



Calibrating Students' Performance in Mathematics: A Rasch Model Analysis

Hasni Shamsuddin, Nordin Abdul Razak and Ahmad Zamri Khairani

School of Educational Studies, 11800 Universiti Sains Malaysia
Corresponding Author Email: emailhasni@gmail.com

Abstract

Rasch model analysis is an important tools in analysing students' performance at item level. As such, the purpose of this study is to calibrate 14 years old students' performance in mathematics test based on the item difficulty parameter. 307 Form 2 students provide responses for this study. A 40-item multiple choice test was developed to gauge the responses. Results show that two of the items need to be dropped since they did not meet the Rasch model's expectations. Analysis on the remaining items showed that the students were most competent in item related to Directed Numbers (mean = -1.445 logits), while they are least competent in the topic of Circle (mean = 1.065 logits). We also provide calibration of the performance at item level. In addition, we discuss how to the findings might be helpful for teachers in addressing students' difficulty in the topics.

Keywords: Rasch model, mathematics, 14 years-old, item difficulty, students' ability

1. Introduction

Malaysia adopts to the strategy of encouraging the use of science and technology as a tool for economic development, improving both physical as well as people's wellbeing and protecting the sovereignty of Malaysia [12]. Through this strategy, technology and science have been identified as having significant function in realizing the aspiration of Malaysia in order for becoming a developed nation. One of steps to achieve this aspiration is through providing outstanding mathematics education, starting from an early age. The aim is to nurture individuals that are not only able to think mathematically, but also can apply mathematical knowledge effectively and responsibly especially to solve problems as well as making appropriate decision [13].

In order to achieve this aspirations, students need to be competent in mathematics. Nevertheless, literature shows that our students show wide range of mathematical difficulties, from retrieving concepts, formulas, facts and procedure [16]

to weakness in problem solving ability (Tarzimah and Tambhy, 2010). In addition, some non-academic factors such as socioeconomic background, attitudes, interest, teaching teachers and peers [1] are also identified as important predictors that influence mathematics performance. Therefore, mathematics educators need to constantly assess their students' performance so that immediate interventions can be implemented for them. This is important since mathematics is widely regarded as a subject that promotes new knowledge from the past ones. Difficulties at any level of mathematical learning will then affect the students' future learning. As such, the purpose of this study is to assess students'

performance at Form 2 grade level to gauge information of the students' difficulty in performing mathematics.

2. Materials and Method

Sample: A total of 307 Form 2 students (14 years-old) students provided responses for this study using a purposive sampling technique since the researcher has the experience of teaching in the school [8].

Instrument: A 40-item multiple choice test was developed to gauge the responses. It contained all topics in the Form 2 syllabus. The table of specifications for the test is given in the Table 1. The weightage of the topics were determine through their importance and coverage in the curriculum specifications. Important topics were given weightage of 2, while other were given weightage

of 1. This includes the topics of Directed Numbers, Algebraic Expressions II, Linear Equations I, Ratios, Rates and Proportions I, Coordinates, Circles I and Transformations 1. In addition, these topics were also popular in the high-stake national-level *Penilaian Tahap 3* testing. The specifications also specifies distribution of easy, moderate and difficult items. In this study, the distribution was set at 45:30:25, of easy moderate and difficult, which was considered as a normal practice. In general, each item measured the knowledge, skills and abilities (KSAs) needed to be mastered by the students. The KSAs were usually predetermined in the curriculum specifications through the learning outcomes for each topic.

Table 1: Table of specifications

Chapter	Tajuk	Weightage	Easy	Moderate	Difficult	Total
1	Directed Numbers	2	1	1	2	4
2	Squares, Square Roots, Cubes and Cube Roots	1	1	0	1	2
3	Algebraic Expressions II	2	1	2	1	4
4	Linear Equations I	2	1	2	1	4
5	Ratios, Rates and Proportions I	2	1	2	1	4
6	Pythagoras' Theorem	1	1	0	1	2
7	Construction	1	2	0	0	2
8	Coordinates	2	2	1	1	4
9	Loci in Two Dimensions	1	2	0	0	2
10	Circles I	2	0	2	2	4
11	Transformations I	2	3	1	0	4
12	Solid Geometry II	1	1	1	0	2
13	Statistics II	1	2	0	0	2
Total		20	18	12	10	40

Data analysis: In this study, a Rasch model analysis was employed for data analysis using the software named Winsteps 3.63. Rasch model analysis involved two important parameters, namely, (1) students' ability,

and (2) item difficulty. The model specified the relationship between these parameters as the following equation [5]:

$$P(\theta_{ni}) = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)} \tag{1}$$

where, $P(\theta_{ni})$ = the probability of a person n with ability β correctly answered the item i ,
 B_n = ability of student n
 δ_i = item difficulty of item i

