Easier Analysis and Better Reporting using Rasch Model to handle Rank Data in Engineering Education Research

Omar, **M.Z** [a], **Rodzo'an**, **N. A**[b], **Saidfudin**, **M** [c] , **Zaharim, A** [d] and **Basri, H** [e]

[a]Assoc.Prof., Centre for Engineering Education Research,
Fac. of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, MALAYSIA
[b] Master (*Candidate*), Dept. of Mathematics, Fac. of Science,
[c] Program Director, Exec.Dip. in Quality Management, UTM SPACE,
University Teknologi Malaysia, 81300 Skudai, MALAYSIA
[d] Professor, Centre for Engineering Education Research, Fac. of Engineering and Built Environment
[e] Professor, Ir; Deputy Vice Chancellor (A) ,Universiti Kebangsaan Malaysia, 43600 Bangi, MALAYSIA

Abstract

It has been a predicament of mankind to comfortably live in the wrong since the famous flat-earthers theory and subsequent Ptolemic believe that the world is the centre of the universe. Despite such widely accepted highly scholastic wrong Greek philosophies, the world does not collapse. Similarly in Engineering Education (EE), we have been doing the traditional academic achievement reporting the classical way believing the number of A's as a measurement of achievement and progress. Such practise is only an inference arising an observation made like the expansion of mercury due to heat energy obtained from the surrounding. However, we need to define the quantum of one degrees centigrade; 1^{o} C before a scale can be meaningful and become of good use. This paper describes a measurement model using Rasch Analysis which provides better fundamentals of measurement to ascertain the subject LO of an engineering subject quantitatively on a uni-dimensional ruler. The ruler serves as the measuring device; an instrument pre-requisite to any measurement. The instrument construct is well founded using the odds of event over expected outcome which give rise to a stochastic probabilistic theorem which is linearised hence the log-odds unit; *logit*. An overview of the measurement model and its key concepts are presented and its application illustrated using the final exam paper given through KKKF1134 – Introduction to Engineering as implemented in the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM). The students performance output were assessed based on Students Observed Learning Outcomes (SOLO) Taxonomy which gives an indication on the student achievement of the subject expected LO i.e. Students' Profiling. The study shows that Rasch model of measurement can classify grades into learning outcomes more accurately especially in dealing with small number of sampling unit.

Keywords: Learning Outcomes, instructional objectives, performance assessment, Quality, continuous improvement.

1. Introduction

Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time. Learning is a complex process. It entails not only what students know but what they can do with what they know; it involves not only knowledge and abilities but values, attitudes, and habits of mind that affect both academic success and performance beyond the classroom. Assessment should reflect these understandings by employing a diverse array of methods, including those that call for actual performance, using them over time so as to reveal change, growth, and increasing degrees of integration. Such an approach aims for a more complete and accurate picture of learning, and therefore firmer bases for improving our students' educational experience [1].

A good assessment recognizes the value of information for the process of improvement but to be useful, information must be connected to issues or questions that people really care about. This implies assessment approaches that produce evidence that relevant parties will find credible, suggestive, and applicable to decisions that need to be made. It means thinking in advance about how the information will be used, and by whom. The point of assessment is not to gather data and return "results"; it is a process that starts with the questions of decision-maker that involves them in the gathering the data and subsequent analysis;

1. How do you assure the correct instrument is used for purpose ? and subsequently;

2. What is the correct method of such data analysis ?

It is of utmost importance on the onset this fundamentals of measurement must be correct. Analysis must be based on valid data and duly interpreted to generate a reliable report with meaningful information for prudent decision making towards continuous improvement of teaching and learning. In an earlier paper, it was shown how academic reporting using Rasch Analysis proved to be more meaningful and make students classification hence better management to improve their achievement in meeting the targeted learning outcomes [2].

2. Overview of Data Types

Fundamentally there are two types of data; quantitative and qualitative type [3]. It was generally perceived as countable and non-countable. Qualitative begins with nominal data which is representative in nature say; race, gender, marital status etc. Such data can counted but not ordered. Normally it can be assigned a code in the form of a number where the numbers are simply labels i.e. Male, 1; Female, 2 or simply M or F.

