# Application of Rasch Model in Evaluating the Reliability and Quality of Examination Paper for Object-oriented Design Course

ZUHAIRA MUHAMMAD ZAIN
Information Systems Department, College of Computer & Information Sciences
Princess Nourah Bint Abdulrahman University
Riyadh
KSA
zmzain@pnu.edu.sa

*Abstract:* - Exam has been used enormously as an assessment tool to measure students' academic performance in most of the higher institutions in KSA. A good quality of a set of constructed items/questions on mid and final exam would be able to measure both students' academic performance and their cognitive skills. We adopt Rasch Model to evaluate the reliability and quality of the first mid exam questions for Object-oriented Design course. The result showed that the reliability and quality of the exam questions constructed were relatively good and calibrated with students' learned ability.

*Key-Words:* - Rasch Model, Item Constructions, Reliability, Quality, Students' Academic Performance, Information Systems, Bloom's Taxonomy

## 1 Introduction

Object-oriented Design (IS223D) is a new course in the Information Systems (IS) Department at the College of Computer and Information Sciences (CCIS) at the Princess Nourah Bint Abdulrahman University (PNU). It is one of the core subject that must be completed by the IS students before they can graduate. The main objective of this course is to introduce various of UML diagrams to the students and they need to apply them in the Object-oriented Software Development environment.

CCIS needs to comply to the American Accreditation Board of Engineering and Technology, 2000 (ABET) program accreditation requirements. One of the ABET criteria is to have a quality student performance measurement. In CCIS, student performance measurement has been essentially dependent on the students' performance in carrying out tasks such as quizzes, assignments, mid examinations, projects and final exams. Measuring students' performance is essential in education in order to monitor the students' learning pattern [1]. A quality task should provide the same level of cognitive thinking skills to all students on what they have learned. In order to increase the students' performance quality, well organized and constructed tasks, which are based on Bloom's cognitive thinking skills [2] and the level of students' ability should be taken into consideration. A reliable and high quality assessment tools in teaching and learning process is required to measure students' understanding and ability. In this paper, the mid examination questions for IS223D for Semester 2 Session 2015/2016 is taken into account as the assessment tool. Furthermore, in the process of constructing these examination questions, it is vital to have fairly distributed examination questions based on Bloom's cognitive thinking skills, the level of students' ability and level of questions/items difficulty. Morales reported that in evaluating the quality of the questions, a discussion of reliability is essential [3]. The reliability is the degree to which an instrument consistently measures the ability of an individual or group. Typically, in CCIS, there is no measurement on reliability of any examination questions. The questions were only checked for their format, spelling, and the relevance of questions by the course specialist. Thus, there is no statistical evidence to prove that a set of examination questions is reliable.

The purpose of this study is to measure the reliability and quality of the mid examination questions of the Object-oriented Design (IS223D) course. This is to evaluate whether these questions calibrated with students' learning abilities and the course contents. It is part of the study to enhance and improve students' cognitive thinking skills and ability in designing UML diagram based on given requirements. The measurement is based on the mid examination questions and the raw marks will be

analyzed by the Rasch Measurement Model. The model fulfill the guidelines that has been emphasized by Wright and Mok [4] that a measurement model must produce linear measures, overcome missing data, provide estimates of precision, detect misfits, and distinguish the parameters of the object being measured from those of the measuring instrument. Thus, it can generate meaningful inferences by transforming an ordinal score into a linear, interval-level variable, through estimating the fit of data to the Rasch model's expectations. The basic principle underlying the Rasch Model is that the probability of a respondent/student successfully verifying a particular item/question is governed by the difference between the item/question's difficulty and respondent/student's ability [5, 6, 7]. The logic underlying this principle is that all respondents/students have a higher probability of answering easier items/questions and a lower probability of answering more difficult items/questions accurately [5]. The model has been used to assess the reliability and quality of examination paper of some Engineering courses in Malaysia [8, 9, 10, 11, 12] yet it has not been applied for Information Systems courses especially in Saudi Arabia. Furthermore, Rasch Model is one of the reliable and appropriate method in assessing students' ability [9, 11, 12]. Aziz et al. stated that the model's Wright Map can give a precise overview of the student's achievement on a linear scale of measurement [13]. Another study by Rashid et al. showed that Rasch Model Wright Map could provide meaningful information on the students' learning effectiveness [14].

