Transferring Raw Data for Rasch Model Analysis

1N. Lohgeswary, 2S. Salmaliza, and 1H. Othman
1Centre of Engineering and Built Environment Education Research, Faculty of Engineering and
Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor Darul Ehsan, Malaysia
2Centre of Engineering Education Research, Faculty of Engineering and Built Environment,
SEGi University, 47810, Kota Damansara, Selangor Darul Ehsan, Malaysia
lohgeswarynagarethinam@gmail.com

Abstract: Assessing students through final exam questions is a common assessment method at tertiary level. The
overall grade will classify the students into categories. Yet, this overall grade does not show how the
student answered the exam questions. The Rasch Model can show the ability of each student answering the
exam questions. It also gives information on which is the most difficult and the easiest question in an
examination. This study illustrates the steps of the exam marks that are processed in the Rasch Model. This
begins by entering raw exam marks in the Excel worksheet. Then, the marks will be normalized and changed to
the Likert scale. Then the data is saved as formatted text. Lastly, it will run in WINSTEPS. Summary statistics
for person, summary statistics for item, fit statistics, item dimensionality and person-item distribution map are
among of the Rasch Model output that can be obtain from WINSTEPS. These outputs show the performance
of the students and the difficulty of the exam questions. The Rasch output can also identify a misfit examination
question. This means the question is extremely hard for the students to answer. This study will be helpful to
a new user to use Rasch Model because it will help him or her to analyze any data, especially, examination
questions.

Key words: Data processing, Rasch Model, Excel, WINSTEPS, exam questions, assessment method

INTRODUCTION

Rasch Model gives a powerful analysis on
determining the performance of students who sit for any
examinations. It can give details on the level of difficulty
of the exam questions. It can show how the students
answered each question in an examination.

The Rasch analysis has been used widely to examine
the reliability of exam questions and the impact on
student’s performance. Student’s performance is not only
dependent on the ability of answering the exam questions
but also the relevance of the questions. Exam questions
need to be arranged from the easiest to the most difficult
in order for the students to have more time to answer the
difficult questions (Nopiah et al., 2012).

Rasch Model has been used to measure student’s
performance in the examination and it is found that, if
student’s performance is higher than the mean item
(question) it means that the student could answer the
questions within the scope of the subject. Otherwise, a
necessary action needs to be taken to improve
student’s understanding on the subject (Aziz et al.,
2013).

Individual (person) reliability is determined by the
summary statistics of individual whereby it shows the
inconsistency of the individual answering the exam
questions. The item which does not fit the whole exam
questions can be determined by the analysis of Rasch
Model.

MATERIALS AND METHODS

Method of research: Raw data which is the exam marks
obtained from the test will be entered into Excel worksheet
and then transferred to the notepad and lastly will be
entered into WINSTEPS to get Rasch analysis.

Below are the steps needed to process the data. The
data taken is from a pilot test conducted on 35 students
from the Engineering Faculty of Universiti Kebangsaan
Malaysia. About twelve students were from the Chemical

Corresponding Author: N. Lohgeswary, Centre of Engineering and Built Environment Education Research, Faculty of Engineering
and Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor Darul Ehsan, Malaysia
Engineering Department, 10 students were from Civil Engineering Department while 13 students were from Electrical Engineering Department. Table 1 shows the distribution of marks for the pilot test questions.

**RESULTS AND DISCUSSION**

**Step 1:** Student’s marks are entered into the Excel sheet as shown in Fig. 1 and 2. CH01 represents the first student from the chemical engineering list, CV01 represents the first civil engineering student and EE01 represents the first electrical engineering student from their respective list. Data ‘1’ (yellow) shows that the chemical engineering students obtained 1 mark for question 3.