Student's ability and item difficulty parameter were estimated based on the ratio of number of correct items to number of incorrect ones. These score were then transformed into equal interval score (call 'measure') using natural log (ln) or 'logits' unit in a procedure called calibration. Measures from Rasch Model calibration is essential in measurement since it possess the equal-interval property as in a thermometer or a ruler. That is, Item A with difficulty measure of 2.0 logits is always two times more difficult to answer than Item B of 1.0 logits. Overall, items with higher measures were more difficult to answer, that in turns, represent more difficult-to-master knowledge, skills or abilities specified in the curriculum specifications (curriculum). In recent years, the study using the Rasch model analysis has been increasingly gaining attention by researchers ([6]; [10]) as it relates to analysis at item level that provide more information compared to at the test-level such as reliability analysis. Even though Rasch model provides important information on the measured construct, its analysis came with two strong assumptions that must be met for the data to have the property of equal-interval. Firstly, the data needs to fit the model's expectation. This quality-control assumption is to ensure that the data collected did not contain too much unintended construct or 'noise'. According to [14], the infit and outfit mean-square (MNSQ) values of 0.7-1.3 are considered reasonable to ensure this assumption is being

fulfilled. Secondly, the data collected must pose the unidimensionality property. This assumption states that the test measures only a single construct [15]. In Rasch analysis, this assumption is assessed using the principle component analysis (PCA) of residuals procedure. This procedure seeks to identify the presence of second construct when the main construct was extracted out. [7] quotes that the issue of unidimensionality is compromised when the unexplained variance from the second construct extracted from the procedure is less than 10%.

3. Results

Rasch model assumption: Assessment of model-data fit showed that Item 9 and Item 15 showed the values of outfit MNSQ of 1.57 and 1.56 respectively, which did not meet the model's expectation. Therefore, these items were deleted and the data was re-analyse. Result showed that all the remaining items demonstrated satisfactory fit of between .78 and 1.30 (Table 2). Meanwhile, the PCA of residuals showed that the unexplained variance from the second construct is only 5.8% (Table 3). As such, both assumptions of Rasch model analysis were met.

Table 3: PCA of residuals

		Table of standardized residual variance (in eigenvalue units)		
			Empirical	Modelled
Total variance in observations	=	53.6	100.0%	100.0%
Variance explained by measures	=	15.6	29.2%	30.1%
Unexplained variance (total)	=	38.0	70.8%	69.9%
Unexplained variance explained by 1 st factor	=	3.1	5.8%	

Table 2: Item difficulty statistics

Item	Raw score	Measure (logits)	SE	Infit MNSQ	Outfit MNSQ
1	104	.28	.13	1.19	1.24
2	259	-2.49	.13	.90	.76
3	220	-1.46	.13	.80	.71
4	249	-2.11	.16	.84	.67
5	200	-1.13	.16	.92	.88
6	189	-.95	.12	.78	.75

7	169	-.66	.12	.82	.81
8	97	.40	.13	1.04	1.03
9			deleted		
10	186	-.94	.12	.89	.90
11	137	-.20	.12	.95	.93
12	32	1.83	.19	1.02	1.08
13	159	-.51	.12	.85	.83
14	133	-.15	.12	1.03	1.02
15			deleted		
16	121	.03	.12	1.11	1.16
17	121	-.13	.12	.92	.90
18	100	.35	.13	1.00	1.05
19	81	.67	.13	1.05	1.15
20	70	.88	.14	1.13	1.30
21	98	.39	.13	.96	.98
22	143	-.31	.12	.89	.90
23	123	-.03	.12	1.12	1.15
24	93	.46	.13	1.04	1.04
25	153	-.44	.12	.97	.98
26	91	.48	.13	1.12	1.18
27	152	-.45	.12	.89	.87
28	99	.36	.13	1.13	1.27
29	78	.73	.14	1.12	1.28
30	20	2.34	.23	1.02	1.11
31	94	.45	.13	1.10	1.18
32	77	.74	.14	1.02	1.06
33	100	.34	.13	1.20	1.28
34	67	.94	.14	1.07	1.28
35	86	.58	.13	1.11	1.20
36	147	-.37	.12	.91	.89
37	154	-.45	.12	1.04	1.04
38	50	1.29	.16	1.09	1.27
39	168	-.67	.12	.92	.90
40	129	-.10	.12	.91	.91
Mean	125.3	.00	.13	1.00	1.02
SD	54.0	.94	.02	.11	.18