A set of data is said to be ordinal if the values or observations belonging to it can be ranked or put in order or perhaps a rating attached. It can be counted and ordered, but not measure, ordinal data. The categories for an ordinal set of data have a natural order, for example, siblings. It is always brothers and sisters but no equal interval exist between them. Suppose a group of people were asked to taste varieties of biscuit and classify each biscuit given a rating of 1 to 5, representing strongly dislike, dislike, neutral, like, strongly like. A rating of 5 indicates more enjoyment than a rating of 4, for example, so such data are ordinal. However, the distinction between neighbouring points on the scale is not necessarily always the same. For instance, the difference in enjoyment expressed by giving a rating of 2 rather than 1 might be much less than the difference in enjoyment expressed by giving a rating of 4 rather than 3.

An interval data is when the distance between any two adjacent units of measurement is the same but the zero point is arbitrary. Scores on an interval scale can be added and subtracted but cannot be meaningfully multiplied or divided. For example, the time interval between the starts of years 1981 and 1982 is the same as that between 1983 and 1984, namely 365 days. The zero point, year 1 AD, is arbitrary; time did not begin then. Other examples of interval scales include the heights of tides, and the measurement of longitude. It can be subtracted and the differences make sense, but it's ratios do not; e.g., 30°-20°=20°-10°, but 20°/10° is not twice as hot!

Ratio data is of the highest order; it can be ordered and has a constant separation which is meaningful i.e., to say that 10 m is twice as long as 5 m is well understood and quantifiable. This ratio hold true regardless of which scale the object is being measured in (e.g. meters or yards). The Scores on an interval scale can be added and subtracted and can also be meaningfully multiplied or divided.

Total marks of a student obtained in an exam gave a rank order but the distant between the next student ability having lower or higher marks is never the same. In reality, crudely speaking we are only counting the number of correct answers. However, it has been grossly misunderstood and treated like a quantitative data which is somehow blatantly added and subtracted and even multiplied or divided.

Modern measurement method as practiced using item response theory with a focus on Rasch measurement model now provides the social sciences with the kind of measurement that characterizes measurement in the natural sciences i.e. the field of metrology [4]. The fundamentals of measurement calls for an instrument to be used for purpose to have specific unit of an agreed standard amount [5]. An instrument must have the correct construct of linear scale which can be zero set and duly calibrated. A valid instrument can then be replicated for use independent of the subject hence measurement taken thereof is therefore a reliable data for meaningful analysis and examination to generate useful information. This information is of utmost importance to be the prime ingredient in a particular decision making.

3. Measurement Method

Responses from the students in an examination, test or quizzes is normally marked against a marking scheme comprising keywords; where when there is a match then the student would be given a mark or otherwise. This is the traditional 'park and mark system'. In theory, at this stage truly the assessors is only counting the number of correct answers which is then added up to give a total raw score. The raw score only give a ranking order which is deemed an ordinal scale that is continuum in nature [6]. It is not linear and do not have equal intervals which contradicts the nature of data fit for the due statistical analysis. It does not meet the fundamentals of sufficient statistics for evaluation [7].

Rasch focuses on constructing the measurement instrument with accuracy rather than fitting the data to suit a measurement model with of errors. In Rasch philosophy, the data have to comply with the principles, or in other words the data have to fit the model. In Rasch point of view, there is no need to describe the data. By focusing on the reproducibility of the latent trait measurement; in this case the students' LO instead of forcing the expected generation of the same raw score, i.e. the common expectation on repeatability of results being a reliable test, hence the concept of reliability takes its rightful place in supporting validity rather than being in contentions. Therefore; measuring LO ability in an appropriate way is vital to ensure valid quality information can be generated for meaningful use; by absorbing the error and representing a more accurate prediction based on Rasch probabilistic model [8].

An attempt of a student to answer a question can be seen as a chance of him being able to get the correct answer or successfully accomplishing a given task. Now, for a given normal score of 7/10 which is normally read as 70%; there is need of a paradigm shift to read it as the odds of success being 70:30; thus a ratio data. A mark of 6/10 shall now be seen as odd of success 60:40 and, so on. After all percentage is statistically recognized only a data summary; which is somehow largely confused as a unit of measurement.

