

# Rasch Model: Quality of Test Instruments to Measure Students' Ability in Algebra

Irmawaty Natsir<sup>1\*</sup>, Dian Mayasari<sup>2</sup>, Anis Munfarikhatin<sup>3</sup>

<sup>1,2,3</sup> Department of Mathematics Education, Universitas Musamus, Merauke, Indonesia

**Abstract.** This research is research using a quantitative approach that aims to analyze the quality of the test instrument by using the Rasch model to measure the ability of students in junior high school mathematics courses, especially in Algebra material. The sample used in the study was 37 students. The instrument used is a multiple-choice test with four answer choices consisting of 30 questions. The results obtained from this study there are 27 valid items and 3 invalid items. The reliability of the test score (reliability) of the estimation results using the Rasch model shows that the reliability is included in the good category with a person reliability value of 0.74, item reliability of 0.84 and Cronbach's Alpha of 0.85. For the level of difficulty or difficulty of the items, 7 items are in the very easy category, 7 items are in the easy category, 11 items are in the difficult category and 5 items are in the very difficult category. Of the 30 questions, there is 1 item that must be discarded because the question is too easy. The level of ability of students with serial numbers 18 and 4 is classified as person misfit in the Rasch model. This is because students with serial numbers 18 and 4 have an unusual pattern of answers, namely not being able to answer correctly on items that are classified as easy, but able to answer correctly on items that are classified as difficult.

**Keywords:** Validity, Reliability, Difficulty Level, Ability, Rasch Model.

## 1 Introduction

Each learning process requires an evaluation process in order to know the abilities possessed by students and the results achieved by educators in the learning process. The process cannot be separated from the role of educators, as well as at the university level. Lecturers play a role in the evaluation process and assess each student's ability in each subject taught. The evaluation process needs to be carried out with good and correct procedures in order to measure the ability of each student according to the actual situation [1]. A good assessment must require a good instrument as well [2]. The assessment carried out in the learning process is a step to improve the learning process and improve the quality of student learning. The results of the assessment become material for reflection for lecturers and students. In order for the purpose of the assessment to be achieved, a planned, gradual and continuous evaluation process is needed to obtain an overview of the development of the learning process of each student.

Assessment in the learning process is a process to collect, analyze and interpret information to determine the level of achievement of learning objectives from time to time [3][4]. The assessment process carried out so far aims to collect information that is used as the basis for knowing the level of progress, development and achievement of student learning and knowing the

success of a lecturer in the learning process carried out. Measurement of students' cognitive abilities can be done using the test method. The selection of the right test method is expected to measure the ability of each student well.

In this study, measurements were carried out using a test instrument. The test instrument is designed to measure the ability of students in Junior High School Mathematics, especially Algebra. The test instrument used has been designed by taking into account the construct component and the descriptive component. The construct component is a question item that can provide an indication of the answers to be obtained, starting from the low stage to the high stage. The questions in this study will produce answers that show a certain qualitative range, namely two alternative answers, true or false, known as a dichotomous response pattern [5]. Furthermore, the descriptive component is a component that explains several points on certain aspects. The aspect that is considered in this study is how the students' ability in working on algebra problems is.

The quality of the test instruments used can be analyzed using the Rasch model. Basically, the Rasch model is a measurement model based on modern test theory by considering the respondent's ability to answer the test or questionnaire given to the level of difficulty of each item making up the instrument [5][6][7]. As a

\* Corresponding author: [natsir\\_fkip@unmus.ac.id](mailto:natsir_fkip@unmus.ac.id)

new measurement system, the Rasch model aims to overcome the limitations of the classical measurement system or Classical Test Theory (CTT) that has been used so far [8][9][10]. In classical measurement, item parameters in the form of analysis results of item difficulty level and item discrimination index are group dependent [11]. In terms of item difficulty level, the classification of item difficulty level will change when given to different sample groups [12], whereas in terms of discrimination index, higher scores tend to be obtained from heterogeneous samples and lower scores are obtained from homogeneous samples [13]. Meanwhile, in modern test theory, item parameters do not change even though they are estimated from different sample groups [14]. This means, modern test theory provides a uniform measurement scale [12], so that sample groups can be tested with a different set of items, according to their ability level and the scores obtained can be directly compared [15].