## 2 Methodology

The data was obtained from the mid examination questions of IS223D course, which was taken by the second year Information Systems students of CCIS, PNU. Data from 59 students were collected and studied. The mid examination consists of 22 questions which was divided into four parts, Part A, Part B, Part C and Part D. Students were required to answers all questions. The questions covering three learning topics in IS223D, UML Overview, UML & Object-oriented Software Development, and Use case modeling. The course learning outcomes for the three learning topics for IS223D expected for the students to achieve is shown in Table 1.

The questions were entered as entry number as shown in Table 2. The item was labelled as Question No., and Taxonomy Bloom Level of Learning, which the students expected to develop three Level of Bloom's Taxonomy, namely Remembering/Understanding (Level 1), Applying/Analyzing (Level 2), Evaluating/Creating (Level 3). Thus, for entry item number 1, the item was coded as "A01_1" (refer to Table 2), where "A01" is equal to Question No., and "_1" is equal to Level of Bloom's Taxonomy.

Score from final examination results were gathered and compiled. As these raw score have different total marks for each question, a standardization method was used. The formula for the standardization is given below [8]:

$$z_{ij} = \frac{x_{ij} - min\, x_j}{\max x_j} \quad (1)$$

where $i$ = the $i$th students ($i$ = 1, 2, ... , 59), $j$ = the $j$th questions ($j$ = 1, 2, ..., 22), $z_{ij}$ = standardized marks for $i$th student and $j$th question, $x_{ij}$ = marks for $i$th student and $j$th question, $min\ x_j$ = minimum marks for $j$th question, and $max\ x_j$ = maximum marks for $j$th question.

Responses from the students' exam results were analysed using rating scale in which the students were rated according to their achievement. From (1),

$$z_{ij} * 10 = A \quad (2)$$

Then, A is classified correspond to the rating scale in Table 3. This grade rating was tabulated in Excel*$prn$ format. Using Rasch software, *Bond&Fox Steps*, this numerical coding is necessary for further evaluation of the students' achievement and also the reliability and the quality of items. The analysis outputs obtained from the *Bond&Fox Steps* were analysed and studied.

Table 1. Course learning outcomes for three learning topics for IS223D

| No | Course Learning Outcomes |
|---|---|
| 1 | Able to analyze requirement. |
| 2 | Able to design use case diagram. |
| 3 | Able to determine Object-oriented modeling fundamental. |
| 4 | Able to explain the purpose of modeling and UML. |

Table 2. Entry number coded for each exam question

| Part | Qs. | Entry No. | Learning Topics |
|---|---|---|---|
| A | 1 | A01_1 | UML Overview |
| | 2 | A02_2 | UML & Object-oriented Software Development |
| | 3 | A03_1 | Use case modeling |
| | 4 | A04_1 | UML Overview |
| | 5 | A05_1 | UML & Object-oriented Software Development |
| | 6 | A06_2 | Use case modeling |
| | 7 | A07_1 | UML & Object-oriented Software Development |
| | 8 | A08_2 | UML Overview |
| | 9 | A09_2 | UML & Object-oriented Software Development |
| | 10 | A10_1 | UML Overview |
| B | 11 | B11_1 | UML Overview |
| | 12 | B12_2 | UML Overview |
| | 13 | B13_2 | UML Overview |
| | 14 | B14_1 | UML Overview |
| | 15 | B15_2 | Use case modeling |
| | 16 | B16_2 | UML & Object-oriented Software Development |
| | 17 | B17_2 | UML Overview |
| | 18 | B18_1 | UML Overview |
| | 19 | B19_2 | UML & Object-oriented Software Development |
| | 20 | B20_2 | UML & Object-oriented Software Development |
| C | 21 | C21_3 | Use case modeling |
| D | 22 | D22_3 | Use case modeling |

Table 3. Marks (*A*) and corresponding rating scales

| Marks (*A*) | 0-1.49 | 1.50-3.49 | 3.50-6.49 | 6.50-8.49 | 8.50-10.00 |
|---|---|---|---|---|---|
| Rating scale | 1 | 2 | 3 | 4 | 5 |

## 3 Data Analysis and Discussions

The summary statistics for the analysis of the sample of 59 students on the 22 scale items comprising the IS223D mid examination items/questions are shown in Fig. 1. The summary fit statistics for items show satisfactory fit to the model. The mean square fit (IMNSQ and OMNSQ) statistics and the z statistics (Infit and Outfit ZSTD) for items are close to its expected values, +1 and 0, respectively. Moreover, the item reliability (Rasch equivalence to Cronbach's alpha) is significantly good, .90. This means that all the items/questions in the IS223D mid examination can be used consistently to measure the ability of the students.

| | RAW SCORE | COUNT | MEASURE | MODEL ERROR | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 236.5 | 59.0 | .00 | .11 | 1.03 | .0 | 1.09 | .1 |
| S.D. | 48.8 | .0 | .38 | .04 | .27 | 1.8 | .70 | 1.5 |
| MAX. | 287.0 | 59.0 | 1.04 | .21 | 1.34 | 3.0 | 3.53 | 3.5 |
| MIN. | 87.0 | 59.0 | -.62 | .07 | .27 | -6.1 | .31 | -3.7 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| REAL RMSE | .12 | ADJ.SD | .36 | SEPARATION | 2.98 | Item | RELIABILITY | .90 |
| MODEL RMSE | .11 | ADJ.SD | .37 | SEPARATION | 3.20 | Item | RELIABILITY | .91 |
| S.E. OF Item MEAN = .08 | | | | | | | | |

Fig. 1. Summary statistics

The Wright map in Fig. 2 establishes the distribution of persons/students on the left, represented by R01-R59, and the distribution of items/questions on the right, represented by the entry number (refer Table 2). The easiest question is A04_1 (UML consists of only two elements, diagram and relationship?) located at -.62 logits (SE .21), while the question that is most difficult to answer is B12_2 (Which of the following is NOT TRUE about OMG?) located at the top of the item/question distribution at +1.04 logits (SE .11). This is because B12_2 is a negative question that might confuse the students. It is recommended that any negative questions should be omitted in any examinations in the future. The person/student distribution confirms the result from the summary statistics. The most capable students are R16, R31, and R33 located at +.84 logits (SE .23), while the least capable student is R20, located at the bottom of the person/student distribution at -.16 logits (SE .13). The mean of the person/student distribution is higher than the mean of the item/question distribution. This indicates that majority of the students involved in the IS223D mid examination have the tendency to answer well most of the questions.

Fig. 3 shows the item statistics in Measure order. The Rasch fit statistics disclose that four items, B12_2, A01_1, A06_2, and A07_1 behaved more erratically than expected with an Outfit MNSQ value > 1.4. However, after confirming that the Infit MNSQ is within the range, it is accepted in this analysis. Other items fit sufficiently to the model, with their Infit and Outfit Mean-square values and Infit and Outfit Z-std values all lying within the acceptable range.

The principal contrast analysis of the Rasch residual variance is shown in Fig. 4. The variance explained by measures is good (50.5%). The uni-dimensionality of the IS223D mid examination instrument is strongly confirmed by having a good unexplained variance in the first contrast (6.1%). Thus, it proved that the questions are only related with the content of this subject.