This enable us to construct a log-odd ruler of probability an event taking place with the odd-of success as shown in Figure. 1 with unit termed as ***logit*** , derived from the term '**log-odd unit**'; as unit of measurement of ability akin to *meter* to measure length or *kilogram* to weight.

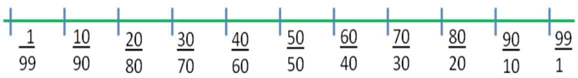


Figure 1. Probabilistic line diagram

In order to achieve an equal interval scale, we can introduce logarithm of the odd probabilistic value. Maintaining the same odd probabilistic ruler as in Figure 1, starting with 0.01 to 100, we can create an equal interval separation between the log odds units on the line, hence the measurement ruler with the *logit* unit. This can be verified by computing the value of $\log_{10}$ 0.01 (10^{-2}) equals to -2.0; value of $\log_{10}$ 0.1 equals to -1; value of $\log_{10}$ 1 equals to 0 and so forth. Figure 2 shows the newly established *logit* ruler as a linear scale with equal interval separation. It is just like looking at a thermometer with '0', as water being ice and 100 as boiling point whilst the negative extreme end as -273ºC, the point where all atoms of any element come to a standstill.

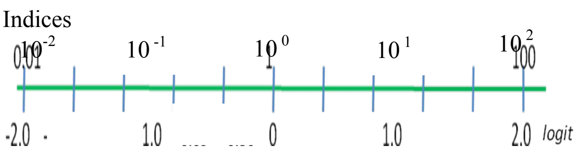


Figure 2. Logit ruler

Thus, we now have a valid construct of an instrument to measure the students ability for each defined LO.

4. Results and Discussion

The test was administered on 1st year Engineering and Architecture students from the Faculty of Engineering and Built Environment, University Kebangsaan Malaysia (UKM) for the course code KKKF1134 – *Introduction to Engineering*. The result from the tests were assessed based on SOLO Taxonomy [9] and ran in *Winsteps v 3.6.8,* a Rasch analysis software; to obtain the *logit* values. Figure 3 shows the Person-Item Distribution Map (PIDM) where the *persons*; i.e. the Students is on the left whilst the *items*; the learning topics were plotted on the right side of the *logit* ruler as in Figure 3. By virtue of the same ruler with the same scale; then the correlation of the *person*, β_n and *item*, δ_i can now be established. In Rasch, the probability of success can be estimated for the maximum likelihood of an event as;

$$P(\theta) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \qquad \text{Equ.(1)}$$

where;

e = base of natural logarithm or Euler's number; 2.7183

β_n = person's ability

δ_i = item or task difficulty

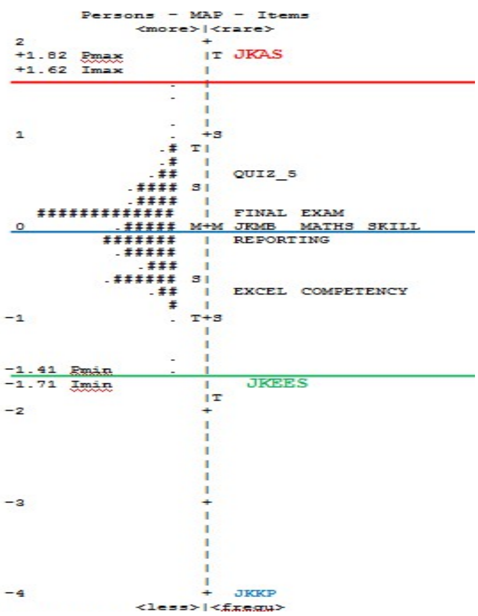


Figure 3 -Person-Item Distribution Map : Item Location

The PIDM Map as in Figure 3 is the heart of Rasch analysis. The vertical dashed line represents the ideal less-to-best continuum of quality. The Items and Students now share the same linear measurement units known as *logits*. On the right hand side of the dashed line, the items are aligned from too easy to too hard, starting from the bottom. The distribution of student positions is on the left side of the vertical dashed line in increasing order of ability; the best naturally being at the top and the poorest student is at the bottom of the rung. Letter "M" denotes the student and item mean, "S" is one standard deviation away from the mean and "T" marks two standard deviations away from the mean. In Rasch Model, since we are interested in the person's ability for a given task, it is most prudent to zero set the scale where the item mean is zero when the ability is deemed 50:50 being the tipping point. Rasch Analysis tabulates the item's location in a very clear graphical presentation which is easy to read and easier to understand. Each item can be coded with attributes of Bloom's Taxonomy that is assessed affecting the students learning process [10]. This will enable in depth analysis of their study pattern to be evaluated meaningfully.

