Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

SYSTEMATIC DEVELOPMENT OF FORMATIVE TEST IN ALGEBRA ACHIEVEMENT AMONG FORM FOUR STUDENTS USING THE RASCH MODEL AT SMK SIKUATI II, KUDAT

Hamzah Malajun[1]
Mad Noor Madjapuni[2]
Tan Choon Keong[3]

[1]Faculty of Psychology and Education, Universiti Malaysia Sabah, Malaysia,
(Email: hamzahmalajun@gmail.com)
[2]Faculty of Psychology and Education, Universiti Malaysia Sabah, Malaysia,
(Email: mdnoormj@ums.edu.my)
[3]Faculty of Psychology and Education, Universiti Malaysia Sabah, Malaysia,
(Email: cktan@ums.edu.my)

Article history

Received date	:	3-8-2021
Revised date	:	4-8-2021
Accepted date	:	14-8-2021
Published date	:	6-10-2021

To cite this document:
Malajun, H., Madjapuni, M. N., & Keong, T. C. (2021). Systematic Development of Formative Test in Algebra Achievement among Form Four Students Using the Rasch Model at SMK Sikuati II, Kudat. *Jurnal Penyelidikan Sains Sosial (JOSSR)*, 4(12), 31 - 41.

Abstract: *A measuring instrumentsis very important to ensure that students can periodically measure their level of achievement in a topic for each lesson that they have learned with systematically. This research measures their level of development in algebra in learning mathematics via online using the self-regulated learning method at SMK Sikuati II, Kudat, Sabah. The researcher has produced a measuring instrument in the form of a 40-item achievement test to periodically measure the development of student command in algebra. Instrument validity is done using the validation of experts in terms of content and construct. The reliability of the test is done by using the Rasch model which can assess the items' level of difficulty and also the level of capability in obtaining correct answers for the next test items, where teachers and students can accurately assess the development of student command. The implication of this research is hoped to provide contributions in the context of development in student topical command assessment, the usage of the Rasch model, improvement of teacher assessment instrument and increase of student achievement.*

Keywords: *Assessment, Mathematics, Algebra, Rasch Model.*

Introduction

Measurement in mathematics is an important aspect to conduct assessment on the increase or development of students in every topic that they have learned. The development of an instrument is important to ensure that the assessment of student achievement is holistic. This is to ensure that the human capital produced from the measurement process is quality assured to fulfil the country's aspiration in accordance with the aim of the Malaysian Education Blueprint 2013-2025, such as in the first forward which is to provide equal access to quality education of an international standard in the field of language, science, and mathematics (Malaysia

Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

Education Blueprint, 2013). This ensures that it is truly effective to be conducted, not just theoretically, but also practically.

Based on the examinations that the researcher has conducted through past studies, there are many instruments created and used for the purpose of measuring mathematics achievement, but the issue of mathematics achievement has not been completely solved. This can be seen from the existence of low achievements in mathematics that is still happening (Tajudin & Chinnappan, 2016; Cleary *et al*., 2017).The implication from this problem will trigger a very concerning scenario because mathematics is an important foundation in various essential sectors that contribute to the country's economy. A suitable measuring instrument for a small topic in mathematics is hard to obtain especially during a pandemic, teachers face constraints to obtain learning materials to be used in online and offline learning.

Hence, this research is in the process of building an instrument using the Rasch model. This model is an approach or a method that can support and verify the effectiveness of the instrument produced. This will enable the teachers and students to assess the progress that they have reached once they have finished a lesson. An inaccurate measurement and a measurement that is not in line with the level student capability will result in inaccurate assessment and will give a wrong picture on the students' achievement (Fives & DiDonato-Barnes, 2013).

An instrument that is too difficult and is beyond the range of the students' capability will lower their level of motivation in learning mathematics. On the other hand, an item that is too easy for the students' capability will give an inaccurate picture and a false hope on the students' progress and achievement projection in centralized assessments such as SPM. Research has found that the quality and skill of constructing questions among teachers still need to be enhanced because most studies are only theoretical without giving emphasis on the practical aspect by explaining the dimensions and perspectives in detail. This issue needs to be highlighted in order to solve problems relating to education especially in observing the development and achievement of students. Thus, the Ministry of Education Malaysia (MOE) needs to give reasonable attention to enhance the education system that is becoming more challenging due to initial problems which is low achievement and latest problems of increasing demands to fulfill needs and requirements of the education and industrial sector.

A testing item that is too difficult will cause students to have problems in understanding mathematics and they will assume that mathematics is more difficult than it should be. On the other hand, an item that is too easy will give a false hope for the teachers and students as well.

