Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia
Social and Behavioral Sciences
Procedia - Social and Behavioral Sciences 60 (2012) 163 – 171

UKM Teaching and Learning Congress 2011

Application of Rasch Measurement Model in Reliability and Quality Evaluation of Examination Paper for Engineering Mathematics Courses

Haliza Othman[a,b,]1*, Izamarlina Asshaari[b], Hafizah Bahaludin[b], Zulkifli Mohd Nopiah[a,b], Nur Arzilah Ismail[b]

[a]Centre for Engineering Education Research, Faculty of Engineering and Built Enviroment, Universiti Kebangsaan Malaysia
[b]Unit of Fundamental Engineering Studies, Faculty of Engineering and Built Enviroment, Universiti Kebangsaan Malaysia

Abstract

Most of undergraduate courses in higher institutions in Malaysia use final exam as assessment tool to measure students' academic achievement. Good constructed items/questions on final exam would be able to measure both students' academic achievement and their generic skills. As such, this study using Rasch Measurement Model to evaluate the reliability and quality of final exam questions Mathematics Engineering III course. The items in the examination paper were studied and items that do not measure up to expectations were identified. The item analysis provides clues to how well the content of the item yielded useful information about student ability. This study focuses on constructed items, where items must retain their relative difficulty on the equal interval scale (*logit*), regardless of the ability of the students that challenges the item. The analysis revealed that even though there are three misfit questions, but overall, the reliability and quality of the exam questions constructed were relatively good and calibrated with students' learned ability.

Keywords: Items construction; Rasch Model; Student's ability; Engineering education; Mathematics Engineering; Bloom's Taxonomy

1. Introduction

Mathematics is imperative to engineering community. It is viewed as fundamental subject for all engineering courses and researches, where mathematical modelling, manipulation and simulation are used extensively and the challenges to teach mathematics to engineers are enormous. Sazhin (1998) mentioned that the objective of teaching mathematics to engineering students is to find the right balance between practical applications of mathematical equations and in-depth understanding of living situation. On the other hand, the impact of teaching mathematical thinking skills on an engineer will enable them to use mathematics in their practice (Cardella, 2008). Based on studies done by Zainuri et.al (2009) and Othman et.al (2010) on the results of Mathematics Pre-Test, which was

* Corresponding author. Tel.: +6-03-8921-6681; fax: +6-03-8921-6960.
E-mail address: haliza@eng.ukm.my.

1877-0428 © 2011 Published by Elsevier Ltd. Selection and/or peer reviewed under responsibility of the UKM Teaching and Learning Congress 2011
doi:10.1016/j.sbspro.2012.09.363

given to the first year students of Faculty of Engineering and Built Environment (FKAB), Universiti Kebangsaan Malaysia (UKM), shows that the engineering students were lack of knowledge in certain important topics in mathematics. These findings agree with Lawson (2003), which describe that there are significant declines in many mathematical skills deemed important by higher education for those undertaking graduate courses with significant mathematical content. Meanwhile, a study conducted by Ma et al. (1999) suggested that the lack of students' cognitive abilities not to blame for the failure some students in mathematics, but rather about the desire to pursue advanced mathematics is identified as a cause. Due to this matter, studies on the assessment methods/tasks should be emphasized alongside with making improvements in teaching and learning methods and CLO for Engineering Mathematics courses. This is not only an effort to recognize the cause of students failure in academic performance and social interaction, but also aid them to achieve academic excellence.

Students' performance measurement mostly dependent on performance in carrying out tasks, such as, series of tests or quizzes, final examination and assignment (Ghulman et al., 2009). A good task must provide the same level of cognitive thinking skills to all students on what they have learned. Well organized and constructed tasks, which are based on Bloom's cognitive thinking skills and also take into account the level of students' ability, contribute to the increase in students' performance. A suitable assessment tools in teaching and learning process is required to measure students' understanding and ability fairly and equally. In this paper, final examination questions for KKKQ2114 (DE) for Semester 1 Session 20102011 is taken into account as assessment tools. Moreover, in the process of constructing these examination questions, it is crucial to have fairly distributed examination questions based on Bloom's cognitive thinking skills, the level of students' ability and level of questions/items difficulty. According to Morales (2009), in evaluating the quality of these questions, a discussion of reliability is essential. The reliability is the degree to which an instrument consistently measures the ability of an individual or group.

