Easier Learning Outcomes Analysis using Rasch Model in Engineering Education Research

SAIDFUDIN, M 1, AZRILAH, A.A 2, RODZO’AN, N. A 3, OMAR, M.Z 4, ZAHARIM, A 5 AND BASRI, H 6

1 Program Director, 2 P.hD, Program Coordinator, Exec.Dip. in Quality Management, UTM SPACE, 3 Master (Candidate), Dept. of Mathematics, Fac. of Science, University Teknologi Malaysia, 81300 Skudai, MALAYSIA 4 Assoc.Prof., 5 Professor Centre for Engineering Education Research, Fac. of Engineering and Built Environment, 6 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°C before a scale can be meaningful and become of good use. 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. 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

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. Assessment approaches should produce evidence that relevant parties will find credible, suggestive, and applicable to decisions that need to be made. 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 data gathering and subsequent analysis;

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

![Fig. 1: Probabilistic line diagram](image)

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 \( \log_{10} 0.01 = \log_{10} 10^{-2} = -2.0 \); value of \( \log_{10} 0.1 = \log_{10} 10^{-1} = -1 \); 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 -273°C, the point where all atoms of any element come to a standstill.

![Fig. 2: Logit ruler](image)
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, $\beta_n$ and item, $\delta_i$ can now be established. In Rasch, the probability of success can be estimated for the maximum likelihood of an event as;

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

where;
- $e$ = base of natural logarithm or Euler’s number; 2.7183
- $\beta_n$ = person’s ability
- $\delta_i$ = item or task difficulty

The PIDM Map as in Figure 3 is the heart of Rasch analysis. On the right hand side of the dashed line, the items are aligned from easy to difficult, 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. 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 topping 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-\(\alpha =0.33\) 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].

Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Item</th>
<th>Score</th>
<th>Count</th>
<th>Measure</th>
<th>Error</th>
<th>Missq</th>
<th>Est</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITEM 1</td>
<td>2</td>
<td>5</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>ITEM 2</td>
<td>1</td>
<td>3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>ITEM 3</td>
<td>0</td>
<td>2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

SUMMARY OF MEASURED Persons

- Raw Score: 4.0
- Item Reliability: 0.64
- Person Reliability: 0.51
- Person MEAN: 0.02

SUMMARY OF MEASURED Items

- Raw Score: 4.0
- Item Reliability: 0.64
- Person MEAN: 0.02

- Person Raw Score-to-Measure CORRELATION = 0.99 (approximate due to missing data)
- Cronbach’s alpha 99-99 RELIABILITY = 0.99 (approx due to missing data)

- Maximum Extreme Item: 1 Items

- Minimum Extreme Score: 1 Items
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 2: Item Measure

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>JKEES</td>
<td>-7.42 logit</td>
<td>Very difficult</td>
</tr>
<tr>
<td>JKKP</td>
<td>-1.71 logit</td>
<td>Difficult</td>
</tr>
</tbody>
</table>

Item measures gave the indication on the level of difficulty the students encountered in attempting 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.

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 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.

Table 3: Item Misfit Order

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>JKEES</td>
<td>-7.42 logit</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>OBSERVED</th>
<th>EXPECTED</th>
<th>ITEM</th>
<th>PERSON</th>
<th>ITEM</th>
<th>PERSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.29</td>
<td>1.81</td>
<td>3.75</td>
<td>2.21</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1.41</td>
<td>2.04</td>
<td>3.25</td>
<td>2.75</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1.24</td>
<td>1.78</td>
<td>3.10</td>
<td>2.17</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1.35</td>
<td>1.95</td>
<td>3.04</td>
<td>2.70</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1.20</td>
<td>1.60</td>
<td>3.94</td>
<td>2.19</td>
<td>4</td>
</tr>
</tbody>
</table>

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:


