Analysis of Students' Mathematical Reasoning Instrument Using Rasch Model

Elfis Suanto[1(✉)], Putri Yuanita[1], Armis Armis[1], and Fadilah Gustin Khainingsih[1]

University of Riau, Pekanbaru, 28296, Indonesia
elfis.suanto@lecturer.unri.ac.id

Abstract. Reasoning ability is a necessary and important ability for students of mathematics education. This ability is required to construct a mathematical thought and provide evidence that the idea is true. Therefore, instruments for measuring mathematical reasoning abilities are very important. The focus of this research was to analyze the instrument for evaluating mathematical reasoning abilities based on the outcomes of the mathematical reasoning ability tests of students of the Mathematics Education Study program in the subject of Initial Value Problems and Boundary Conditions. The goal of this study was to analyze the instrument for assessing mathematical reasoning abilities. This research is quantitative research. Data were analyzed using the Winstep 5.2.3.0 application to obtain Rasch Model Analysis (RMA) item fit and polarity based on PTMEA CORR, misfit items, item and person separation and reliability, and a person-item map. The findings demonstrated that the instrument for evaluating students' mathematical reasoning skills was valid and reliable because the items matched the criteria for measurement. Therefore, it can be said that the test can be used to evaluate pupils' aptitude for mathematical reasoning.

Keywords: Mathematical Reasoning, Instrument, Rasch Model.

1 Introduction

Students must have mathematical reasoning abilities, especially mathematics education students. This is in line with the vision of mathematics to meet future needs. Learning mathematics aims to give possibilities for the development of logical thinking skills, an understanding of the value of mathematics, increasing self-confidence, and an objective and open mindset to face a constantly changing future [1]. The assertion demonstrates that in order to establish a mathematical idea and provide evidence that it is true, reasoning is required. Reasoning becomes important in life, especially in mathematics. Because mathematics contains active, dynamic, and generative processes that are carried out by actors and users of mathematics [2]. Therefore, reasoning ability is the capacity of a person to offer a logical conclusion based on a single legitimate piece of data and a methodical approach to problem-solving. The capacity for reasoning consists of a number of skills, such as analysis, generalization, synthesis, justification, and non-routine problem solving [3].

J. Warmansyah et al. (eds.), *Proceedings of the International Conference on Social Science and Education (ICoeSSE 2023)*, Advances in Social Science, Education and Humanities Research 789,
https://doi.org/10.2991/978-2-38476-142-5_10

In the curriculum of the Mathematics Education study program University of Riau (2013), the distribution of subjects in the scientific and skills study group consists of 80 credits covering subjects related to the field of mathematics which are further divided into 5 groups of fields of study namely analysis, algebra, geometry, statistics, and applied mathematics. One of the subjects in the applied mathematics group is Initial Value Problems and Boundary Conditions. Initial Value Problems and Boundary Value is the application of Differential Equations to various simple real problems such as waves, heat flow and others. In this course eight major topics are discussed. To master all of the topics, the main competence is required in the form of mathematical reasoning abilities. Students' mathematical reasoning abilities can be measured through test instruments on Initial Value Problems and Boundary Conditions material. The exam is one technique to determine someone's talent level in an indirect manner, specifically through how they respond to various questions. A high-quality test has the traits of reliable test objects and machinery [4].

Several studies have been conducted regarding the ability of mathematical reasoning and measurement. Among them is research that Nurharyanto has done on the analysis of students' mathematical reasoning. In his research, it was found that from the results of the tests conducted, students' mathematical reasoning skills still need to be trained. Nurharyanto also said that test instruments to measure mathematical reasoning ability need to be tested for reliability first. Thus an instrument needs to be analyzed for reliability first before being used [5]. Unlike the research conducted by Nurharyanto, research conducted by Purwati shows that test questions that are worth using need to be analyzed for difficulty index, differentiating power and reliability. In his research, empirical analysis of items was carried out using the classical test theory (CTT) approach [6].