Research Objectives

Thus, as a platform to achieve the research goals, the following research objectives are formed:

i) Studying the validity and reliability of the algebra testing item using the Rasch model
ii) Studying the formative development of the command of algebra among students

Research Questions

i) How is the validity and reliability of the algebra item using the Rasch model?
ii) How is the formative developmentof the command of algebra among students?

Literature Review

Undeniably, there are many literature that examines the importance of assessment in teaching and learning mathematics. A good measurement instrument is very important to assess the effectiveness of the learning activities that have been done(Sjoe *et al*., 2019). In the West, studies regarding validity and reliability in mathematical testing inthe frame of Item Response Theory or Rasch model have been used to ensure that the items in the mathematical testing instrument are suitable with the level of the students (Alkhadim *et al*., 2021;Confrey *et al*., 2020). In Malaysia, the Rasch model has also been used in ensuring suitable testing instrument; one of them is a study among higher institution university students conducted by Asshaari *et al*., (2012) who uses the Rasch model to produce an item testing instrument forengineering mathematics subject which is divided according to the learning content and cognitive field that is to be tested based on theTable of Specifications of Bloom's Taxonomy.

Besides that, a research among pupils in a nursery aims to examine how far the testing instrument in the science and technology domain can be developed based on the testing standard of early childhood development and the Zone of Proximal Development (ZPD) framework using the Rasch model. The findings indicate that there is a high value of validity and reliability which successfully provide a method to facilitate the construction of the testing instrument for the early childhood development stage in Malaysia (Mashitah *et al*., 2017).There is also a research conducted by Samsudin *et al.*(2020),which consists of Form 2 students in Malaysia where they are used to create a computer-aided testing that caters more to individual needs and save more time, which is known as the Computer Adaptive Testing (CAT), to fulfil the assessment standard in today's time. The 30 items used were adapted from the testing items in Trends in International Mathematics and Science Study (TIMSS) 2003-2015 which were recalibrated using the Rasch model. The findings from the analysis show that the fit analysis, polarity analysis, unidimensional analysis, item measure, and Person-Map-Item fulfil the testing standard and are suitable to be applied in the assessment instrument constructed.

Furthermore, a research (Chan *et al*., 2014) which was conducted among secondary school students in Malaysia the topic of algebra also used the Rasch model to ensure that the test items used have a high reliability before analysis to answer the research question constructed. In a research conducted in Germany regarding the usage of measurements suitable for the level of student capability, it was found that for students who went through the method of measurement that is suitable with their level will influence their achievement in the self-assessment conducted (Spoden *et al*., 2020). A research conducted in the USA found that quality instrument items influence the increase of student command by giving a positive impact that shows a continuous increase in the level of comprehension in a learning field(Yang *et al*., 2018; Crawford *et al*., 2020). This shows that the usage of quality measurement instruments have undergone a validity, reliability, and suitability test using the Rasch model, which gives a significant impact in terms of achievement as well as the level of motivation compared to the usage of testing items taken from normal reference books.

The result from the researcher's examination found that in developed countries as well as in various places and education institutions in Malaysia, research regarding the importance of the construction of suitable testing instruments for science and mathematics learning using the Rasch model has already been conducted using the Winstep 4.8 software. The application of this method enables the instrument to have a suitable periodic testing due to the decalibration with the logit scale which indicates that the level of item difficulty and level of student capability can be produced easier and more accurate.

Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

Research Methodology

This research employs the mixed method which utilises both a quantitative and qualitative approach that combines the experiment using tests and after that, interviews to elaborate the data from the test that has been conducted.

The quantitative phase is suitable with the purpose of the study which is to identify the effects of the usage of the algebra assessment instrument using the Rasch model in constructing test papers compared with the normal methods in learning mathematics at SMK Sikuati II, Kudat, Sabah.

This school are chosen for this research because there are sufficient sample from two classroom consist of more than 30 students who are taking algebra class who has about the same marks in latest mathematic test. Thus, ideal for conducting quasi experiment pretest-posttest design consist of experimental and control group (Chua, 2011). Without any systematic test instrument and self regulated approaches the achievement of the form four student in algebra was not satisfactory. The active roles and cooperation from students, teachers and administrator towards teaching innovations has been evident in the past facilitate smooth implementation of this study.

The implementation of this research comprises of two phases, where the first phase is the quantitative item construction in algebra according to the KSSM syllabus; the selection of samples is based on the purposive sampling method (Samsudin *et al.*, 2020) and is conducted among 30 form 4 students who took mathematics in SMK Sikuati II, Kudat,Sabah, Malaysia.

