Pertanika J. Soc. Sci. & Hum. 25 (S): 399 - 408 (2017)

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Engineering Education Final Examination Assessment using Rasch Model

Nurul Syuhadah Khusaini, Nor Hayati Saad*, Nurul Hayati Abdul Halim, Salmiah Kasolang and Bulan Abdullah

[1]*Mechanical Engineering Faculty, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

ABSTRACT

The bridge between teaching and learning is different for each student, hence an assessment method is needed to gauge understanding. For Engineering Education, the assessment must be designed based on the Engineering Accreditation Council (EAC) requirements, which fulfills the Outcome-based education (OBE) learning approach. Rasch Model offers an approach that provides empirical evidence on each student's true ability with respect to the difficulty of the final examination questions. This paper explores the wonders of Rasch Model to assess the students' performance in the final examination, as well as the validity of the assessment instrument for a core course in Mechanical Engineering.

Keywords: Assessment, engineering education, mechanical engineering, OBE measurement, Rasch Measurement Model

ARTICLE INFO

Article history:
Received: 15 September 2016
Accepted: 30 December 2016

E-mail addresses:
nurulsyuhadah2089@salam.uitm.edu.my (Nurul Syuhadah Khusaini),
norhayatisaad@salam.uitm.edu.my (Nor Hayati Saad),
hayatihalim@salam.uitm.edu.my (Nurul Hayati Abdul Halim),
salmiahk@salam.uitm.edu.my (Salmiah Kasolang),
bulan_abd@salam.uitm.edu.my (Bulan Abdullah)
* Corresponding author

INTRODUCTION

It is known from experience that every student learns differently, and the outcome of those learning experiences can be assessed through examinations. In Engineering Education (EE) in Malaysia assessment is based on the Outcome-based Education (OBE) learning development (Aziz, Masodi, Ibrahim, Omar, & Zaharim, 2013).

Assessment may be conducted in many ways. There are quizzes, assignments, and of course, intermediate examination and final examination. Walvoord quoted

ISSN: 0128-7702

assessment as "the systematic collection of information about student learning, using the time, knowledge, expertise, and resources available, in order to inform decision about how to improve learning" (Walvoord, 2004). However, Saidfudin defined assessment in an interesting way, where an assessment should have the value of information for a process to be improved, and should provide evidence which is credible, suggestive, and applicable to decisions that need to be made by relevant parties (Saidfudin, et al., 2010).

METHODS

This paper focuses on the performance for 467 third semester students in the final examination of a course named AA. This course is among the five courses with a high failure rate of fifty-two percent. This study serves to assist the management's action plan to overcome this alarming trend. AA is a core subject in Mechanical Engineering and is assessed through assignments, intermediate exams, and a final examination. However, in this paper, the students' performance on COs for the course looks at the final examination assessment.

There are two COs involved in the final examination, namely CO1 and CO2. The description for each CO is shown in the Table 1.

Table 1
Description of each CO

Course Objectives	Description
CO1	Apply the basic concepts and fundamental principles in dynamics of particles and rigid bodies
CO2	Analyze and solve kinematics and kinetics of particles and rigid bodies' problems.

For the AA COURSE final examination, there are five questions, each question has two sub-questions namely (a) and (b). For the sub question (a), the students will be tested on CO1, while for the sub question (b), the students will be tested on CO2. A simple and conventional Continuous Quality Improvement (CQI) analysis has

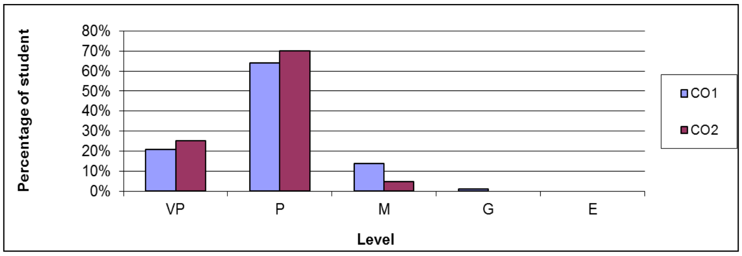

Figure 1. Histogram for students' performance for each category

been conducted for the final examination performance, where students are grouped to five categories, based on their performance. The five categories are: Excellent (E); Good (G); Mediocre (M); Poor (P); and Very Poor (VP). Figure 1 and Table 2 highlight the students' performance. From Figure 1 and Table 2, clearly it can be seen that a total of four hundred and forty-nine students did not achieve the satisfactory level for CO1 and a total of five hundred and four students did not achieve the satisfactory level for CO2.

