The use of Rasch measurement model for the validity and reliability

Sharifah Nurulhuda Tuan Mohd Yasin¹, Mohd Fauzi Mohd Yunus², Izwah Ismail³
¹² Politeknik Sultan Mizan Zainal Abidin
³ Politeknik Ungku Omar

Article Info

Article history:
Received Nov 7th, 2018
Revised Nov 9th, 2018
Accepted Des 2nd, 2018

Keyword:
Validity;
Reliability;
Rasch Measurement Model

ABSTRACT

This study was conducted to produce empirical evidence about the validity and reliability of the questionnaire II (Student) using Rasch Measurement Model. A pilot study was conducted at the Department of Mechanical Engineering, Polytechnic Kota Kinabalu, Sabah on 38 students in their final semester of a diploma program in Mechatronic Engineering. Validity and reliability of the questionnaire II (Students) were measured using Rasch Measurement Model Winsteps Version 3.69.1.11. Rasch analysis showed respondents reliability index is 0.97 and reliability index is 0.91. From the point of polarity items, each item can contribute to the measurement because the PTMEA CORR each item above 0.30, which is between 0.30 to 0.81. Appropriateness test shows 23 items in the questionnaire should be dropped and the second item MNSQ repaired because the outfit has to be outside the range between 0.6 to 1.40. These findings provide the dimensions of the construct validity of the questionnaire, which constructs the implementation of the teaching and learning process, use of facilities and equipment for teaching and learning and assessment. Rasch Measurement Model has shown that the questionnaire II (Students) have a high validity and reliability for use in measuring the evaluation of Diploma in Mechatronics Engineering Program implementation at Polytechnic College in regards to the Industrial Requirement in Malaysia.

© 2018 The Authors. Published by Redwhitepress. This is an open access article under the CC BY-NC-SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/).

Corresponding Author:
Sharifah Nurulhuda Tuan Mohd Yasin,
Politeknik Sultan Mizan Zainal Abidin
Email: shnurulhuda@gmail.com

Introduction

A pilot study was conducted to test the reliability of the instruments that have been developed and detect weaknesses. Popham (1990) and Gallagher (1998) states that the pilot study was conducted to improve the quality of the item and confidence in interpreting the data. The researchers conducted a pilot study by distributing questionnaires to 38 respondents drawn from population studies based on the recommendations Chua (2006).

Respondents were 38 students in their final semester Polytechnic Kota Kinabalu and they were given a questionnaire containing 109 items Likert scale. The results of this pilot study were analyzed using the software Winsteps Version 3.69.1.11 using Rasch measurement model. The researchers conducted a functional assessment of the reliability and item sorting item-respondents, polarity and fitness items as well as the correlation of the standardized residuals.

However, before carrying out a pilot study using instruments built, the instrument must be determined in advance of the face and content validity. All respondents involved in this research was not chosen in the actual study. Researchers have confirmed three things done during this pilot process is
carried out as in (1) to ensure complete instrument given to students; (2) observing the behavior of students while answering instruments; and (3) require students to provide comments and suggestions on the instrument in terms of clarity of meaning and clarity of the items. Referring to Rasch (1980), the validity and reliability of the items in a research instrument can also be determined using the Rasch measurement model. Rasch measurement model is a model that is formed as a result of considerations include the ability or the ability of each candidate or respondents who answered questionnaires, tests or instruments and difficulty of each test item or items Rasch (1980).

Methodology

Sample and Data Collection

This study is a survey research conducted at the Department of Mechanical Engineering, Polytechnic Kota Kinabalu. A total of 38 students in the sixth semester students who take the program Diploma in Mechatronics Engineering, involved in this study. The good cooperation extended by course lecturer and researcher is allowed to enter the working class on time. Researchers introduce themselves and explain the purpose of the study before the questionnaires were distributed. Each respondent was asked to read the instructions before answering the questionnaire and answered individually by each of opinion without the help of other students. Respondents were given a period of about 30 minutes to fill out the questionnaire. Questionnaires were answered collected and verified to ensure that the sample according to the instructions properly and give a complete answer before data analysis.