In phase two, the qualitative method utilises the semi-structured interview method (Brown & Danaher, 2019) for teachers who have implemented the teaching and testing of the level of command using the testing instrument in algebra and the analysis on the progress of student achievement in the classroom. A teacher with more than ten years of experienced was interviewed to obtain information regarding the measurement accuracy of the level of student command of themultiple choice test for the algebra topic that was tested.

Next, the interview data and the data from the analysis of the level of student command using the Rasch model were triangulated to increase the accuracy of the findings (Fusch *et al*., 2018). The tirangulation was conducted manually as the total number of respondents needed is low to achieve saturation (Low, 2019).

Findings and Discussion

The findings and discussion were done to answer both research objectives. Information in this section details the results of the research findings obtained by the researcher in two forms of research which are quantitative and qualitative research.

Validity of the Algebra Achievement Test Using the Rasch Model

The validity of the items in the algebra achievement test is analyzed using the Rasch Measurement Model as it can produce a more accurate and detailed measurement based on the relationship between the items and the respondents' capability (Sumintono & Widhiarso, 2014). For this purpose, a pilot study was analyzed using the Winstep version 4.8 software with the Rasch measurement approach which contain various diagnoses to test the validity of an instrument. However, for this research, the four diagnoses that was conducted was from the four aspects which are the validity of item and respondent separation, identifying item polarity to measure construct, suitability of items (item fit)to measure the construct, and determining

Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

the level of item difficulty and respondent capability. The explanation for each test conducted on the functional are as follows.

Item Validity and Separation

The result from the pilot study conducted found that the value of validity obtained based on Cronbach Alpha is 0.84 for the item and 0.88 for the respondent. This shows that the instrument is suitable to be used in the actual research as it has a high level of consistency due to a Cronbach Alpha reliability value that is within the range of 0.70-0.89, which is considered good and can be accepted according to interpretation (Bond & Fox, 2007). The analysis of the separation index shows that the item and respondent values are 2.26 and 2.69 respectively. This result means that there are more than two types of difficulty level for the item whereas the respondent has almost the same category of difficulty level as can be seen in Figure 1. This is considered good as a respondent and item separation index that is more than two is considered good (Linacre, 1999). The high value of reliability and separation index indicate a good measurement quality.

```
 Calculating Fit Statistics
>===================================<
Standardized Residuals N(0,1)  Mean: .01 S.D.: 1.00
Time for estimation: 0:0:0.484
 Processing Table 0
Ujian Pra algebra 40 item 30 responden.xlsx
-------------------------------------------------------------------------------
| PERSON      30 INPUT       30 MEASURED              INFIT          OUTFIT    |
|          TOTAL      COUNT      MEASURE   REALSE    IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN      23.5       40.0          .74      .52     1.00     .0   1.00     .0|
| S.D.       9.1         .0         1.82      .36      .17     .8    .37     .9|
| REAL RMSE    .63 TRUE SD    1.71  SEPARATION  2.69  PERSON RELIABILITY  .88|
|------------------------------------------------------------------------------|
| ITEM       40 INPUT       40 MEASURED              INFIT          OUTFIT    |
|          TOTAL      COUNT      MEASURE   REALSE    IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN      17.6       30.0          .00      .50      .99     .0   1.00     .1|
| S.D.       5.8         .0         1.25      .06      .26    1.2    .52    1.1|
| REAL RMSE    .51 TRUE SD    1.15  SEPARATION  2.26  ITEM   RELIABILITY  .84|
-------------------------------------------------------------------------------
Output written to C:\Users\user\Desktop\ZOU486WS.TXT
CODES= ABDC
```

Figure 1: The Reliability Index and Separation Index

Item Polarity to Measure Construct

The Point Measure Correlation (PTMEA CORR) value test was used to detect item polarity. This can determine whether a construct really measures what it intends to measure. The values obtain from the test are all positive as can be seen in Figure 2 which shows that the item can measure the construct. The item is considered unable to measure the construct if it has a negative value (Bond & Fox, 2007).

Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

NPUT: 30 PERSON 40 ITEM REPORTED: 30 PERSON 40 ITEM 2 CATS WINSTEPS 3.73

RSON: REAL SEP.: 2.69 REL.: .88 ... ITEM: REAL SEP.: 2.26 REL.: .84

ITEM STATISTICS: MEASURE ORDER

NTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
34	7	30	2.39	.57	.68	-1.0	.41	-.9	.76	.63	89.3	85.2	AK
36	7	30	2.39	.57	1.30	.9	1.60	1.0	.50	.63	82.1	85.2	AM
35	7	30	2.39	.57	.81	-.5	.59	-.5	.71	.63	89.3	85.2	AP
12	8	30	2.09	.53	1.12	.5	1.28	.6	.56	.62	82.1	82.7	O
16	10	30	1.57	.49	.81	-.7	.60	-1.0	.69	.60	89.3	77.6	S
18	10	30	1.57	.49	.97	.0	.99	.1	.61	.60	75.0	77.6	U
17	11	30	1.34	.47	.93	-.2	.85	-.3	.62	.59	82.1	76.1	T
19	11	30	1.34	.47	1.57	2.2	1.55	1.4	.37	.59	53.6	76.1	V
25	13	30	.91	.45	.49	-2.9	.40	-2.4	.78	.56	92.9	75.0	AB
37	13	30	.91	.45	.88	-.5	.87	-.3	.61	.56	78.6	75.0	AN
10	14	30	.71	.45	.96	-.1	.94	-.1	.57	.55	82.1	74.6	M
24	14	30	.71	.45	1.02	.2	1.03	.2	.54	.55	75.0	74.6	AA
35	14	30	.71	.45	1.11	.6	1.38	1.2	.48	.55	75.0	74.6	AL
3	16	30	.32	.44	1.14	.8	2.28	3.2	.39	.52	75.0	73.2	F
30	16	30	.32	.44	1.57	2.5	1.63	1.8	.27	.52	46.4	73.2	AG
33	16	30	.32	.44	.88	-.6	.79	-.6	.58	.52	82.1	73.2	AJ
7	17	30	.13	.44	1.24	1.2	1.25	.8	.40	.50	64.3	72.6	J
9	17	30	.13	.44	.74	-1.4	.63	-1.2	.62	.50	78.6	72.6	L
23	18	30	-.06	.44	.62	-2.3	.52	-1.5	.66	.48	89.3	72.3	Z
29	18	30	-.06	.44	1.10	.9	1.12	.4	.42	.48	67.9	72.3	AF
31	18	30	-.06	.44	.76	-1.3	.67	-.9	.59	.48	89.3	72.3	AH
38	18	30	-.06	.44	1.00	.1	1.57	1.5	.45	.48	75.0	72.3	AO
40	18	30	-.06	.44	1.68	3.1	1.76	1.9	.18	.48	53.6	72.3	AQ
20	19	30	-.26	.44	1.16	.9	2.53	2.9	.34	.47	71.4	71.9	W
22	19	30	-.26	.44	1.08	.5	1.09	.4	.42	.47	78.6	71.9	Y
26	20	30	-.45	.45	1.00	.1	.79	-.4	.46	.44	53.6	71.4	AC
32	20	30	-.45	.45	.88	-.6	.75	-.5	.51	.44	75.0	71.4	AI
11	22	30	-.87	.46	.87	-.6	.66	-.5	.47	.40	85.7	73.8	N
13	22	30	-.87	.46	1.17	.9	.92	.0	.35	.40	64.3	73.8	P
21	22	30	-.87	.46	.85	-.8	.72	-.4	.47	.40	78.6	73.8	X
15	23	30	-1.09	.48	.72	-1.4	.58	-.8	.51	.37	82.1	76.2	R
6	24	30	-1.33	.50	.79	-.8	.53	-.6	.46	.35	82.1	79.3	I
14	24	30	-1.33	.50	.71	-1.3	.47	-.7	.49	.35	82.1	79.3	Q
27	24	30	-1.33	.50	1.45	1.7	1.67	1.0	.12	.35	75.0	79.3	AD
28	24	30	-1.33	.50	.84	-.6	.76	-.1	.41	.35	82.1	79.3	AE
5	25	30	-1.59	.53	.90	-.3	.56	-.4	.39	.32	78.6	82.4	H
2	26	30	-1.89	.57	.97	.0	.80	.1	.30	.28	85.7	85.7	E
4	26	30	-1.89	.57	.88	-.2	.57	-.2	.35	.28	85.7	85.7	G
8	26	30	-1.89	.57	.76	-.6	.44	-.4	.41	.28	85.7	85.7	K
1	27	30	-2.26	.64	.94	.0	1.71	.9	.22	.25	89.3	89.2	D
MEAN	17.6	30.0	.00	.48	.99	.0	1.00	.1			77.6	76.9	
S.D.	5.8	.0	1.25	.05	.26	1.2	.52	1.1			10.9	5.1	

Figure 2: Item Point Measure Correlation

Suitability of Item Measuring Construct

The result of the pilot study obtained was measured in terms of suitability by looking at the infit and outfit Mean square (MNSQ) values; if all outfit and infit values obtained are within the range of 0.6-1.4, thus, based on Bond and Fox’s (2007) interpretation, all items are suitable to measure the construct.

