) Test TIMSS 2011 Test included 356 items in mathematics. The test time was one and a halfhours (45 minutes per subject). Each subject test was divided into 14 clusters including multiple-choice items and constructed-response items. Test formulation was orginated by multi-country curriculum synthesis. Each cluster consisted of subject contents and learning behaviors for evaluation. There were 14 tests. Each contained 22-29 items. Systematic random was applied. Thus, the students whose number orders were close would never receive the same tests. Each students were to start the test simultaneously.

Table 3 Percentage of Evaluation Content, Grade 4

Content	Percentage
1. Number	50
2. Geometry and Measurement	30
3. Data	15

Table 4 Learning Topics for Each Evaluation Content, Grade 4

Content	Topic
Number	- Integer - Fraction and Decimal - Number Sentence - Pattern and Relation
Geometry and Measurement	- Point, Line and Angle - 2D and 3D Geometry
Data	- Analysis and Interpretation of Data - Data Collection and Presentation

The content topics for TIMSS 2011 covered all of the participant countries' grade 4 curricula. The details were shown in Table 4.

Section 3 Data Analysis

The research on differential item functioning and differential distractor functioning in mathematics through characteristic classification analysis was divided into two steps as follows;

Step 1 Organizing data according to the study factor

The data analysis of descriptive statistics was conducted with SPSS 19 analysis program for windows for analyzing fundamental data by means of descriptive statistics i.e. frequency, percentage.

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

To investigate the differential item functioning (DIF) and differential Distractor functioning (DDF) by DDFS 1.0 (Penfield, 2010), which was applied for analysis by Mantel-Haenszel indicated the formula as below:

Mantel-Haenszel common odds ratio for an item at score level j

$$\alpha_j = \frac{p_{Rj}}{q_{Rj}} \Big/ \frac{P_{Fj}}{q_{Fj}} = \frac{a_j d_j}{b_j c_j}$$

Where

p_{Rj} = number of persons in Reference group in score interval j who answered correctly

q_{Rj} = number of persons in Reference group in score interval j who answered incorrectly

Notation F relates to the focal group

Mantel-Haenszel common odds ratio for item i

- For the slice j $\alpha_j = \dfrac{a_j d_j}{b_j c_j}$
- Across all slice $\hat{\alpha}_{MH_i} = \dfrac{\sum_{s=1}^{s} a_j d_j / n_j}{\sum_j b_j c_j / n_j}$
- The logarithm of common log-odds

ratio is normally distributed and is used as effect size measure

$$\lambda_{MH} = \log(\hat{\alpha}_{MH})$$

As the analysis of Mantel-Haenszel common log-odds ratio was to compare the response of Reference group with Focal group, criteria for considering differential functioning and differential distractor functioning of items were obtained from LOR Z value and MH LOR value according to the process of interpreting as below:

1) Considering LOR Z value

1.1 LOR Z value > 2 or LOR Z value < -2 indicates that can consider evidence of presence of DIF or DDF and MH LOR value in the step 2 shall be considered.

1.2 -2 ≤ LOR Z ≤ 2 indicates no evidence of presence of DIF or DDF.

2) Considering MH LOR value for the question with DIF and DDF between groups

2.1 Positive (+)

For DIF **-** MH LOR value indicates that item favor reference group.

For DDF **-** MH LOR value indicates that the focal group has chance to choose distractor than reference group.

2.2 Negative (-) MH LOR value indicates that item favor focal groups.

For DIF **-** MH LOR value indicates that item favor focal groups.

For DDF **-** MH LOR value indicates that the reference group has chance to choose distractor than focal group.

V. RESEARCH CONCLUSION

Item Content; There were three item contents – number, geometry and data. The analysis result was shown in details as follows;

Gender: The result was classified by 14 tests. There were 24 items on differential item functioning/differential distractor functioning from the total of 356. The most common content was number – 11 items on differential item functioning/differential distractor functioning, accounted for 45.83 percent. The second one was geometry – 8 items on differential item functioning/differential distractor functioning, accounted for 33.33 percent. The least common one was data – 5 items on differential item functioning/differential distractor functioning, accounted for 20.83 percent.