SE = Standard error, SD = Standard deviation

Rasch calibration: Results from Rasch model calibrations showed that the topic of Circle 1 was the most difficult for the students since the mean measure was the highest (1.065 logits). This is followed by the topic of Pythagoras’ Theorem (mean = 0.775 logits). The topic of Solid Geometry (mean = 0.42 logits) completed the top three of the most difficult topics for the students. Meanwhile, the Directed Numbers was the easiest topic since based on its lowest item difficulty measure (-1.445 logits). This was followed by the topic of Square, Square Root, Cube and Cube Root (mean = -1.04 logits) and Algebraic Expressions II (mean = -0.4 logits). With regards to individual item, Item 30

(Figure 1) was the most difficult item (measure = 2.34 logits) followed by Item 12 (measure = 1.83 logits, Figure 2) and Item 38 (measure = 1.29 logits, Figure 3). Meanwhile Item 2 (Figure 4) was the easiest (mean = -2.49 logits), followed by Item 4 (-2.11 logits, Figure 5) and Item 5 (-1.13 logits, Figure 6). Since the item difficulty measure possessed equal interval property, the following observation can be made: The most difficult item, Item 30 (measure = 2.34 logits) was about three times more difficult than Item 32 (measure = 0.74 logits). And, Item 32 was about two times more difficult than Item 28 (measure = 0.36 logits).

30

In the diagram, OAB and ODC are straight lines and O is the common centre of arcs AD and BC. The perimeter of the shaded region, in cm, is

(Take $\pi = \frac{22}{7}$)

A 24 B 25 C 27 D 28

Fig. 1: Item 30

12

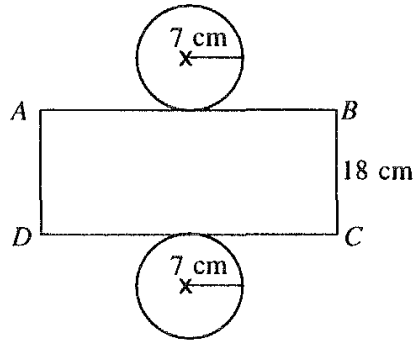
When 4 is subtracted from x , the result is 3

The linear equation which represents the above information is

- A $x - 4 = 3$
 B $x - 3 = 4$
 C $x + 4 = 3$
 D $x + 3 = 4$

Fig 2: Item 12

38



Given that the net in the diagram forms a cylinder when folded, find the surface area of the solid in cm^2 .

(Use $\pi = \frac{22}{7}$)

- A 894 B 926 C 1024 D 1110

Fig. 3: Item 38

2

The initial temperature of a liquid is 19°C . When the liquid is cooled down, its temperature drops by 78°C . The final temperature of the liquid is

- A 15°C B 15.5°C C 16°C D 18.5°C

Fig. 4: Item 2

4

A student purchased $25m$ pencils at $8p$ sen each and $20n$ books at $10p$ sen each. The total amount paid for the purchases in sen is

- A $2mp + 2np$ B $8mp + 10np$ C $25mp + 20np$ D $200mp + 200np$

Fig. 5: Item 4

5

Find the value of $\sqrt{2} \times \sqrt{32}$

- A 6 B 7 C 8 D 9

Fig. 6: Item 5

4. Discussions

The purpose of this study is to calibrate students' performance on mathematics test using the Rasch model analysis. Results showed that the students were able to master knowledge, skills and abilities related to the topics of Directed Numbers, Square, Square Root, Cube and Cube Root and Algebraic Expressions II. However, they were found to have difficulties in the topics of Circle 1, Pythagoras' Theorem and Solid Geometry. It should be noted that, while the position of Directed Numbers and Square, Square Root, Cube and Cube Root are expected, the same cannot be said for Algebraic Expressions. This is because, local studies [2]; [3] show that students usually have difficulties in answering algebraic items. This is probably due to the fact that algebra often involves concepts that are abstract and requires students' deep understanding. As rightly observed by researchers such as [11], the inability to grasp the meaning of the concepts is considered

one of the contributing factors why students didn't see algebra as interesting.

In addition, algebra is also difficult because of the non-visual nature, and is difficult to be represented by pictorial terms. Pictorial terms is one way to address an item and this technique has been proven to be easily understood by the students. For example, in the topic of Directed Numbers of Item 2 (*The initial temperature of a liquid is 19°C . When the liquid is cooled down, its temperature drops by 78°C . The final temperature of the liquid is...*) can be easily solved by drawing diagram that models the actual situation. Meanwhile, Item 5 (*Find the value of $\sqrt{2} \times \sqrt{32}$*) from the topic of Square, Square Root, Cube and Cube Root is rather not

unexpected since students are usually comfortable with arithmetic items that require straightforward procedures [4].

One possible explanation why Item 4 (*A student purchased 25m pencils at 8p sen each and 20n books at 10p sen each. The total amount paid for the purchases in sen is*) is considered easy despite its algebraic nature is that the students are familiar with this type of items. They may have been exposed to this type of item during classroom teaching, or as part of homework. As such, they are able to recall back the procedure to solve the problem without really have the understanding of the concept of algebraic terms.

With regards to the difficult items, Item 12 is algebra-related item while both Item 30 and Item 12 require the students to make connections. For example, in Item 30, the students need to not only plan a strategy to find the perimeter of the shaded area. The strategy requires them to make connections between the related sections in the diagram. The lack of ability to make connections can be traced back to their weakness in memorize and recall concepts and facts in mathematics. A study by [16] had found that students have major difficulty in retrieving concepts, formulas, facts and procedure. Therefore, the students are not able to bring meaning to the problem presented by the item. Note that this finding needs to be investigate further since making connections is important to effective mathematics learning and teaching [9].

Apart from providing information on the difficulty of the items, the Rasch model calibration also provide information on how much is the difference. For example, it is perhaps important for the teachers to investigate further why Item 30 (Topic: Circle 1) is about three times more difficult than Item 32 (Topic: Transformation 1) so that appropriate diagnostics and interventions can be carried out for remedial purposes. Note that, this information will be revealed through the conventional raw score measurement framework that does not have the equal interval property.

As a conclusion, the present study provides empirical evidence on the adequacy of Rasch model analysis to provide useful information on students' performance. The appropriate information will help teachers to plan their teachings of mathematics to achieve its aim to produce individuals that are able to think mathematically and to be able to apply mathematical knowledge effectively and responsibly.

Acknowledgement

The authors would like to thank School of Educational Studies, Universiti Sains Malaysia; Education Sponsorship Division, Ministry of Education and Principal, Sekolah Menengah Sains Kepala Batas for making this research possible.

References

- [1] Abu Bakar, Z., M.I. Kamaruddin and M.Y. Tan, 2009. Pengaruh sikap, minat, pengajaran guru dan rakan sebaya terhadap pencapaian matematik pelajar. Project Report. RMC, UTM Johor.
- [2] Ahmad Zamri, K. and S. Hasni (2016). Assessing Item Difficulty and Discrimination Indices of Teacher-Developed Multiple-Choice Tests. In S.F. Tang and L. Logonathan (eds.), *Assessment for Learning within and beyond the Classroom*, 417-426.
- [3] Ahmad Zamri, K. and A. R. Nordin (2012). Advance in educational measurement: A Rasch model analysis of Mathematics Proficiency Test. *Int. J. Soc. Sc. Human.*, 2:248-251.
- [4] Ahmad Zamri, K. and N. M. Sahari (2011). The development and construct validation of the mathematics proficiency test for 14-year-old students. *Asia Pac. J Educ. and Educ.*, 26:33-50.
- [5] Bond, T. G. and C. M. Fox (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Lawrence Erlbaum, Mahwah, NJ.
- [6] Boone, W. J. (2016). Rasch model analysis: What, when, and how? *CBE Life Sci. Educ.*, 15:1-7.
- [7] Eakman, A. M. (2012). Measurement characteristics of the engagement in meaningful activities survey in an age-diverse sample. *Am J Occup. Ther.*, 66:20-29.
- [8] Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational Research* (8th ed.). Pearson Prentice Hall, New Jersey.
- [9] Lappan, G., J. T. Fey, W. Fitzgerald, S. N. Friel, and E. D. Phillips (2002). *Connected Mathematics Series*, Prentice Hall, Glenview, IL.
- [10] Ling, M. (2016). The development and validation of an instrument to measure leadership competency among secondary-school students: A pilot study. *TPM-Testing*, 23:53-63.
- [11] Martinez, J. G. R. (2002). Building conceptual bridges from arithmetic to algebra. *Math. Teach. Middle Sch.*, 7:326-332.
- [12] Ministry of Education (2012). *Laporan strategi mencapai dasar 60:40 aliran sains/teknikal:sastera*, Ministry of Education, Putrajaya.
- [13] Curriculum Development Centre (2002). *Curriculum Specifications for Mathematics Form 2*. Curriculum Development Centre, Putrajaya.
- [14] Wright, B. D. and J. M. Linacre (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8:370.
- [15] Wright, B. D. and G. N. Masters (1982). *Rating Scale Analysis*, MESA Press, Chicago, IL.
- [16] Zahrah, A. K., B. Jamaliah, N. Rohana, M. Badariah, and K. Jaafar (2003). Meningkatkan kemahiran membentuk ungkapan kuadratik. *Jurnal Penyelidikan Pendidikan*, 7:12-1