Different from previous studies, the instrument analysis carried out in this study using the Rasch model and the material of instrument is Initial Value Problems and Boundary Conditions. An instrument to produce accurate information requires proper analysis. There are two analytical approaches that can be done, namely the classical test theory approach and modern theory or grain response theory. The theory of grain response can be done using the Rasch Model. Rasch's model evolved quickly to overcome the CTT's limitations. The Rasch model can: 1) accurately estimate students' abilities; 2) offer a linear assessment between the test's format and students' abilities; 3) locate missing data; and 4) identify students' learning misconceptions [7].

A single instrument's reliability can be assessed using its individual reliability and items as well as its individual separation indexes and items according to the Rasch Analysis Model. In the meanwhile, item fit, item polarity, and unidimensionality can all be used to gauge an instrument's validity [8]. A single instrument's reliability can be assessed using its individual reliability and items as well as its individual separation indexes and items according to the Rasch Analysis Model. In the meanwhile, item fit, item polarity, and unidimensionality can all be used to gauge an instrument's validity. Validity refers to the accuracy of the measurement. To evaluate the reliability of the estimated metrics, fit statistics are used. The proportion of reproducible observed reactions is referred to as reliability. Both people and items have dependability estimates. How well we can distinguish individuals based on their assessed

visual ability is estimated by the person measuring dependability. How well objects may be differentiated from one another based on their level of difficulty is shown by the item measure dependability. Reliability is a number between 0 and 1. The less fluctuation in the measurement that can be attributed to measurement error, the closer the reliability is to 1.0 [9].

The focus of this research was to analyze the instrument for evaluating mathematical reasoning abilities based on the outcomes of the mathematical reasoning ability tests of students of the Mathematics Education Study program in the subject of Initial Value Problems and Boundary Conditions. The goal of this study was to analyze the instrument for assessing mathematical reasoning abilities.

2 Methods

The validity and reliability of the subject of Initial Value Problems and Boundary Conditions (IVP&BC) instruments used in this study to assess students' mathematical reasoning skills are measured quantitatively. The respondents used were students in semester 5 of the UNRI Mathematics Education, Initial Value Problems and Boundary Conditions course in 2022. The number of respondents was 34 students. The instrument consists of 5 items (I1, I2, I3, I4, I5) with indicators presented in Table 1.

Table 1. Indicators of Mathematical Reasoning on test

Indicators	Test Description
Analysis	Description In mathematical settings, identify or make use of relationships between variables or objects, and draw reliable conclusions from the data given in IVP&BC problems.
Generalization	Expanding the domain so that the results of mathematical thinking and problem solving on IVP&BC problems can be applied more generally and broadly
Synthesis	Make connections between different knowledge elements with related representations. Combining facts, concepts, and procedures in determining results and combining these results to determine further results.
Justification	Present evidence that is guided by known results or characteristics of the IVP&BC
Non-routine Problem-Solving	Solving problems in the context of IVP&BC with the aim that students are accustomed to solving similar problems and applying facts, concepts, and procedures in unfamiliar or complex contexts.

The instrument used is in the form of questions that have been developed by being analyzed using Rasch Model Analysis (RMA). The analysis was carried out by utilizing the Winstep 5.2.3.0 application. The Rasch Model was used to analyze the data's fitness and reliability. In order to determine the validity of the instrument using the output value, fit statistics of the items to the model were evaluated in order to provide a fit score that indicated if the items and individuals' behaviors were compatible with those of the model. The sufficiency of separation indices and the items' and people's dependability of the constructs were assessed [10]. According to the item fit based on

point measure correlation, misfit items, item and person separation and dependability, and person-item map, the validity and reliability of the instrument were examined. The Criteria of validity and reliability presented in Table 2.

