

Daya Matematis : Jurnal Inovasi Pendidikan Matematika

Volume, 9 Nomor 2 July 2021 Hal. 155 - 163 p-ISSN: 2541-4232 dan e-ISSN: 2354-7146

# Assessment Of The Difficulty Of Mathematics High-Level Reasoning Using Focus Group Discussion And Program Item Approaches

## Rukli<sup>1</sup>, Ma'rup<sup>2</sup>

<sup>1</sup>Pendidikan Matematika Universitas Muhammadiyah Makassar Email: <u>rukli@unismuh.ac.id.</u>, ma'rup@unismuh.ac.id

(Received:27-03-2021; Reviewed: 2-05-2021; Revised: 11-05-2021; Accepted: 20-05-2021; Published: 13-07-2021)

© 2021–Daya matematis: Jurnal inovasi pendidikan matematika. This article open acces by licenci CC BY-NC-4.0 (https://creativecommons.org/licenses/by-nc/4.0/)

#### Abstract

This study aims to estimate the difficulty level of high-level reasoning math problems by comparing the estimated difficulty level of the Focus Group Discussion (FGD) approach and the Iteman program. The estimation of the FGD approach uses a semantic differential scale on a scale of 1-7, while the Iteman program uses version 4.0. The study used a quantitative-comparative approach involving 79 FGDs of students and teachers of SMP/MTs in Soppeng Regency. Each FGD consisted of four people, one teacher and three students of class VIII. The comparative of difficulty level using Scheffe test with a significance level of 0.05. The results showed that the average level of difficulty in the FGD approach had significant similarity with the average level of difficulty in the output of the Iteman program approach. This means that the estimated level of difficulty of the FGD examinees has a similarity with the output of the Iteman program from the response data of the FGD examinees. This shows that examinees from FGD in assessing the level of difficulty of high-level reasoning questions can be the teacher's choice in schools other than the Iteman program.

**Keywords**: Difficulty Level; High-Level Reasoning Questions, Focus Group Discussion, iteman program

#### INTRODUCTION

Multiple-choice questions have several characteristics according to the scope of the test theory approach. In the classical test theory approach, multiple-choice questions have three characteristic names: the Difficulty Level (DL), the discriminatory power, and distractors. This is different from modern test theory in terms of assumptions and implementation (Raykov, Dimitrov, Marcoulides, & Harrison, 2019) and (Raykov & Marcoulides, 2016). These three characteristics determine good character, not a question to be tested on students. However, due to the continuous line linking the ability level of the group of examinees, the DL of the questions has special priority (Zubairi, 2006), (Jabrayilov, Emons, & Sijtsma, 2016).

The DL of the questions is related to the continuum of the ability of the examinee group. In addition, the DL of the questions empirically is the easiest to determine. However, sometimes teachers do not do this, so the questions are sometimes not suitable for measuring tools. This task is the teacher's obligation as part of the learning process. However, for teachers, estimating the DL of the questions will add to the burden so that there needs to be another way without taking time, energy and thought, and does not need trials and other analyzes. The results of observations and interviews with several teachers at the school showed that the multiple-choice questions did not test the questions. There are several reasons: feeling lazy after working all day and not understanding how to determine it. There is another way: the teacher can use a computer program, but it adds to the teacher's burden to calculate

the DL of the questions. As a result, the calculation of the difficulty level of the questions was never carried out for the purpose of daily exams, midterm exams, and final class exams.

The calculation of the DL of the questions has been using computer programs, such as the Iteman program. The program uses a classical test theory approach (Martinková et al., 2017) and (Juškaite, 2018). Although this program is easy to apply at the class level, it still requires additional understanding of several other applications when entering data into the program, namely the excel, word, and notepad programs to transform txt type data in order to use the program.

Likewise, the Iteman program requires certain specifications in terms of the computer operating system. This requires additional software to be compatible. Therefore, it is necessary to have another simple approach for teachers in the classroom but can be a solution to determine the DL of the questions.