According to differential item functioning (DIF), it was found that when both genders had equal competence in mathematics, females students gained more advantage and opportunity to choose correct answers over male students for geometry and data while males students gained more advantage and opportunity to choose correct answers over female students for number.

According to differential distractor functioning (DDF), it was found that when both genders had same level of competency in mathematics, males were more likely to choose distractors for number while females were more likely to choose distractors for data. For geometry, both genders had equal tendency to choose distractors.

School Size: The result was classified by 14 tests. There were 54 items on differential item functioning/differential distractor functioning from the total of 356. The most common content was number – 23 items on differential item functioning/differential distractor functioning, accounted for 45.59 percent. The second one was geometry – 21 items on differential item functioning/differential distractor functioning, accounted for 38.89 percent. The least common one was data – 10 items on differential item functioning/differential distractor functioning, accounted for 18.52 percent.

According to differential item functioning (DIF), it was found that when students from small and large schools had same level of competency in mathematics, students from small schools gained more advantage and opportunity to choose correct answers over students from large schools for number and geometry while students from large schools gained more advantage and opportunity to choose correct answers over students from small schools for data.

According to differential distractor functioning (DDF), it was found that when students from small and large schools had same level of competency in mathematics, students from large schools were more likely to choose distractors for number and geometry while students from both schools had equal tendency to choose distractors for data.

VI. RESEARCH DISUSSIONS

From the study on differential item functioning and differential distractor functioning in mathematics through characteristic classification analysis, the research discussed the overall of primary school level through two points as follows;

1. To study differential item functioning and differential distractor functioning in mathematics, primary school level through charactertistic classification analysis in gender and school size.

Differential item functioning and differential distractor functioning were examined by gender and school size. The analysis result was classified by 14 tests in mathematics, primary school level. The results showed that there were 24 items on differential item functioning and differential distractor functioning, accounted for 6.74 in term of gender. there were 54 items on differential item functioning and differential distractor functioning, accounted for 15.16 in term of school size. The results are consistent with Yisoonsri (2001) which stated that English item resulted in differential item functioning for gender and geographical site and mathematics item resulted in differential item functioning for gender and both subjects led to the most differential item functioning. Likewise, Sumalee Kleawtanong (2004) stated that the result revealed that the causes of differential item functioning in Thai language were content interest and item language in term of gender differential item functioning while the causes of differential item functioning in social studies, religion and culture were content interest and culture/tradition content in term of gender differential item functioning.

2. To study To study differential item functioning and differential distractor functioning in mathematics, primary school level through charactertistic classification analysis in gender and school size through item content.

Differential item functioning and differential distractor functioning were examined by gender and school size, classified by item content. The results showed that there were three item contents – number, geometry and data. In term of gender, DIF analysis result revealed that when both genders had equal competence in mathematics, female students gained more advantage and opportunity to choose right answers over male students for geometry and data while males students gained more advantage and opportunity to choose right answers over female students for number. DDF analysis result revealed that when both genders had equal competence in mathematics, males were more likely to choose distractors for number while females were more likely to choose distractors for data. For geometry, both genders had equal tendency to choose distractors. In term of school size, DIF analysis result revealed that when students from small and large schools had equal competence in mathematics, students from small schools gained more advantage and opportunity to choose right answers over students from large schools for number and geometry while students from large schools gained more advantage and opportunity to choose right answers over students from small schools for data. DDF analysis result revealed that when students from small and large schools had equal competence in mathematics, students from large schools were more likely to choose distractors for number and geometry while students from both schools had equal tendency to choose distractors for data. Items with different contents may lead to differential item functioning and differential distractor functioning, classified by gender and school size, with consistent with the research from Katherine and Meichu (1996) which analyzed differential item functioning for mathematics multiple-choice item considering gender as variable. The research result found that algebra and calculation were easier for females. Geometry was male-oriented and arithmetic was female-oriented.

VII. SUGGESTIONS

The research on differential item functioning and differential distractor functioning in mathematics through characteristic classification analysis provided a lot of results beneficial in many circumstances. The researcher presented the following suggestions.