The Rasch model has the advantage of predicting missing data based on a systematic response pattern which cannot be done in classical theory [16]. The Rasch model is based on two principles, namely the ability of the subject and the relationship between the ability of the subject. The ability of the subject referred to in this case is the ability of students to a question that can be predicted by using a set of factors called traits. Traits is a dimension of individual ability consisting of cognitive, psychomotor and verbal abilities. Where, the correlation that occurs between the ability of the subject, in this case what is meant is the ability of students on a question or question that is related to other abilities and can be described in a grain characteristic curve [17]. In the Rasch model, students with high abilities have a greater chance of answering questions correctly than other students. On the other hand, students with low abilities have a smaller opportunity to answer questions correctly where the questions have a higher level of difficulty [18].

In the Rasch model, the score generated from the test instrument aims to determine Outfit MNSQ, Outfit ZSTD, Point Measure Correlation, Item Reliability and Apha Cronbach. The MNSQ outfit is useful for seeing the suitability of the data with the model used. The expected Mean Square value is 1 (one). If the mean-square value on the outfit is greater than one, then the variation of the instrument is more than the prediction made by the Rasch model. If the outfit value is less than 1, then the variation in the instrument is less when compared to the predictions made by the Rasch model [1].

Referring to the explanation above, the purpose of this study is to determine the quality of the instruments used to measure students' abilities on the competencies tested on algebraic material using the Rasch model approach.

## 2 Research Methods

The approach used in this study is a quantitative approach, with a total sample of 37 students who have attended Junior High School Mathematics courses. The

instrument used in this study was a multiple-choice test with four answer choices containing algebraic material that had been obtained by students at the junior secondary level. The number of test items used in this study were 30 items with easy, medium and difficult item difficulty levels. The test results obtained are in the form of scores that will be analyzed using Winsteps Rasch software [19] which aims to determine fit and misfit items. In addition, to determine the value of Cronbach's Alpha which is the result of the overall item reliability test. The value of Outfit MNSQ, Outfit ZSTD and the correlation value of the items as a whole shows the limit of items that are declared fit with the model. If the Outfit MNSQ value is  $0.5 < \text{MNSQ} < 1.5$ ; Outfit ZSTD value ranges from  $-2.0 < \text{ZSTD} < 2.0$ ; and the correlation value of the total score (Point Measure Correlation) ranges from  $0.4 < \text{Pt M Corr} < 0.85$  [18][20]

## 3 Results and Discussion

### 3.1 Validity Test Result

Analysis of the quality of the ability test instrument on algebraic material seen from the aspects of validity, reliability, level of problem difficulty and ability. In the Rasch Model, to see the quality of the item from the aspect of validity, it is if it meets the following criteria [18].

Outfit MNSQ (Mean Square) value received is  $0,5 < \text{MNSQ} < 1,5$

Outfit ZSTD value received is  $-2,0 < \text{ZSTD} < 2,0$

Pt Measure Corr (Point Measure Correlation) value is  $0.4 < \text{Point Measure Corr} < 0.85$

After testing for the validity aspect on each item, it can be seen in Table 1 and Table 2.

Based on the result of the analysis of students' ability test instruments on algebraic material in items of item validity using the Rasch model, 27 items were said to be valid and 3 items were said to be invalid because they did not meet the requirements for MNSQ outfit, ZSTD Outfit, and Point Measure Correlation (Pt Measure Correlation).

**Table 1.** Validity between items

Entry Number	Measure	Infit		Outfit	
		Mnsq	Zstd	Mnsq	Zstd
6	3,05	1,47	1,8	2,25	2,1
17	2,85	1,44	1,8	1,99	1,9
22	0,95	1,39	2,4	1,77	3,1
13	1,76	1,22	1,4	1,65	2,5
30	-0,49	1,40	1,5	1,57	1,2
3	-2,84	1,10	0,4	1,40	0,7
8	-1,60	1,24	0,6	1,25	0,6
29	1,62	1,24	1,6	1,16	0,8
10	-0,96	1,09	0,4	0,79	-0,2
11	-2,84	1,01	0,3	0,45	0,0
12	-1,60	1,01	0,2	0,51	-0,4
2	0,06	0,98	0,0	0,89	-0,2
16	-0,49	0,97	0,0	0,78	-0,4
21	-0,11	0,96	-0,1	0,77	-0,5
5	1,76	0,95	-0,3	0,83	-0,7
23	0,38	0,95	-0,2	0,79	-0,7
24	0,22	0,91	-0,5	0,74	-0,8
4	-0,11	0,88	-0,5	0,71	-0,7
26	0,22	0,88	-0,6	0,71	-0,9
25	0,95	0,81	-1,3	0,86	-0,6
28	-1,25	0,86	-0,2	0,51	-0,6
9	0,06	0,85	-0,7	0,66	-1,0
20	-0,71	0,85	-0,5	0,55	-0,8
1	-2,08	0,83	-0,1	0,45	-0,2
14	0,38	0,80	-1,2	0,78	-0,7
7	0,06	0,79	-1,1	0,61	-1,2
15	0,52	0,79	-1,3	0,66	-1,4
27	0,95	0,78	-1,6	0,69	-1,6
19	-0,71	0,75	-0,8	0,47	-1,1

**Table 2.** The results of the analysis of the validity of the item

No	Result	Item Number
1	Valid	1, 2, 3, 4, 5, 7,8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30
2	Invalid	6, 17, 18

**3.2 Item Reliability**

The criteria in determining the value of Item Reliability and Person Reliability in the Rasch model are based on the following criteria [18].

- a. > 0,94 : Special
- b. 0,91 – 0,94 : Very good
- c. 0,81 – 0,90 : Good
- d. 0,67 – 0,80 : Enough
- e. < 0,67 : Weak

The results of the reliability analysis of students' ability test instruments on algebraic material can be seen in Table 3.

**Table 3.** The results of the instrument reliability analysis

	Reliability
Person	0,74
Item	0,84

The reliability of the test score estimation results indicate that the reliability is included in the good category. It is based on two reliability coefficients including item reliability and person reliability. Both coefficients are interpreted to have the same provisions in determining reliability in classical measurements by looking at the Cronbach Alpha value [19]. The value of the reliability of the test score obtained, if it is close to the value of 1, it shows a more internally consistent measur [20]. In this study, the value of person reliability and item reliability are 0.74 and 0.84, respectively.

**3.3 Item Difficulty Level**

The level of difficulty of the items indicates the possibility of how many respondents can answer an item correctly. In the Rasch model, the difficulty level of the items is categorized based on the Logit Measure and the Logit Item Standard Deviation (SD) value and is divided into four categories [18] as follows.

- Measure logit < – SD logit : Very easy items
- SD logit ≤ Measure logit ≤ 0 : Easy items
- 0 ≤ Measure logit ≤ SD logit : Difficult items
- Measure logit > SD logit : Very difficult items

The results of the analysis of students' ability test instruments on algebraic material for the difficulty of each it item can be seen in Table 4 dan Tabel 5.

**Table 4.** Item difficulty level

Entry Number	Measure	Infit		Outfit	
		MNSQ	ZSTD	MNSQ	ZSTD
6	3,05	1,47	1,8	2,25	2,1
17	2,85	1,44	1,8	1,99	1,9
22	0,95	1,39	2,4	1,77	3,1
13	1,76	1,22	1,4	1,65	2,5
30	-0,49	1,40	1,5	1,57	1,2
3	-2,84	1,10	0,4	1,40	0,7
8	-1,60	1,24	0,6	1,25	0,6
29	1,62	1,24	1,6	1,16	0,8
10	-0,96	1,09	0,4	0,79	-0,2
11	-2,84	1,01	0,3	0,45	0,0
12	-1,60	1,01	0,2	0,51	-0,4
2	0,06	0,98	0,0	0,89	-0,2
16	-0,49	0,97	0,0	0,78	-0,4
21	-0,11	0,96	-0,1	0,77	-0,5
5	1,76	0,95	-0,3	0,83	-0,7
23	0,38	0,95	-0,2	0,79	-0,7
24	0,22	0,91	-0,5	0,74	-0,8
4	-0,11	0,88	-0,5	0,71	-0,7
26	0,22	0,88	-0,6	0,71	-0,9
25	0,95	0,81	-1,3	0,86	-0,6
28	-1,25	0,86	-0,2	0,51	-0,6
9	0,06	0,85	-0,7	0,66	-1,0
20	-0,71	0,85	-0,5	0,55	-0,8
1	-2,08	0,83	-0,1	0,45	-0,2
14	0,38	0,80	-1,2	0,78	-0,7
7	0,06	0,79	-1,1	0,61	-1,2
15	0,52	0,79	-1,3	0,66	-1,4
27	0,95	0,78	-1,6	0,69	-1,6
19	-0,71	0,75	-0,8	0,47	-1,1

