Journal of Engineering Science and Technology
Special Issue on PEKA 2016, November (2016) 31 - 39

EVALUATING THE RELIABILITY OF PRE-TEST DIFFERENTIAL EQUATIONS QUESTIONS USING RASCH MEASUREMENT MODEL

N. LOHGHESWARY[1], Z. M. NOPIAH[1], E. ZAKARIA[2]

[1]Faculty of Engineering and Built Environment,
[1]Centre for Engineering Education Research,
[2]Faculty of Education,
Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia
*Corresponding Author: futuredzines@gmail.com

Abstract

A good exam questions should be able to gauge student's understanding and achievement related to Course Outcome (CO), Bloom's Taxonomy level and Programme Outcome (PO). To achieve this, a set of pre-test questions were prepared to evaluate the pre achievement level among the students related to CO, PO and the Bloom's Taxonomy level. In this study, a pre-test for Differential Equations (KKKQ2123) was given to 100 second year students from the department of Electrical, Electronic and Systems Engineering. The level of Bloom's Taxonomy measured consists of level 1 (knowledge) to level 6 (creation). Rasch Measurement Model was applied to analyse the reliability of the pre-test questions. The analysis revealed that all the pre-test questions were reliable and no questions were found unsuitable. Prior assessment (pre-test) is important in the preparation of final exam questions as it would indicate the level of student's understanding in a particular topic that relates to the CO and PO of the programme.

Keywords: Differential equation, Course outcome (CO), Programme outcome (PO), Bloom's taxonomy, Rasch measurement model.

1. Introduction

Tests, assignments, and final examinations are the common instruments to assess students' performance. The construction of these instruments must take into account the Course Outcome (CO), Programme Outcome (PO) and the level of Bloom’s Taxonomy.

Rasch analysis can be applied to assessments in a wide array of disciplines, these includes education, health studies, psychology, marketing, economics and social sciences. Most assessments in these disciplines involve a well-defined group of people responding to a set of items for assessment. Azrilah et al. [2] stated Rasch Measurement Model is not focused on developing one ‘best fit line’. It focuses on constructing a reliable measurement instrument with accuracy.

Aziz et al. [1] applied an entry test to new students for the engineering program. The study explored the students’ performance on basic engineering mathematics. The Rasch Measurement Model was used to measure the students’ ability with item difficulty. The findings suggested that extra attention should be given to respective students who scored low in the pre-test. Nopiah et al. [9] studied the effectiveness of the pre-test mathematics questions in predicting the performance of the students in the subsequent engineering mathematics courses. The analysis showed a low correlation between pre-test and Vector Calculus and Linear Algebra result. The study seeks for further improvement in the construction of pre-test questions. Fuaad et al. [5] designed a pre-test exam to gather early information on student’s ability in mathematics to aid the lecturers to monitor the students. The study identified the item difficulties in pre-test situations Rasch Measurement Model. Students who completed Malaysian High School Certificate (STPM) performed better compared to students from a matriculation background.

According to Fuaad et al. [6], students lack the basic knowledge in some important topics in Mathematics. Final exam questions of Engineering Mathematics II (Linear Algebra) were analyzed using Rasch Measurement Model and the study revealed that there exists a huge gap between achievement in the Blooms’ cognitive skills. Osman et al. [10] measured the students’ performances in terms of Course Outcomes (COs) based on students’ mark entries together with Rasch Measurement Model. Comparative analysis was conducted against the conventional distribution marks and Rasch Measurement Model gave close to the same results on the students’ achievement. Nopiah et al. [8] validated linear algebra examination questions using Rasch Measurement Model. The results revealed that questions were constructed correctly without misfit questions. The results proved that a well-constructed examination question should commensurate with the level of the intended knowledge.

