How to Cite:
Lohgheswary, N., Salmaliza, S., Lun, A. W., Jedi, A., & Nopiah, Z. M. (2022). Evaluating reliability of final exam questions via rasch model. *International Journal of Health Sciences*, *6*(S3), 1065–1074. https://doi.org/10.53730/ijhs.v6nS2.5064

Evaluating Reliability of Final Exam Questions Via Rasch Model

N. Lohgheswary
Department of Electrical and Electronics Engineering, XIAMEN University Malaysia, Sepang, 43900, Selangor Darul Ehsan, Malaysia

S. Salmaliza
Centre of Engineering Education Research, SEGi University, Kota Damansara, 47810 Selangor Darul Ehsan, Malaysia

A. Wei Lun
Centre for Sustainable Process Technology, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor Darul Ehsan, Malaysia

A. Jedi
Centre of Engineering and Built Environment Education Research, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor Darul Ehsan, Malaysia

Z. M. Nopiah
Centre of Engineering and Built Environment Education Research, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor Darul Ehsan, Malaysia

***Abstract*---**Students grade in a subject is determine mainly from the final examination marks. The final exam questions need to be reliable in order to measure the performance of students in the particular subject. Rasch model able to measure the reliability of an instrument, the final exam questions. A total of 114 students from Mechanical Engineering department sat for the Engineering Statistics subject. Marks were entered in excel and transferred as *prn format. Marks were analyzed against Rasch model, WINSTEPS. The summary statistics for person, summary statistics for item, item statistics, item dimensionality and item correlation are among the output being analyzed. One item is identified as misfit item. The Engineering Sattistics exam questios was very reliable and able to give insight view of the questions. A corrective action is taken to review the misfit item and rephrase the item.

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
***Corresponding author**: Lohgheswary, N.; Email: lohgheswary.nagarethinam@xmu.edu.my*
Manuscript submitted: 27 Nov 2021, Manuscript revised: 09 Feb 2022, Accepted for publication: 18 March 2022

***Keywords*---**reliability, engineering statistics, course outcome, final exam, rasch model.

Introduction

Students grade in a subject is determine mainly from the final examination marks. This is because final examination contributes 60% of the total course grade and the remaining 40% is determine from the coursework. The coursework includes quizes, assignment and mid term test. The exam questions must be reliable as that will ensure the Course Outcome of a subject is achieved. The selection for final exam questions will be base on the Course Outcome. Thus the Course Outcome will be achieved if the questions are reliable enough to measure the performance of students. Engineering Mathematics has been a foundation subject for all the undergraduate engineering programme. This is because the strong foundation in understanding the mathematics concept and applying those concepts to solve engineering problems is the aim of incorporating Engineering Mathematics subjects in engineering courses. Vector Calculus, Linear Algebra, Ordinary Differential Equations, Numerical Analysis, Statistics and Probability are the major component subjects in Engineering Mathematics. Although some universities may have different names or lesser subjects, but the content of those subjects will be the same for any undergraduate engineering subjects. The reliability of exam questions can be examine via Rasch model. Polytomous Rasch model is a generalization model from Dichotomus Rasch model. It is a measurement model that measures attribute or character through response for item in integer. Polythomous Rasch model was derived from [1], [2] and [3]. Reference [4] used Polythomous Rasch model to explore the students' ability in solving statistics problem and also to identify the difficult topic in Engineering Statistics. The study resulted that finding confidence interval and p-value problems are the difficult topics for students.

Methodology

The final examination for Engineering Statistics was conducted in session 2017 / 2018. After the students undergo a teaching of 14 weeks in a semester, the final exam was conducted. The total percentage of the mid semester test, assignment, quiz and final examination will determines the grade of a student for the Engineering Statistics subject. A total of 7 subjective questions were designed and validated by a group of lecturers. The duration of the final examination is 3 hours and the total marks are 100. A total of 114 students from Mechanical Engineering department sat for the final exam of Engineering Statistics. Table 1 list the Course Outcome for Engineering Statistics subject. There are a total of 5 Course Outcomes. They are understand basic concepts of probability distributions, able to perform appropriate hypothesis testing, able to perform simple linear regression and correlation analysis, able to use different tools in quality control and able to solve and interpret engineering problems using appropriate statistical methods. Table 2 shows the details entry of each of the final exam questions. This includes the details of Course Outcome, the level of Bloom Taxonomy and the description of Bloom Taxonomy for each question.