Before delving any further, it is best to look at the analysis Summary Statistics as in Table 1. The prime information we are looking for in this table is the validity of this assessment. The value of Cronbach-α =0.33 is disturbingly low which is well below the acceptable level 0.6 and, in normal statistical analysis this test evaluation would have been disregarded. However, Rasch analysis offer a better evaluation where it shows the two components of the test; the Person and the instrument, i.e. item reliability. Rasch found the Person Reliability rather low at 0.31 and a very high Item Reliability of 0.99. This conclude that the students need further scrutiny and yet we can proceed with the analysis as the instrument has a very high reliability in measuring what is supposedly to be measured. This is where Rasch has the major strength as the better model is making measurement [11].

SUMMARY OF 243 MEASURED Persons

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	28.9	9.0	-.03	.36	1.03	.0	1.06	.0
S.D.	3.8	.2	.48	.02	.58	1.2	.84	1.1
MAX.	40.0	9.0	1.62	.52	3.13	3.1	7.00	4.7
MIN.	16.0	8.0	-1.77	.34	.16	-2.9	.18	-2.1

REAL RMSE	.40	ADJ.SD	.26	SEPARATION	.66	Person RELIABILITY	.31
MODEL RMSE	.36	ADJ.SD	.32	SEPARATION	.89	Person RELIABILITY	.44
S.E. OF Person MEAN = .03							

VALID RESPONSES: 99.9%
Person RAW SCORE-TO-MEASURE CORRELATION = .99 (approximate due to missing data)
CRONBACH ALPHA (KR-20 RELIABILITY = 0.33 (approx due to missing data)

SUMMARY OF 8 MEASURED Items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	726.8	242.2	.00	.07	1.10	.1	1.06	-.2
S.D.	215.9	2.0	0.94	.02	.42	4.4	.34	3.7
MAX.	1105.0	243.0	1.82	.10	1.91	5.9	1.60	4.3
MIN.	329.0	237.0	-1.71	.06	.53	-7.3	.58	-6.4

REAL RMSE	.08	ADJ.SD	.94	SEPARATION	11.67	Item	RELIABILITY	.99
MODEL RMSE	.07	ADJ.SD	.94	SEPARATION	13.33	Item	RELIABILITY	.99
S.E. OF Item MEAN = .36								

MINIMUM EXTREME SCORE: 1 Items

Table 1. Summary Statistics

The Summary of 8 measured items gave a measurement of $Maximum_{item}$=+1.82*logit* and $minimum_{item}$= -1.71*logit*. One item is identified to be classified as minimum extreme score. Close study revealed in Table 2 –Item Measure shows the item to be JKKP= -7.42*logit*.

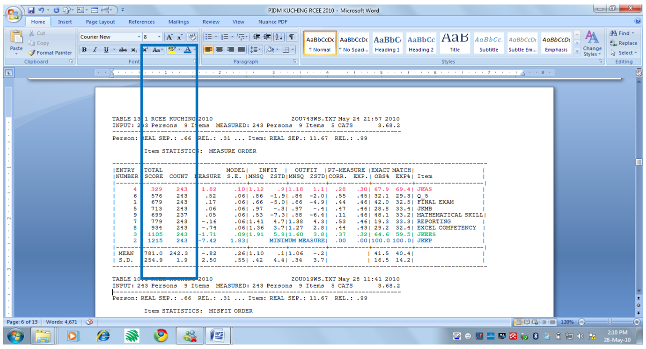


TABLE 13.1 RCEE KUCHING 2010 ZOU743WS.TXT May 24 21:57 2010
INPUT: 243 Persons 9 Items MEASURED: 243 Persons 9 Items 5 CATS 3.68.2
Person: REAL SEP.: .66 REL.: .31 ... Item: REAL SEP.: 11.67 REL.: .99

