Journal of Applied Science and Agriculture, 9(12) Special 2014, Pages: 7-12

AENSI Journals

Journal of Applied Science and Agriculture

ISSN 1816-9112

Journal home page: www.aensiweb.com/JASA

Construct Validity: A Rasch Measurement Model Approaches

[1]Abdul Aziz, A., PhD, [2]Jusoh, M.S., PhD, [2]Omar, A.R., PhD, [2]Amlus, M.H., PhD, [3]Awang Salleh, T.S

[1]*Infrastructure University Kuala Lumpur (IUKL), Faculty of Information and Computing Technology, 43000 Kajang, Selangor, Malaysia*
[2]*University Malaysia Perlis (UniMAP), School of Technopreneurship and Business Innovation, 01000 Kangar, Perlis, Malaysia*
[3]*University Kuala Lumpur Malaysian French Institute, Technical Foundation, 43650 Bandar Baru Bangi, Selangor, Malaysia*

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form
8 July 2014
Accepted 10 August May 2014
Available online 30 August 2014

Keywords:
Construct validity, Test evaluation, Student's achievement, Rasch measurement model, Subjective items

ABSTRACT

Academic assessment circles around assessing the achievement level of the students and the teaching staff. Assessment enables staff to assess and evaluate their own practice. On the other hand, assessments of students' performance within and between modules enable evaluation and alignment of learning outcomes on the students. The students are given grades for their achievement in the subjects and thus reflecting their level of achievement. It is hence crucial to investigate on the construct validity of the test in ensuring that the test truly reflect the students' achievement, not just a test that students have to take in fulfilling their coursework. Often than not, our "expert judgment" are used to indicate the difficulty level of each item involved in the test, during construction of the test question. Truly, these judgments also influenced the students' achievement when the grading is granted on them. Judgment will be subjective depending on the judges' expectation on the students. It would be better if the 'judgment' can be done systematically with clear evidence. The students' achievement should provide indication on the ability of the students considering the difficulty of the test items. This paper demonstrate an evaluation on the construct of the test items, in ensuring that the test or the instrument truly able to measure the ability of the students. The sample comprised of engineering students from selected Malaysian higher learning institution on an engineering Mathematics subject that comprised of multiple choice and subjective questions. The construct validity is assessed using Rasch measurement model which enable clear indication of each item reliability and thus leads to construct validity of the instrument.

To Cite This Article: Abdul Aziz, A., PhD, Jusoh, M.S., PhD, Omar, A.R., PhD, Amlus, M.H., PhD, Awang Salleh, T.S., Construct Validity: A Rasch Measurement Model Approaches. ***J. Appl. Sci. & Agric.,*** *9(12): 7-12, 2014*

INTRODUCTION

Academic institutions are trying their very best effort in making sure that the program they offered are of quality and meeting industrial requirement. Ministry of Education (MOE), as in the case in Malaysia, provided guidelines and framework on how best to tailor the academic programs in achieving quality outcomes. MOE thus enforce these guidelines for all the academic programs before granting accreditations, and entrusted the Malaysian Qualifications Agency (MQA) as a regulatory entity (MQA, 2013). This is seen as a good move towards controlling the accreditation program thus ensures programs are of quality standards by all the academics institutions. It also provide framework for the institutions in standardizing their programs among other academic institutions, local and foreign.

Program are developed based on intended program outcome and hence in line with the learning outcome (LO) for each of the syllabus within the program. Development of the LO pretty much determine what is expected on the students at the end of the semester for each syllabus. However, the monitoring of the students' performance towards achieving the outcome may not only happen at the end of the semester. Progressive monitoring would add value to the progression of the students' learning process. Students' learning process would be made more effective whenever academician understand what and where to enhance. Students' performance would give indication on the students' level of mastery for a particular syllabus, in fact, in micro level may determine the mastery for specific area or topic. However, there are many factors effecting students' performance (Carole, 1992; Durlak & DuPre, 2008).

In making sure that correct performance measures are conducted (Fuchs *et al*., 1999; Figlio & Lucas, 2004; Banta, 2007) and valuable inference can be made (Wright & Mok, 2004), it is crucial that the measurement tool is reliable and valid (Baghaei, 2008) in measuring the students' ability in mathematics. In the case of learning

Corresponding Author: Abdul Aziz, A., Infrastructure University Kuala Lumpur (IUKL), Faculty of Information and Computing Technology, 43000, Kajang, Selangor.
Tel: +603-89266993/Ext: 683; E-mail: azrilah@gmail.com,

Journal of Applied Science and Agriculture, 9(12) Special 2014, Pages: 7-12

achievement, test papers, quizzes, and project work are used and the marks obtained from those tests are used to infer on the students' achievement. The higher the scores obtained it infer the higher the students' performance. However, cumulative raw scores are only counts of discrete observations, and it is not measuring students' performance (Wright & Mok, 2004). Rasch measurement model satisfies the criteria of a measurement tool where it provides a measurement unit in measuring latent traits in logits, provides equal interval on its measurement ruler, overcome missing data, allows repeatability, and able to predict future events (Wright & Mok, 2004).