An adjustment approach is an approach to solving a particular problem with minimal requirements (Stone, Glass, Munn, Tugwell, & Doi, 2020). With the help of a semantic differential scale or other scales, students and teachers can determine the level of difficulty. This method can use a Focus Group Discussion (FGD) approach. FGD is a way that can be done to solve a problem in a short but accurate time (Richard A. Krueger; Mary Anne Casey, 2014) and (Lafferty, 2004). It was further revealed that FGDs were efficient when using smaller groups (Lafferty, 2004) and (Guest, Namey, & McKenna, 2017). The results of the study using consensus theory by involving experts obtained that the characteristics of the questions were not different from the results of computer program analysis (Kozierkiewicz-Hetmańska & Poniatowski, 2014). Furthermore, if using the teacher\_siswa group to estimate the characteristics of the questions with small sample size, it shows no difference with the output of the Iteman program (Rukli, Ma'rup, Bahar, & Ramdani, 2021). Therefore, is there a significant difference in the characteristics of the DL of the HLRQ from the estimated results of the FGD and the Iteman program output?

#### Question Difficulty Level

The characteristics of the DL of the questions refer to the continuous line of the ability. The DL of the question is one of the characteristics of a good question guide. There are several other things, such as discriminatory, distracting, validity, and test reliability, including scoring (Applegate, Sutherland, Becker, & Luo, 2019a). Distinguishing power refers to how the question can distinguish groups of examinees with high and low abilities. Distinguishing power is a combination of the proportions of the upper and lower groups of examinees. Ability is not considered here, which refers to the term differential power.

Distractors refer to multiple-choice questions, sometimes two, three, or more choices. Multiple-choice tests open up for examinees to guess if certain conditions occur. For example, the difficulty level of the questions is higher than the examinees' ability at that time. The distractor has uniformity of choice or is homogeneous (Applegate et al., 2019a). It is intended that low-ability examinees choose this option. Likewise, the validity and reliability of the test are quite important, but the DL of the questions is most often complained of by students, teachers, and even parents.

These complaints can be understood theoretically and empirically (experience), although sometimes comments appear about not being difficult at the beginning of the test and at the end of the test being difficult. The proportion of the number of students who answered the questions correctly compared to the number of examinees refers to the term DL of the questions. The formula for the DL of the questions is as follows.

 $DL = \frac{\text{Number of Examinee Who Answered the Test of True}}{\text{Number of Examinee}}$ 

The above equation certainly guarantees the teacher to determine the proportion of questions. The proportion of not difficult and extremely difficult questions can be refined into very difficult, difficult, moderate, easy, and very easy questions. Kindergarten balance in a fair test, although it is difficult to touch the existing substance. The practice is often done because the giving of test questions during the exam uses a group approach so that assembling the test requires art more than science (Applegate, Sutherland, Becker, & Luo, 2019b).

## High-Level Reasoning Questions

High-Level Reasoning Questions (HLRQ) are adapted from Trends International Mathematics and Science Study (TIMSS) questions. TIMSS is an international level competition to see trends in math and science abilities. This study was initiated by the International Association for the Evaluation of Education Achievement (IEA) based at Lynch School of Education, Boston College, USA. One of the activities is to test the math and science skills of the fourth grade VIII students.

The material exam uses two forms of questions, namely multiple-choice and description, using four answer choices for class IV and five answer choices for class VIII. For the description questions, there are brief descriptions and graded descriptions. The problem is based on the framework for assessing the ability of the mathematical field that is tested in terms of dimensions and domains.

Assessment framework the assessment is divided into two dimensions, namely the content dimension and the cognitive dimension. The content dimension assessment for class VIII consists of the following four domains. First, the number 30% contains the topics of numbers, fractions, decimals, ratio proportions, and percents. Second, 30% of algebra contain the topics of patterns, algebraic expressions, equations and functions. Third, 20% of geometry contain the topic of geometric shapes, geometric measurements, location and motion. Fourth, 20% of data and opportunities contain the topic of data organization and representation, interpretation and opportunity. All of these domains are associated with the assessment of cognitive dimensions. The results of the TIMSS Indonesia are ranked below the average of other participating countries(Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, 2016). The results of Indonesia's achievements in mathematics at the 2015 TIMSS are not much different from the results of the 2011 TIMSS(Martin, Mullis, Foy, & Stanco, 2012).