Table 2
Data for students' performance for each category

Rating	Range	CO1		CO2	
VP	0 – 19%	110	21%	134	25%
P	20 – 49%	339	64%	370	70%
M	50 – 69%	74	14%	25	5%
G	70 – 84 %	6	1%	0	0%
E	85 – 100%	0	0%	0	0%

However, the analysis for the CQI conducted only limited to analysis such as count, percentage, median and mode, since the data is in the form of raw score. Assessment using raw score fails to portray the real story of what is going on behind those marks as it works by adding the marks obtained from all correct answers, and the total marks treated as the measure of performance, or maybe for some, it is a measure of ability. However, this is not true because the ability of students depends on their responses towards the difficulty level of the testing questions.

Table 3
Example from students' result database

STUDENT	1(a)	1(b)	2(a)	2(b)	3(a)	3(b)	4(a)	4(b)	5(a)	5(b)	TOTAL
M01	0.0	4.0	0.0	3.0	0.0	9.0	0.0	7.0	4.0	1.0	28.0
F19	4.0	3.0	2.0	4.5	0.0	1.0	0.0	11.5	0.0	2.0	28.0
M78	2.0	3.0	5.0	5.0	0.0	4.0	0.0	5.0	1.0	3.0	28.0

In every test and assignment, the questions were constructed based on levels of difficulty, and the raw score method fails to identify what is the student's ability with respect to different questions. Table 3 documented the example of the data extracted from the complete students' result database. It can be seen that, all three students have the same total marks, which is 28.0. But the marks obtained from each question is different. This resulted in misinterpretation of student's ability, since the difficulty level for each question is not clearly determined.

There will always be students who obtain good marks in theoretical questions, but do badly in calculation. How are these students to be graded? Rasch Model on the other hand, offers a better and thorough analysis of the final examination assessment, and students true ability is assessed concurrently with the final examination's difficulty level. Rasch Model originated with a Danish mathematician named Georg Rasch circa 1952 to 1960 (Tymms, 2012). Rasch Model is now acceptable worldwide by the medical and food and beverages industry among others (Dougherty, Nichols, & Nichols, 2011; Khusaini, Jaffar, & Yusof, 2014). Rasch Model works based on two theorems. The two underpinning theorems are (Khusaini, et al., 2014; Bond & Fox, 2015);

1) Easy items are most likely to be answered correctly by any person of any given ability;
2) High ability students have greater likelihood to answer all items correctly, or to complete any given task.

With respect to the two theorems, Rasch Model arranges the responses based on item (question) difficulty, and person's ability. For item's difficulty, it is arranged in columns, while for person's ability, it is arranged in rows. The two theorems permit further analysis to be in the form of probability, hence Rasch Model is also known as a probabilistic model (Saidfudin, et al., 2010).

Rasch Model sees "a turn of an event as a chance, or a likelihood of happenings". For instance, what is the chance of a student with an unknown ability to answer a question correctly? It will be 50:50. However, what is the chance of a high ability student to answer a question correctly? The chance will certainly increase, say 70:30 (Saidfudin, et al., 2010). This statement reflects the essence of ratio data (Stevens, 1946). Converting an ordinal or nominal data in hand, to ratio data is crucial, since it allows statistical analysis such as mean and standard deviation to be conducted (Stevens, 1946; Aziz et al., 2013).

RESULTS AND DISCUSSION

The final examination results were analyzed using Winstep, which is one of the Rasch Model software. The focus on the analysis is on the Summary Statistics and Person-Item Distribution Map.

Summary Statistics

The Summary Statistics describe the overall reliability of both respondents (students) and the questions (items) for the AA course final examination, called the instrument. A total of 467 students' responses were analyzed. The Cronbach-Alpha value of this test is 0.52, which is lower than the minimum requirement of 0.7 (Aziz, Masodi, & Zaharim, 2013). This indicates that this instrument may face hurdles in assessing the students' performance. It is normal to discard the data if the Cronbach Alpha value does not meet the requirement. However, Rasch Model offers another two paradigms of reliability, which is the item and person reliability.