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
4	329	243	1.82	.10	1.12	.9	1.18	1.1	.28	.30	67.9	69.4	JKAS
6	576	243	.52	.06	.86	-1.9	.84	-2.0	.55	.45	32.1	29.3	Q_5
1	679	243	.17	.06	.66	-5.0	.66	-4.9	.44	.46	42.0	32.5	FINAL EXAM
5	713	243	.06	.06	.97	-.3	.97	-.4	.47	.46	28.8	33.4	JKMB
9	699	237	.05	.06	.53	-7.3	.58	-6.4	.11	.46	48.1	33.2	MATHEMATICAL SKILL
7	779	243	-.16	.06	1.41	4.7	1.38	4.3	.53	.46	19.3	33.3	REPORTING
8	934	243	-.74	.06	1.36	3.7	1.27	2.8	.44	.43	29.2	32.4	EXCEL COMPETENCY
3	1105	243	-1.71	.09	1.91	5.9	1.60	3.8	.37	.32	64.6	59.5	JKEES
2	1215	243	-7.42	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	JKKP
MEAN	781.0	242.3	-.82	.26	1.10	.1	1.06	-.2			41.5	40.4	
S.D.	254.9	1.9	2.50	.55	.42	4.4	.34	3.7			16.5	14.2	

TABLE 10.1 RCEE KUCHING 2010 ZOU019WS.TXT May 28 11:41 2010
INPUT: 243 Persons 9 Items MEASURED: 243 Persons 9 Items 5 CATS 3.68.2
Person: REAL SEP.: .66 REL.: .31 ... Item: REAL SEP.: 11.67 REL.: .99

Item STATISTICS: MISFIT ORDER

Table 2. Item Measure

Item measures gave the indication on the level of difficulty the students encountered in attempting a given task. The *logit* scale; -ve means they are to the left of the 'thermometer', i.e. easy task, +ve means they are located to the right of the scale, a difficult task, the lesser the probability of success a student to get it right in accomplishing a given task. Now we can sense and have a better appreciation if the students have trouble or not since now their performance is duly measured on sound metrology principles hence JKAS is the most difficult task whilst JKEES is the easiest. JKKP point measure correlation = 0.00 with extreme measure, a match of 100% means the item cannot discriminate between a good and a poor student.

Generally, the item separation, G=11.67 is a big value which indicates that there is a very good differentiation of item difficulty to separate the students into distinct difficulty levels. So, if sample separation is 2, then strata are (4*2+1)/3 = 3, means;

Separation= 2: The test is able to statistically distinguish between high and low performers.

Strata= 3: The test is able to statistically distinguish between very high, middle and very low performers.

Thus, a student separation G=11.67 was computed into the strata formula which yielded a distinct 15.89 strata. This indicate there is a large separation between a very easy question and a very difficult question. This call for a review of the assessment done to close the gap. Rasch item characteristic curves (ICC) gives a simple and clear graphic presentation to show if it is an easy or a difficult

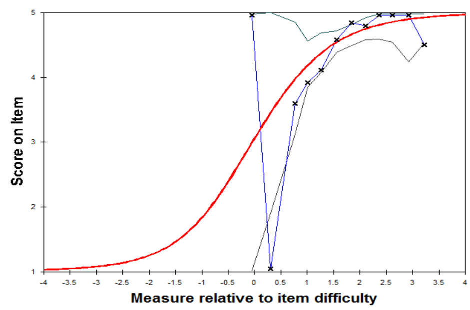


Figure 5a. Item Characteristic Curve –JKEES Easy item. So, item 3-JKEES being at the lowest of the PIDM; being is reflected in Figure 5a whilst JKAS at the topmost hence most difficult is graphically shown in Figure 5b. Note the respondents in Figure 5a is located to the top right, being easy and, Figure 5b respondents is to the left bottom, being difficult.