This study used Rasch Measurement Model to evaluate the reliability and the quality of final examination questions for KKKQ2114 (DE) course. Rasch (1960) described that Rasch Model is one of the reliable and appropriate method in assessing students' ability. Ghulman et al. (2009) mentioned that Rasch Measurement Model useful with its predictive feature to overcome missing data. A study done by Masodi et al. (2010) shows that this model can classify grades into learning outcomes more accurately especially in dealing with small number of sampling units. Aziz et al. (2008b) applied bio-based Rasch Model in an attempt of paradigm shift in testing and validating the construct of measurement instrument. It follows that in Aziz et al. (2008a), this model was used as a new paradigm in assessing competency of Information Professionals. Meanwhile, Aziz et al. (2007) stated that Person and Items Distribution Map (PIDM) can give a precise overview of the student's achievement on a linear scale of measurement. Rashid et al. (2007) also mentioned that Rasch Model PIDM could provide meaningful information on the students' learning effectiveness.

This paper focuses on using Rasch Measurement Model to evaluate the reliability and quality of the final exam questions of KKKQ2114 (DE) course and 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 solving mathematics problems. Therefore, the engineering students' performance in mathematics courses at FKAB, UKM can be improved significantly.

2. Methodology

The data was obtained from the final examination questions of KKKQ2114 Engineering Mathematics III (DE) course, which was taken by second year engineering students of FKAB, UKM. Data from 218 students from Department of Civil and Structural Engineering, Department of Electric, Electronic and System Engineering, Department of Chemical and Process Engineering and Department of Mechanical and Materials Engineering were collected and studied. The final examination consists of 30 questions which was divided into three parts, which are Part A, Part B and Part C. Students are required to answers all questions in Part A and B, while Part C is an optional question. Covering most of the learning topics in KKKQ2114 such as first and second order Differential Equations, Laplace Transformation, Fourier Series and Partial Differential Equation. Rasch Measurement Model used in this study is assumed fit to measure the learning ability of students. The course outcomes for KKKQ2114 expected for the students to achieve is shown in Table 1.

Table 1. Course outcomes for KKKQ2114

No.	Course Outcomes
1	Understand the basic concepts of differential equations and their solutions.
2	Able to solve first and second order ordinary differential equations.
3	Able to determine the Laplace transforms and the inverse Laplace transforms of elementary functions.
4	Able to build and solve a differential equations model of problems involving half-life, mixing problem, spring-mass system and electric circuits.
5	Able to determine the Fourier series, integrals and transforms of simple functions.
6	Know the types of partial differential equations and their applications in engineering.

Table 2. Topics coded for each examination question

Part	Qs.	Entry No.	Learning Topic
A	1a	A01_C	Definition and Terminology
	1b	A02_K	Solution curve
	1c	A03_K	Solution curve
	2ai	A04_P	Homogeneous equation
	2aii	A05_P	Homogeneous equation
	2bi	A06_C	Variations of parameter
	2bii	A07_P	Variations of parameter
	3ai	A08_K	Laplace Transforms
	3aii	A09_P	Laplace Transforms
	3b	A10_P	Inverse Laplace Transforms
	4a	A11_P	Series Solution
	4bi	A12_C	Fourier Series
	4bii	A13_C	Fourier Series
	4c	A14_C	Heat Equation
B	a	B15_K	Definition and Terminology
	b	B16_P	Homogeneous Equation
	c	B17_A	Particular Solution using Undetermined Coefficient General Solution for *RLC* circuit
	d	B18_A	Initial Value Problem
	e	B19_P	Steady State Solution for *RLC* circuit
	f	B20_A	Solution for *RLC*
C	1a	C21_P	Population Growth
	1b	C22_A	Limiting value of Population Growth
	2a	C23_P	Damping Force
	2b	C24_C	Equation of Motion for Spring Mass
	2c	C25_P	Equilibrium Position
	3a	C26_C	Inverse Laplace
	3bi	C27_C	Unit Step Function
	3bii	C28_P	Unit Step Function
	3ci	C29_A	*RLC* circuit in Laplace
	3cii	C30_P	Unit Step Function in *RLC* circuit

The questions are entered as entry number as shown in Table 2. The item is labelled as Question No., Learning Topic and Taxonomy Bloom Domain, which the students expected to develop four out of six Level of Bloom's Taxonomy, namely Knowledge (K), Comprehension (C), Application (P) and Analysis (A). Thus for entry item number 1, the item is coded as QA01_C (refer to Table 2).

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

$$z_{ij} = \frac{x_{ij} - \min x_j}{\max x_j} \tag{1}$$

where i = the ith students (i = 1, 2, ... , 218), j = the jth questions (j = 1, 2, ..., 30), z_{ij} = standardized marks for ith student and jth question, x_{ij} = marks for ith student and jth question, min x_j = minimum marks for jth question, and max x_j = maximum marks for jth question.

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

$$z_{ij} \times 10 = A \tag{2}$$

Then, A is classified correspond to the rating scale in Table 3:

Table 3. Marks (A) and Correspond Rating Scale

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

This grade rating is tabulated in Excel**prn* format. Using Rasch software, *Winstep,* 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 *Winstep* were analysed and studied.

3. Data Analysis and Discussion

An overall explanation on how well the questionnaire were constructed and whether student's ability levels exist or otherwise, can be read from the summary statistics as depicted in Table 4. The first statistic that we refer to is called separation, which is the index of spread of item positions.

If the index reads 1.0 or below, the item may not have sufficient breadth in position, which will further cause item redundancy. In that case, we may wish to reconsider the rating scale that has been applied in this study.

The item separation is 6.6, an even broader continuum than a person. This large index can be expected from the good item spread value of 2.6*logits*. This separation index translates to about five levels of item difficulties e.g. very easy, easy, moderate, difficult and very difficult. Next, with the reliability index of person valued at 0.98 (analogous to the traditional Cronbach's alpha), it indicates that the items are in line with consistently reproducing a participant's score. In parallel to this, the item reliability of 0.98 indicates that a similar item hierarchy along the variable is highly reproducible in a similar sample from the population. This means good reliability at which items measuring students' learning abilities.

Table 4. Summary Statistics for Item

SUMMARY OF 30 MEASURED Items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	488.3	199.6	.00	.07	1.02	-.4	1.01	-.3
S.D.	183.8	44.1	.54	.02	.43	4.4	.40	3.4
MAX.	987.0	218.0	.83	.15	2.00	8.3	2.01	8.8
MIN.	80.0	40.0	-1.77	.06	.41	-8.0	.51	-5.6

REAL RMSE	.08	ADJ.SD	.53	SEPARATION	6.60	Item	RELIABILITY	.98
MODEL RMSE	.07	ADJ.SD	.53	SEPARATION	7.31	Item	RELIABILITY	.98
S.E. OF Item MEAN = .10								

UMEAN=.000 USCALE=1.000
Item RAW SCORE-TO-MEASURE CORRELATION = -.80 (approximate due to missing data)
5988 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 14883.21 with 5738 d.f. p=.0000

3.1. Person-Item Distribution Map

Item difficulty and person ability were mapped side to side on the same measurement scale (vertical line with *logit* unit) as depicted in Figure 1. The scale is made up of samples ranging from 0.85 to -2.2 where the most difficult item and the most able test takers were laid out on top of the scale. On the person distribution area, both symbols "." and "X" represent one and two test taker(s). "S" marks one standard deviation away from the mean. The right hand side illustrates test items which are represented by the letter A, B or C and this is followed by the number of question and cognitive level of Blooms' Taxonomy. For instance, A12_P represents the $12th$ question of Part A and Application (cognitive level) in Blooms' Taxonomy.

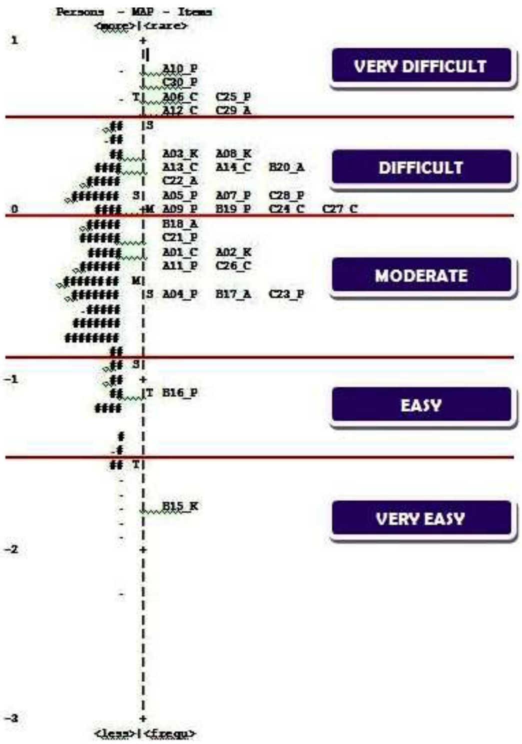

Figure 1. Person-Item Distribution Map

On top of this, students in general felt that the given test set was tough since the person mean had fallen below the mean of items (-0.41 against 0.0). What is interesting in this map is that almost 90% of items were located above the person mean. This indicates that the test is not able to measure the ability of half of the class. Strong evidence was found when those samples had high possibility (more than 50% of chance) to answer only two up to five questions correctly. In addition, 82 students were found to not have the ability to solve even one question from Part C (the easiest item of Part C was located slightly below the person mean). Further revision on the item's structure; e.g. language styles should be immediately performed in way to investigate the cause of the problem. Figure 1 also demonstrates that three students are positioned below the easiest item (B15_K).

Another important finding was that redundancies on the item measured appeared in all participated locations within $\mu \pm S$ except on C21_P and C22_A. This situation gives us room to analyze and further replace or drop these redundancies, thus the instrument would spread out wider and it may reduce the sample's standard deviation value. There are, however, in case of unrelated topics or different levels of Bloom's taxonomy, no replacements are needed. In future, in order to gather additional information, the instrument should consider extra items if possible. Other issues that emerge from the item's map distribution is the existence of a distinct gap which is located between item B16_P and a row of three items (A04_P, B17_A, C23_P) should be examined closer. It could be suggested that the items would be fitted to the Rasch Measurement Model by relocating one or two item(s) from the row into the space.

3.2. Fit Statistics

To determine which item does not fit the Rasch Measurement Model, a three-step comparison procedure was performed. Starting with a point measure correlation value followed by an outfit MNSQ and finally concluded with the outfit standardized value, those criteria are sequentially compared with a specific acceptable region. An item is labeled as misfit if all controls cannot be met. A point measure correlation x calculates the index of the item discrimination where the item with greater value might be too good to other items. In the Rasch analysis, inconsistency responses in items such as a less ability student answering difficult items correctly can be measured by an outfit index. Two statistics namely the mean square (MNSQ) and z-value were used to compute the item outfit. As proposed by Rasch experts, an acceptable region for each control is given as follows: $0.4 < x < 0.8$, $0.5 < MNSQ < 1.5$ and $-2.0 < z < 2.0$. Table 5 presents 30 items and these were sorted in the descending order with respect to a 'Measure' column. Several items (A01_C, A02_K and A03_K) were found to have fallen outside the acceptable regions. Further analysis on those misfit items should be taken as part of enhancing the instrument. Two actions might be considered such as rephrasing or deleting the item.

Table 5. Item Measure for Fit Statistics

Person: REAL SEP.: 2.44 REL.: .86 ... Item: REAL SEP.: 6.60 REL.: .98

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
10	335	218	.83	.08	1.19	1.4	1.04	.3	.33	.35	55.5	54.5	A10_P
30	293	187	.78	.09	.99	-.1	.76	-1.4	.52	.34	54.5	52.8	C30_P
6	363	218	.65	.07	1.01	.1	1.36	2.2	.01	.38	38.1	45.7	A06_C
25	346	203	.63	.08	.88	-1.0	.73	-1.9	.59	.38	46.3	43.6	C25_P
29	328	188	.56	.08	.55	-4.4	.61	-3.0	.55	.38	47.3	41.0	C29_A
12	381	218	.56	.07	.41	-6.7	.62	-3.1	.26	.40	59.6	42.0	A12_C
3	423	218	.37	.06	1.68	5.7	1.70	4.7	.38	.43	27.1	35.6	A03_K
8	[illegible]	218	.32	.06	[illegible]	-5.7	.74	[illegible]	.43	.44	[illegible]	[illegible]	A08_K
20	444	218	.28	.06	1.19	1.9	.98	-.1	.61	.45	27.1	33.8	B20_A
14	451	218	.25	.06	2.00	8.3	1.82	5.7	.42	.46	17.9	33.3	A14_C
13	455	218	.24	.06	.92	-.9	.87	-1.2	.50	.46	36.7	33.3	A13_C
22	80	40	.19	.15	1.61	2.2	1.49	1.6	.49	.54	27.5	37.7	C22_A
5	492	218	.10	.06	.67	-4.2	.66	-3.6	.51	.48	45.0	31.0	A05_P
7	503	218	.06	.06	1.21	2.4	1.18	1.6	.41	.49	30.3	31.1	A07_P
28	440	188	.04	.06	.52	-6.4	.64	-3.8	.53	.48	39.4	30.3	C28_P
9	521	218	.00	.06	1.47	4.9	1.48	4.2	.47	.50	18.8	30.1	A09_P
24	498	203	-.01	.06	.81	-2.3	.77	-2.4	.64	.49	27.6	29.6	C24_C
19	527	218	-.02	.06	.95	-.6	.87	-1.3	.69	.50	26.1	30.1	B19_P
27	459	188	-.03	.06	.44	-8.0	.51	-5.6	.58	.49	45.2	30.0	C27_C
18	536	218	-.05	.06	.60	-5.6	.60	-4.7	.66	.50	39.0	28.4	B18_A
21	97	40	-.17	.14	.63	-2.0	.76	-1.0	.63	.58	45.0	30.3	C21_P
2	592	218	-.22	.06	1.33	3.8	1.49	4.7	.34	.52	22.0	27.4	A02_K
1	606	218	-.27	.06	1.78	8.1	2.01	8.8	.17	.53	21.1	27.1	A01_C
11	[illegible]	218	-.33	.06	1.15	1.7	1.13	1.4	.44	.53	22.0	26.7	A11_P
26	555	188	-.37	.06	1.17	2.0	1.20	2.1	.30	.53	30.3	25.8	C26_C
23	636	203	-.47	.06	.66	-4.7	.66	-4.2	.66	.54	32.5	25.3	C23_P
17	683	218	-.50	.06	.95	-.6	.91	-1.0	.68	.54	23.9	25.5	B17_A
4	692	218	-.53	.06	.56	-6.7	.59	-5.4	.49	.54	41.3	25.3	A04_P
16	861	218	-1.10	.06	1.02	.3	.92	-.7	.66	.53	34.9	33.7	B16_P
15	987	218	-1.77	.09	1.75	4.3	1.30	1.6	.45	.43	65.1	62.5	B15_K
MEAN	488.3	199.6	.00	.07	1.02	-.4	1.01	-.3			36.4	34.6	
S.D.	183.8	44.1	.54	.02	.43	4.4	.40	3.4			12.4	9.2	

As can be seen, three items are misfit based on scalograms that represent in Figure 2. Thirty five top excellence respondents/students were taken as a reference. Misfit items are questions A01_C (item 1) and A02_K (item 2), which are categorized as ‘Moderate’ (perhaps all excellent students can answer these questions easily) and A03_K as ‘Difficult’ questions, see Figure 1. From Figure 2, scalograms show that students (person) 39, 38, 37 and 35 fails to score these questions, even though they are top excellence students. From Table 1, although question A01_C is regarding the definition and terminology of the logistics differential equation, but it's requires critical thinking to solve this problem. Hence, it is obvious that these engineering students yet to have critical thinking skills. Meanwhile, A02_K and A03_K are on the solution curve of the logistics differential equation. Each of these questions are advised to be split into two different questions in order to improve the reliability and quality of these questions.