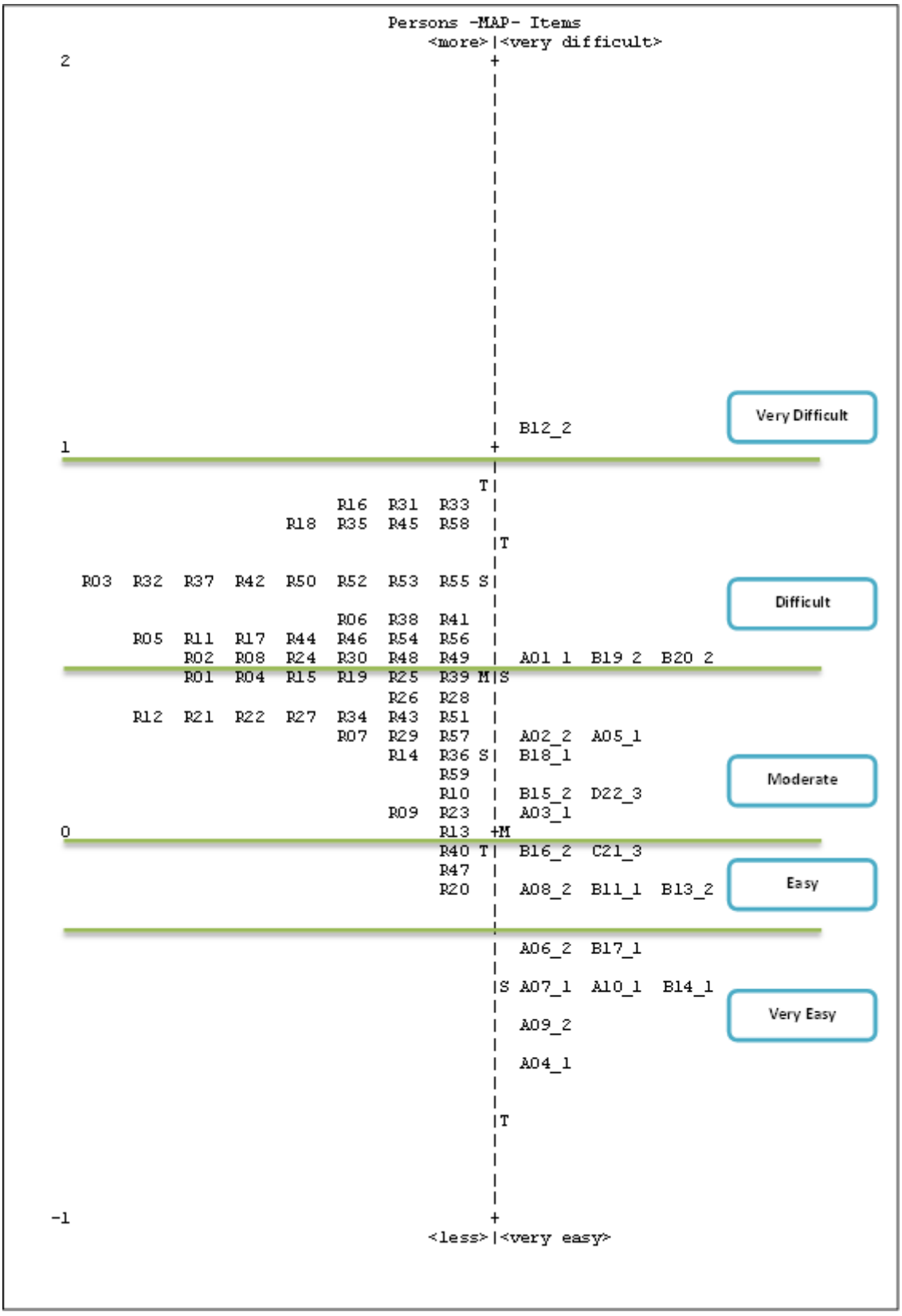


Fig. 2. Wright Map

| ENTRY NUMBER | RAW SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD | PTMEA CORR. | EXACT MATCH OBS% | EXACT MATCH EXP% | Item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 87 | 59 | 1.04 | .11 | 1.32 | 1.1 | 3.53 | 3.5 | -.19 | 61.0 | 64.0 | B12_2 |
| 1 | 171 | 59 | .47 | .07 | 1.33 | 3.0 | 1.45 | 3.0 | .11 | 3.4 | 6.7 | A01_1 |
| 19 | 171 | 59 | .47 | .07 | 1.20 | 1.9 | 1.21 | 1.5 | .25 | 5.1 | 6.7 | B19_2 |
| 20 | 175 | 59 | .45 | .07 | .98 | -.2 | .98 | -.1 | .46 | 3.4 | 5.2 | B20_2 |
| 5 | 207 | 59 | .27 | .08 | .86 | -1.1 | .81 | -1.2 | .59 | 11.9 | 12.8 | A05_1 |
| 2 | 211 | 59 | .25 | .08 | 1.14 | 1.1 | 1.06 | .4 | .34 | 11.9 | 13.0 | A02_2 |
| 18 | 219 | 59 | .20 | .08 | .96 | -.2 | .87 | -.6 | .50 | 11.9 | 13.3 | B18_1 |
| 15 | 235 | 59 | .10 | .08 | .88 | -.7 | .84 | -.6 | .54 | 23.7 | 25.0 | B15_2 |
| 22 | 236 | 59 | .09 | .08 | .27 | -6.1 | .31 | -3.7 | .48 | 50.8 | 25.1 | D22_3 |
| 3 | 239 | 59 | .07 | .09 | 1.14 | .8 | 1.29 | 1.1 | .29 | 22.0 | 25.3 | A03_1 |
| 16 | 251 | 59 | -.03 | .09 | .96 | -.1 | 1.04 | .2 | .45 | 33.9 | 34.3 | B16_2 |
| 21 | 256 | 59 | -.07 | .10 | .36 | -3.6 | .39 | -2.1 | .25 | 49.2 | 44.3 | C21_3 |
| 8 | 263 | 59 | -.15 | .11 | 1.11 | .5 | .85 | -.2 | .35 | 59.3 | 57.5 | A08_2 |
| 11 | 263 | 59 | -.15 | .11 | 1.01 | .1 | .72 | -.6 | .44 | 61.0 | 57.5 | B11_1 |
| 13 | 263 | 59 | -.15 | .11 | 1.09 | .4 | 1.15 | .5 | .29 | 55.9 | 57.5 | B13_2 |
| 6 | 275 | 59 | -.31 | .13 | 1.34 | .9 | 1.91 | 1.4 | .00 | 78.0 | 77.8 | A06_2 |
| 17 | 275 | 59 | -.31 | .13 | .91 | -.1 | .49 | -.9 | .51 | 83.1 | 77.8 | B17_1 |
| 7 | 279 | 59 | -.38 | .15 | 1.28 | .7 | 2.28 | 1.7 | .05 | 83.1 | 81.8 | A07_1 |
