Measurement of Rasch Analysis towards Requirement Engineering Education: Industry Perspective

NOR AZLIANA AKMAL JAMALUDIN
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia (UTM)
Skudai, 81310 Johor
MALAYSIA
norazliana@unisel.edu.my http://www.unisel.edu.my

SHAMSUL SAHIBUDDIN
Advanced Informatics School (AIS), UTM International Campus
Universiti Teknologi Malaysia (UTM)
Jalan Semarak, 54100 Kuala Lumpur
MALAYSIA
shamsul@utm.my http://www.utm.my

Abstract: - Software development industry identifies that human-based give a significant problem in Requirement Engineering. To that reason, education gives a substantial impact in delivering a skill worker and should be a medium to reduce the problem. Survey question was distributed among 206 Shared Sharing Services company under MSC status. Malaysia. 15.53% (N = 32) respondent successfully return their respond back. The result shows that only 27 person is analyzed regarding to misfit data provided by Rasch Measurement Model. The unidimensionality, person-item map and misfit data are discussed. Research objective to identify the undergraduate problem in Requirement Engineering education is achieved. Future work will be discussed on further analysis on expert recommendation using QSR NVIVO v8 to enhance employability skill.


1 Introduction

Requirement Engineering is a fundamental process to meet stakeholder’s needs in software development project. Failure in meeting stakeholder satisfaction will contribute to delay software development project, waste of time, energy, resources and poor quality [22]. The argument of this matter was widely discussed relying on the industry perspective. Industry was spent lots of money in research and development to identify the problem occurs in establishing a strong rapport within stakeholder.

Human-based give a weighty problem to the Requirement engineering. The classification of requirements problems include 1) lack of customer, user and developer 2) lack of communication 3) lack of training 4) lack of define responsibility 5) unstable workforce (low staff retention) 6) inappropriate skills 7) poor time and resource allocations [22].

To that reason, education should be a medium to reduce the human-based problem. Requirement Engineering education can be a weapon to sharpen the human-based skill particularly. Five evidence shows that Prophet Muhammad S.A.W. is an educator that takes into account individual differences in delivering teaching that consist of [1]: 1) provide appropriate advice in accordance with the difference that each seeking some advice 2) give a different answer to the same question that tailored to the individual who asked 3) behave and be different accordingly to the suitability of the mix therewith 4) deliver and legal adapted to the ability for the person to receive it and 5) implement and receive behavior but a person nor receive from someone else because of the different situations.

Each Higher Learning Education (HLE) has set objectives differently. Towards CGPA or student result, it goes the same meaning for all HLE. Cumulative Grade Point Average or CGPA is used widely in a developing country. Many people assuming that the highest CGPA student will have the highest performance to show their competency [9, 10, 11].

However, the validity of Cumulative Grade Point Average (CGPA) is purely the mean of raw scores, lack precision and linearity [8] to meet criteria for measure human-based skill should be revised back. The objective of this paper is to identify the problem in the current
Requirement Engineering education practice using Rasch Measurement Model.

2 Methodology
The survey was administered randomly from 206 Shared Service Outsourcing organizations personnel of Malaysia Status Company (MSC) that registered under Multimedia Development Corporation (MDeC). The services sector is a vital contributor to the growth of the Malaysian economy and functions include Information Technology (IT) services, shared services and business process outsourcing (BPO), regional headquarters, research and development (R&D), training and environmental management. Malaysia has a vibrant ICT and Services industry that is world-class, confirmed by the AT Kearney 2004 and 2005 Offshore Location Attractiveness Index, which ranks Malaysia as the world’s third most attractive Shared Services & Outsourcing (SSO) location [16].

3 Validity of Instrumentation
The questionnaire is revision based on the Cognitive Domain [4], Bloom’s Revised Taxanomy [17] for undergraduate practice [10].

Content [7,12], construct [3] and predictive [13] validity is crucial in getting the sufficient result of the study. Survey questionnaire in this research used two types of scale. First, the dichotomous scale [18] that comprise of 1 – 2 scale for ‘Yes’ and ‘No’. The scale is based on management style of decision making. Second, Likert [23] scale (1 – 2 – 3 – 4 – 5) used for descriptive response categories (never - rarely - sometime - often - always) as a means of partitioning the underlying latent quantitative continuum into successively increasing (or decreasing) amounts of the variable [2].