Table 2. Summary of item validity and reliability using Rasch Model

Criteria	Statistic Info	Result
validity	Item fit and polarity	$0.4 < \text{PTMEA CORR} < 0.85$ [11]
	Misfit	$0.5 < \text{MNSQ} < 1.5$ [12]
Reliability		$-2.00 < \text{ZSTD} < +2.00$ [12]
	Item and person Separation	Value > 2.00 [13]
	Item and person Reliability	Value > 0.80 [14]

3 Results and Discussion

Students' acquired mathematical reasoning skills were used in the data analysis utilizing the Rasch Model to determine the test instrument's validity and reliability.

3.1 Item Fit Based on Point Measure Correlation

In order to determine how well constructs were created, the Point Measure Correlation (PTMEA CORR) was examined to look for polarity items. The item was not developed to assess the constructs to be measured if the value in the PTMEA CORR is negative (-). The item does not lead to the question, hence it has to be reevaluated [15]. The PTMEA CORR index for item category is extremely good for values greater than 0.40, good for values between 0.30 and 0.39, and fair for values between 0.20 and 0.29. The query is unable to distinguish between items if the values are in the range of 0 and 0.19 [16].

According to Table 2, there are no items with negative values in the PTMEA CORR, indicating that the item is measuring the target construct. According to the measure construct, all of the items' appropriateness falls into the "ex-tremely good" group. All of the tasks can be used to assess student reasoning mathematics abilities based on these PTMEA CORR indices.

Table 3. PTMEA CORR Index

Item	PTMEA CORR	Category
I5	0.60	Extremely good
I3	0.64	Extremely good
I4	0.72	Extremely good
I1	0.73	Extremely good
I2	0.81	Extremely good

3.2 Misfit Item

The outfit and infit Mean-Square (MNSQ) values show if an item is suitable for measuring the construct. To guarantee that the item is appropriate for measuring the

construct, that value should be between the range of 0.50 and 1.50. If the MNSQ index value is greater than 1.50, the item is considered to be mismatch, whereas an index value of 0.50 suggests that pupils can predict the item with ease. If the MNSQ value does not fall within those ranges, it is required to examine the derived Z score, also known as the Z Standardized (ZSTD) value, which must fall within the range of -2.00 and +2.00 to indicate that the data are reasonably predictable. The ZSTD index can be disregarded if the outfit and infit MNSQ are acceptable [13–15].

The infit MNSQ value ranges from 0.54 to 1.05, according to an analysis in Table 3, while the outfit MNSQ ranges from 0.37 to 1.25. There is only one item outside the range of the outfit MNSQ and no things outside the range of the infit MNSQ. I4 is the item. With 0.37, that item is below 0.50. However, outfit ZSTD value ranges from -1.73 to 0.30, and there are no items outside of this range, indicating that each item is suitable for assessing students' mathematical reasoning skills.

Table 4. Misfit Items According to Outfit and Infit MNSQ

Item	INFIT		OUTFIT		PTMEA CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
I2	1.05	0.30	1.25	0.99	0.81
I1	1.13	0.55	1.18	0.48	0.73
I3	1.18	0.72	1.11	0.54	0.64
I5	0.98	0.07	0.67	0.21	0.60
I4	0.54	-1.70	0.37	-0.13	0.72

3.3 Item and Person Separation and Reliability

According to [12] indices of item and person separation are a crucial contribution to the assessment of the performance of measuring tools. The separation index reveals the degree of difficulty for both the thing and the person. That index suggests that the tool can classify objects and people into distinct groups [14]. The ratio of the measurement inaccuracy is referred to as separation. The accepted minimum standard of performance is a separation of 2.00. How much trust can be placed in the constancy of the estimates of these weights is then determined by item and person reliability [13]. Category of the item-and-person dependability index Excellent for values greater than or equal to 0.94, Very Good for values in the range of 0.91 to 0.94, Good for values in the range of 0.81 to 0.90, Sufficient for values in the range of 0.67 to 0.80, and Low for values below 0.67 [16].