Table 4
Item reliability

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	2153.2	471.0	.00	.01	.96	-1.1	.99	-.5
S.D	834.0	.0	.15	.00	.43	5.4	.39	4.8
MAX	3612.0	471.0	.31	.02	2.00	9.9	1.96	9.9
MIN	834.0	471.0	-.22	.01	.40	-9.9	.49	-8.0
REAL	RMSE .01 TRUE SD		.15	SEPARATION 10.07	Item	RELIABILITY	.99	
MODEL	RMSE .01 TRUE SD		.15	SEPARATION 10.61	Item	RELIABILITY	.99	
S.E. OF ITEM MEAN = .05								

As seen in Table 4, the item reliability marks an impressive value of 0.99. This value shows that the questions in this instrument are sufficient in terms of size, as well as having different difficulty level to assess the students' ability. In laymen's term, a good instrument should have ranges of item difficulty, to assess students for any ability. An item reliability of 0.99 fulfils the minimum value requirements of 0.7.

Table 5
Person reliability

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	46.1	10.0	-.18	.10	1.03	.1	.99	.0
S.D	16.6	.0	.16	.03	.47	.9	.46	.9
MAX	93.0	10.0	.14	.51	3.17	2.5	3.53	2.7
MIN	2.0	10.0	-1.42	.08	.20	-3.1	.21	-2.9
REAL	RMSE.11 TRUE SD		.11	SEPARATION 1.03	Person RELIABILITY		.51	
MODEL	RMSE.10 TRUE SD		.12	SEPARATION 1.22	Person RELIABILITY		.60	
S.E. OF ITEM MEAN = .01								

On the other hand, in Table 5, the person reliability is recorded at a poor 0.51, which is lower than the requirement, which is 0.7. The person reliability is an indicator of response consistency. A reliability of 0.51 shows that incongruity exists in the response pattern. It is worth noting that the person separation is at 1.03. Person separation is related to respondents' classification. A low value for person reliability and person separation may imply that the instrument is not able to distinguish between high and low performers. The analysis should be able to differentiate and categorize high and low performers; hence, the Person Separation should be at minimum of 2. However, there are insufficient good items to critically differentiate the respondents into significant groups based on their ability. In Table 5, the person mean is logged with negative value, which is -0.18 logit. A negative person mean shows that the students did not do well in this examination, which means, they perceived the final examination to be difficult. .

Thus, it can be seen that, while the Person Reliability positions at a value of 0.51, the Item Reliability is at an impressive value of 0.99. It can be deduced that; the items are sufficient to measure what is supposedly to be measured. The only aspect that needs further justification, is indeed the person (students). This is also proved that a low value of the Cronbach Alpha does not necessarily mean that the testing instrument is at stake, it may also cause by the students responses.

Person-Item Distribution Map (PIDM)

Figure 2 and Figure 3 represents the unique Person-Item Distribution Map (PIDM). Figure 2 arranges the item according to CO, while Figure 3 arranges the item according to question number. The PIDM is constructed in the form of a measurement ruler (the dashed-line). The measurement ruler measures the person ability, and