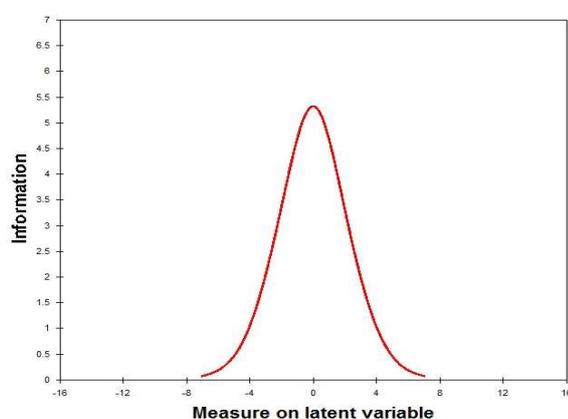
**Table 5.** The results of the distribution of item difficulty levels using the Rasch Model

Measure Value (logit)	Interpretation of Item Difficulty	Number of Items
Measure logit < - SD logit	Very easy items	7
- SD logit ≤ Measure logit ≤ 0	Easy items	7
0 ≤ Measure logit ≤ SD logit	Difficult items	11
Measure logit > SD logit	Very difficult items	5

Based on table 4 and 5, the level of difficulty or difficulty of the items in the Rasch model is seen from the measure value in the logit unit of each item. A total of 7 questions are in the very easy category, namely points 18, 11, 03, 01, 12, 08, 28. A total of 7 questions are in the easy category, namely points 10, 20, 19, 30, 16, 21, 4 A total of 11 items are in the difficult category, namely items 2, 7, 9, 24, 26, 14, 23, 15, 22, 25, 27. A total of 5 items are in the very difficult category, namely points 6, 17, 5, 13, 29.

### 3.4 Ability Level

In the Rasch model, an analysis of the ability level is carried out which aims to distinguish the abilities of students who are able to answer questions and those who are unable to answer questions. The results of the analysis for the level of capability can be seen in figure 1.



**Fig.1.** Test information function

Figure 1 shows that students with serial numbers 18 and 4 are classified as person misfit. This is because the student has an unusual pattern of answers, namely students are not able to answer correctly on items that are classified as easy, but are able to answer correctly on items that are classified as difficult. Students with lower levels of ability do not have the opportunity to solve problems that are classified as difficult, so it can be concluded that the answers given by students with serial numbers 18 and 4 are probably just guessing and the guess happens to be true or known as lucky guessing., the answer from the student is probably the result of cheating. This is in line with the opinion [21] and [22] who stated that there are 5 things that cause person

misfit, namely cheating or copying answers from friends/other exam participants. Such actions refer to unfair behavior in the answers obtained because the actual item questions cannot be answered correctly. Careless responding occurs when someone answers correctly on question items that are classified as difficult/difficult and cannot answer correctly on question items that are classified as easy. Lucky guessing occurs when someone works on a given problem by guessing and the answer is correct, but he doesn't know why the answer is correct. Creative responding only happens to someone who has a high ability when responding incorrectly to a question item that is actually quite easy, this is because they interpret the item in a unique and creative way. And, random responding refers to a situation where someone chooses an answer option randomly on a question regardless of whether the selected option is true or false. Furthermore, person misfit can also occur if someone answers all items correctly which causes extreme scores so that the fit statistics cannot be measured (over fit). According to [23], person fit measurement not only identifies an impossible response pattern, but also identifies a very likely response pattern, so that if too much uncertainty is predicted, it indicates a constraint on the response.

### 4 Conclusion

The results showed that the quality of the questions about students' abilities in algebraic material through the Rasch model approach was generally categorized as good quality questions with 27 valid items out of 30 items. Reliability analysis is in good category with a person reliability value of 0.74, item reliability of 0.84 and Alpha Crobach of 0.85. Analysis of the level of difficulty or difficulty of the questions showed 7 items belonging to the very easy category, 7 items belonging to the easy category, 11 items belonging to the difficult category and 5 items belonging to the very difficult category. Of the 30 questions, there is 1 item that must be discarded because the question is too easy. The level of ability of students with serial numbers 18 and 4 is classified as person misfit in the Rasch model. This is because the student has an unusual response pattern, namely being able to answer correctly on items that are classified as difficult and answer incorrectly on items that are classified as easy

### Acknowledgments

The Author Thanks to The Rector Of Universitas Musamus And From Lecturers In The Department Of Mathematics Education, Who Have Given Feedback In This Research.