The dependability of the item and person for the test instrument, as well as their separability index, are shown in Table 5's data analysis. The item reliability value achieved is 0.99 (excellent) and the person reliability value is 0.86 (good) based on the Rasch Analysis Model. The values of the person separation index and item separation index are 2.44 and 10.98, respectively. Separation between item and person is more than 2.00. The test instrument has an adequate item and person separation index as well as a solid acceptance of item and person reliability. The analysis produced a result with a Cronbach Alpha reliability of 0.75. According to [16], the mean value of the instrument used is acceptable.

Table 5. Item and Person Separation and Reliability

Criteria	Separation	Reliability	Cronbach Alpha
Item	10.98	0.99	0.75
Person	2.44	0.86	

3.4 Person-Item Map

Rasch Model Analysis uses a logarithmic function to determine the relationship between human (student) ability and item difficulty. The range and position of the item measure distribution (left side of Figure 1) were compared to the range and position of the person measure distribution (right side of Figure 1) using the person-item map. In order to assess differences that are relevant, items should be placed at each point on the scale. If we want to assess everyone's capacity for mathematical thinking, the items must cover the entire ruler [9]. The student ability and item difficulty are shown in Figure 1 according to a logit scale that ranges from -2 to 1. Students with weak mathematical reasoning skills are depicted in the top left, while those with strong mathematical reasoning skills are shown in the bottom left. The person-item map's right side displays the item difficulty.

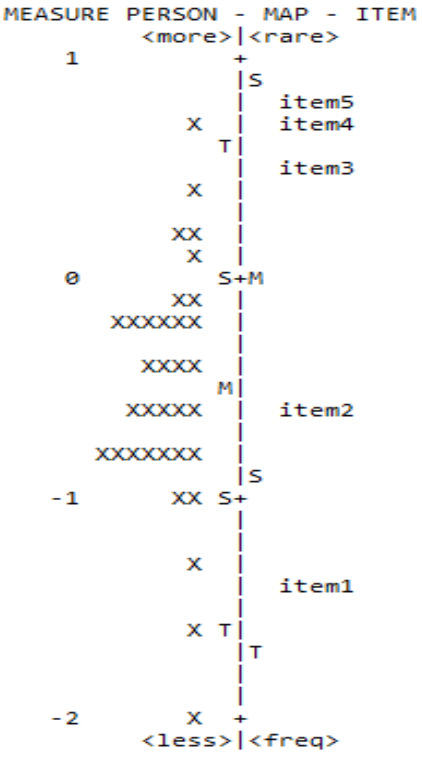

Fig. 1. Person-Item Map

4 Conclusion

Test instrument use in this study is built based on five indicators of reasoning mathematics ability. The indicators are analysis, generalization, synthesis, justification, and non-routine problem-solving. There are five items. Besides that, the analysis based on Rasch Model Analysis (RMA) indicates that the instrument for measuring students' mathematical reasoning abilities was valid and reliable because the items corresponded to what had to be measured on the instrument. So it can be concluded that the instrument can be used to measure the mathematical reasoning abilities of mathematics education students.

However, if additional researchers are interested in using the test instrument for various targeted populations, further analysis of the test instrument's validity and reliability is required. This research may also aid Community College lecturers in bettering several facets of the teaching and learning process, particularly in the areas of assessment, diagnosis, and intervention. To provide insightful results, it is imperative to develop a high-quality test. As a result, professors can utilize this test instrument to get feedback from students on the topics of Initial Value Problems and Boundary Value while also developing an intervention strategy that will work for the students.