Apart from using the right measurement tool, reliable instrument is crucial in determining what is meant to measure is measured. Thus, this study applies Rasch measurement model to analyze the instrument reliability and validity, hence provide valuable findings. The test is to evaluate students' performance on engineering mathematics for engineering students in one of the Malaysian academic higher learning institution. The test is to evaluate the proficiency of the students on their basic mathematics skill and knowledge in order for them to successfully be able to apply the logic in solving engineering problem.

Research Methodology:

The test comprised of three (3) sections, in which section A is on multiple choice type of questions, both section B and C are of subjective type of questions, but on section C, the students need to answer only two (2) out of three (3) questions. This paper will only focus on section B and C, of subjective questions of the test.

Section B is the subjective questions that has seven (7) questions which the students has to answer all questions. Each question is further divided into subsections, which will be treated as an item by itself. Therefore, section B has eleven (11) items. In section C, there are 3 questions which students were asked to answer two out of the three (3) questions. Similarly, each of the questions is further divided into subsections and will be treated as individual items for convenient of measurement and analysis. Altogether, there will be 19 items in the test which will be referred to as the instrument.The sample comprised of 125 engineering students. The students are from the class of September-January 2013 intake. The students were coded according to their respective student matrix number.

The items are keyed in according to their marks. However, to avoid mistreated to each of the questions due to their difference of base-marks, the marks are converted to have same base-mark of 10 for all the questions or from now on will be referred to as items. The data then is ran into WinSteps analysis software, one of the software for Rasch measurement model.

Findings:

The summary statistics revealed that the Cronbach alpha value is 0.52 which indicates that the raw data consistency in low. Further investigation showed that the person reliability yield 0.61 indicating that the students' sample involved in this study has fair ability spread when tested against the test items. The average ability is at negative (-ve) 0.30 (-0.30) logit, demonstrated that the ability of the students is below the average difficulty of the tested items. The average difficulty of the items, mean item, is set at 0.00 logit. Rasch theorised that a person have a chance of 50:50 in succeeding a given task at difficulty 0.00 logit. As the item gets harder, the odd of success is reduced; getting less chance to success (Rasch, 1960; Rasch, 1961). Therefore in the case of this test, the items are more difficult and the chance of success for the student is low (Wright & Stone, 1979; Abdul Aziz, 2009).

The summary statistics, refer to Table I, also revealed that there are two (2) students having minimum extreme score. This indicates that the students are having the lowest logit measures among the other students that are involved in this study. Practically, they did not managed to answer all the subjective questions given in the test.

Referring back to the summary statistic, Table II on the summary statistic of Items, it showed that the item reliability is at 0.93 indicating a good item difficulty spread of the items used in the test. Basically, this is to show that the items have good difficulty measurement in measuring the students' mathematics ability. The difficulty of those items are highly likely will have similar difficulty logit measurement, when given to another group of students to test their mathematical ability (Andrich, 1998).

The item and person reliability has no sufficient evidence in showing how good is the test items, but only provide indication on the difficulty and ability spread.

The low Cronbach alpha value hinted that the instrument has some disturbance, which further indicated by the low person reliability. Therefore, there is a reason to investigate further on the validity of the test items, in better revealing the reason behind the low performance of the students. The question is whether the students really having problem in their mathematics ability or it might be from the ambiguity of the test construct (Baghaei, 2008).

Table I: Summary statistic for person.

SUMMARY OF 123 MEASURED (NON-EXTREME Person									
	TOT			MODEL	INFIT		OUTFIT		
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	40.3	16.1	-0.30	0.08	0.74	-0.30	0.75	-0.20	
S.D	62.2	0.7	0.15	0.05	0.44	0.90	0.42	0.70	
MAX	481.0	19.0	-0.09	0.23	2.26	2.30	2.58	2.60	
MIN	3.0	14.0	-0.81	0.01	0.11	-2.20	0.22	-1.40	
REAL RMSE	0.09	TRUE SD	0.12	SEPARATION		1.26	Person RELIABTY .61		
MODEL RMSE	0.09	TRUE SD	0.12	SEPARATION		1.28	Person RELIABTY .62		
S.E. OF Person MEAN = .01									
MINIMUM EXTREME SCORE: 2 Person									

Table II: Summary statistic for item.

