IJER Indonesian Journal of Educational Review
Available online at IJER Website: http://pps.unj.ac.id/journal/ijer
Indonesian Journal of Educational Review, Vol. 1 (1), 2014, 33-38

COMPARING THE USE OF FIVE-PACKAGE OF MATH IN THE 2011 NATIONAL EXAMINATION BETWEEN SMA AND MADRASAH ALIYAH: A CASE STUDY IN JAKARTA

Wardani Rahayu
Universitas Negeri Jakarta
wardani9164@yahoo.com

ABSTRACT

This study aims to compare five question packages used in the 2011 National Examination, specifically for Maths for Social Studies subject between public senior high school (SMA) and Islamic senior high school (MA) in Jakarta. It employed simple random sampling as its data collection technique. The independent variables used were the category of high schools and the Five Question Packages used in the National Examination. The schools were divided into two: regular senior high school (SMA) and Islamic senior high school or Madrasah Aliyah (MA). The packages chosen to be analyzed were package number: 12, 25, 39, 46 and 54. The data was analyzed by two-factor ANOVA for its 11 Differential Item Function (DIF) and 29 non Differential Item Function (non-DIF) items. The items used for this study are the items that fit the three-parameter logistic model. This study found that no difference between the scores on both the DIF and non-DIF items in either for SMA or Ma is found.

Keywords*: national examination in math, five packages, DIF, SMA, MA.*

The Indonesian government began implementing national examinations for junior and senior high schools in 1969. The name for national examination has been changed for several times. Between 1982-2002 it was named as the National Final Learning Evaluation (EBTANAS). Between 2003-2004 it was named as the National Final Exam (UNAS) and between 2005-2013 as the National Examination (UN). The national examination aim to measure the students competency in certain subjects. It is conducted simultaneously at the end of the final year of junior and senior high schools in order to assess the achievement of national education standards. The implementation is expected to improve the quality of education in Indonesia.

The Ministry of National Education and Culture is in charge in ensuring the quality the national examination and improving the implementation. One of the improvements made was the use of the Five Question Packages in the 2011 National Examination. The packages were implemented to improve the security measures against cheating for it is found difficult to detect cheating during the examination. The Five Question Packages are comprised of the same questions with different order. They were developed based on the Graduate Competency Standards (SKL) compiled by the National Education Standard Agency (BSNP).

Each package was numbered. As for packages of math for social science, the packages numbers are 12, 25, 39, 46 and 54. Each test package was distributed to the students randomly by the proctors in each school. For example, for a class consisted of 20 students, there would be four proctors responsible in invigilating students who do the same question package. This makes students difficult in trying to cheat as the packages were randomly distributed.

The most frequent question raised in the 2011 National Examination particularly by schools and parents was whether the five question packages distributed would produce similar scores of participants with equal ability. The items on the Five Question Packages are the same but they are put in different order. This means that the questions will not be ordered based on the difficulty levels. For example, one package may have easy-level question in the beginning while other packages may have difficult-level question. This may affect the psychological state of participants when solving the questions and affect their final score.

Beside the problem above, the implementations of the national examinations also met other problems. The diversity of ethnicity, religion and culture in Indonesia becomes a challenge as each region has different socio-cultural condition in conducting the national examination. Schools in the urban areas are much more advanced than those in rural areas. The students in the urban areas could also meet the eight education standards compared to those who live in the rural areas. Due to this, I argue that the focus of education in the urban areas is in motivating the students to achieve better in schools. As for the less advanced regions, the primary focus is still in encouraging students to go to school. These issues should be put in consideration in preparing for qualified testing instrument.

The items in the testing instrument should undergo proper validation process. Validation itself is the most important aspect in determining the quality of the measurements in the fields of psychology and education (ETS, 2002, p.?). In the validation process, however, if the testing instruments applied to test participants with equal ability which produced different results, the instruments will be considered as bias. This bias is called as Differential Item Functioning (DIF). Naga (1992, p.?) stated that DIF reflects the purpose of bias detection method. This is conducted by identifying items which have different functions for different groups of participants. In alignment with this, Hidalgo-Montesinos and Lopez Pina (2002, p.?) added that Mantel-Haenszel method can be applied as non-parametric statistical approach for detecting DIF after matching the ability of certain participants. The Mantel-Haenszel procedure compares the odds for success between groups after assessing the participants' abilities.

Studies on national examinations and DIF have been carried out by Kartowagiran (2005), Effendi (2011) and Sudaryono (2012). They compared several methods of DIF detection. Another research conducted by Rahayu (2010) examined the accuracy of linking methods on DIF detection based on the number of false positive items and found....????. In relation to the research above, this study is different as it is aimed at identifying DIF and non-DIF items and seeing the difference in scores between SMA and MA in Jakarta.