| 10 | 279 | 59 | -.38 | .15 | 1.25 | .7 | 1.03 | .3 | .19 | 83.1 | 81.8 | A10_1 |
| 14 | 279 | 59 | -.38 | .15 | 1.02 | .2 | .61 | -.5 | .41 | 84.7 | 81.8 | B14_1 |
| 9 | 283 | 59 | -.48 | .17 | 1.12 | .4 | .47 | -.6 | .38 | 89.8 | 87.2 | A09_2 |
| 4 | 287 | 59 | -.62 | .21 | 1.20 | .5 | .77 | .1 | .27 | 94.9 | 91.6 | A04_1 |

Fig. 3. Item Measure

Note: Acceptable range for Infit and Outfit Mean-square is between 0.6 to 1.4 [15] and acceptable range for Infit and Outfit Z-std is between -2 to +2 [5]

STANDARDIZED RESIDUAL VARIANCE SCREE PLOT

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

| | | Empirical | | | Modeled |
|---|---|---|---|---|---|
| Total variance in observations | = | 44.4 | 100.0% | | 100.0% |
| Variance explained by measures | = | 22.4 | 50.5% | | 55.6% |
| Unexplained variance (total) | = | 22.0 | 49.5% | 100.0% | 44.4% |
| Unexplned variance in 1st contrast | = | 2.7 | 6.1% | 12.3% | |
| Unexplned variance in 2nd contrast | = | 2.1 | 4.7% | 9.6% | |
| Unexplned variance in 3rd contrast | = | 1.8 | 4.0% | 8.1% | |
| Unexplned variance in 4th contrast | = | 1.7 | 3.8% | 7.6% | |
| Unexplned variance in 5th contrast | = | 1.5 | 3.4% | 6.9% | |

Fig. 4. Principal Contrast Analysis

Note: Variance explained by measures should be ≥ 50% and unexplained variance in the first contrast should be ≤ 15% [16]

## 4 Conclusion

This study discovered that the questions of the mid examination paper for IS223D is reliable and in a good quality to measure students' academic performance. However, any negative questions should be omitted in any exams in the future. This findings can be future references for questions construction of other Information Systems courses.