SUMMARY OF 19 MEASURED (NON-EXTREME Item									
	TOT			MODEL	INFIT		OUTFIT		
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	260.7	106.1	0.00	0.03	0.85	-0.40	0.72	-0.80	
S.D	248.8	32.6	0.18	0.03	0.37	1.20	0.23	0.90	
MAX	1005.0	125.0	0.43	0.15	1.88	2.40	1.31	0.80	
MIN	8.0	34.0	-0.16	0.01	0.21	-2.90	0.42	-2.80	
REAL RMSE	0.05	TRUE SD	0.17	SEPARATION		3.77	Person RELIABTY .93		
MODEL RMSE	0.05	TRUE SD	0.17	SEPARATION		3.77	Person RELIABTY .93		
S.E. OF Item MEAN = .04									

Profiling of test items:

The Wright map or the map managed to show the distribution of all person (students) and test items on the logit measurement ruler. Refer to figure 1 for the distribution of all students and all test items involved in this study.

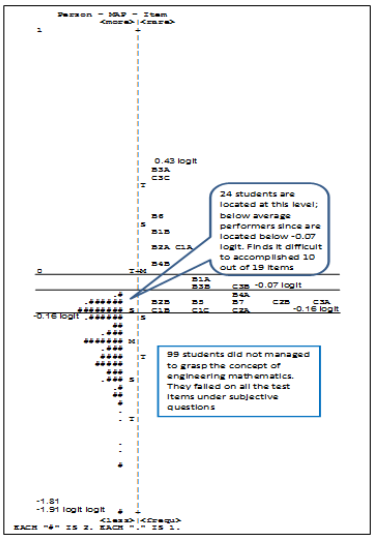


Fig. 1: Wright Map-Distribution of Person and Item.

Ten (10) items are considered very difficult for the students to accomplish since all the ten items are located above the highest persons' ability logit. This simply indicates that even the smartest student in this cohort are not able to perform those 10 items successfully. The difficulty level of the 10 items are above the ability logit of the smartest student at -0.09 logit.

In investigating the test items, to see whether the items are 'behaving' as expected, Rasch provide fit statistics. The three (3) parameters to investigate in making sure that the test items are fit. Fit statistics are used in checking the unidimensionality of the instrument (Wright & Linacre, 1994; Baghaei, 2008), whether the instrument is measuring what it is supposed to measure. Fit statistics also managed to reveal the direction of skewness of the responses in determining the proximity of those responses against the ideal model of human behavior (Linacre, 2004).

Table III listed all the items involved in the test in descending order of their respective difficulty logit measures. The highest logit measure or the most difficult item is at +0.43 logit for item B3A, where else the lowest measured item is at -0.16 logit for item C2A, which is the easiest item.

Table III: Item measure table.

ENTRY	MEASURE	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MSQ	OUT.ZSTD	PTME	NAME
20	0.43	30	0.07	0.82	-0.44	0.67	-0.83	0.12	B3A
34	0.38	8	0.15	0.50	-0.67	0.42	-0.70	0.13	C3C
25	0.23	64	0.05	0.68	-1.64	0.54	-2.01	0.19	B6
17	0.15	87	0.04	0.73	-1.68	0.58	-2.13	0.21	B1B
27	0.11	100	0.04	0.80	-1.35	1.01	0.13	0.00	C1A
18	0.09	113	0.04	0.80	-1.52	0.84	-0.83	0.10	B2A
23	0.03	145	0.04	0.67	-2.95	0.60	-2.81	0.27	B4B
16	-0.02	173	0.03	0.75	-0.58	0.66	-1.43	0.24	B1A
21	-0.07	198	0.02	0.62	0.08	0.89	-0.07	0.20	B3B
33	-0.07	42	0.04	0.21	-0.39	0.52	-0.93	0.31	C3B
22	-0.11	235	0.01	0.47	-0.87	0.63	-0.67	0.29	B4A
24	-0.13	381	0.01	0.80	-0.83	0.57	-0.98	0.16	B5
26	-0.13	345	0.01	1.01	0.13	0.69	-0.59	0.16	B7
32	-0.13	105	0.02	0.65	-0.26	0.49	-1.14	0.21	C3A
19	-0.14	411	0.01	1.19	0.90	1.31	0.78	0.12	B2B
31	-0.14	297	0.01	0.88	-0.44	0.57	-0.69	0.15	C2B
29	-0.15	556	0.01	1.18	0.83	0.72	-0.55	0.20	C1C
28	-0.16	1005	0.01	1.88	2.39	1.05	0.25	0.25	C1B
30	-0.16	658	0.01	1.54	1.67	0.96	0.08	0.22	C2A

The fit parameters are those columns highlighted in the square box in Table III. The point measure correlation (PTME) revealed the direction of the responds' pattern, the outfit mean square (OUT.MSQ) and the outfit z-standard (OUT.ZSTD) [16], [17]. In this study, all the point measure correlation value is of positive value, indicating that all the items are behaving as expected. The outfit mean square should fall within 0.5 to 1.5 and thus, all the items are within the range. The outfit z-standard should be within +2.0 to -2.0, showing only two (2) items out of the range; item B1B and B4B. However, in order for the items to be considered as misfit item, all the three (3) fit parameters should be out of the fit ranges, in which they are not in this case. Thus, the two (2) items can be considered fit, but the items need to be closely monitored during class session in making sure that students be able to master them.

Profiling of respondents:

Fit statistics can revealed the behavior of the test items, subsequently, be able to reveal the behavior of the students too. In this case the three (3) fit parameter value of the person measure table is investigated. Refer to Table IV of the person measure table.

The person measure table which lists out all the person in descending order of their respective logit measures. Table IV list out those person with suspected fit parameter value, out of the rest of the students involved in this study.

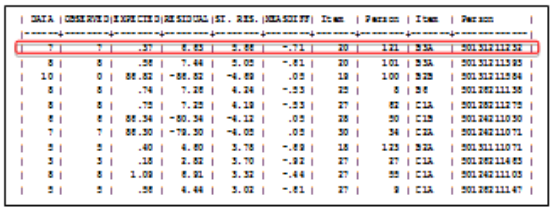
Table IV: Person measure table.

ENTRY	MEASURE	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MSQ	OUT.ZSTD	PTME	NAME
100	-0.09	481	0.01	2.13	1.29	1.85	0.96	0.58	50131211584
34	-0.11	280	0.01	2.26	2.09	1.94	1.12	0.41	50124211071
36	-0.11	303	0.01	2.13	1.81	1.67	0.91	0.41	50124211482
50	-0.11	217	0.01	2.01	2.25	1.91	1.12	0.35	50124211030
52	-0.12	132	0.01	1.77	1.93	1.77	0.98	0.21	50124211327
24	-0.13	121	0.01	1.86	2.03	1.65	0.89	0.19	50126211355
101	-0.17	33	0.02	0.22	-0.36	2.30	1.15	**-0.38**	50131211393
121	-0.28	23	0.09	1.04	0.24	**2.58**	**2.59**	**-0.18**	50131211232
8	-0.31	20	0.10	1.22	0.80	1.72	1.41	**-0.03**	50126211138
55	-0.33	19	0.10	1.37	1.20	1.46	1.04	**-0.06**	50124211103
14	-0.41	14	0.11	0.63	-0.99	0.72	-0.38	**-0.02**	50126211214
9	-0.5	8	0.14	0.68	-0.46	0.90	0.11	**-0.02**	50126211147
123	-0.61	6	0.16	0.70	-0.25	1.04	0.35	**-0.06**	50131111071
27	-0.81	3	0.23	0.54	-0.15	0.94	0.36	**-0.18**	50126211463

All the 14 person or students' pattern of responds are suspected to be misfit, due to exceeded the acceptable value of outfit mean square, or z-standard or having negative point measure correlation value. Those are highlighted with circle. However, it has to disqualify all three (3) parameters in order for the responds to be considered misfit, hence, only person coded '50131211232' is considered misfit. This student is not responding as expected which is read from his negative point measure correlation value of -0.18, his outfit mean square value of 2.58 which higher than 1.5 and z-standard value of 2.59 which exceed positive (+ve) 2.00.

Rasch provide option to investigate further on which item does the student misbehaving for him to be considered misfit. Table of most unexpected responses, refer to Table V, revealed that he managed to respond to the most difficult item in the test, item B3A at difficulty +0.43 logit. Ironically, he is of mediocre ability of -0.28 logit but managed to accomplish the most difficult item.

Table V: Most Unexpected Responses.



DATA	OBSERVED	EXPECTED	RESIDUAL	ST. RES.	MEASURE	Item	Person	Item	Person
7	7	.37	6.63	5.66	-.71	20	121	B3A	501312112 52
8	8	.56	7.44	5.05	-.61	20	101	B3A	501312113 93
10	0	86.62	-86.62	-4.69	.05	19	100	B2B	501312113 84
8	8	.74	7.26	4.24	-.53	25	6	B6	501282111 58
8	8	.75	7.25	4.19	-.53	27	82	C1A	501282112 75
8	8	86.34	-80.34	-4.12	.05	26	50	C1B	501242110 50
7	7	86.30	-79.30	-4.05	.05	30	34	C2A	501242110 71
5	5	.40	4.60	3.76	-.69	16	125	B2A	501311110 71
3	3	.16	2.82	3.70	-.92	27	27	C1A	501282114 65
8	8	1.09	6.91	3.52	-.44	27	55	C1A	501242111 05
5	5	.56	4.44	3.02	-.61	27	9	C1A	501282111 47

The possible explanation for this scenario is, he might have like the topic so much that he can accomplished it. This would give an opportunity for the lecturer to investigate the matter further. Thus, Carole, 1992 and Durlak & DuPre, 2008 remarks on the many factors influencing the students' performance is reflective here.

Conclusion:

Monitoring of performance managed to provide inference to the students' level of mastery in engineering mathematics during their engineering technology program. It is crucial in making sure that the instrument used truly measuring the ability of the students, and not only summarizing the total marks for each students. By taking only total raw marks into the consideration of performance might not truly reflect the true ability of the student. Assessment and evaluation should also consider the difficulty of each task given during the assessment, then only the evaluation is deemed worthy and true.

On top of including difficulty of tasks in the evaluation process, the instrument used in assessing the students is also critical to ensure correct assessment is made. Thus, by applying Rasch measurement model and analysis, it provides a convenient and systematic approach of evaluation on the reliability and validity of the instrument used. It provides further means of investigating in finding out the specific issue to address when any irregular pattern or misbehaving of either the item or the person.

This study revealed that the test items are difficult for the students to accomplish and their mean person ability is only at -0.30 logits. The students' ability measure are much lower than the mean difficulty of the items. The Wright map showed the distribution of the person and items on the logit measurement ruler, making it simpler to visualize the persons' ability and difficulty of tasks.

The findings revealed that the students find it difficult to understand the domain of a piece-wise function. They face difficulty to differentiate the function because it is given in a composite function form. Even computing a composition function is a problem to most of the students. Apart from that, quotient of function, differentiation of a quotient, simplifying a trigonometry expression is always a problem.

REFERENCES

Abdul Aziz, A., 2009. *Rasch Model Fundamentals: Scale Construct and Measurement Structure*. ISBN:978-967-5772-00-9 CIP:519.54 : Azrilah

Andrich, D., 1998. *Rasch Models for Measurement*. Beverly Hills: Sage Publications.

Baghaei, P., 2008. The Rasch Model as a Construct Validation Tool. *Rasch Measurement Transactions*, 22(1):1145-1146.

Banta, T.W., 2007. *A Warning on Measuring Learning Outcomes*. Inside Higher Education.

Carole, A., 1992. Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3): 261-271. doi:10.1037/0022-0663.84.3.261.

Durlak, J.A. and E.P. DuPre, 2008. Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology*, 41: 327-350.

Figlio, D.N. and M.E. Lucas, 2004. Do High Grading Standards Affect Student Performance? *Journal of Public Economics*, 88: 1815-1834.

Fuchs, L.S., D. Fuchs, K. Karns, C.L. Hamlett and M. Katzaroff, 1999. Mathematics Performance Assessment in the Classroom: Effects on Teacher Planning and Student Problem Solving. *American Educational Research Journal*, 36(3): 609-649.

Linacre, J.M., 2004. Test Validity and Rasch Measurement: Construct, Content, etc. *Rasch Measurement Transactions*, 18(1): 970-971.

Linacre, J.M., 2006. *A User's Guide to Winsteps Rasch-Model Computer Programs*.

Malaysian Qualifications Agency (MQA), 2013. *Official Portal of Malaysian Qualifications Agency* [Online]. Available: http:// http://www.mqa.gov.my/

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

Rasch, G., 1961. On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, IV, pp. 321-334. Berkeley, California: University of California Press.

Smith, R.M., 1996. Polytomous mean-square fit statistics. *Rasch Measurement Transactions*, 10(3): 516-517.

Wright, B.D. And J.M. Linacre, 1994. Reasonable Mean-square Fit Values. *Rasch Measurement Transactions*, 8(3): 370.

Wright, B.D. and M.C. Mok, 2004. *An Overview of the Family of Rasch Measurement Models*. A chapter in the Introduction of Rasch Measurement. Chapter, 1: 1-24. Downloaded from www.jampress.org/irmch1.pdf on 15 Mac, 2013.

Wright, B.D. and M.H. Stone, 1979. *Best test design Rasch Measurement*. Chicago,IL: Mesa Press.