**Step 2:** This step is to normalize all the marks ‘over 100’ using the given equation:

\[
\text{Normalized mark} = \frac{\text{Student’s mark}}{\text{Total marks of the question}} \times 100
\]

For example, data ‘1’ (yellow) is normalized to 17 using the following method:

\[
\frac{1}{6} \times 100 = 17
\]

**Step 3:** Normalized data is changed to the Likert scale using the following mathematical equation:

- IF(K3 = "","", IF(K3 = "xx","x", IF(K3 ≥ 70.5,IF(K3 ≥ 60,4,IF(K3 ≥ 50,3, IF(K3 ≥ 40,2,IF(K3<39,1)))))

Table 2 shows the range of marks for the Likert scale. For example, data ‘17’ (yellow) was given Likert scale 1. Figure 3 shows the changing of data transformed into Likert scale.

**Step 4:** Delete the data from column B to column O. Delete row 1 and 2. Next at column A, right click and choose ‘column width’ and type ‘4’. For the data from column B to G, right click and choose ‘column width’ and type 1. Figure 4 shows the adjustment of the column width.

**Step 5:** ‘Save as’ the file as formatted text (Space delimited) as shown in Fig. 5.

**Algorithm 1:** Open the file. Add the following information:

```plaintext
&INST
TITLE = "Pilot Test"
PERSON = Person; persons are ...
ITEM = Item; items are ...
NITEM = 5; column of response to first item in data record
NI = 6; number of items
NAME1 = 1; column of first character of person label
NAMELEN = 4; length of person identifying label
XWIDE = 1; number of columns per item response
CORDER = 12345; valid codes in data file
UMEAN = 0; item mean for local origin
USCALE = 1; user scaling for logits
UDECIM = 2; reported decimal places for user scaling
MISSCORE = -1
LINELENGTH = 50
&END
```

END LABELS.
Fig. 2: Normalized data

Fig. 3: Changing of data to Likert scale

Fig. 4: Adjustment of the column width

Title is the name of the file. ITEM1 = 5 means the first item (data) placed at column 5. NI = 6 means number of item (questions). NAME 1 – 4 means the length of CV01
Fig. 6: Save as the file formatted text

Fig. 7: Output files

is 4. XWIDE = 1 means the number of column for one piece of data. CODES = 12345 means that the Likert scale has been used. Then, we save the file as shown in Fig. 6.

Step 7: Close the file run the file in WINSTEPS. Figure 7 shows the output files from WINSTEPS. Below are some of the outputs from the Rasch Model. Figure 8 shows
SUMMARY OF 35 MEASURED (EXTREME AND NON-EXTREME) Person

<table>
<thead>
<tr>
<th></th>
<th>TOTAL</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOTAL</td>
<td>MODEL</td>
<td>INFIT</td>
<td>OUTFIT</td>
</tr>
<tr>
<td>Score</td>
<td>8.6</td>
<td>6.0</td>
<td>-1.32</td>
<td>.77</td>
</tr>
<tr>
<td>Count</td>
<td>4.0</td>
<td>6.0</td>
<td>-3.00</td>
<td>1.61</td>
</tr>
<tr>
<td>Measure</td>
<td>5.0</td>
<td>6.0</td>
<td>2.09</td>
<td>0.37</td>
</tr>
<tr>
<td>(S.E. of Person)</td>
<td>5.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 8: Summary statistics for person

SUMMARY OF 5 MEASURED (NON-EXTREME) Item

<table>
<thead>
<tr>
<th></th>
<th>TOTAL</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOTAL</td>
<td>MODEL</td>
<td>INFIT</td>
<td>OUTFIT</td>
</tr>
<tr>
<td>Score</td>
<td>53.2</td>
<td>35.0</td>
<td>.06</td>
<td>.32</td>
</tr>
<tr>
<td>Count</td>
<td>17.2</td>
<td>10.0</td>
<td>.08</td>
<td>.11</td>
</tr>
<tr>
<td>Measure</td>
<td>45.0</td>
<td>35.0</td>
<td>1.08</td>
<td>.75</td>
</tr>
<tr>
<td>(S.E. of Item)</td>
<td>5.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 9: Summary statistics for item