It leads wrong contextualize the result if we only using ordinal data to achieve the sufficient result. As a solution, a reliable instrument is needed. Rasch shows differently. Rasch transform an ordinal data which is qualitative data into ratio data. Rasch Analysis is deployed vigorously. It used to achieve an effective instrument construct of precision. Rasch able to sieve the instrument clean from any item misfit hence potential data defects [19, 20].

4 Rasch Measurement Model
Rasch Model is used to analyze data. Application of the Rasch model through software such as Winstep [14] and other Rasch software provide estimates of person and threshold locations on the latent variable scale. The software also yields indices of item and person fit to show that the requirement of unidimensionality is met.

Rasch answer on how to have the right measurement with valid instrument. Instrument is extremely crucial if involve human life. Based on Linacre (2011), things would change appreciably when you want a thermometer fit for an open heart surgery. Certainly we will need a more precise measurement instrument. Life is at stake, so it is necessary to have the correct instrument in place. Cost is no more the issue; precision and reliability overrides all. So it goes in an instrument construct; the Standard Error of Measurement and Item Reliability matters most that ought to be given priority when it comes to high stake measurement.

4.1 Rasch Analysis
The normal solution is to apply the regression approach. It shows the best fit line that inline with the points as best as possible. Then, it can be used to make the required predictions by interpolation or extrapolation [8] as necessary as shown in Figure 1.

\[ y = \beta_0 + \beta_1 m \]  

Equ.. (1)

In obtaining the best fit line, there exist differences between the actual point; \( y \) and the best line, the predicted point; \( \hat{y} \). The difference is referred to as error; e.

\[ y_i - \hat{y}_i = e_i \]  

Equ .. (2)

By accepting the fact that there is always error involved in the prediction model, the deterministic model of equation: 1) can be transformed into probabilistic model by including the prediction error into the equation; Equ. 3) Rasch moves the concept of reliability from establishing "best fit line" of the data into producing reliable repeatable measurement instrument. Rasch focuses on constructing the measurement instrument rather than fitting the data to suit the measurement model. \[ y = \beta_0 + \beta_1 m + e \]  

Equ .. (3)

![Figure 1. Best fit line: Linear Regression Model](image-url)
5 Discussion

Only 15.53% (N = 32) personnel were successfully returned the survey question. Rasch help evaluate small sample size that give 95% confidence level.

5.1 Summary Statistic

Result shows in Figure 2 that 84% (N = 27) of respondents is a valid response after clean the misfit data. Besides, 77% of an item is measured after clean the invalid item. A total of 2295 data points arising from 27 respondents on 85 items was analyzed. It yields a Chi-Square value of 3493.40. Cronbach-α value is 0.90, which contribute high reliability of raw score for the instrument in measuring the undergraduate problem.

The optimal categorization [15] in which provides the best construct definition, best separates respondents along the variable, and produces the best fit of data to model. Targeting is at 0.81 logit (S.E. = 0.13) which refering to Meanperson – Meanitem (0.81 logit – 0.00 logit). Targeting is less than 1 logit. Based on a rating scale instrument quality criter ia; if targeting < 1 error, then it is good targeting. So, the instrument is on target

<table>
<thead>
<tr>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCREE COUNT MEASURE ERROR MNSQ ZSTD</td>
<td>MEAN 85.0</td>
<td>85.0</td>
<td>81</td>
</tr>
<tr>
<td>S.D. 3.26</td>
<td>3.26</td>
<td>.00</td>
<td>.66</td>
</tr>
<tr>
<td>MAX. 154.0</td>
<td>154.0</td>
<td>-2.8</td>
<td>19</td>
</tr>
</tbody>
</table>

Figure 3 Summary of Measured 27 Persons

Meantem from Figure 3 is set to an arbitrary 0.00. The instrument where ‘zero-setting’ all items is at 50:50 situation. Error 0.37 ( > 0.25) is slightly high. MeanPerson give a value of +0.81 logit; Person meet expectation. The person-item map in Figure 6, reveals that there is one poor person is located below Meantem and has low ability with -0.28 logit. The Excellent person is at 2.79 logit, which above the highest located practice at 1.94 difficulty logit. Inspite of the good reliability, more difficult items. However, the items need to be introduced for that gap of 0.95 logit.