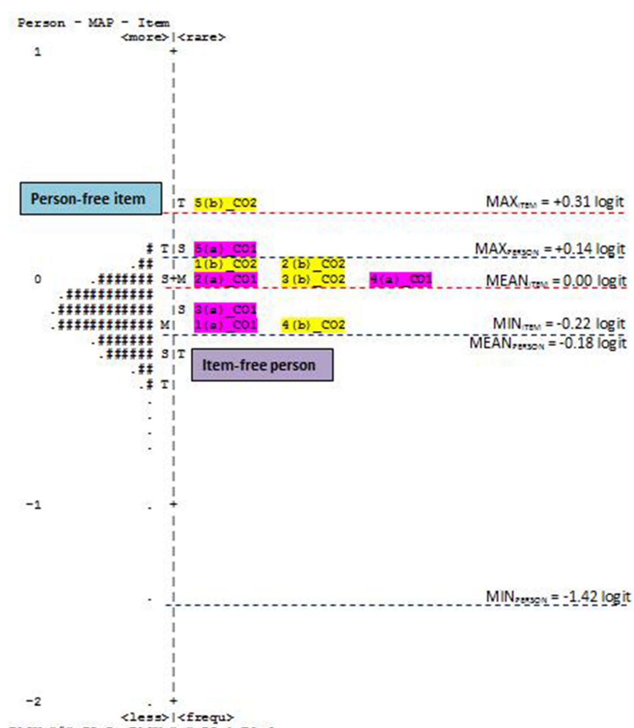

Figure 2. The PIDM according to CO

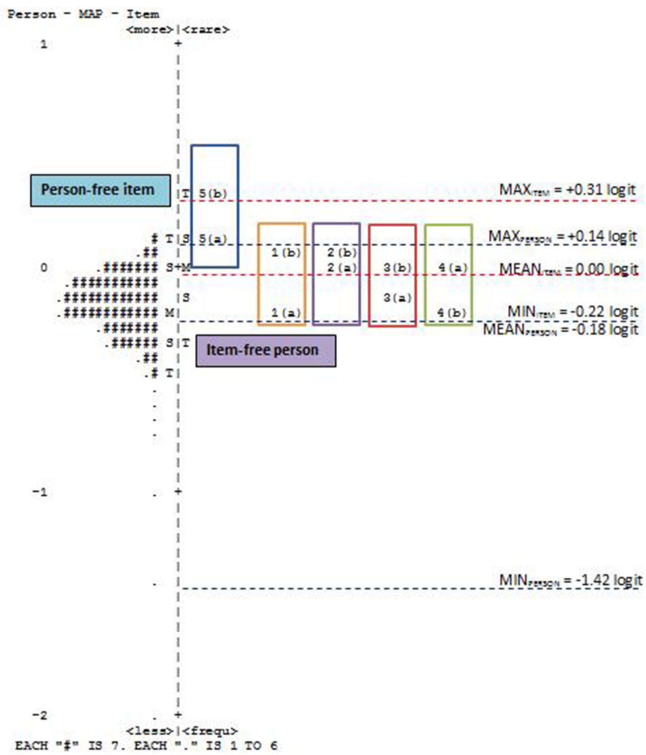

Figure 3. The PIDM according to question number

item difficulty at the same time, using the Logit measurement unit. The PIDM also displays the location of each item based on the difficulty level, and the location of the respondents based on the ability level. The person's data is on the left side of the ruler, while the item's data will be on the opposite side. The least able person / easiest item will be at the bottom of the map, while the most able person / hardest item will be at the top.

The maximum person measure is recorded at +0.14 logit, while the minimum person measure is recorded at -1.42 logit. In laymen's term, it can be said that, the smartest student is as smart as +0.14 logit for this particular examination. The similar goes to the poorest student. He/she is as poor as -1.42 logit. From Figure 2, looking at the CO1 items, the easiest CO1 item is item 1(a) with a measure of -0.20 logit (S.E = 0.01), while the hardest CO1 item is item 5(a) with a measure of +0.14 logit (S.E = 0.02). Also, it can be seen from Figure 2, item 2(a) and item 4(a), have the same level of difficulty. Good items should have different level of difficulty to measure different ability, hence item 2(a) and item 4(a) should be scrutinized and redesigned.

For item CO2, the easiest item is item 4(b) with a measure of -0.22 logit (S.E

= 0.01), while the most difficult item for CO2 is item 5(b) with a measure of +0.31 logit (S.E = 0.02). It is also worth noting that item 1(b) and 2(b) have the same level of difficulty. These items too need further scrutiny.

On the other hand, item 5(b) is not measuring any of the students' ability, since everyone perceived this question as extremely difficult and could not be answered correctly. This item can be regarded as Person-free Item. Hence, this item should be investigated further to justify students' responses towards this questions. Also, there is Item-free Person, where there is no item to measure the students' ability at the bottom of the PIDM. The overall question for course AA should be redesigned to measure all range of students' ability.