Survey of Research

The questionnaire submitted to obtain feedback from respondents, the students in connection with the Program Implementation Evaluation Diploma in Mechatronics Engineering at the Polytechnic towards Industrial Requirements in Malaysia. Section A questionnaire (student) contained items related to the background of samples such as gender, technical colleges and courses. Section B relates dimensional constructs implementation process that involves teaching and learning process which contains 59 items. For section C, 45 items were related constructs use of facilities and equipment for teaching and learning Diploma in Mechatronics Engineering. Part D is a construct that has a 5-item assessment.

Likert scale of five options used for students stated degree of agreement with each item submitted. Students were asked to state their degree of agreement on the item submitted either Very High (5), High (4) Average (3), Low (2) or Very Low (1). The researchers used the model for reliability Rash obtained using the Rasch model is more accurate than the Cronbach alpha test. In addition, the Rasch model is able to calculate the score of each respondent in the form of interval data, even data collected using a 5 point Likert scale.

Data Analysis

Data for this study is based on student scores on the items in the questionnaire II students. The scoring is based on the Likert scale as stated in the questionnaire survey for parts B, C and D. Data were checked manually before Winsteps software were used to analyze the items in the questionnaire II (students) to obtain the construct validity of the questionnaire based on Rasch measurement model.

Findings and Discussions

The Validity and Content

To ensure the validity of the survey instrument, the questionnaire II (students) are reviewed to ensure the content of the survey items were appropriate and meet the objectives of the study to be studied. Checking specialist is necessary to ensure the accuracy and clarity of content constructs (Kline, 2005; andHulse, 2006). According to the Gay & Air Asian (2003), review and validation of the expert is needed to ensure that the instrument can meet the needs of researchers based on the objectives that have been set.

In this study, researchers have conducted two types of authentication interface authentication and content validation performed by three experts. Verification done is to check the delivery terms of language use and layout of the entire set of questionnaires. In terms of content validation, Aiken (2003) stated that it intended to examine the extent to which the ability of a measuring instrument to measure what should be measured. Content validity refers to the situation where the contents of a measuring instrument that
The use of Rasch measurement model for the validity and reliability successfully elicited responses that can represent the entire domain, global skills, knowledge and behaviors that should be measured by the device.

Reliability and Isolation Index

According to Bond & Fox (2007), the Apha Cronbach (α) is acceptable reliability is between 0.71 - 0.99 where it is at its best (71% - 99%) described in Table 1.

Table 1. Interpretation of Cronbach Alpha-Score

<table>
<thead>
<tr>
<th>Alpha-Cronbach Score</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9 - 1.0</td>
<td>Very good and effective level of consistency</td>
</tr>
<tr>
<td>0.7 - 0.8</td>
<td>Good and acceptable</td>
</tr>
<tr>
<td>0.6 - 0.7</td>
<td>Acceptable</td>
</tr>
<tr>
<td>&lt;0.6</td>
<td>Item need to be repaired</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Items need to be dropped</td>
</tr>
</tbody>
</table>

Results from the analysis of the pilot study, the reliability of which refer to the Alpha (α) is 0.97. According to Bond & Fox (2007), the mean value of the instruments used in a very good and effective with a high level of consistency and can be used for actual research. The results of the pilot study also showed the reliability of the respondent (person reliability), namely 0.97 in which the reliability can be accepted. The index value of isolation (separation index) obtained also showed high results of 5.81, which to Linacre (2005), good isolation index value is greater than the value of 2.0.