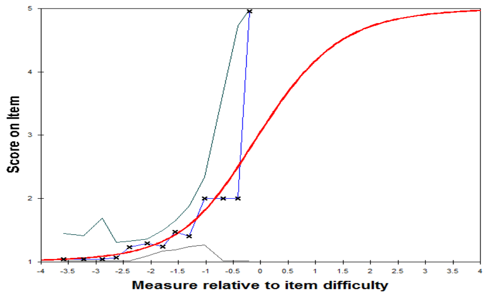


Figure 5b. Item Characteristic Curve –JKAS Difficult

Rasch has a unique ability in recognizing the students development based on the students responses. Table 3 shows the Item Misfit Order. This table gives an indication the validity of the person responses whether it fits the model or not. The topmost being worst where the data provided are outfit to the model thus multi-dimensionality.

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
3	1105	243	-1.71	.09	1.91	5.9	1.60	3.8	A .37	.32	64.6	59.5	JKEES
7	779	243	-.16	.06	1.41	4.7	1.38	4.3	B .53	.46	19.3	33.3	REPORTING
8	934	243	-.74	.06	1.36	3.7	1.27	2.8	C .44	.43	29.2	32.4	EXCEL COMPETENCY
4	329	243	1.82	.10	1.12	.9	1.18	1.1	D .28	.30	67.9	69.4	JKAS
5	713	243	.06	.06	.97	-.3	.97	-.4	d .47	.46	28.8	33.4	JKMB
6	576	243	.52	.06	.86	-1.9	.84	-2.0	c .55	.45	32.1	29.3	Q_5
1	679	243	.17	.06	.66	-5.0	.66	-4.9	b .44	.46	42.0	32.5	FINAL EXAM
9	699	237	.05	.06	.53	-7.3	.58	-6.4	a .11	.46	48.1	33.2	MATHEMATICAL SKILL
MEAN	781.0	242.3	-.82	.26	1.10	.1	1.06	-.2			41.5	40.4	
S.D.	254.9	1.9	2.50	.55	.42	4.4	.34	3.7			16.5	14.2	

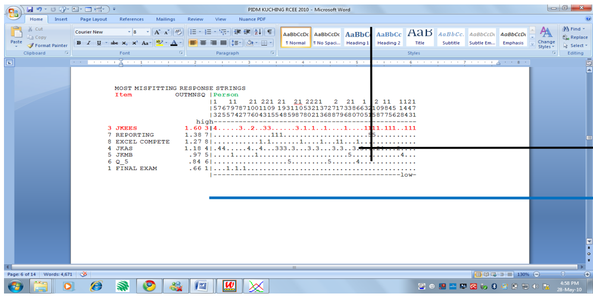
Table 3. Item Misfit Order

Item3 –JKEES measured at -1.71*logit* which is approaching +2.00*logit* thus overly easy item has been identified to have the most misfitting responses. All 243 students attempted the task where a very high N=187 (77%) get an 'A'(5) and N=31(13%) obtained a 'B' giving a strong bias of 90% of the total students getting a very good grade. The outfit MNSQ=1.60 is above 1.5 upper limit showing there are far items that need review.

Rasch examine item or person fit by looking at two types of fit values known as infit and outfit Rasch typically examine 'outfit' which is less threatening to measurement and easier to manage. Hence, we look at "outfit MNSQ" where the mean square (MNSQ) outfit for the item is expected to be near 1.0. Acceptable MNSQ outfit shall be between 0.5 and 1.5.

Closer examination of the scalogram pattern response in Table 4 shows that Person 158 onwards to the end are somehow not well assessed and need review. Those rating 1 could have been more appropriately be higher value.

Table 4 . Scalogram of Item Misfit Response String



Further analysis of the expected value is shown in Table 5 –Item Most unexpected Response Prediction. The item Z-STD =3.80 is beyond the upper limit +2.00. It can be generalized that the item has been under rated and Rasch would ask the researcher to identify the reasoned argument 'why' does this happen. One possible conclusion is that these cohort could have been careless in attempting their works which lead to such a grossly under rated works.