Several factors can cause this low TIMSS result. One of the contributing factors, among others, is that students in Indonesia are not trained in solving contextual questions, demanding reasoning, argumentation, and creativity to solve these questions where these questions are characteristic of TIMSS questions. Several studies show that the percentage of mathematics learning time in Indonesia is mostly used to discuss or discuss low-complexity questions, which is 57% and to discuss high-complexity questions, it only takes about 3% less time, while model questions TIMSS includes questions that have medium and high complexity, and require reasoning to solve.First, the multiple-choice type favoured Indonesian students than about the type of construct the response to all content domain and cognitive domain. Second, cognitive domain knowledge is preferred over an application. It is preferable to the second reasoning type of item. Third, algebra is least controlled by students, especially about the type of construct responses (Rukli, 2012).

From these results, it can be concluded that Indonesian students are not used to working on the TIMSS model questions. For this reason, it is very important to multiply the TIMSS model questions that contain mathematical reasoning in learning. In this case, it is important to socialize to teachers about what and how the characteristics of the TIMSS model questions are to be implemented in the learning process in the classroom in the form of direct participation, for example, estimating the characteristics of TIMSS questions because by estimating students or teachers working on these questions directly and then deepening to carry out in-depth and independent assessments.

### Focus Group Discussion

Focus Group Discussion (FGD) relates to the method of collecting data from a study. The use of FGDs is to obtain data/information from various sources and relevant preferences in a discussion group. This approach explores a specific issue/phenomenon of discussion of a group of individuals involved in making joint decisions. These activities interact and argue, assess, justify, comment, and argue on the basis of knowledge and experience.

Researchers widely use the FGD approach to explore a range of life experience phenomena throughout the human life cycle through their social interactions in groups. The main objective is to obtain data interaction from a discussion of a group of examinees /respondents and to increase the depth of information to reveal various aspects of a life phenomenon. The phenomenon can be defined and explained(Monique M. Hennink, 2012). The data from the group discussion's interaction results can focus on or emphasize the similarities and differences of experience and provide solid information/data about a perspective. This resulted from the results of the group discussion.

FGDs have the following characteristics. First, FGD is a data collection method for this type of qualitative research where the data generated comes from the exploration of social interactions that occur during the discussion process carried out by the informants involved. Second, the implementation of FGD activities is carried out objectively and externally. Third, FGDs require trained and reliable facilitators/moderators to facilitate discussions so that interactions that occur between examinees are focused on problem-solving (Monique M. Hennink, 2012). Fourth, FGDs used semi-structured interviews with a group of individuals with a moderator who led the discussion in an informal setting and aimed at collecting data or information on a particular topic. Fifth, FGDs have the characteristics of a number of individuals that are quite varied for one discussion group. One discussion group may consist of 4 to 8 individuals (Kitzinger, 1996; Twin, 1998) or 6 to 10 individuals (Howard, Hubelbank, & Moore, 1999).

## METHODS

## **Types of research**

This study uses a quantitative-comparative type of research by comparing the estimated DL of the questions using the FGD approach compared to the output of the DL using the Iteman program.

#### Population and Research Sample

The population of this research is SMP and MTs in Soppeng Regency, South Sulawesi Province. Sampling was done randomly, namely SMP and MTs in Soppeng Regency. The results of randomization selected several SMP and MTs in the study. The research sample was 79 FGDs spread across several SMP and MTs in Soppeng Regency. Each FGD consists of 4 members, where one of them is a mathematics teacher while the other is a grade VIII student.

#### Data analysis

The analysis of the data processed by the Iteman program and the results of the FGD used descriptive and inferential statistics, namely the percentage, mean, median, mode, standard deviation, slope, kurtosis, and the test with the Schefe test approach with a significance level of 0.05. Data processing using SPSS and Excel programs.

#### **Research Flow Chart**

The flow chart follows the following steps. Tracking questions from the TIMSS website, Adapting questions, analyzing qualitatively, pre-testing questions, determining the DL of questions, analyzing questions qualitatively, measuring the impact of student involvement in the form of post-tests.



Figure 2. Research Flowchart

## **Research Instruments**

The study used two instruments, namely 1) a semantic differential scale using a scale of 1-7, where 1 question was the easiest and seven questions were the most difficult. The instrument is intended to measure the level of difficulty of each HLRQ, and 2) a multiple-choice test, namely the TIMSS adaptation question. Some of these questions were then analyzed qualitatively by involving experts and teachers. The results of the qualitative analysis were then cloned as many as 40 questions to be assembled into a test.