In Table 3 all questions have the same trend except, question 4. All of the students perceived question 1(a), 2(a), 3(a), and 5(a) as easier than question 1(b), 2(b), 3(b) and 5(b). However, for question 4, the students' responses show that 4(b) is easier than 4(a), which can be questioned since 4(a) is related to CO1 (theoretical based question), while 4(b) is related to CO2 (problem based learning/ calculation). Clearly, the students are having problems in understanding the theoretical part (CO1) of that particular question / chapter.

It is also noteworthy that more than 120 students were unable to answer the easiest item in the final examination, thus calling for more interventionist programs to be made available.

CONCLUSION

The AA course is one of the core courses in Mechanical Engineering. To ensure the competency of each student in the subject a reliable assessment is needed.

Rasch Model is a powerful analysis method to measure not only the students' ability but also the means of assessment as well. From our analysis it was found around 25 percent of students with a Mean of -0.18 logit are unable to answer even the easiest items while none were able to answer question 5(b) correctly. The unexpected students responses should be further investigated to identify the root causes to educators to identify innovative teaching and learning approaches.

There is also a discrepancy in pattern response, since the person reliability is only at a poor reading value of 0.55. The ability of the students is more or less the same (Person Separation = 1.03). The smartest student is at a measure of +0.14 logit, while the poorest student is at -1.42 logit. With a small range of logit i.e. 1.56, the students' ability could not be distinguished clearly. It is important to identify different range of students to reflect on their achievement towards the course outcome (CO) and to gauge the success of the teaching and learning approaches.

Looking at the item, it possesses a reliability index of 0.99 and it is considered excellent. The easiest item is item 1(a) and item 4(b) but somehow, there are quite a number of student fails to answer it correctly. In conclusion, the management should identify the root causes of the overall analyses behaviour. It is important for management to design a programme for students' performance improvement, upgrading the teaching and learning approaches and to restructure the assessment tool to consider various range of students' ability.

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of the Research Management Centre, Universiti Teknologi MARA Shah Alam in providing the Research Grant (Project Number: 600-RMI/DANA/5/3/ ARAS (40/2015)) to undertake this research.

REFERENCES

Aziz, A. A., Masodi, M. S., & Zaharim, A., (2013). A*sas Model Pengukuran Rasch : Pembentukan Skala & Struktur Pengukuran*. 1st ed. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Aziz, A. A., Masodi, M. S., Ibrahim, F. M., Omar, M. Z., & Zaharim, A. (2013). Insights into engineering education learning outcome's assessment with rasch model. *Research Journal of Applied Sciences, Engineering and Technology, 6*(19), 3520-3526..

Bond, T., & Fox, C. M., (2015). *Applying the Rasch Model.* 3rd ed. s.l.:Routledge.

Dougherty, B. E., Nichols, J., & K.Nichols, K., (2011). Rasch Analysis of the Ocular Surface Disease Index (OSDI). *In*vestigative Ophtalmology and Visual Science, 52(12), 8630-8635.

Khusaini, N., Jaffar, A., & Yusof, N., (2014). A Survey on Lean Manufacturing Practices in Malaysian Food and Beverages Industry. *Applied Mechanics and Materials*, 564, 632-637.

Khusaini, N., Jaffar, A., & Yusof, N., (2014). A Survey on Lean Manufacturing Tools Implementation in Malaysian. *Advanced Materials Research*, 845, 642-646.

Saidfudin, M., Azrilah, A. A., Rodzo'An, N. A., Omar, M. Z., Zaharim, A., & Basri, H. (2010). Easier learning outcomes analysis using rasch model in engineering education research. In *Proceedings of the 7th WSEAS international conference on Engineering education* (pp. 442-447). World Scientific and Engineering Academy and Society (WSEAS).

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, New Series*, 103(2684), 677-680.

Tymms, P. (2012). Early History of Rasch Measurement in England. *Rasch Measurement Transaction, 26*(2), 1365.

Walvoord, B. E. (2004). *Assessment Clear and Simple : A PracticalGuide for Institutions, Departments and General Education*. 2nd ed. s.l.:John WIley & Sons.

Wright, B. (1997). Fundamental Measurement for Outcome Evaluation. *Physical Medicine and Rehabilitation :State of the Art Reviews, 11*(2), 261-288.