Results and Discussion

As a first step, grades were compiled in the EXCEL *prn format. The grades were transferred using Bond and Fox [10] known as WINSTEPS. It is a Rasch analysis software used to obtain the logit values. Figure 1 shows the summary statistics for person. Person represents the students who took Engineering Statistics final examination. The person summary reveals a strong reliability of Cronbach Alpha = 0.78 and person reliability = 0.67. The result of separation was 1.42 indicating that the students can be divided into two groups. Figure 2 shows the summary statistics for the 13 items involved in this study. "Items" represents the questions tested on the final examination. The item summary summarises very high reliability of 0.97 and item separation = 5.74. The value of the item separation indicates that the final exam questions can be grouped into five groups. The value for the mean item is 0.

Fig. 3 shows the item statistics for Engineering Statistics subject. There are 3 steps of examination is done in order to find the misfit item. Thery are Point-Measure Correlation, the Outfit MNSQ and Outfit *z*-Standard. The standard range for Point-Measure Correlation is $0.4 < x < 0.8$ [6] and [7]. Item 4 is out the Point-Measure Correlation range. The range for Outfit MNSQ is $0.5 < MNSQ < 1.5$ [8]. Item 4 is out of range of Outtfit MNSQ. The range for Outfit *z*-Standard is $-2 < z < 2$ [5]. Item 4 is out of the range of Outfit *z*-Standard. Since item 4 is out of range of Point-Measure Correlation, Outfit MNSQ and Outfit *z*-Standard therefore item 4 is a misfit item for the final examination. Question 4 require the students to write a report on the lateness in a factory's work. Question 4 contribute to 10 marks. This question need to breakdown into few parts.

Fig. 4 shows the item dimensionality for Engineering Statistics subject. The raw variance explained by measures from Fig. 4 is 47.7%. This value fulfill the need for Rasch measurement dimensional which is 40% [9]. There is a sufficient reliability in measuring problem solving skills in Engineering Statistics subject. The measured noise level for item or known as unexplained variance in 1st contrast is 7.4% whereby the value is in a good control as it is within the range of 5% - 15% [10] and [11]. These values show that the items or the final exam questions are unidimensionality. Fig. 5 shows the item correlation and to check if there is any item is dependent on other item. An item is said to be dependent on another item if the correlation value is more than 0.7. From Fig. 5, no item is more than 0.7. Thus all the item are independent.

Table 1
Course Outcome for Engineering Statistics subject

Course Outcome	Description
1	Understand basic concepts of probability distributions.
2	Able to perform appropriate hypothesis testing.
3	Able to perform simple linear regression and correlation analysis.

Course Outcome	Description
4	Able to use different tools in quality control.
5	Able to solve and interpret engineering problems using appropriate statistical methods.

Table 2
Entry number for final questions

Q	Course Outcome	Bloom Taxonomy Level	Description
1(a)	CO1	3	Application
1(b)	CO1	3	Application
2(a)	CO2	2	Understand
2(b)	CO2	5	Synthesis
3	CO2	5	Synthesis
4	CO4	3	Application
5	CO5	5	Synthesis
6(a)	CO3	3	Application
6(b)	CO3	4	Analysis
6(c)	CO3	4	Analysis
6(d)	CO3	5	Synthesis
7(a)	CO5	3	Application
7(b)	CO5	4	Analysis

```
SUMMARY OF 114 MEASURED (EXTREME AND NON-EXTREME) Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      47.2      13.0         .42     .26                                |
|  SEM       1.0        .0         .05     .02                                |
| P.SD      10.1        .0         .58     .19                                |
| S.SD      10.2        .0         .59     .19                                |
| MAX.      65.0      13.0        2.68    1.57                                |
| MIN.      25.0      13.0        -.60     .20                                |
|-----------------------------------------------------------------------------|
| REAL RMSE    .34 TRUE SD     .48  SEPARATION  1.42  Person RELIABILITY  .67 |
|MODEL RMSE    .32 TRUE SD     .49  SEPARATION  1.51  Person RELIABILITY  .69 |
| S.E. OF Person MEAN = .05                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .94
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78  SEM = 4.80
```