<table>
<thead>
<tr>
<th>Item Statistics</th>
<th>MEASURE ORDER</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Total Total</td>
<td>Model</td>
<td>INFIT</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>35</td>
<td>1.81</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>35</td>
<td>1.08</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>35</td>
<td>.88</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>35</td>
<td>.97</td>
</tr>
<tr>
<td>1</td>
<td>35</td>
<td>35</td>
<td>-1.3</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>35</td>
<td>.92</td>
</tr>
<tr>
<td>1</td>
<td>35</td>
<td>35</td>
<td>.82</td>
</tr>
<tr>
<td>Mean</td>
<td>50.2</td>
<td>35.0</td>
<td>.51</td>
</tr>
<tr>
<td>S.D.</td>
<td>17.1</td>
<td>.00</td>
<td>.93</td>
</tr>
</tbody>
</table>

Fig. 10: Fit statistics

the summary statistics for person. Person represents the students who sit for an examination. Summary statistics for person gives the mean person, person separation and person reliability (Lohgheswary et al., 2016).

Figure 9 shows the summary statistics for item. Item means the questions in an examination. Summary statistics for item gives the mean item, item separation and item reliability value (Lohgheswary et al., 2017a).

Figure 10 shows the fit statistics. Fit statistics is also known as item statistics. It is able to identify the item (question) which does not fit the examination. This is done by inspecting the point correlation (1), outfit MNSQ (2) and outfit ZSTD (3) (Lohgheswary et al., 2017b). Figure 11 shows the item dimensionality. Unidimensionality means that the instrument is measuring in one dimension. Raw variance explained by measures and unexplained variance in 1st contrast determines whether or not the instrument is unidimensional (Lohgheswary et al., 2017c). Figure 12 shows the person-item distribution map. This map is also known as Wright MP. One side on the map shows the ability of students in answering the exam questions while the other side of the map shows the difficulty of the exam questions (Lohgheswary et al., 2018).
Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

<table>
<thead>
<tr>
<th></th>
<th>-- Empirical --</th>
<th>-- Modeled --</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw variance in observations</td>
<td>9.1 100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw variance explained by measures</td>
<td>4.1 45.0%</td>
<td>40.8%</td>
</tr>
<tr>
<td>Raw variance explained by persons</td>
<td>1.3 28.5%</td>
<td>26.9%</td>
</tr>
<tr>
<td>Raw Variance explained by items</td>
<td>3.8 42.3%</td>
<td>39.2%</td>
</tr>
<tr>
<td>Raw unexplained variance (total)</td>
<td>5.0 55.0% 100.0%</td>
<td>59.2%</td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast</td>
<td>2.0 22.2%</td>
<td>40.5%</td>
</tr>
<tr>
<td>Unexplained variance in 2nd contrast</td>
<td>1.4 15.3%</td>
<td>27.5%</td>
</tr>
<tr>
<td>Unexplained variance in 3rd contrast</td>
<td>1.0 10.8%</td>
<td>19.7%</td>
</tr>
<tr>
<td>Unexplained variance in 4th contrast</td>
<td>0.6 6.8%</td>
<td>12.0%</td>
</tr>
<tr>
<td>Unexplained variance in 5th contrast</td>
<td>0.0 0.1%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

Fig. 11: Item dimensionality

Fig. 12: Person-item distribution map

CONCLUSION

Rasch Model is a powerful tool that examines the performance of students in the examination by providing the details on how each student attempts to answer the exam questions. This study provides the detailed steps on how to process the raw data of the exam questions. The data was transferred to Excel and then normalized.
Furthermore, the Likert scale is given to the normalized data. Then, the data is saved as formatted text. Next some information is added and then the file in formatted text is run into WINSTEPS to obtain the Rasch Model analysis. Summary statistics for person, summary statistics for item, fit statistics, item dimensionality and person-item distribution map are some of Rasch Model output which is also shown in this study. The details of the steps will ensure that one will get clear picture on how to analyze the data via the Rasch Model. This procedure will be very helpful for a beginner who is starting to use the Rasch Model to analyze the data.

ACKNOWLEDGEMENT

The researchers wish to express gratitude towards SEGi University and University Kebangsaan Malaysia for supporting the research.

REFERENCES