Kamsuriah et al. [7] stated since there is not any specified method to measure the actual performance of each Co, a modern measurement method was developed. This method used item response theory and collaborated with Rasch Measurement Model. Felder et al. [4] commented that lower performance by students was not only because of the students themselves but also due to the difficulty of the questions. For example, questions constructed not at par with what the student had been taught. Questions should be constructed to meet the students’ ability. Questions should be embedded at a higher-level of skill. Instructors should include a higher-level of learning objectives by giving some examples in assignments. Draugalis et al. [3] evaluated students and item performance strengths and weaknesses using 65-item multiple-choice examination. The curriculum strength was parallel with the expected learning outcomes of the college.

2. Methodology

The Differential Equations (KKKQ 2123) is offered on the third semester. The pre-requisite for this subject is Engineering Mathematics I (KKKQ1123 /Calculus Vector) and Engineering Mathematics 2 (KKKQ1223/Linear Algebra). The pre-test consists of 6 questions and a time frame of 2 hours was given to students to complete the test. A total of 100 students from the department of Electrical, Electronic and Systems Engineering of Faculty of Engineering and Built Environment (FKAB), Universiti Kebangsaan Malaysia (UKM) participated in this test.

This pre-test was constructed subjective questions which contribute to a total of 60 marks and was validated by 2 internal lecturers who teach this subject for semester 1 2015 / 2016. Course Outcome (CO) and Programme Outcome (PO) are used as guide line to construct the questions. Each CO is tested in each of the pre-test questions/items. PO1 (The ability to acquire and apply the knowledge of mathematics, science and engineering) and PO2 (The ability to perform identification, formulation and solution of engineering problems) were considered in the pre-test.

Table 1 shows the lists of CO-PO related coding and description for Differential Equation. In the first row, the CO is 1 and the respective code is 1C2. 1C2 indicates that 1 is for PO1 while 2 indicates the highest level of Bloom's Taxonomy that can be tested for this CO. The Bloom's Taxonomy levels are as follows: level 1 (Knowledge), level 2 (Comprehension), level 3 (Application), level 4 (Analysis), level 5 (Evaluation) and level 6 (Creation).

Table 2 shows the pre-test questions with the marks allocated for each sub-question. Each question was constructed to take into account the requirement of the CO, PO and the Bloom's Taxonomy level.

Table 3 shows the details regarding pre-test questions as entry number, CO, PO, code, Bloom's Taxonomy level and description. For question 4, the highest Bloom's Taxonomy that is tested is 3 (application). This is tested in question 4 (iii). For question 4(i) and (ii) first and second level of Bloom's Taxonomy are examined respectively.

Students' results were tabulated in Excel using **prn* format which was then run in *Winstep*, a Rasch analysis software.

Table 1. CO for differential equations subject.

CO	Code	Description
1	1C2	Understand the basic concepts of differential equations and their solutions.
2	1C3	Able to solve first and second order of differential equations.
3	2C4	Able to perform step-by-step analysis to model the simple engineering problem using differential equations and to solve the differential equations using an appropriate technique.
4	2C3	Able to evaluate the Laplace transform for solving ordinary differential equations.
5	1C3	Able to use Fourier Series to solve partial differential equations.

Table 2. Pre-test questions.

Q	Description	Marks
1	Verify the function $y'' - 2y' + y = 0; \quad y = xe^x$ is a solution of the given differential equation on the interval $(-\infty, \infty)$.	4
2	Solve the differential equation by integrating factor $\cos x \frac{dy}{dx} + (\sin x) y = 1$.	5
3	Solve $(x^2 + y^2)dx + (x^2 - xy)dy = 0$.	7
4 (i)	A series circuit consists of a resistor with $R = 40\Omega$, an inductor $L = 1H$, a capacitor with $C = 16 \times 10^{-4} F$ and the impresses voltage with $E(t) = 100\cos 10t$.	2
(ii)	Find the general solution of the equation.	12
(iii)	If the initial charge and current are both 0, find an expression for the charge at time t.	4
(iv)	What happens to current as $t \to \infty$?	3
5	Find $L^{-1}\left\{\frac{1}{4s^2+1}\right\}$.	5
6	The solution of heat equation: $k\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad 0 < x < L, \quad t > 0$ $u(0,t) = 0 \quad u(100,t) = 0$ $u(x,0) = f(x) = \begin{cases} 0.8x & 0 \le x \le 50 \\ 0.8(100 - x) & 50 \le x \le 100 \end{cases}$ is given by the infinite series $u(x,t) = \sum_{n=1}^{\infty} A_n e^{-k\left(\frac{n\pi}{L}\right)^2 t} \sin\frac{n\pi}{L}x$	
(i)	Find A_n.	14
(ii)	Determine the first two terms of $u(x,t)$ about $x = 10$.	4