Fig. 1 Summary statistics for person

SUMMARY OF 13 MEASURED (NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	414.2	114.0	.00	.08	1.01	-.05	1.04	.11
SEM	26.9	.0	.15	.01	.06	.45	.08	.34
P.SD	93.0	.0	.50	.02	.20	1.56	.29	1.17
S.SD	96.8	.0	.52	.02	.21	1.62	.30	1.21
MAX.	542.0	114.0	.83	.13	1.26	2.02	1.69	2.25
MIN.	250.0	114.0	-.89	.07	.70	-2.64	.70	-1.92

REAL RMSE .09 TRUE SD .50 SEPARATION 5.74 Item RELIABILITY .97

MODEL RMSE .08 TRUE SD .50 SEPARATION 6.04 Item RELIABILITY .97

S.E. OF Item MEAN = .15

Fig. 2 Summary statistics for item

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
8	542	114	-.89	.13	1.12	.48	.85	-.17	.28	.25	79.5	79.2	6a
13	518	114	-.58	.10	1.01	.13	.93	-.06	.34	.34	62.5	62.3	7b
5	515	114	-.55	.10	1.26	1.27	1.11	.42	.35	.35	62.5	61.9	3
6	465	114	-.20	.08	1.26	1.78	1.69	2.25	.35	.45	40.2	40.9	4
12	488	114	-.34	.08	.81	-1.25	1.05	.26	.37	.41	47.3	49.7	7a
1	457	114	-.16	.07	1.23	1.69	1.45	1.64	.39	.47	33.9	36.9	1a
3	416	114	.05	.07	1.09	.75	1.26	1.22	.42	.52	25.0	25.5	2a
4	328	114	.45	.07	1.24	2.02	1.18	1.05	.54	.58	20.5	22.6	2b
7	434	114	-.03	.07	.70	-2.64	.76	-1.07	.58	.50	35.7	31.9	5
9	357	114	.32	.07	1.04	.35	.99	.00	.62	.57	16.1	21.1	6b
10	250	114	.83	.07	.70	-2.44	.79	-.91	.63	.60	42.0	38.3	6c
2	355	114	.33	.07	.80	-1.84	.70	-1.92	.66	.57	24.1	21.7	1b
11	260	114	.77	.07	.87	-.97	.74	-1.26	.67	.60	38.4	36.0	6d
MEAN	414.2	114.0	.00	.08	1.01	-.1	1.04	.1			40.6	40.6	
P.SD	93.0	.0	.50	.02	.20	1.6	.29	1.2			17.8	17.3	

Fig. 3 Item statistics

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	24.8437	100.0%		100.0%
Raw variance explained by measures	=	11.8437	**47.7%**		47.5%
Raw variance explained by persons	=	4.4316	17.8%		17.8%
Raw Variance explained by items	=	7.4121	29.8%		29.7%
Raw unexplained variance (total)	=	13.0000	52.3%	100.0%	52.5%
Unexplned variance in 1st contrast	=	1.8460	**7.4%**	14.2%	
Unexplned variance in 2nd contrast	=	1.5647	6.3%	12.0%	
Unexplned variance in 3rd contrast	=	1.3618	5.5%	10.5%	
Unexplned variance in 4th contrast	=	1.1781	4.7%	9.1%	
Unexplned variance in 5th contrast	=	1.1317	4.6%	8.7%	

Fig. 4 Item dimensionality

LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT Item

CORREL-ATION	ENTRY NUMBER It	ENTRY NUMBER It
.16	8 6a	9 6b
-.33	11 6d	13 7b
-.30	7 5	10 6c
-.28	2 1b	10 6c
-.26	1 1a	9 6b
-.25	1 1a	4 2b
-.22	4 2b	11 6d
-.22	2 1b	9 6b
-.22	11 6d	12 7a
-.21	2 1b	3 2a
-.21	4 2b	6 4
-.21	9 6b	12 7a
-.19	1 1a	10 6c
-.18	3 2a	5 3
-.18	2 1b	8 6a
-.18	5 3	11 6d
-.17	10 6c	12 7a
-.14	3 2a	11 6d
-.14	5 3	13 7b
-.14	2 1b	6 4

Fig. 5 Item correlation

Fig. 6 shows item distribution map. On the left side of the dotted line, the students were represented by ‘x’. The right side illustrates the test items which was represented by numbers. The item distribution map categorize the exam question into 4 groups. They are ‘difficult’, ‘mediocre’, ‘easy’ and ‘very easy’. Question 6c and question 6d are the difficult questions in Engineering Statistics final examination. Question 1b, question 2a, question 2b and question 6b are categorized in mediocre group. This means the questions are average level for the students to answer. Question 1a, question 3, question 4, question 5 and question 7a are grouped as easy questions. Question 6a and question 7b were the most easiest to answer in the Engineering Statistics final examination. The students can be divided into 2 groups. This is shown in Fig 7. They are ‘above average’ and ‘below average’. There are more students in the ‘above average’ category compared to ‘below average’ category. This means the final exam questions are considered answerable by the students.

```
MEASURE Person - MAP - Item
           <more>|<rare>
    2        XX  +
                 |
                 |
                 |
                 |
                 |
                 |
          XXXXX  |
                 |
                T|
             XX  |
                 |
            XXX  |
                 |
              X  |
    1     XXXXX  +T
             XX  |
             XX S|
            XXX  |  6c     6d
      XXXXXXXXX  |
           XXXX  |     DIFFICULT
            XXX  |
         XXXXXX  |S
         XXXXXX  |  2b
          XXXXX M|
          XXXXX  |  1b     6b
       XXXXXXXX  |     MEDIOCRE
              X  |
       XXXXXXXX  |
             XX  |  2a
    0    XXXXXX  +M
           XXXX  |  5
        XXXXXXX S|  1a
            XXX  |  4  EASY
            XXX  |
            XXX  |  7a
             XX  |
              X  |
              X  |S 3
             XX T|  7b
                 |
                 |     VERY EASY
                 |
                 |  6a
                 |
   -1            +T
           <less>|<freq>
```

Fig. 6 Item distribution map

```
MEASURE Item - MAP - Person
         <rare>|<more>
    2          +  79   100
               |
               |
               |
               |
               |
               |
               |  39   54   81   103  114
               |
               |T
               |  74   88
               |
               |  5    72   89
               |
               |  34
    1         T+  41   62   85   90   112
               |  43   84
               |S 70   87
           XX  |  51   60   69
               |  16   22   25   33   44   45   80   102  107
               |  61   65   76   101
               |  38   94   108
              S|  3    21   28   52   59   93
            X  |  17   40   47   63   68   78
               |M 30   31   46   48   106
           XX  |  18   19   37   64   77
               |  26   50   53   71   82   105  111  113
               |  24
               |  23   27   32   57   66   75   99   110
            X  |  49   55
    0         M+  29   56   67   73   86   104
            X  |  20   58   91   97
            X  |S 11   35   42   83   92   95   98
            X  |  2    6    96
               |  13   36   109
            X  |  1    14   15
               |  4    8
               |  7
            X S|  12
            X  |T 9    10
               |
               |
               |
            X  |
               |
   -1         T+
         <freq>|<less>
```