As a conclusion, Rasch Measurement Model can be an effective tool in evaluating the reliability and quality of any assessment tools for Information Systems courses. Hence, this study showed that, by using Rasch Measurement Model, the result accurately classified the questions according to students learning ability and their cognitive thinking skills. It allows each question (item) to be evaluated discretely and calibrated with what students have learned. It also precisely classified the students according to their capability to answer the questions.

For future work, the other assessment tools such as quizzes, project and final examination questions for other IS courses will be evaluated to check for their reliability and quality in order to comply to the ABET program accreditation requirements.

*References:*

[1] A. A. Aziz, N. Khatimin, A. Zaharim, & T. S. A. Salleh, Evaluating multiple choice items in determining quality of test. *Teaching Assessment and Learning for Engineering (TALE)*, 2013, pp. 565–569.

[2] L. W. Anderson, D. R. Krathwohl, & B. S. Bloom, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, Allyn & Bacon, 2001.

[3] R. A. Morales, Evaluation of Mathematics Achievement Test: A Comparison between CTT and IRT, *The International Journal of Educational and Psychological Assessment*, Vol.1, No.1, 2009, pp. 19-26.

[4] B. D. Wright, & M. M. C. Mok, An Overview of the Family of Rasch Measurement Models, In E. V. Smith & R. M. Smith (Eds.), Introduction to Rasch Measurement, Maple Grove, MN: JAM Press, 2004, pp. 1-24.

[5] T. V. Bond, & C. M. Fox, *Applying The Rasch Model: Fundamental Measurement in the Human Sciences (2nd ed.)*, New Jersey: Lawrence Erlbaum Associates, 2007.

[6] G. Rasch, *Probabilistic models for some intelligence and attainment test*, University of Chicago Press, Chicago, 1960.

[7] B. D. Wright & M. H. Stone, *Best Test Design*. Chicago, IL: MESA Press, 1979.

[8] H. Othman, I. Asshaari, H. Bahaludin, Z. M. Nopiah, and N. A. Ismail, Application of Rasch Measurement Model in Reliability and Quality Evaluation of Examination Paper for Engineering Mathematics Courses, *Procedia Social and Behavioral Sciences*, Vol. 60, 2012, pp. 163-171.

[9] R. F. M. Said, Application of Rasch Measurement Model in Evaluating Student Performance for Foundation of Computing II. In 7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings, Springer Singapore, 2016, pp. 251-259.

[10] H. Othman, N. A. Ismail, I. Asshaari, F. M. Hamzah, & Z. M. Nopiah, Application of Rasch Measurement Model for Reliability Measurement Instrument in Vector Calculus Course, *Journal of Engineering Science and Technology*, Vol. 10, No. 2, 2015, pp. 77-83.

[11] A. A. Aziz, A. Zaharim, N. F. A. Fuaad, & Z. M. Nopiah, Students' performance on engineering mathematics: Applying Rasch measurement model, *Information Technology Based Higher Education and Training (ITHET)*, 2013, pp. 1–4.

[12] S. A. Osman, M. A. Khoiry, W. H. W. Badaruzzaman, & A. Mutalib, Measurement of students' understanding in final examination of statics and dynamics course using Rasch measurement model, *Teaching Assessment and Learning for Engineering (TALE)*, 2013, pp. 805–810.

[13] A. A. Aziz, A. Mohamed, N. H. Arshad, S. Zakaria, & S. Masodi, Appraisal on Course Learning Outcomes Using Rasch Measurement: A Case Study in Information Technology Education, *International Journal of Systems Application, Engineering and Development*, 2007, pp. 164-171.

[14] A. R. Rashid, A. Zaharim, and S. Masodi, Appication of Rasch Measurement in Evaluation of Learning Outcome: A Case Study inElectrical Engineering, *Regional Conference on Engineering Mathematics, Mechanics, Manufacturing & Architecture (EMARC)*, 2007, pp. 151-165.

[15] B. D. Wright, M. Linacre, J.-E. Gustafsson, & P. Martin-Loff, Reasonable Mean-square Fit Values. *Rasch Measurement Transactions*, Vol.8, No.3, 1994, pp. 370.

[16] W. P. Jr. Fisher, Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions*, Vol.21, No.1, 2007, pp. 1095.