Analysis on the reliability of the item (item reliability) registered 0.91 points. This shows the high reliability in accordance with the Bond & Fox (2007) which states that a value greater than 0.80 is a value that indicates high reliability and received are strong. In addition, the isolation index item (separation of items) recorded 3.11 points, which means a good isolation against the difficulty level according to the item as Linacre (2005), the separation of the good is the greater of 2.0.

Table 2. Analysis Reliability respondents & item respondents and item sorting index

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Person</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person RAW SCORE TO MEASURE CORRELATION</td>
<td>0.97</td>
<td>0.91</td>
</tr>
<tr>
<td>Cronbach Alpha (KR-20)</td>
<td>5.81</td>
<td>3.11</td>
</tr>
<tr>
<td>Person RAW SCORE RELIABILITY</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>

Item Polarity

Examination of the polarities item is intended to test the extent of construction construct achieve its goals and the relationship among the items that are built with the respondents. According to Bond & Fox (2007), to determine whether the item that you want to measure the constructs measured, the value shown on the PT-Measure Corr must be in the positive value (+). If the value obtained is negative (-), it means the item does not measure the construct developed to be measured and it need to be dropped because it is difficult or not leading to questions (out of focus). Based on the analysis, the PT-Measure Corr, no items show a negative value and should not be dropped because all items conform to the concept of determining a predetermined polarity item. (Appendix 1)

Item Fit

Examination of the conformity of the constructed item refers to the value recorded in the index MNSQ outfit. Researchers have to observe the value of this index to determine whether the item developed appropriate (item fit) to measure a latent variable or construct. Based on the Bond & Fox (2007), a study to determine the suitability item built, the outfit MNSQ should be in the range between 0.6 to 1.4. The results of this pilot study analysis found that there are 25 items that are not in the specified range and 23 items have been dropped, and 2 items repair. These items are item CS21, CS 20, CS 37, CS
38, CS36, CS 25, CS 19, CS 17, CS 34, CS 15, CS 02, CS 16, BS13, BS 35, BS 46, BS 52, BS 32, BS 27, BS 06, BS 09, BS 23, BS 17, BS 10(dropped) and CS 14 , CS 33(repair).

**Standardized Residual Value Measurement Correlation**

Standardized measurement of residual correlation value is to determine whether there are items that overlap. High residual correlation for the two items showed the item is not independent, either because the item has the same characteristics among each other or for both combining several other dimensions that are shared. According Linacre (2005), if the correlation value of the two items above 0.7, it shows a high correlation value and only one item have to be maintained while the other items should be dropped. In the analysis of this research, it was found that 10 pairs of overlapping items based on the correlation established. Table 3 shows the dropped items. However, if the same item was dropped in the ranking of suitability item (item fit), the item is automatically aborted first.

Table 3. Largest standardized residual correlation used to identify defendant item.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Entry Number item</th>
<th>Entry number item</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>.92</td>
<td>54 CS01</td>
<td>55 CS02</td>
<td>CS01 accepted</td>
</tr>
<tr>
<td>.88</td>
<td>57 CS04</td>
<td>58 CS05</td>
<td>CS04 &amp; TCS05 were dropped</td>
</tr>
<tr>
<td>.88</td>
<td>72 CS19</td>
<td>73 CS20</td>
<td>CS19 accepted</td>
</tr>
<tr>
<td>.86</td>
<td>56 CS03</td>
<td>58 CS05</td>
<td>CS03 accepted</td>
</tr>
<tr>
<td>.86</td>
<td>73 CS20</td>
<td>74 CS21</td>
<td>CS20 &amp; CS21 were dropped</td>
</tr>
<tr>
<td>.83</td>
<td>56 CS03</td>
<td>57 CS04</td>
<td>CS03 accepted</td>
</tr>
<tr>
<td>.83</td>
<td>69 CS16</td>
<td>70 CS17</td>
<td>CS16 accepted</td>
</tr>
<tr>
<td>.82</td>
<td>80 CS27</td>
<td>81 CS28</td>
<td>CS28 accepted</td>
</tr>
<tr>
<td>.79</td>
<td>89 CS36</td>
<td>90 CS37</td>
<td>CS36 accepted</td>
</tr>
<tr>
<td>.79</td>
<td>9 BS09</td>
<td>10 BS10</td>
<td>BS09 accepted</td>
</tr>
</tbody>
</table>