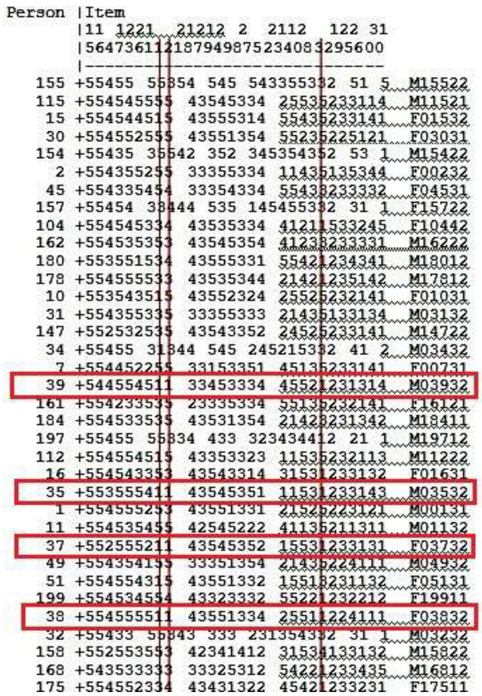

Figure 2. Scalograms

These are the misfit questions:

Consider the logistic differential equation $\frac{dP}{dt} = 0.08P\left(1 - \frac{1}{1000}P\right) - 15.$

(a) *Suppose* $P(t)$ *represent a fish population at time t, where t is measured in weeks. Explain the meaning of the term [-15].* {QA01_C}

(b) *Find the equilibrium solutions and phase portrait for this differential equation.* {QA02_K}

(c) *Classify each critical point as asymptotically stable, unstable or semi-stable. Illustrates the typical solution curves determined by the graphs of the equilibrium solutions.* {QA03_K}

The question is misfit due to the deficiency in the questions. The question is valid for this subject based on item dimensionality test. Suppose item dimensionality must be greater than 40%, so it proves that question is measured in one dimension. Basically, one dimension means the question only related with the content of the subject such as the subtopic of vector cannot be asked in this subject. Figure 3 shows that the result of item dimensionality test is 58.3%, where it is greater than 40%. Thus, it proved that the questions are only related with the content of this subject.

Of the 30 questions (items) considered in the final exam for KKKQ2114 (DE), only 3 or 10% come up to be misfit items, but these questions should not be rejected since its Pt-Measure in Table 5 are within acceptable range. These questions (QA01_C, QA02_K and QA03_K) need to be reviewed and revised. In these particular questions, it is advisable to split the questions into two parts and students have to have critical thinking ability to tackle these problems. This improvement will enhance the reliability and the quality of the final exam of KKKQ2114, subsequently improved students' academic achievement and performance.

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --			Modeled
Total raw variance in observations	=	51.4	100.0%		100.0%
Raw variance explained by measures	=	21.4	41.7%		39.2%
Raw variance explained by persons	=	3.9	7.5%		7.1%
Raw Variance explained by items	=	17.6	34.2%		32.1%
Raw unexplained variance (total)	=	30.0	58.3%	100.0%	60.8%
Unexplned variance in 1st contrast	=	3.9	7.6%	13.0%	
Unexplned variance in 2nd contrast	=	2.4	4.7%	8.1%	
Unexplned variance in 3rd contrast	=	2.0	3.9%	6.7%	
Unexplned variance in 4th contrast	=	1.7	3.3%	5.7%	
Unexplned variance in 5th contrast	=	1.7	3.2%	5.6%	