Table 5 . Item Most Unexpected Response Prediction

OBSERVED	EXPECTED	RESIDUAL	ST. RES.	MEASDIFF	Item	Person	Item
4	1.14	2.86	6.95	-2.46	4	97	JKAS
3	1.19	1.81	3.75	-2.21	4	180	JKAS
3	1.19	1.81	3.75	-2.21	4	60	JKAS
3	1.22	1.78	3.41	-2.09	4	79	JKAS
4	1.47	2.53	3.35	-1.49	4	14	JKAS
3	1.25	1.75	3.18	-2.00	4	78	JKAS
3	1.26	1.74	3.10	-1.97	4	230	JKAS
3	1.26	1.74	3.10	-1.97	4	211	JKAS
4	1.55	2.45	3.02	-1.36	4	17	JKAS
3	1.35	1.65	2.54	-1.73	4	214	JKAS
3	1.35	1.65	2.54	-1.73	4	198	JKAS
3	1.35	1.65	2.54	-1.73	4	19	JKAS
4	1.75	2.25	2.41	-1.08	4	65	JKAS
2	1.12	.88	2.32	-2.59	4	2	JKAS
3	1.40	1.60	2.28	-1.61	4	5	JKAS
1	4.60	-3.60	-5.44	1.68	3	58	JKEES
1	4.54	-3.54	-5.04	1.56	3	222	JKEES
1	4.54	-3.54	-5.04	1.56	3	76	JKEES
2	4.73	-2.73	-5.03	2.04	3	206	JKEES
1	4.48	-3.48	-4.67	1.44	3	36	JKEES
1	4.33	-3.33	-3.99	1.20	3	135	JKEES
1	4.33	-3.33	-3.99	1.20	3	21	JKEES
1	4.33	-3.33	-3.99	1.20	3	15	JKEES
1	4.24	-3.24	-3.67	1.07	3	208	JKEES
4	4.93	-.93	-3.51	3.33	3	153	JKEES
3	4.77	-1.77	-3.49	2.17	3	82	JKEES
1	4.13	-3.13	-3.36	.94	3	187	JKEES
1	4.13	-3.13	-3.36	.94	3	145	JKEES
1	4.13	-3.13	-3.36	.94	3	56	JKEES
3	4.69	-1.69	-2.91	1.92	3	213	JKEES
3	4.69	-1.69	-2.91	1.92	3	201	JKEES
1	3.85	-2.85	-2.78	.65	3	144	JKEES
3	4.60	-1.60	-2.42	1.68	3	218	JKEES
1	3.46	-2.46	-2.20	.31	3	243	JKEES
1	3.46	-2.46	-2.20	.31	3	171	JKEES
1	4.23	-3.23	-3.66	1.07	8	14	EXCEL
1	4.13	-3.13	-3.37	.94	8	201	EXCEL
1	3.91	-2.91	-2.88	.70	8	107	EXCEL

Table 5 shows a very interesting finding where the most difficult item; Item 4 -JKAS was found to be the reversed where it is observed to be generally over rated with quite low point measure correlation of 0.28. Conversely, for this difficult item suspects could probably have special interest or knowledge on the topic. On the other hand, they could have a very kind hearted assessor who is gave away marks rather easily. Rasch has this particular predictive properties embedded in the model to make it a very reliable validation model.

5. Conclusion

Rasch Model provides a sound platform of measurement equivalent to natural science which matches the SI Unit measurement criteria where it behaves as an instrument of measurement with a defined unit and therefore replicable. It is also quantifiable since it's linear. Rasch Model has made it very useful with its predictive feature to overcome missing data [12].

The logit ruler has been developed with purpose to measure ability; in this case students learning ability of specific learning outcomes. It can define the students profile and most important we are now able to validate a question construct on line. It is a noble innovation where the ability 'ruler' can transform ordinal data into measurable scale. It's graphical output is great which gives better clarity for quick and easy decision making [13].

The measurement conducted reveals the true degree of cognitive learning abilities of the Engineering undergraduates[14]. Previously, lack of such measurement in Engineering Education has made the necessary corrective actions in the form of skills development, education and competency training difficult to formulate. This major problem faced by Engineering Education Administrators in an IHL to design the necessary curriculum to mitigate the going concern is therefore resolved. Rasch has all the capabilities to rigorously analyse examination results more accurately thus making evaluation clearer to read and easier to understand. This method of reporting was found to consistent with research done in other countries treating ordinal data the correctly by application of Rasch Analysis to obtain a more meaningful information of the item validity hence prudent LO measurement [15].

Acknowledgements

The authors wish to acknowledge the financial support received from the Centre for Engineering Education Research, University Kebangsaan Malaysia as research grant in the effort of improving the quality of teaching and learning in engineering education.

References

[1] Astin. A.W, et. al. 9 Principles of Good Practice for Assessing Student Learning, The American Association for Higher Education, 1991, Stylus Publishing, LLC, 2005

[2] Saidfudin, M, Azlinah M , Azrilah AA, Nor Habibah, A. & Sohaimi Z, "Appraisal of Course Learning Outcomes using Rasch measurement: A case study in Information Technology Education", *International Journal of Systems Applications, Engineering & Development*; Issue 4, vol.1, University Press, UK. pp.164-172, July 2007

[3] Leedy, PD and Ormond JE., *Practical Research: Planning and Design (9th Edition)*. **ISBN-13: 9780137152421 :** Pearson, 2009

[4] Saidfudin, M., and Azrilah, A.A., , "Structure of Modern Measurement", Rasch Model Workshop Workbook v.1, ILQAM, UiTM, Shah Alam. 2009. Retrievable at http://www.ilqam.uitm.edu.my/wp-content/uploads/2009/06/0.1-Rasch-Workshop-Booklet_Structure-of-Measurement.doc accessed: 28 May 2010

[5] Saidfudin, M. and Ghulman, H . A; "Modern measurement paradigm in Engineering Education: Easier to read and better analysis using Rasch-based approach", *International Conference on Engineering Education*, ICEED2009, Dec. 9-10, Shah Alam

[6] Sick, J. Rasch Measurement in Language Education Part 3: The family of Rasch Models, Shiken: *JALT Testing & Evaluation SIG Newsletter* Vol. 13 No. 1 Jan. 2009 (p. 4 - 10) [ISSN 1881-5537]

[7] Wright B.D. Rasch Model from Counting Right Answers: Raw Scores as Sufficient Statistics. Rasch Measurement Transactions, 1989, 3:2 p.62

[8] B. D. Wright and M. M. C. Mok, "An overview of the family of Rasch measurement models," in *Introduction to Rasch Measurement: Theory, Models, and Applications*, J. Everett V.Smith and R. M.Smith, Eds., 2004, p. 979

[9] ATHERTON J S (2009) *Learning and Teaching; SOLO taxonomy* [On-line] UK: Available at: http://www.learningandteaching.info/learning/solo.htm accessed: 18 December 2009

[10] B. S. Bloom, M. D. Englehart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *Taxonomy of Educational Objectives: Handbook 1: Cognitive Domain*.New York: Longman, 1956.

[11] Saidfudin, M, Azlinah M , Azrilah AA, NorHabibah, A; Hamza A Ghulman & Sohaimi Z, "Application of Rasch Model in validating the construct of measurement instrument", in *International Journal of Education and Information Technologies,* Issue 2, Volume 2,. pp. 105-112; May 2008

[12] Saidfudin, M, Rozeha, A , Razimah A. & Hamza A Ghulman, " Application of Rasch-based ESPEGS Model in Measuring Generic Skills of Engineering Students: A New Paradigm", in *WSEAS Transactions on Advances in Engineering Educationn,* Issue 8 Vol.5, WSEAS Press. pp. 591-602, August 2008.

[13] Saidfudin, M., Azrilah, A.A., Azlinah, M.,Nor Habibah, A., Zakaria, S., and H.A. Ghulman, "Development of Rasch-based Descriptive Scale in profiling Information Professionals' Competency", in *IEEE XPLORE* indexed in INSPEC; *IEEE IT Simposium*. K Lumpur,2008 pp.329-333, Aug 2008.

[14] Saidfudin, M. Invited Paper, "Intelligent Students' Learning Ability Measurement System: Easier to understand and clearer analysis using Rasch Model", *Proceedings of 2009 IEEE International Conference on Antennas, Propagation and Systems* (INAS 2009), 3-5 Dec. 2009, Johor, Malaysia

[15] Doig, B. and Groves, S. Easier Analysis and Better Reporting: Modeling Ordinal Data in Mathematics Education Research. *MERJ* 2006, Vol. 18, No. 2, pp.56-76