#### **Iteman Program**

The Iteman program is an analysis of items with a classical test theory approach. The program has undergone several revisions to date. The version used is the DOS version, but because it is no longer compatible with Windows 10, it requires a program DOSBox0.74-win32-installer.exe. The stages of analyzing the examinee's response data using the Iteman Program are as follows. The initial stages of creating a data file are as follows. The first line is the control line to describe data about the number of questions, answer types, and the number of examinees. The second line contains a list of answer keys for each item. The third line contains the number of options for each item where the number corresponds to the number of options for each question. The fourth line contains the status of the question. The fifth line and so on contain the examinee's data and the student's answer choices without spaces.

The output of the Iteman program is divided into two, namely the main characteristics of each question, namely item statistics and alternative statistics. In addition to these statistical values, there are other Iteman program outputs on a secondary basis, including mean, variance, standard deviation, skew, kurtosis, alpha reliability coefficient, standard measurement error, and scale intercorrelation are correlation indexes between examinees scores obtained from each subtest/subscale (Guyer, R., & Thompson, 2006).

## **RESULTS AND DISCUSSION**

### Result

1. Centralized Data on Student\_Guru Estimates and FGDs, as well as the Iteman Program

Table 1 shows the centralized values of the mean, median, mode of groups estimated for the DL of the FGD and the DL of the output of the Iteman program. The statistical values of the two groups each have a difference of .0641 for the mean, a difference of .0731 for the median, and 0.188 for the mode. Furthermore, the difference between the two statistical values of the mean, and group mode is small enough.

Table 1. Descriptive Statistical Values DL of Problem

Statistics						
ECD	Iteman					
гоD	Program					
.1467	.2108					
.1561	.2292					
.023	.211					
	Statistics   FGD   .1467   .1561   .023					

## 2. Scattered Data Estimated FGD and Iteman Program

From the research shows the data distribution for the two groups of data where the difference in the standard error of the mean is .016, the standard deviation is .111, while the variance is .032 with a difference in the data range.184. These data indicate that there is an opportunity that the characteristics of the data spread are not too far in the range of the data.

## 3. Comparison of the Estimated Results of FGDs and the Iteman Program

Table 2 shows that sig.=.992 is greater than 0.05. This means that the interpretation of the FGD group is not significantly different from the results of the Iteman program output from the FGD group.

Table 3. Comparison of Data on the DL of Questions

Multiple Comparison							
Difficult	y level						
Scheffe							
	(1)				95% Cor	nfidence	
(I) E	(J) E	Maaa	C 4 1		Interval		
Г	F	Mean	Sta.		Lower	Upper	
а	а	Diff	E		B	R	
с	с	ere	rr	Sig.	D	D	
t	t	nce	0	-	0	0	
0	0		r		u	u	
0	0	(15)	1		n	n	
r	r				d	d	
1	2	1376	.02841	.992	22793	.19452	
*. The mean difference is significant at the 0.05 level.							

## Discussion

The results of previous studies showed that the data were not normally distributed, so that the Scheffe test was used. The Scheffe test results show no significant difference between the estimated DL of the HLRQ and the output of the processed response data of the examinees using the Iteman program. The results of this study further refine the description of the DL of the questions. The teacher uses the characteristics of the level of difficulty of the questions to be limited to easy questions, medium questions, and difficult questions.

Referring to the category of DL, the normal curve is related. The medium category questions are at most about 40%, while the bottom is 30% and the top is 30%. If it is detailed again, of course, the proportion will be smaller so that it looks smoother in the range of examinees' abilities, which are essentially partitions, not continuous.

If they follow such proportions in practice at school, it will be easy to simulate proportions. Comparisons between easy, medium, and difficult questions can be made in a 30%-40%-30% balance pattern. That is, the test has 30% questions in the easy category, 40% questions in the medium category, and another 30% questions in the difficult category or other, more subtle patterns, for example, 1%-2%-4%-2%-1%, and so on. This method can be used to make questions in class in stages of DL. Although there are still weaknesses, it is quite helpful in interpreting. However, this method is still shallow in relation to the characteristics of the examinees' abilities which are essentially continuous in the range of real numbers.