METHOD

The method applied was the method of ex post facto. The 2011 Examination Scores was applied as the dependent variable, while the independent variables were the question numbers 12, 25, 39, 46, and 54, DIF and non-DIF items, and the school categories (SMA and MA).

The design used in this study was a 5 x 2 x 2 factorial design. The data used were the score results from the participants of the Maths for Social Science in the

2011 National Examination in Jakarta region. The sample was determined through random sampling technique, while the data analysis technique used was the two-way ANOVA.

Table 1 Number of Test Participant Responses as Samples

Type of Schools	National Examination Package Number (A)									
	12		25		39		46		54	
	DIF	Non DIF	DIF	Non DIF	DIF	Non DIF	DIF	Non DIF	DIF	Non DIF
SMA	300		300		300		300		300	
MA	300		300		300		125		125	

This study was limited to the National Examination scores derived from the items which were model-fit to the three-parameter logistic model and DIF detection deploying the Mantel-Haenszel method. The study undertook several procedural steps. Firstly, the study began with the retrieval of data, i.e. the score results of the Maths for Social Science subject (Jakarta) in the 2011 National Examination from the Ministry of National Education's Center of Assessment and Testing. The data collected were in binary format, zero-and-one-shaped score of 40 items. The testing instrument are the Five Question Packages. The second step of the procedure includes selecting the 40 items that were model-fit to the three-parameter logistic model (L3P). Thirdly, the DIF items would be classified based on gender as in the number 12, the total number of male and female participants were 3100. Fourthly, the number of responses would be randomly selected in both SMA and MA groups. Lastly, the responses from the test participants would be summed up and compared to different groups of students. They would be transformed using T-Score formula: $T = 10\ z + 50$.

RESULTS AND DISCUSSIONS

By looking at the model-fit to the parameter logistic model on package 12, it was found that 12 questions were not model-fit while 28 questions were model-fit. The questions that were model-fit were questions numbers 1, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 27, 29, 31, 32, 33, 34, 35, 37, 38, 39, and 40. The DIF detection is conducted by using the Mantel Haenszel method.

The average scores for both SMA and MA groups for all five packages for the Maths for Social Science subject were almost the same, ranging from 49.95 to 50.02. This is similar to the score distribution of the 2011 National Examination in which the same homogeneity in the group of SMA and MA for the five-packages took place.

Results of testing with a two-way ANOVA found that there was no difference on the exam scores on SMA and MA groups, no difference on all the Five Question Packages, no difference on the scores of DIF items, non-DIF items. It was also found that no correlation between the packages, the school categories to the exam scores; between the packages, DIF items, non-DIF items and the whole items to the exam scores; between the packages, the school category, DIF items, non-DIF items and the whole items to the exam scores existed.

The inferential test results of the average value of the exam scores in SMA and MA groups for all Five Question Packages found that all are quite similar, nearing to the score of 50. The score distribution homogeneity was quite similar.

The non-DIF items were identified as those with difficulty level between 1.194 to 1.199. According to Hambleton and Swaminathan (1990, p.?), test items are considered easy when its difficulty level value of bj is approaching -2, while test items are considered difficult when the difficulty level of bj is approaching 2. Out of 28 questions analyzed, 26 questions were model-fit, i.e. approaching the value of -2, while the other two questions is approaching the value of 2. Therefore, it can be concluded that 26 questions were easy and 2 questions were difficult. Question numbers 35 and 40 were the most difficult test items. They belong to non-DIF item. The DIF item that was identified to be the easiest one was question number 1, while the easiest non-DIF item was question number 20.

Table 2 Mean and Standard Deviation Based on Type of UN Package and Type of School

Item	School	Package 12		Package 25		Package 39		Package 46		Package 54	
		Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
Non DIF	SMA	50.02	9.99	50.04	10.00	49.96	10.02	50.02	10.02	49.99	9.99
	MA	50.01	9.95	49.96	9.99	49.99	10.01	49.87	9.96	50.02	10.01
DIF	SMA	50.00	10.02	49.99	9.99	50.01	10.01	50.00	10.00	50.00	10.00
	MA	50.00	10.02	50.00	10.00	50.01	10.00	50.00	10.00	50.00	10.00
Total	SMA	49.96	10.02	49.95	10.01	50.00	10.00	49.99	9.98	49.99	9.99
	MA	50.00	10.02	50.00	9.99	50.01	10.00	50.00	10.00	50.01	9.98