Table 3. Entry number for each question.

Question No.	CO	PO	Code	Bloom's Taxonomy level	Description
1	1	1	1C2	2	Comprehension
2	2	1	1C3	3	Application
3	2	1	1C3	3	Application
4(i)	3	2	2C4	1	Knowledge
4(ii)	3	2		2	Comprehension
4(iii)	3	2		3	Application
4(iv)	3	2		4	Analysis
5	4	2	2C3	3	Application
6(i)	5	1	1C3	3	Application
6(ii)	5	1		3	Application

3. Results and Discussion

Figure 1 shows summary statistics for individuals (person) that represent students who took the pre-test. The Cronbach alpha α shows 0.66 reflecting acceptable internal consistency of a raw responses pattern. The summary statistics for person reveal a fair spread of abilities logit of students involved in this study, at 0.65. The highest item on the difficulty scale is +1.05 while the lowest item is at -2.16. The person separation index at 1.36 is able to segregate the students into two groups: students with average performance and those with poor performance in Differential Equations.

According to Wright [11], Logits form an equal interval linear scale. Logit scale is constructed by deducing a theory that produces equal interval, linear measures and derive a method for applying that theory. The theory is

$$\log\left(\frac{\text{Probability of success}}{\text{Probability of failure}}\right) \equiv \text{Ability-Difficulty}$$

Since all elements can be represented as fixed positions along one straight line therefore this is a linear model. In games of chance, the (Probability of Success)/(Probability of Failure) is called "odds of success". $Log_e[(\text{ProbabilityofSuccess})/(\text{ProbabilityofFailure})]$ is called log-odds. The units of measurement constructed by this theory are called "Log-odds units" or "logits".

Subsequently, the Rasch Measurement Model provides item reliability at 0.88 which shows excellent item difficulty spread within the items. Item refer to pre-test questions. The maximum item on the logit ruler is located at +4.38 and the minimum item on the ruler is located at -1.99 logit. A separation 2.67 indicates that the items can be grouped into three groups: difficult, moderate and easy. The summary statistics for item that shows the performance of questions is given in Fig. 2.

To identify whether an item fits the instrument, a three step calculation is examined. They are Point-Measure Correlation value, Outfit Mean Square (MNSQ) and Outfit z-Standard (ZSTD). An item is classified as non fit if it fails to meet all three criteria. The standard range for Point-Measure Correlation, MNSQ, and z-Standard are as follows: $0.4 < x < 0.8$, $0.5 < MNSQ < 1.5$ and

$-2 < z < 2$. Since no value falls outside these regions therefore all the items in the pre-test are acceptable. Figure 3 shows the fits statistics for the items.

SUMMARY OF 100 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	25.6	10.0	-.14	.33	1.06	.0	.96	.2
S.D.	6.3	.0	.65	.05	.79	1.1	1.51	.7
MAX.	39.0	10.0	**1.05**	.54	4.16	2.7	9.90	3.6
MIN.	12.0	10.0	**-2.16**	.27	.16	-2.2	.09	-.7

| REAL RMSE .39 TRUE SD .53 **SEPARATION 1.36 Person RELIABILITY .65**
| MODEL RMSE .33 TRUE SD .56 SEPARATION 1.70 Person RELIABILITY .74
| S.E. OF Person MEAN = .07

Person RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .66

Fig. 1. Summary statistics for individual (person).