The categorization of the DL of the questions makes it easier to interpret, for example, easy, medium, and difficult questions. A lot of literature refers to this interpretation by making other criteria (Singh et al., 2021) and (Turkmen & Caner, 2020). The DL is classified as follows. First, questions with a level of difficulty in the range 0 - 0.30 are questions in the difficult category. Second, questions with a level of difficulty in the range of 0.31 - 0.70 are questions in the medium category. Third, questions with what DL in the range 0.71 - 1.00 are questions in the easy category. This means that the higher the DL, the easier the question is. On the contrary, the lower the level of difficulty, the more difficult the question. The classification is an example that can be revised according to the purpose of the test. On this basis, it is appropriate that the DL of the questions has a strategic and main position from the characteristics of the questions to be used as a reference for making tests or providing adaptive tests or other tests at all levels of education.

The difficulty level of HLRQ using the FGD approach follows the pattern of real numbers with limits [0,7] according to the semantic differential scale range. The DL of mathematics HLRQ shows that the estimated FGD approach has similarities with the results of the Iteman program analysis. The DL is not partition but continuous. These results indicate that the FGD results can be used as a reference choice in determining the DL of mathematics HRLQ, although it is still limited to classical test theory. Therefore, the results of this study contribute to the assessment of the DL of multiple-choice questions using FGD as an alternative for teachers in addition to computer programs (Rukli et al., 2021). This method is applicable in the field, especially for teachers whose basic computer skills are limited. In addition, the involvement of teachers and students determines the DL. There will be added value in the interaction between teachers and students who are more familiar with discussing and asking questions so that students are expected to be more active in learning. This is a further study. In addition, further studies use the item response theory approach. Especially for the response data for true-false questions with the item response theory approach or Rasch measurement theory, the response data analysis can use the Bilog\_MG program (Rukli, 2018).

## CONCLUSION AND SUGGESTION

#### Conclusion

The estimated level of difficulty of the high-level reasoning questions from the FGD group was not significantly different from the output of the Iteman program analysis.

## Suggestion

- 1. This study only uses the classical test theory approach, so it is recommended to use modern test theory or the Rasch measurement model
- 2. The results of the study only use a quantitative approach, but if there are parties who will follow up to conduct a qualitative study.
- 3. Students' involvement in the assessment process requires further studies concerning student achievement, student responses, and student activities in classroom learning.
- 4. The teacher can apply the results of this study in the classroom to determine the DL of the questions.

## ACKNOWLEDGEMENTS

This research was funded by a Ministry of Education Grant of Culture, Research, and Technology for Fiscal Year 2021 with Contract Number 1868/E4/AK.04/2021 dated May 18, 2021, through a research contract for the Single Year of Applied Research between Universitas Muhammadiyah Makassar and LLDIKTI Region IX on April 6, 2021. Therefore, the team The researcher would like to thank the Minister of Education. Culture, Research, and Technology, the Head of LLDIKTI Region IX, and the Chancellor of Unismuh Makassar through the Head of LPE3M Unismuh Makassar for the support for these funds.

## REFERENCES

- Applegate, G. M., Sutherland, K. A., Becker, K. A., & Luo, X. (2019a). The Effect of Option Homogeneity in Multiple-Choice Items. *Applied Psychological Measurement*, 43(2), 113–124. https://doi.org/10.1177/0146621618770803
- Applegate, G. M., Sutherland, K. A., Becker, K. A., & Luo, X. (2019b). The Effect of Option Homogeneity in Multiple-Choice Items. *Applied Psychological Measurement*, 43(2), 113–124. https://doi.org/10.1177/0146621618770803
- Guest, G., Namey, E., & McKenna, K. (2017). How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. *Field Methods*, 29(1), 3–22. https://doi.org/10.1177/1525822X16639015
- Guyer, R., & Thompson, N. A. (2006). User's Manual for the ITEMAN<sup>TM</sup> Conventional Item Analysis Program. Retrieved from http://smp.ypk.or.id/mediaguru/upload/download/ ITEMAN Manual.pdf
- Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement*, 40(8), 559–572. https://doi.org/10.1177/0146621616664046
- Juškaite, L. (2018). *PAPER-BASED AND ONLINE TESTS IN NATURAL. 1*, 293–303. Retrieved from http://journals.rta.lv/index.php/SIE/article/view/3438/3178
- Kozierkiewicz-Hetmańska, A., & Poniatowski, R. (2014). An Item Bank Calibration Method for a Computer Adaptive Test. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 8397 LNAI (pp. 375– 383). https://doi.org/10.1007/978-3-319-05476-6 38
- Lafferty, M. I. (2004). Focus group interviews as a data collecting strategy. *Journal of Advanced Nursing*, 48(2), 187–194.
- Martin, M., Mullis, I. V. S., Foy, P., & Stanco, G. (2012). *TIMMS 2011 International Result in Science*. Retrieved from http://timssandpirls.bc.edu/timss2011/downloads/T11\_IR\_