**Implications Of Findings.**

In this study, using the Rasch measurement model, the researchers have obtained the high reliability of the test reliability. Reliability test items and questionnaire respondents also indicated that examined valid and reliable to measure the dimensions of the construction process of teaching and learning process, use of facilities and equipment for teaching and learning and assessment. However, respondents involved in this study are students from Borneo. In addition, questionnaires were administered carefully by the respondent at the time and hence there was no issue and item pitching motions of respondents (over 50% fit) were found during the process of data analysis. One of the advantages of modern psychometric methods as described in the introductory part, including Rasch modeling method is its ability to identify the formula items and misfit respondents. For example, students who are very good at answering questions should know how to answer a simple and easy question. This method can identify the level of difficulty of the items and the ability of respondents (Bond & Fox 2007). Thus the findings obtained in connection reliability and construct validity of the survey questionnaire is acceptable to answer research questions.
The use of Rasch measurement model for the validity and reliability

76% of respondents give low approval rating to construct

- Item groups that are difficult to be agreed

- Group items that are easy to be agreed

Summary

Validity and reliability of each item in the questionnaire is important. Data were also important and it is necessary to ensure data accuracy and the entry is as defined as it contributes to the validity and reliability of the results. If the reliability or validity of the questionnaire was high, the questionnaire is reliable and valid. Although a questionnaire that was used by previous researchers have been tested for validity and reliability, however, the questionnaire is still need to be tested again because of the inference that the pilot obtained only suitable for the purpose and specific study sample. The Winsteps findings were recorded reliability index and the index of high reliability respondents and structure reliability items hierarchically according to the difficulty level of the item. This clearly shows that the questionnaire II (students) are reliable and legitimate to measure of students’ feedback related to the evaluation of Diploma in Mechatronics Engineering Program implementation at Polytechnic College in regards to the Industrial Requirement in Malaysia.

Acknowledgments

This work was supported in part by the financial support of Politeknik Sultan Mizan Zainal Abidin

References


<table>
<thead>
<tr>
<th>ENTRY NUMBER</th>
<th>TOTAL SCORE</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>MODEL</th>
<th>INFIT ZSTD</th>
<th>OUTFIT ZSTD</th>
<th>PT-MEASURE</th>
<th>CORR.</th>
<th>Exact Match</th>
<th>Item Statistics: Correlation Order</th>
<th>Item</th>
<th>Item Polarity</th>
<th>Input: 3 Person 91 Item</th>
<th>Measured: 3 Person 91 Item</th>
<th>5 CATS WINSTEPS 3.6.9.1.11</th>
<th>Appendix 1</th>
</tr>
</thead>
</table>

### Item Statistics: Correlation Order

<table>
<thead>
<tr>
<th>Entry Number</th>
<th>Total Score</th>
<th>Count</th>
<th>Measure</th>
<th>Model</th>
<th>Infit ZSTD</th>
<th>Outfit ZSTD</th>
<th>Pt-Measure</th>
<th>Corr.</th>
<th>Exact Match</th>
<th>Item Statistics: Correlation Order</th>
<th>Item</th>
<th>Item Polarity</th>
<th>Input: 3 Person 91 Item</th>
<th>Measured: 3 Person 91 Item</th>
<th>5 CATS WINSTEPS 3.6.9.1.11</th>
<th>Appendix 1</th>
</tr>
</thead>
</table>

### Item Polarity

*Input: 3 Person 91 Item* | *Measured: 3 Person 91 Item* | *5 CATS WINSTEPS 3.6.9.1.11* | *Appendix 1*