Table 3 Analysis Results of the Two-Way ANOVA

Source	**Df**	**Mean Square**	**F**	**Sig.**
Corrected Model	29	.175	.002	1.000
Intercept	1	1.984E7	1.984E5	.000
Package	4	.024	.000	1.000
School	1	.015	.000	.990
Bias	2	.126	.001	.999
Package * School	4	.153	.002	1.000
Package * Item Bias	8	.116	.001	1.000
School * Item Bias	2	.573	.006	.994
Package * School *Item Bias	8	.253	.003	1.000

Items that were categorized as difficult are the ones related to probability. Question number 35 was about the expected frequency of three coins tossed together 600 times to produce at least two heads and two tails. Question number 40 was about the probability for random distribution. Probability is considered as difficult topic for Marths for Social Science in both SMA and MA as it requires more time to learning it outside the classroom.

The overall information function of the National Examination for both SMA and MA is at 7.3092 with test participants' ability −1.−Each item information function is between 0.0142 and 0.6012. Question number 20 had the biggest item information function at 0.6012, while question number 4 has the smallest information function. Item information function refers to the power of an item in the testing instrument. It is critical in determining the items selected for exams. The higher the information function, the better the test is. Item analysis results showed that there were 28 good items which were model-fit to the three-parameter logistic model (L3P) in the Maths for Social Science subject of the 2011 National

Examiantion. The high value of information function found in the examination indicates that the item validation process of the national examinations has been well-implemented. However, more studies on item bias is required for national examinations, particularly those related to gender, region, ethnicity throughout Indonesia.

Further studies are recommended, especially in the form of comparative studies. It is recommended to compare the scores of other regions, such as the western and eastern parts of Indonesia in order to obtain a comprehensive look in the difficulties in assessing students' competence in Maths. Future studies can be conducted to detect items with non-parametric approach without being restricted by items model-fit in IRT models.

CONCLUSION

There was no difference between in the results of Five Question Packages in Maths for Social Science subject of the 2011 National Examination in SMA and MA in Jakarta. No difference is also found on the DIF and non-DIF items in the packages. Most of the questions were model-fit to the three-parameter logistic model -- easy and difficult levels. This due to the random sequences in the Five Question Packages. However, the bias on the validity of national examination must continuously be kept at minimum considering the ethnic, religious and cultural diversity in Indonesia.

REFERENCES

Camili, G., and Lorrie, S. (1994). *Methods for Identifying Biased Test Items*. London: Sage Publications.

Clauser, B., Mazor, K., Hambleton, R. K. (1993). *The Effects of Purification of Matching Criterion on The identification of DIF Using the Mantel-Haenszel Procedure*. Applied Measurement in Education, Vol. 6 (4), 269-279.

Effendi (2011). *Detecting Crossing Differential Item Function (CDIF): Based on Item Response Theory*. Jurnal Evaluasi Pendidikan Vol. 2(2).

Hambleton, R. K. and Swaminathan, H. (1990). *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff Publishing.

Hidalgo-Montesinos, M. D. & Lopez-Pina, J. A. (2002). *Two-Stage Equating In Differential Item Functioning Detection Under The Graded Response Model With The Raju Area Measures And The Lord Statistic*. Educational and Psychological Measurement, Vol. 62(1), 32-44.

Kartowagiran, B. (2005). *Perbandingan Berbagai Metode Untuk Mendeteksi Bias Butir*. Unpublished dissertation. Yogyakarta: Universitas Gadjah Mada.

Naga, D. S. (1992). *Pengantar Teori Sekor Pada Pengukuran*. Jakarta: Besbtas.

Rahayu, W. & Ridho, A. (2008). *Analisis Data Indonesian National Assessment Program (INAP) Tahun 2007*. Jakarta: Pusat Penilaian Pendidikan Badan Penelitian dan Pengembangan DIKNAS.

Rahayu, W. (2010). *Linking Method and Flase Positive Item on DIF Detection Based on Item Respons Theory*. Jurnal Penelitian dan Evaluasi Pendidikan, Vol. 14(1).

Republika (2011). *Lima Paket Soal UN Akan Tekan Kecurangan*. Published 21 February 2011 by Republika. Accessed on 17 April 2013 from http://www.republika.co.id/berita/pendidikan/berita/11/02/21/165282-lima-paket-soal-un-akan-tekan-kecurangan.

Suara Merdeka (2011). *Lima Paket Soal UN untuk Cegah Kecurangan*. Published 7 April 2011 by Suara Merdeka. Accessed on 17 April 2013 from http://m.suaramerdeka.com/index.php/read/news/2011/04/07/82358.

Sudaryono, S. (2012). *Sensitivity of Differential Item Functioning (DIF) Detection Method*. Jurnal Evaluasi Pendidikan Vol. 3(1).