### References

- [1] Azisah, Sapti Wahyuningsih, Penggunaan Model Rasch Untuk Analisis Instrumen Tes pada Matakuliah Matematika Aktuaria, vol.3, (2020), pp. 45-50. DOI : <https://doi.org/10.30598/jupitekvol3iss1ppx45->

- 50.
- [2] I Natsir, A Munfarikhatin, AR Taufik, Development of Student Worksheet Based on Blended Learning Oriented to Multiple Intelligences in Algebra, *International Joint Conference on Social Science*, vol.473, (2021), pp. 438-442.
- [3] Ratna Wulan, E & Rasdiana, A, *Evaluasi Pembelajaran*, Pustaka Setia, (2015).
- [4] Dunn, L., Morgan, C., O'Reilly, M., & Parry, S, *The student assessment handbook : New directions in traditional and online assessment*, Routledge, (2003).
- [5] Thissen, D., Nelson, L., Rosa, K., et al, *Item Response Theory for Items Scored in More than Two Categories* dalam D. Thissen & H. Wainer, *Test Scoring*, New Jersey : Lawrence Erlbaum Associates Publishers, (2001), pp. 141-184.
- [6] Rasch, G, *Probabilistic models for some intelligence and attainment test*, Chicago, IL: University of Chicago Press, (1980).
- [7] Rozeha, A. R., Azami, Z. & Mohd Saidudin, M, *Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering*, *Regional Conference on Engineering Mathematics, Mechanics, Manufacturing & Architecture*, (2007).
- [8] Ashraf, Z.A., & Jaseem, K, *Classical and modern methods in item analysis of test tools*, *International Journal of Research and Review*, vol.7, 2020, pp. 397-403.
- [9] Meyer, J.P., & Zhu, Shi, *Fair and equitable measurement of student learning in MOOCs: an introduction to item response theory, scale linking, and score equating*, *Journal of Research and Practice in Assessment*, vol.8, (2013), pp. 26-39.
- [10] Yilmaz, H.B, *comparison of IRT model combinations for assessing fit in a mixed format elementary school science test*, *International Electronic Journal of Elementary Education*, vol.11, (2019), pp. 539-545. DOI: <https://dx.doi.org/10.26822/iejee.2019553350>
- [11] Fan, X, *Item response theory and classical test theory: An empirical comparison of their item/person statistics*, *Education and Psychological Measurement*, vol.58, pp. 357-381.
- [12] Magno, C, *Demonstrating the difference between classical test theory and item response theory using derived test data*, *The International Journal of Educational and Psychological Assessment*, vol.1, (2009), pp. 1-11.
- [13] Bichi, A.A, *Classical test theory: An introduction to linear modelling approach to test and item analysis*, *International Journal for Social Studies*, vol.2, pp. 27-33.
- [14] Rezaee, R., Shafiayan, M., Jafari, P., & Zarifsanaiey, N, *Invariance of item difficulty parameter estimates based on classical test theory and item response theory*, *Journal of Advance Pharmacy Education and Research*, vol.8, pp. 156-161.
- [15] Anastasi, A. & Urbina, S, *Psychological testing*, Prentice Hall: New York, (2002).
- [16] Aziz, R, *Aplikasi model Rasch dalam pengujian alat ukur kesehatan mental di tempat kerja*. *Psikoislamika: Jurnal Psikologi dan Psikologi Islam*, vol. 12, pp. 29-39.
- [17] Hambleton, R.K., Swaminathan, H., & Rogers H.J, *Fundamental of item response theory*. London: Sage Publication, (1991).
- [18] Sumintono, B., & Widhiarso, W, *Aplikasi permodelan rasch pada assessment pendidikan*. Cimahi: Trim Komunikata, (2015).
- [19] Linacre, J.M, *A user's guide to WINSTEPS*. Chicago, IL: Winsteps, (2009).
- [20] Boone, W.J., Staver, J.R., & Yale, M.S, *Rasch analysis in the human sciences*. Dordrecht: Springer, (2014).
- [21] Meijer, R.R, *Person-fit research: an introduction*. *Applied Measurement in Education*, vol.9, (1996), pp. 3-8.
- [22] Karabatsos, G, *Comparing the aberrant response detection performance of thirty-six person-fit statistics*. *Applied Measurement in Education*, vol.16, (2003), pp. 277-298.
- [23] Meijer, R.R., & Sitsma, K, *Person fit statistic: What is their purpose?.* *Rasch Measurement Transactions*, vol.15, (2001).