The results obtained can be referred to Figure 3 which shows that there are 15 items that did not fit the range of the outfit and infit MNSQ as shown in Table 1. This indicate that the higher the value of the logit item from 1.4, means the more it is less homogeneous with other items in the same scale measurement. A logit value that is lower than 0.6 means that there is an overlap with another item and that item is too easy to be assumed by the respondent. Besides that, the outfit ZSTD and infit ZSTD values are also supposed to be within the range of -2 to 2 (Bond & Fox, 2007). However, if the outfit and infit MNSQ values are accepted, the ZSTD index can be ignored (Linacre, 2004).

If these conditions are not met, the item can be considered to be improved or omitted Based on the consideration of the MNSQ values,8 items which are W (2.53),F (2.28),AQ (1.76), D (1.71), AD (1.67), AB (0.40), AK (0.41), and K (0.44) will be omitted in the actual study and items AG (1.63), AM (1.60), AO (1.57),Q (0.47),R (0.50),Z(0.52), I (0.53), AO (1.57), AM (1.60), and AG (1.63) will be improved.

Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

PERSON: REAL SEP.: 2.69 REL.: .88 ... ITEM: REAL SEP.: 2.26 REL.: .84

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
34	7	30	2.39	.57	.68	-1.0	.41	-.9	.76	.63	89.3	85.2	AK
36	7	30	2.39	.57	1.30	.9	1.60	1.0	.50	.63	82.1	85.2	AM
39	7	30	2.39	.57	.81	-.5	.59	-.5	.71	.63	89.3	85.2	AP
12	8	30	2.09	.53	1.12	.5	1.28	.6	.56	.62	82.1	82.7	O
16	10	30	1.57	.49	.81	-.7	.60	-1.0	.69	.60	89.3	77.6	S
18	10	30	1.57	.49	.97	.0	.99	.1	.61	.60	75.0	77.6	U
17	11	30	1.34	.47	.93	-.2	.85	-.3	.62	.59	82.1	76.1	T
19	11	30	1.34	.47	1.57	2.2	1.55	1.4	.37	.59	53.6	76.1	V
25	13	30	.91	.45	.49	-2.9	.40	-2.4	.78	.56	92.9	75.0	AB
37	13	30	.91	.45	.88	-.5	.87	-.3	.61	.56	78.6	75.0	AN
10	14	30	.71	.45	.96	-.1	.94	-.1	.57	.55	82.1	74.6	M
24	14	30	.71	.45	1.02	.2	1.03	.2	.54	.55	75.0	74.6	AA
35	14	30	.71	.45	1.11	.6	1.38	1.2	.48	.55	75.0	74.6	AL
3	16	30	.32	.44	1.14	.8	2.28	3.2	.39	.52	75.0	73.2	F
30	16	30	.32	.44	1.57	2.5	1.63	1.8	.27	.52	46.4	73.2	AG
33	16	30	.32	.44	.88	-.6	.79	-.6	.58	.52	82.1	73.2	AJ
7	17	30	.13	.44	1.24	1.2	1.25	.8	.40	.50	64.3	72.6	J
9	17	30	.13	.44	.74	-1.4	.63	-1.2	.62	.50	78.6	72.6	L
23	18	30	-.06	.44	.62	-2.3	.52	-1.5	.66	.48	89.3	72.3	Z
29	18	30	-.06	.44	1.16	.9	1.12	.4	.42	.48	67.9	72.3	AF
31	18	30	-.06	.44	.76	-1.3	.67	-.9	.59	.48	89.3	72.3	AH
38	18	30	-.06	.44	1.00	.1	1.57	1.5	.45	.48	75.0	72.3	AO
40	18	30	-.06	.44	1.68	3.1	1.76	1.9	.18	.48	53.6	72.3	AQ
20	19	30	-.26	.44	1.16	.9	2.53	2.9	.34	.47	71.4	71.9	W
22	19	30	-.26	.44	1.08	.5	1.09	.4	.42	.47	78.6	71.9	Y
26	20	30	-.45	.45	1.00	.1	.79	-.4	.46	.44	53.6	71.4	AC
32	20	30	-.45	.45	.88	-.6	.75	-.5	.51	.44	75.0	71.4	AI
11	22	30	-.87	.46	.87	-.6	.66	-.5	.47	.40	85.7	73.8	N
13	22	30	-.87	.46	1.17	.9	.92	.0	.35	.40	64.3	73.8	P
21	22	30	-.87	.46	.85	-.8	.72	-.4	.47	.40	78.6	73.8	X
15	23	30	-1.09	.48	.72	-1.4	.50	-.8	.51	.37	82.1	76.2	R
6	24	30	-1.33	.50	.79	-.8	.53	-.6	.46	.35	82.1	79.3	I
14	24	30	-1.33	.50	.71	-1.3	.47	-.7	.49	.35	82.1	79.3	Q
27	24	30	-1.33	.50	1.45	1.7	1.67	1.0	.12	.35	75.0	79.3	AD
28	24	30	-1.33	.50	.84	-.6	.76	-.1	.41	.35	82.1	79.3	AE
5	25	30	-1.59	.53	.90	-.3	.56	-.4	.39	.32	78.6	82.4	H
2	26	30	-1.89	.57	.97	.0	.80	.1	.30	.28	85.7	85.7	E
4	26	30	-1.89	.57	.88	-.2	.57	-.2	.35	.28	85.7	85.7	G
8	26	30	-1.89	.57	.76	-.6	.44	-.4	.41	.28	85.7	85.7	K
1	27	30	-2.26	.64	.94	.0	1.71	.9	.22	.25	89.3	89.2	D
MEAN	17.6	30.0	.00	.48	.99	.0	1.00	.1			77.6	76.9	
S.D.	5.8	.0	1.25	.05	.26	1.2	.52	1.1			10.9	5.1	

*TABLE 13.3 Ujian Pra algebra 40 item 30 responde ZOU486WS.TXT Jun 16 3:04 2020
INPUT: 30 PERSON 40 ITEM REPORTED: 30 PERSON 40 ITEM 2 CATS WINSTEPS 3.73

Figure 3: Infit and Outfit Mean Square (MNSQ)

Table 1: Items That Did Not Fit the MNSQ Range

Items with an MNSQ value of less than 0.6	Items with an MNSQ value of more than 1.4
AK (0.41)	AM (1.60)
AB (0.40)	F (2.28)
Z (0.52)	AG (1.63)
R (0.50)	AO (1.57)
I (0.53)	AQ (1.76)
Q (0.47)	W (2.53)
K (0.44)	AD (1.67)
	D (1.71)

Level of Item Difficulty and Respondent Capability

Result of the analysis found that all items are spread and directed to the level of capability according to the variety of respondents. The respondents with high competency are at the top position of the scale, whereas respondents with low competency are at the bottom part of the scale as can be seen in Figure 4. Item map shows that T is an item that can easily be answered by the students and it is at the bottom part of the scale which indicate that this item is constantly answered correctly by the students. The items at the top part of the scale in terms of difficulty are items AK, AM, and AP.

Volume: 4 Issues: 12 [September, 2021] pp. 31 - 41]
Jurnal Penyelidikan Sains Sosial (JOSSR)
eISSN: 2637 -0956
Journal Website: www.jossr.com

The result of the findings discussed found that the Rasch model approach in the construction of achievement test question items and questionnaire items can increase the accuracy and achievement of students in algebra which gives a positive effect and benefit towards the progress and performance of students in mathematics, hence, able to solve problems in teaching and learning mathematics systematically and effectively.

The qualitative approach can describe the instrument scientifically to avoid doubts towards the accuracy of the empirical data obtained from the analysis of the quantitative research, which is the findings of the first objective obtained using the Rasch model software. A researcher conducting an empirical study to prove every infromation can be verified scientifically. In this research, the researhcer will describe the process and development in the implementation of instrument construction which is tested and supported by analysis using the Rasch model. In this research, the experience attained by the teachers is used as evidence that all instrument testing processes are conducted systematically.

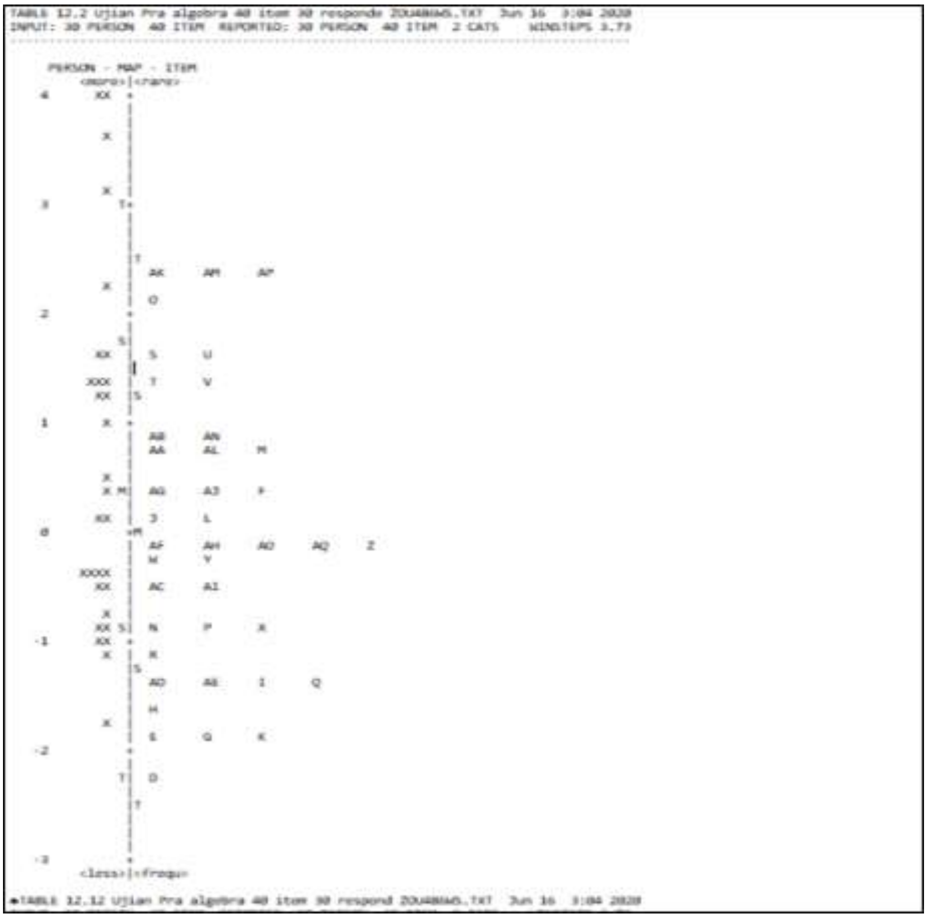

Figure 4: Item Map Algebra Achievement Testing

Interview Findings with Teachers

The results of the interview analysis have successfully answered two research objectives which are how is the validity and reliability of the algebra testing items using the Rasch model and how is the formative developmentof the command of algebra among students. For the findings of the first research question, it was found that the usage of the Rasch model in constructing multiple choice questions measures the actual capability of the students based on their probability in answering difficult and easy items. The probability, which was succesfully placed in a new scale based on the logit, have increased the validity and reliability of the items in terms of measuring the actual level of student command.

Next, for the second research finding, the results of the interview indicate that there is an increase in achievement through the usage of the test items that were analysed with the Rasch model and student self-efficacy. This is due to the accurate assessment on the level of capability and it was sensed that the items used are at the level of difficulty where most of them can be answered well by the students and only a few of them are outside of the students' capability to be understood and solved.

Summary

The research findings obtained indicate that students who utilised the instrument constructed using the validity and reliability via the Rasch model display a significant increase. The result of the students who utilised the instrument using the Rasch model also display a more consistent trend compared to the group of students who utilised an instrument that did not use the Rasch model to determine item validity and reliability. Empirical test results alone are not enough to determine the suitability of the instrument items used. To ensure that the instrument is also assessed scientifically, interview with teachers who have used the instrument is also analysed to obtain detailed information.

The result of the analysis of the teacher interview, which assessed every item and answer given by the students throughout the period of intervention,will provide practical confirmation regarding the actual application in the classroom compared to the theory or model used. This is in line with the research conducted by Cascella *et al*., (2020) who uses a mixed method in a study regarding the construction of a mathematics test instrument which uses both quantitative and qualitative approach.

There are obviously many research conducted about the Rasch model. However, there is still a huge gap which requires a deeper study, which employs a qualitative approach to look deeper and not only looking at the scores obtained but also the dimensions of research aspects that have not been explored regarding the research instrument used. This is to ensure the continuation of quality, exactitude, and consistency which can inspire a more effiecient and effective teaching and learning.

Thus, this research provides an innovation in terms of analysis on the method of preparing a mathematics test which has been tested quantitavely using the Rasch model and assessed by teachers based on their observation towards students in the classroom. This is to ensure item accuracy and suitability in the test questions thus, providing a more accurate assessment.

Conclusion

Overall, significantly application of the usage of validity and reliability of the algebra achievement test using the Rasch model at SMK Sikuati II, Kudat, Sabah can provide an opportunity and a space for students and teachers to conduct effective lessons and this model can be applied for learning in other fields. This implication can contribute in the development of the method of construction of assessment tests and systematically holistic increase in students achivement.

Rujukan

Alkhadim, G. S., Cimetta, A. D., Marx, R. W., Cutshaw, C. A., & Yaden, D. B. (2021). Validating the Research-Based Early Math Assessment (REMA) among rural children in Southwest United States. *Studies in Educational Evaluation*, *68*(November 2020), 100944. https://doi.org/10.1016/j.stueduc.2020.100944

Asshaari, I., Othman, H., Bahaludin, H., Ismail, N. A., & Nopiah, Z. M. (2012). Appraisal on Bloom's Separation in Final Examination Question of Engineering Mathematics Courses using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, *60*(2009), 172–178. https://doi.org/10.1016/j.sbspro.2012.09.364

Bond, T. G., & Fox, C. M. (2007). *Fundamental measurement in the human sciences*. Institute for Objective Measurement.

Brown, A., & Danaher, P. A. (2019). CHE principles: Facilitating authentic and dialogical semi-structured interviews in educational research. *International Journal of Research & Method in Education*, *42*(1), 76–90.

Cascella, C., Giberti, C., & Bolondi, G. (2020). An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks. *Studies in Educational Evaluation*, *64*(November 2018), 100819. https://doi.org/10.1016/j.stueduc.2019.100819

Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. *Procedia - Social and Behavioral Sciences*, *129*, 133–139. https://doi.org/10.1016/j.sbspro.2014.03.658

Chua, Y. P. (2011). *Kaedah dan statistik penyelidikan: kaedah penyelidikan*. Mcgraw-Hill Education.

Cleary, T. J., Velardi, B., & Schnaidman, B. (2017). Effects of the Self-Regulation Empowerment Program (SREP) on middle school students' strategic skills, self-efficacy, and mathematics achievement. *Journal of School Psychology*, *64*, 28–42. https://doi.org/10.1016/J.JSP.2017.04.004

Confrey, J., Toutkoushian, E., & Shah, M. (2020). Working at scale to initiate ongoing validation of learning trajectory-based classroom assessments for middle grade mathematics. *Journal of Mathematical Behavior*, *60*(January), 100818. https://doi.org/10.1016/j.jmathb.2020.100818

Crawford, A. R., Johnson, E. S., Zheng, Y. Z., & Moylan, L. A. (2020). Developing an understanding procedures observation rubric for mathematics intervention teachers. *School Science and Mathematics*. https://doi.org/10.1111/ssm.12393

Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research, and Evaluation*, *18*(1), 3.

Fusch, P., Fusch, G. E., & Ness, L. R. (2018). Denzin's paradigm shift: Revisiting triangulation in qualitative research. *Journal of Social Change*, *10*(1), 2.

Linacre, J. M. (1999). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of Outcome Measurement*, *3*, 382–405.

Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus*, *52*(2), 131–139.

Malaysia Education Blueprint, M. (2013). Malaysia Education Blueprint 2013 - 2025. *Education*, *27*(1), 1–268. https://doi.org/10.1016/j.tate.2010.08.007

Mashitah, M. R., Nor, M. M., Masnan, A. H., & Masran, M. (2017). Evaluation of an early science and mathematics standard instrument: task development, scoring set and psychometric issues. *Advanced Science Letters*, *23*(3), 2097–2100.

Samsudin, M. A., Chut, T. S., Ismail, M. E., & Ahmad, N. J. (2020). A calibrated item bank for computerized adaptive testing in measuring science TIMSS performance. *Eurasia Journal of Mathematics, Science and Technology Education*, *16*(7).

https://doi.org/10.29333/EJMSTE/8259

Sjoe, N. M., Bleses, D., Dybdal, L., Tideman, E., Kirkeby, H., Sehested, K. K., Nielsen, H., Kreiner, S., & Jensen, P. (2019). Short Danish Version of the Tools for Early Assessment in Math (TEAM) for 3–6-Year-Olds. *Early Education and Development*, *30*(2), 238–258. https://doi.org/10.1080/10409289.2018.1544810

Spoden, C., Fleischer, J., & Frey, A. (2020). Person misfit, test anxiety, and test-taking motivation in a large-scale mathematics proficiency test for self-evaluation. *Studies in Educational Evaluation*, *67*(July), 100910. https://doi.org/10.1016/j.stueduc.2020.100910

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Trim Komunikata Publishing House.

Tajudin, M., & Chinnappan, M. (2016). Relationship between scientific reasoning skills and mathematics achievement among Malaysian students. *Geografia : Malaysian Journal of Society and Space*, *12*(1), 96–107.

Yang, Y., He, P., & Liu, X. (2018). Validation of an Instrument for Measuring Students' Understanding of Interdisciplinary Science in Grades 4-8 over Multiple Semesters: a Rasch Measurement Study. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-017-9805-7