Figure 3. Item Dimensionality Test

4. Conclusions

This study revealed that the items of the final examination paper for KKKQ2114 should to be revised and improved in effort to improved students' academic performance. This findings can be future references for items construction of other Engineering Mathematics courses. As a conclusion, Rasch Measurement Model can be an effective tool in evaluating the reliability and quality of any assessment tools for Engineering Mathematics courses. Therefore, this study revealed that, by using Rasch Measurement Model, the result more accurately classified the questions according to students learning ability and their cognitive thinking skills. It enables each question (items) to be evaluated discretely and calibrated with what students have learned. It also accurately classified the students according to their observed achievements. For further work, overlapping items tests have yet to be analyzed for the redundancy of the questions.

Acknowledgements

We would like to thank UKM for providing the research grant (PTS-2011-020).

References

Aziz, A. A., Mohamed, A., Arshad, N. H., Zakaria, S. & Mosadi, S. (2007). Appraisal on Course Learning Outcomes Using RaschMesurement: A Case Study in Information Technology Education. *International Journal of Systems Application, Engineering and Development,* 164-171.

Aziz, A. A., Mohamed, A., Arshad, N. H., Zakaria, S., Ghulman, H. A., & Mosadi, S. (2008a). Development of Rasch-based Descriptive Scale in profiling Information Professionals' Competency, in *IEEE XPLORE indexed in INSPEC; IEEE IT Simposium (ITSim KL), August 2008,* 184-191.

Aziz, A. A., Mohamed, A., Arshad, N. H., Zakaria, S., Zaharim, A., Ghulman, H. A., & Mosadi, S. (2008b). Application of Rasch Model in Validating the Construct of Measurement Instrument, *International Journal of Education and Information Technologies*, (2)2, 105-112.

Cardella, M.E. (2008), Which Mathematics Should We Teach Engineering Students? An Empirically Grounded Case for A Broad Notion of Mathematical Thinking, *Teaching Mathematics and Its Applications*, 27(3), 150-159.

Ghulman, H.A. & Masodi, M.S. (2009). Modern measurement paradigm in Engineering Education: Easier to read and better analysis using Rasch-based approach. *2009 International Conference on Engineering Education (ICEED 2009), December 7-8, 2009, Kuala Lumpur, Malaysia*, 1-6.

Lawson, D. (2003). Changes in Student Entry Competencies 1991-2001, *Teaching Mathematics and its Applications,* 22(4), 171-175.

Ma. X., & Willms, J.D. (1999). Dropping Out of Advanced Mathematics: How Much Do Students and Schools Contribute to the Problems? *Educational Evaluation and Policy Analysis*, 21(4), 356-383.

Morales, R. A. (2009). Evaluation of Mathematics Achievement Test: A Comparison between CTT and IRT, *The International Journal of Educational and Psychological Assessment*, 1(1), 19-26.

Othman, H., Ariff, F.H.M, Ismail, N.A., Asshaari, I., Zainuri, N.A, Razali, N., Nopiah, Z.M. (2010). Engineering students` performance in mathematical courses: The case study of Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia. *Proceeding of The 1st Regional Conference on Applied and Engineering Mathematics Proceedings (RCAEM),* 5(15), 512-516.

Rasch G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press, Chicago, (Reprinted 1980)

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

Masodi, S., Aziz, A. A, Rodzo'an, N.A, Omar, M.Z, Zaharim, A and Basri, H. (2010). Easier Learning Outcomes Analysis using Rash Model in Engineering Education Research. *EDUCATION'10 Proceedings of the 7th WSEAS International Conference on Engineering Education,* 442-447.

Sazhin, S.S. (1998), Teaching Mathematics to Engineering Students, *International Journal Engineering Education,* 14(2), 145-152.

Zainuri, N.A., Nopiah, Z.M., Razali, N., Asshaari, I., & Othman, H. (2009). The Study on The Weaknesses of Mathematical Foundation in The First Year Engineering Students, UKM, *Prosiding Seminar Pendidikan Kejuruteraan & Alam Bina (PeKA09),* 226-233.