Science\_FullBook.pdf

- Martinková, P., Štěpánek, L., Drabinová, A., Houdek, J., Vejražka, M., & Štuka, Č. (2017). Semi-realtime analyses of item characteristics for medical school admission tests. *Proceedings of the* 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017, 11, 189–194. https://doi.org/10.15439/2017F380
- Monique M. Hennink. (2012). Focus Group Discussions (Understanding Qualitative Research). In *Oxford University Press*.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). Timss 2015 International Results in Science Saved. Retrieved from http://timss2015.org/timss-2015/science/studentachievement/distribution-of-science-achievement/
- Raykov, T., Dimitrov, D. M., Marcoulides, G. A., & Harrison, M. (2019). On the Connections Between Item Response Theory and Classical Test Theory: A Note on True Score Evaluation for Polytomous Items via Item Response Modeling. *Educational and Psychological Measurement*, 79(6), 1198–1209. https://doi.org/10.1177/0013164417745949
- Raykov, T., & Marcoulides, G. A. (2016). On the Relationship Between Classical Test Theory and Item Response Theory. *Educational and Psychological Measurement*, 76(2), 325–338. https://doi.org/10.1177/0013164415576958
- Richard A. Krueger; Mary Anne Casey. (2014). Casey Focus Groups\_A Practical Guide for Applied Research (Edition 5, ed.). SAGE Publications Ltd.
- Rukli, R. (2012). INTERNATIONAL CONFERENCE ADRI -5 " Scientific Publications toward Global Competitive Higher Education " 384 CHARACTERISTICS ANALYSIS OF MATH ITEMS TYPE ON DATA TREND IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY Rukli. 384– 389. Retrieved from http://eprints.unm.ac.id/5808/54/49-Rukli.pdf
- Rukli, R. (2018). ANALYSIS OF MATHEMATICAL PROBLEM WITH GROUP WISE FOR COMPUTERIZED ADAPTIVE TESTING. Jurnal Daya Matematis, 6(1), 96. https://doi.org/10.26858/jds.v6i1.5600
- Rukli, R., Ma'rup, M., Bahar, E. E., & Ramdani, R. (2021). The estimation of test item difficulty using focus group discussion approach on the semantic differential scale. *Kasetsart Journal of Social Sciences*. https://doi.org/10.34044/j.kjss.2021.42.3.22
- Singh, R., Timbadia, D., Kapoor, V., Reddy, R., Churi, P., & Pimple, O. (2021). Question paper generation through progressive model and difficulty calculation on the Promexa Mobile Application. *Education and Information Technologies*. https://doi.org/10.1007/s10639-021-10461-y
- Stone, J. C., Glass, K., Munn, Z., Tugwell, P., & Doi, S. A. R. (2020). Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. *Journal of Clinical Epidemiology*, 117, 36–45. https://doi.org/10.1016/j.jclinepi.2019.09.010
- Turkmen, G., & Caner, S. (2020). The investigation of novice programmers' debugging behaviors to inform intelligent e-learning environments: A case study. *Turkish Online Journal of Distance Education*. https://doi.org/10.17718/TOJDE.762039
- Zubairi, A. M. (2006). Classical and Rasch analyses of dichotomously scored reading comprehension test items. *Malaysian Journal of ELT Research*, 2(March), 1–20. Retrieved from http://journals.melta.org.my/index.php/majer/article/view/168/82