1. Suggestion for Research Application

Item is a development tool for evaluation. A lot of basic educational institutions have been improving their items. Generally, level of difficulty and classification power are used to determine items. Differential item functioning and differential distractor functioning are another ways to create item reliability.

2. Suggestion for Future Works

Differential item functioning (DIF) and differential distractor functioning (DDF) are used for TIMSS 2011. Future works should do in-depth study on item formulation indicators leading to higher benefits.

BIBLIOGRAPHY

1. IPST (2009). Trends in International Mathematics Study. Bangkok: Institute for the Promotion of Teaching Science and Technology.
2. IPST (2011). Sample Examination for Assessment of Mathematics Learning Results. Bangkok: Institute for the Promotion of Teaching Science and Technology.
3. Ittirith Phongpiyaratana. (2008). An Item Analysis and An Investigation of Differential Item Functioning: A Multilevel Analysis. Doctoral Thesis, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University.
4. 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, Faculty of Education, Chulalongkorn University.
5. Ruangdech Sirikit (2011). Comparative Analysis of the Model of a Quality Assessment of Mathematics Subject: An Application of Differential Item Functioning and Differentia Distractor Functioning. Doctoral Thesis, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University.
6. Sirichai Kanjanawasi. (2005). Classical Test Theory. 5th Edition. Bangkok: Chulalongkorn University Press.
7. Sirichai Kanjanawasi. (2007). Modern Test Theories. 3rd Edition. Bangkok: Chulalongkorn University Press.
8.
9. Camilli G., and Shepard L. A. (1994). Method for identifying biased test items.
10. California: SAGE.
11. 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.
12. Chiu,M.M., Xihua,Z. (2008). Family and motivation effect on mathematics achievement :Analyses of students in 41 countries. Learning and Instruction 18: 321-336.
13. 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.
14. Hambleton K. R., Swaminathan H., and Rogers H J. (1991). Fundamentals

of item response theory. California: SAGE.
15. Koon S. (2010). A Comparison of methods for detecting differential distractor functioning. Doctoral dissertation, Florida State University.
16. Lord F. M. (1980). Applications of item response theory to practical testing problems. hillsdale. New Jersey: Lawrence Erlbaum Association.
17. Mellenbergh G. J. (1994). Generalized linear item response theory. Psychological Bullentin 115(2): 300-307.
18. Middleton K., and Laitusis C. C. (2007). Examining test items for differential distractor functioning among students with learning disabilities. Princeton, NJ.: Education Testing Service.
19. Monahan O. Patrick, and co. (2007). Odds Ratio, Delta, ETS classification, and standardization measures of DIF magnitude for binary logistic regression. Journal of Educational and Behavioral Statistics. 32(1): 92–109.
20. Narayanan P. and Swaminathan H. (1994). Performance of the Mantel-Haenzel and simultaneous Item bias proceders for detecting differential item functioning. Applied Psychological Measurement 18: 15-328.
21. Oshima, T. C., Raju, N. S., and Nanda, A. O. (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. Journal of Educational Measurement 43: 1-17.
22. 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.
23. Penfield, R. D. (2010). DDFS 1.0 Differential item functioning analysis system: User’s manual. Unpublished manuscript.
24. Potenza M. T., and Dorans N. J. (1995, March). DIF assessment for polytomously scored item: A framework for classification and evaluation. Applied Psychological Measurement 19(1): 23-37.
25. Raju N. S. (1990). Determining the significance of estimated signed and unsigned areas between two Item response functions. Applied Psychological Measurement. 19(4): 353-368.
26. Raju N. S., Fortmann-Johnson K. A., Kim W., Morris S. B., Nering M. L., and Oshima T.C. (2009). The item parameter replication method for detecting differential functioning in the polytomous DFIT framework. Applied Psychological Measurement 33: 133-147.
27. Shealy R., and Stout W. T. (1993). A Model-base standardization approach that separates true Bias/DIF as well as item Bias/DIF. Psychometrika 58(2): 159-194.
28. TIMSS. (2011). User guide for the International database. Publisher: TIMSS & PIRLS International Study Center. Lynch School of Education, Boston College.
29. Trautwein, Ludtke, Schnyder and Niggli. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. Journal of Educational Psychology 98: 438-456.