ABOVE AVERAGE

BELOW AVERAGE

Fig. 7 Person distribution map

Table 3
Summary of the item distribution map

Q	Course Outcome	Bloom Taxonomy Description	Item Difficulty Level
1(a)	CO1	Application	Easy
1(b)	CO1	Application	Mediocre
2(a)	CO2	Understand	Mediocre
2(b)	CO2	Synthesis	Mediocre

Q	Course Outcome	Bloom Taxonomy Description	Item Difficulty Level
3	CO2	Synthesis	Easy
4	CO4	Application	Eaay
5	CO5	Synthesis	Easy
6(a)	CO3	Application	Very Easy
6(b)	CO3	Analysis	Mediocre
6(c)	CO3	Analysis	Difficult
6(d)	CO3	Synthesis	Difficult
7(a)	CO5	Application	Easy
7(b)	CO5	Analysis	Very easy

Table 3 summarises the item distrution map from Fig. 6. It shows the question, Course Outcome, Bloom Taxonomy level and item difficulty level from the item distribution map. Question 6c and question 6d are the difficult question in Engineering Statistics. Course Outcomme 3 is the difficult Course Outcome for this subject. Course Outcome 3 is able to perform simple linear regression and correlation analysis. In the upcoming semester, the lecturer can emphasis more in the correlation testing topic.

Conclusion

Rasch model able to evaluate the reliability of Engineering Statistics exam questions. The summary statistics for person, summary statistics for item, item statistics, item dimensionality, item correlation, item distribution map and person distribution map are the outputs from rasch model. Item 4 is identified as misfit item. The item dimensionality shows that the final exam questions is unidimensional. The item distribution map showed that questions can be group into 4 groups. The person distribution map classified the person into 2 groups. Course Outcome 3 which is able to perform simple linear regression and correlation analysis is the difficult Course Outcome for Engineering Statistics subject. The final exam questions for Engineering Statistics is very reliable.

Acknowledgment

The authors wish to express gratitude towards XIAMEN University Malaysia (XMUMRF/2022-C9/IECE/0026), Universiti Kebangsaan Malaysia (GUP-2018-151) and SEGi University (SEGiIRF/2018-14/FOEBE-21/84) for supporting the research.

References

1. G. Rasch, “Probabilistic model for some intelligence and attainment tests”, The University of Chicago Press, 1960.
2. E.B. Anderson, “Sufficient statistics and latent trait models”, Psychometrika, 42, pp. 69-81, 1977.
3. D. Andrich, “A rating formulation for ordered responses categories”, Psychometrika, 43, pp. 561-573, 1978.

4. N.A.C. Musa, Z. Mahmud and N. Bahrun, "Exploring students' perceived and actual ability in solving statistical problem based on Rasch measurement tools", IOP Conference series: Journal of Physics, 2017.
5. T.G. Bond and C. M. Fox, "Applying the Rasch model: Fundamental measurement in the human sciences", 2nd Edition, New Jersey: Lawrence Elbaum Inc. Publishers, 2007.
6. A.A. Azrilah, N. Azlinah, H.A. Noor, A.G. Hamzah, Z. Sohaimi and M. Saifudin, "Application of Rasch model in validating the construct measurement instrument", International Journal of Education and Information Technologies, 2, pp. 105-112, 2008.
7. A.A. Azrilah, S.M. Mohd and Z. Azami, "Asas model pengukuran Rasch, Bangi, Penerbit Universiti Kebangsaan Malaysia, 2013.
8. J.M. Linacre, A user's guide to WINSTEPS, Chicago: winsteps.com, 2005.
9. XjJ.M. Linacre, Data variance explained by measure, Rasch Measurement Transactions, 20(1), pp. 1045, 2006.
10. J.W.P. Fisher, "Rating scale instrument quality criteria", Rasch Measurement Transcation, 21(1), pp. 1095, 2007.
11. A.M. Eakman, "Measurement characteristics of the engagement in meaningful activities survey in an age-diverse sample", American Journal of Occupational Therapy, 66(2), pp. 20-29, 2012.