5.2 Person and Item Fit

To fit the item into the model, we should first identify the sum of Mean and Standard Deviation (SD) to clean the data based on Point Mean Correlation (PTMEA), Mean Square (MNSQ) and z-standard (ZSTD). If the data is indicated high z-std that bigger than 1.29 logit, it is person misfit. Figure 3 shows person 20, 27, 31, 32 and 13 are misfit with MNSQ > 1.25 logit and z-std > ±2 logit. Item whose MNSQ is nearer to 1 and z-std nearer to 0 is deemed a better fit. However, Point of Mean (PTMEA) Correlation allow the negative response because the study is to identify the competence graduates students that get low CGPA but has a high skill. Ignore if in between the range of 0.5 < MNSQ < 1.5 and ZSTD ±2 logit.

<table>
<thead>
<tr>
<th>ENTRY TOTAL</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTEMA</th>
<th>EXACT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER SCREE COUNT MEASURE ERROR MNSQ ZSTD</td>
<td>MEAN 85.0</td>
<td>85.0</td>
<td>80</td>
<td>.99</td>
<td>.00</td>
</tr>
<tr>
<td>S.D. 16.6</td>
<td>16.6</td>
<td>.00</td>
<td>.66</td>
<td>.02</td>
<td>1.9</td>
</tr>
<tr>
<td>MAX. 154.0</td>
<td>154.0</td>
<td>-2.8</td>
<td>19</td>
<td>.53</td>
<td>-4.8</td>
</tr>
</tbody>
</table>

Figure 3 Person Fit

It goes the same to item fit, from 110 item constructed, 25 item deleted. Again, the suggestion is based on Point Mean Correlation (PTMEA), Mean Square (MNSQ) and z-standard (ZSTD). Item whose MNSQ is nearer to 1 and z-std nearer to 0 is deemed a better fit. Remain
the same if the item is in between the range of 0.5 < MNSQ < 1.5 and ZSTD ±2 logit. Remain the same to this item whose MNSQ is a measure the same but different group of the item to be measured because of content validity is preserved in Figure 4.

### Table 4 Consolidated Item Misfit

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>TOTAL</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PCTM</th>
<th>EST</th>
<th>BIAS</th>
<th>DIFFERENT</th>
<th>EXP</th>
<th>ITEM</th>
<th>ITEM</th>
<th>DIFFERENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>55</td>
<td>32</td>
<td>.26</td>
<td>.77</td>
<td>.87</td>
<td>.04</td>
<td>.28</td>
<td>53.1 62.4</td>
<td>73.13</td>
<td>33.1</td>
<td>53.1</td>
<td>33.13</td>
</tr>
<tr>
<td>31</td>
<td>45</td>
<td>32</td>
<td>.26</td>
<td>.77</td>
<td>57</td>
<td>.34</td>
<td>.28</td>
<td>53.1 62.4</td>
<td>73.13</td>
<td>33.1</td>
<td>53.1</td>
<td>33.13</td>
</tr>
<tr>
<td>43</td>
<td>45</td>
<td>32</td>
<td>.26</td>
<td>.77</td>
<td>57</td>
<td>.34</td>
<td>.28</td>
<td>53.1 62.4</td>
<td>73.13</td>
<td>33.1</td>
<td>53.1</td>
<td>33.13</td>
</tr>
</tbody>
</table>

#### 5.3 Unidimensionality

To ensure the measurement is measuring the specific objective, thus, unidimensionality is crucial. Rasch Analysis applies the Principal Component Analysis (PCA) of the residuals to know on how much variance of the instrument measuring that supposedly to measure.

The raw variance explained by measures is 24.8% closely match the expected 23.7%. However, the analysis shows that only 20% of unidimensionality requirement minimum. Rasch cut-low point of 40% is not achieved [5, 19]. Nevertheless, the unexplained variance in the 1st contrast of good 11.4% is obtained as tabulated in Figure 5.