SUMMARY OF 10 MEASURED (EXTREME AND NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	255.6	100.0	.44	.27				
S.D.	128.7	.0	1.62	.49				
MAX.	486.0	100.0	**4.38**	1.75				
MIN.	100.0	100.0	**-1.99**	.08	.75	-1.7	.47	-1.2

REAL RMSE .57 TRUE SD 1.51 **SEPARATION 2.67 Item RELIABILITY .88**
MODEL RMSE .56 TRUE SD 1.51 SEPARATION 2.68 Item RELIABILITY .88
S.E. OF Item MEAN = .54

Item RAW SCORE-TO-MEASURE CORRELATION = -.85
900 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1315.31 with 789 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): 1.0704

Fig. 2. Summary statistics for items (questions).

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL 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
7	100	100	4.38	1.75	**MAXIMUM MEASURE**				**.00**	.00	100.0	100.0	4(iv)
10	132	100	1.16	.13	1.10	.5	**.81**	**-.2**	**.36**	.34	76.0	75.0	6(ii)
6	134	100	1.13	.12	1.13	.6	**.47**	**-1.1**	**.45**	.35	79.0	75.1	4(iii)
5	192	100	.55	.09	.75	-1.7	**.63**	**-1.2**	**.61**	.53	49.0	47.5	4(ii)
3	205	100	.45	.08	1.08	.6	**.94**	**-.1**	**.55**	.56	44.0	43.9	3
9	223	100	.33	.08	1.12	.8	**1.31**	**1.1**	**.44**	.59	34.0	38.6	6(i)
2	260	100	.11	.08	.83	-1.3	**.89**	**-.4**	**.65**	.62	33.0	31.4	2
8	384	100	-.63	.08	.99	.0	**.71**	**-1.0**	**.66**	.63	49.0	47.1	5
1	440	100	-1.10	.11	1.55	2.2	**2.09**	**2.1**	**.46**	.57	61.0	60.7	1 CO1
4	486	100	-1.99	.20	1.32	.8	**1.20**	**.5**	**.35**	.37	91.0	89.7	4(i)
MEAN	255.6	100.0	.44	.27	1.10	.3	1.00	.0			57.3	56.6	
S.D.	128.7	.0	1.62	.49	.23	1.1	.46	1.0			19.5	18.5	

Fig. 3. Item measure for fit statistics.

Students' problem solving skills and item difficulty were mapped in the Person-Item Distribution Map (PIDM). This map is also known as Wright Map. PIDM is shown in Fig. 4. The performance of all six questions spread on a logit scale from 1.16 to -1.99 where the most difficult item and the simplest item were laid out on top of the scale.

On the left of PIDM, each student is named by their metric number. The right of PIDM are the question number for the pre-test. 4 (iii) shows question 4 part (iii). From Fig. 4, 55% (n=55) students were measured above item mean, μ_{item}, while 45% (n=45) students were under μ_{item}.PIDM shows that pre-test questions / items can be divided into three categories, namely easy, moderate and difficult. Question labelled by 4 (iv) were the most difficult item to be answered by students while question 4(i) is the most easy question to be solved.

During correlations and comparisons between person (student) and item, it is noticed that none of the students can answer the most difficult item 4 (iv), and one student A150457, could not answer the easiest question, 4 (i). There exist gaps between the two questions, 4 (i) and 1 shown by the vertical arrow which lies on the 'easy' category. The gap indicates that the students find it hard answering these two questions in the pre-test.

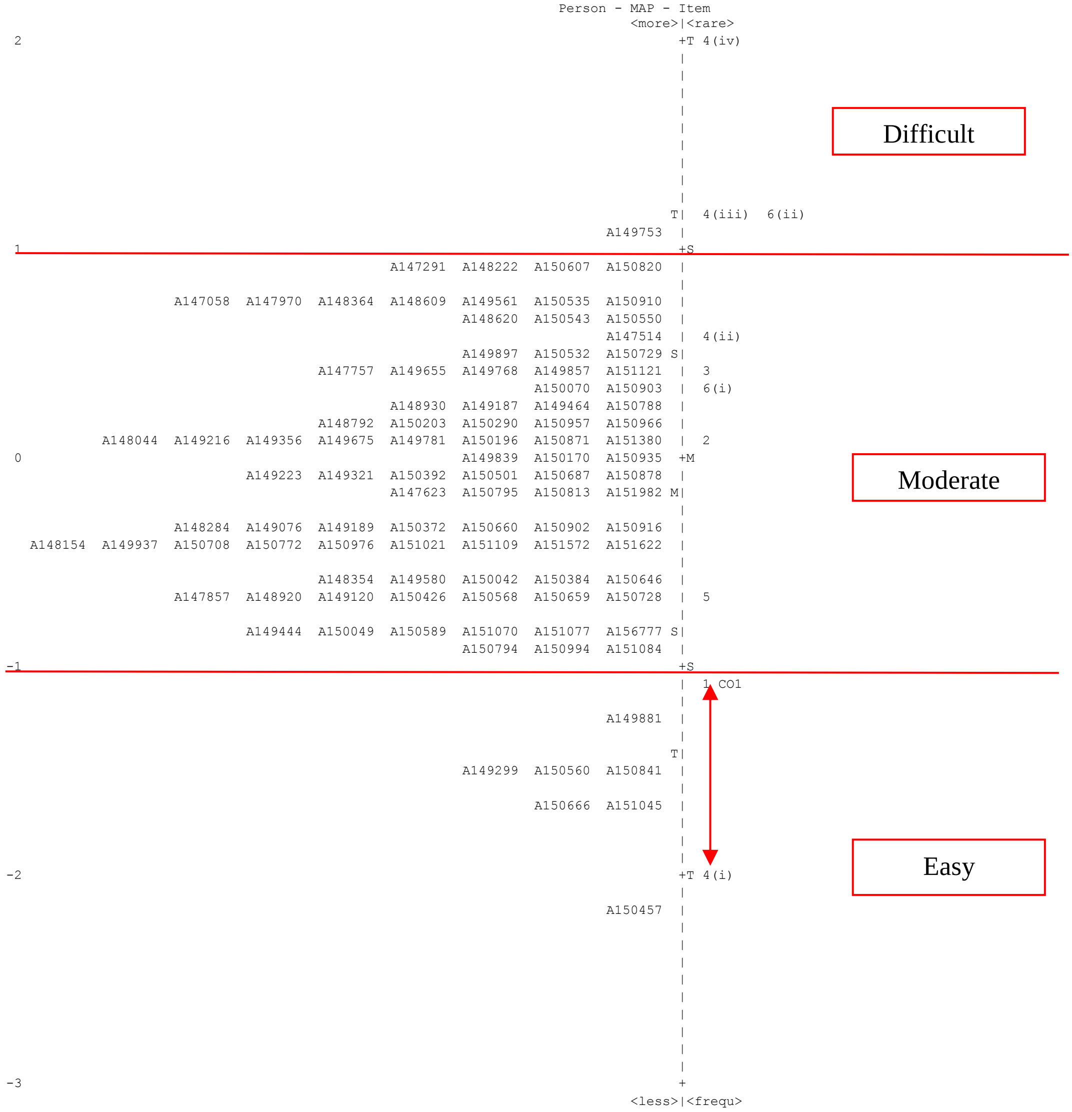

Fig. 4. Person-item distribution map.

4. Conclusions

From the Rasch analysis, the pre-test questions/items for KKKQ2123 Differential Equations were well designed and considered reliable as there is no question/item labelled as non-fit item. The pre-test questions are aimed to explore and to validate the need for student appropriate test questions. The individual separation manages to divide students into two groups, namely students with average performance and students with poor performance. The items can be segregated into three groups difficult, moderate and easy. No items fall out of the acceptable region of point-measure correlation, outfit mean square and outfit z-standard. Therefore, all the pre-test questions are acceptable. All the CO and PO have tested well in each of the questions. In conclusion, Rasch Measurement Model is vital for testing reliability of an instrument, as in this study the pre-test questions of Differential Equations. For further research, it is recommended to compare the pre-test questions with the final questions of KKKQ2123 Differential Equations.

Acknowledgement

The authors wish to express gratitude towards Universiti Kebangsaan Malaysia (UKM) under the grants AP-2015-015, STEM-2014-007 and PTS-2014-036 for supporting the research

References

1. Aziz, A.A.; Zaharim, A; Fuaad, N.F.A.; and Nopiah, Z.M. (2013). Students' performance on Engineering Mathematics applying Rasch measurement model. *Information Technology Based Higher Education and Training (ITHET)*, 1- 4.
2. Aziz, A.A.; Mohamed, A.; Zaharim, A.; Zakaria, S.; Ghulman, H.A.; and Masodi, M.S. (2008). Evaluation on information professionals competency face validity test using Rasch model. *Proceeding of the 5th WSEAS/IASME International Conference on Engineering Education (EE'08)*, 396-403.
3. Draugalis, J.R.; and Jackson, T.R. (2004). Objective curricular evaluation: Applying the Rasch model to a cumulative examination. *American Journal of Pharmaceutical* Education, 68(2), 1-12.
4. Felder, R.M.; and Brent, R. (2004). The ABC's of engineering education: ABET, Bloom's taxanomy, cooperative learning, and so on. *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition.*
5. Fuaad, N.F.A.; Nopiah, Z.M.; Tawil, N. M.; Othman, H.; and Asshaari, I. (2014). Identifying item difficulties of pre-test using Rasch measurement model: A case study. *Proceedings of the 5th International Conference on Education and Educational Technologies, Computation and Technology in Modern Education.* Kuala Lumpur, Malaysia, 98-101.
6. Fuaad, N.F.A.; Nopiah, Z.M.; Tawil, N.M.; Othman, H.; Asshaari, I.; Osman, M.H.; and Ismail, N.A. (2014). Blooms' separation of the final exam of

Engineering Mathematics II: Item reliability using Rasch measurement model. *Proceedings of the 3rd International Conference on Mathematical Sciences*. Kuala Lumpur, Malaysia, 508-513.

7. Ahmad, K.; Ali, N.M.; and Zainudin, S. (2011). An improved course assessment measurement for analyzing learning outcomes performance using Rasch model. *Procedia-Social and Behavioral Sciences*, 18, 442-449.
8. Nopiah, Z.M.; Jamalluddin, M.H.; Ismail, N.A.; Othman, H.; Asshaari, I.; and Osman, M. H. (2012). Reliability analysis on examination questions in a mathematics course using Rasch measurement model. *Sains Malaysiana*, 41 (9), 1171-1176.
9. Nopiah, Z.M.; Fuaad, N.F.A.; Rosli, N.S.; Ismail, N.A.; and Othman, H. (2013). Predicting the performance of the Diploma Engineering students using the pre-test method. *Proceeding of the 6th International Forum on Engineering Education (IFEE 2012), Procedia-Social and Behavioral Sciences*, 102, 153-157.
10. Osman, S.A.; Naam, S.I.; Jaafar, O.; Badaruzzaman, W.H.W.; and Rahmat, R.A.A.O.K. (2012). Application of Rasch model in measuring students' performance in Civil Engineering Design II course. *International Conference on Teaching and Learning in Higher Education (ICTLHE 2012) in conjunction with RCEE & RHED 2012, Procedia-Social and Behavioral Sciences*, 56, 59-66.
11. Wright B.D. (1993). Logit is? *Rasch Measurement Transaction*, 7(2), 288.