References

1. U. Sumarmo, Berpikir dan Disposisi Matematik serta Pembelajarannya. Bandung: Jurusan Pendidikan Matematika FPMIPA UPI Bandung. Bandung: Jurusan Pendidikan Matematika FPMIPA Bandung, 2016.
2. A. . Schoenfeld, Mathematical Thinking and Problem Solving. UK: Lawrence Erlbaum Associates, 2014.
3. S. Y. Wulandari and D. U. Wutsqa, "A study of junior high school students reasoning skill in mathematics," 2019, doi: 10.1088/1742-6596/1320/1/012059.
4. F. Alfarisa and D. N. Purnama, "ANALISIS BUTIR SOAL ULANGAN AKHIR SEMESTER MATA PELAJARAN EKONOMI SMA MENGGUNAKAN RASCH MODEL," J. Pendidik. Ekon., vol. 11, no. 2, pp. 366–374, 2019.
5. D. W. Nurharyanto, "ANALISIS PENALARAN MATEMATIS MAHASISWA PGSD," vol. 4, no. 1, pp. 1–6, 2023.
6. H. Purwati, H. Retnawati, J. Jailani, and T. H. Retnowati, "Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik," J. Sains dan Edukasi Sains, vol. 4, no. 2, pp. 46–51, 2021, doi: 10.24246/juses.v4i2p46-51.
7. S. Samritin et al., "Matching the student ' s ability and their math test using Rasch analysis," 2019, doi: 10.1088/1742-6596/1318/1/012059.
8. M. F. Ridzuan, H. L. Lim, F. A. Ahmad Fozee, and S. N. A. Mohd Nasser, "Rasch Analysis Model: Reliability and Validity of Superitem Test Instrument," Int. J. Acad. Res. Progress. Educ. Dev., vol. 9, no. 4, 2020, doi: 10.6007/ijarped/v9-i4/8166.
9. J. Stelmack et al., "Use of Rasch person-item map in exploratory data analysis:," vol. 41, no. 2, pp. 233–241, 2004.
10. N. E. Ismail, N. S. Jimam, M. L. P. Dapar, and S. Ahmad, "Validation and Reliability of Healthcare Workers ' Knowledge , Attitude , and Practice Instrument for Uncomplicated Malaria by Rasch Measurement Model," vol. 10, no. January, pp. 1–8, 2020, doi: 10.3389/fphar.2019.01521.

11. I. Natsir, D. Mayasari, and A. Munfarikhatin, “Rasch Model : Quality of Test Instruments to Measure Students ’ Ability in Algebra,” vol. 01033, pp. 6–10, 2022.
12. W. J. Boone, J. R. Staver, and M. S. Yale, Rasch Analysis in the Human Sciences. Netherlands: Spinger, 2014.
13. M. S. Visser et al., “Validation, test–retest reliability and norm scores for the Dutch Catquest-9SF,” Acta Ophthalmol., vol. 95, no. 3, pp. 312–319, 2017, doi: 10.1111/aos.13287.
14. B. Sumintono and W. Widhiarso, Aplikasi pemodelan RASCH pada assessment pendidikan. Cimahi: Trim Komunikata Publishing House, 2015.
15. R. M. Yasin, F. A. N. Yunus, R. C. Rus, A. Ahmad, and M. B. Rahim, “Validity and Reliability Learning Transfer Item Using Rasch Measurement Model,” Procedia - Soc. Behav. Sci., vol. 204, no. November 2014, pp. 212–217, 2015, doi: 10.1016/j.sbspro.2015.08.143.
16. R. Rosli et al., “Student Awareness of Space Science: Rasch Model Analysis for Validity and Reliability,” World J. Educ., vol. 10, no. 3, p. 170, 2020, doi: 10.5430/wje.v10n3p170.
17. B. Setiawan, M. Panduwangi, and B. Sumintono, “A Rasch analysis of the community’s preference for different attributes of Islamic banks in Indonesia,” Int. J. Soc. Econ., vol. 45, no. 12, pp. 1647–1662, 2018, doi: 10.1108/IJSE-07-2017-0294.
18. T. G. Bond and C. M. Fox, Applying The Rasch Model: Fundamental Measurement in the Human Sciences. London: Mahwah, New Jersey, 2007.