#### 6 Person-Item Map

##### 6.1 Item Analysis

Figure 5 depict the location of the undergraduate practice based on industry personnel experience during their undergraduate study in Requirement Engineering. The location of practice is according to industry personnel ability and difficulty logit. Twenty-nine out of eighty-five items discussed in this paper is referring to identify problem in undergraduate study based on industry personnel experience.

Industry personnel experienced on difficulty which involved seventeen (0.59%) items. They are hardly endorsing this item that above ItemMean. Eight items in between ItemMax and PersonMean which can be considered as the most difficult item and seventeen industry personnel not experience those items. Final Examination and Motivation from Lecturer measured inline MeanItem 0.00 logit. This two items show normality of undergraduate Requirement Engineering study. All industry personnel agreed, to easily endorse ten items that below ItemMean.

##### 6.2 Person Analysis

There are six female and twenty-one male industry personnel randomly picked from Sharing Services Outsourcing was evaluated. The person-item map shown there is five A’s person, eighteen B’s person and two C’s person. We can generalize that Sharing Services Outsourcing hire more industry personnel from B’s person rather than A’s and C’s person.

The most excellent industry personnel is at highest ability 2.79 logit and seen out of the target. The off target person is the one without corresponding items. The poor industry personnel are at -0.28 logit. The difference between PersonMax and PersonMin is 2.51 logit. The difference is slightly over Standard Deviation (SD) of 0.60 logit. This shows the There are 0.56% (N = 15) industry personnel located above MeanPerson and 0.04% (N = 1) is below than minimum MeanPerson. In between A’s person, located a C’s industry personnel above from MeanPerson.

The highest industry personnel at 2.79 logit is female, 23 - 27 years old, hold a first degree from Software Engineering. Earn CGPA 3.50-3.67 and get requirement grade is in between A- to A. Working as software developer with experience five years and less in software industry. Motivates in successfully finishing project are self-esteem, token from company and responsibility. She involved in 6 to 10 projects in HLE and software development that involved web-based and multimedia project. She had an experience as developer and system designer.

Other A’s person that close to ItemMax is male. His age is 33 and above. Hold Master in Information Technology (IT). He earn CGPA 3.50-3.67 during his first degree with grade A- to A in Requirement Engineering subject. He had an experience 10 years and above which involved web-based, networking, stand-alone system, others project software development. Experience in doing a software development project is 10 and above project at HLE and software development project. He had an experience as developer, system engineer, system developer, documenter and project manager.

There is B’s person below PersonMin is male, age in between 33 and above. He holds First Degree in Computer science. He manages to get CGPA 2.50-2.99 with grade D- to D in Requirement Engineering. He has five and less experience in networking. He had 5 and less experience in HLE and one to five project at industry as a developer.
Figure 6 Person-Item Map for Requirement Engineering Education (REE)
C’s person that range in between A’s person is male, age 33 and above, diploma, network, 2.00-2.49, C- to C, system engineer, experience 5 years and less, manufacturing, motivates in successfully finishing project is token from company, rate current learning is average, involved web-based, 5 and less project in HLE, 1-5 in working environment, need training after HLE, need tool to capture requirement, average in using internet to finish the project.

Contradict with C’s person above PersonMin and B’s person is male, age 33 and above, diploma, computer science, 2.00-2.49, C- to C+, software engineer, 5-10 year experience, multimedia project, 5- less project in HLE, 5-10 project experience, experience as developer only.

However, all persons (A’s, B’s and C’s person) agree that they need training after Higher Learning Education (HLE). In increasing the skill among undergraduate students, tool is much recommended to capture requirement from stakeholder. Internet is very useful to finish the project.

7 Conclusion
As a conclusion, Rasch help to identify missing data. Person and item misfit is managed with careful manner. Content, construct and predictive validity are maintained. Based on Person-Item Map in Figure 5, the unidimensionality is achieved. Research objective to identify the problem during undergraduate study based on industry personnel is easily achieved and successfully discussed using Rasch Measurement Model. Future work, will be discussed on further analysis on expert recommendation using QSR NVIVO v8 to enhance employability skill for Software Engineering undergraduate HLE

References:


