Sunan Kalijaga International Journal on Islamic Educational Research (SKIJIER)
Vol. 7, No. 1, 2023, Page-57-67
DOI: https://doi.org/10.14421/skijier.2023.71.06

THE USE OF RASCH MODEL IN ANALYZING GRADE V LEMENTARY SCHOOL MATHEMATICS ITEMS

Debbi Silvia Aqida[1]
Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
22204085003@student.uin-suka.ac.id[1]

Destia Khairunisa[2]
Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
22204085019@student.uin-suka.ac.id[2]

Rovika Meisya[3]
Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
22204085014@student.uin-suka.ac.id[3]

Raudhatul Jannah[4]
Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
22204085018@student.uin-suka.ac.id[4]

Abstract: This study was conducted to analyze the items used to measure students' mathematical abilities in a private elementary school in Pekalongan. There were 22 fifth-grade students involved in the study. The test items consisted of 15 multiple-choice questions related to the topics of spatial calculations and data presentation. The Rasch model was used to determine the fitness of these test items. The analysis was carried out using the Winsteps software. Test items were considered fit for the model in the Winsteps program if the Outfit MNSQ value ranged from 0.5 to 1.5, the Outfit ZSTD value was between -2 and 2, and the Pt-measure Corr value was positive, indicating that the item was a good fit for the model (Sumintono & Widiharso, 2015:98). The Winsteps program's output revealed that 15 test items were in accordance with the Rasch model, with an Outfit MNSQ value of 1.03. The Outfit ZSTD value was 0.0. Additionally, 11 test items had a positive Pt-measure Corr, while 4 test items had a negative Pt-measure Corr, indicating that these 4 items did not fit the model.

Keywords: Item analysis, Rasch Model, Winstep.

Received: 21-02-2023	Revised: 30-05-2023	Accepted: 17-06-2023

INTRODUCTION

Education is an effort to enhance the quality of education for the progress of the nation, enabling Indonesia to compete in the era of globalization. The role of educators in the learning process is crucial in determining the quality of education. Educators' roles are not confined to designing and implementing learning activities but also encompass assessing the process and learning outcomes.[1]

In the context of primary school education, teachers play a crucial role in evaluating and assessing the students' abilities in the subjects they teach. The evaluation process should be carried out effectively to measure the true abilities of the students. Effective assessment requires well-constructed test instruments. The assessment of learning outcomes conducted by teachers can be a significant step in improving the learning process. Furthermore, it serves to enhance the quality of learning for the students. To achieve these goals, the evaluation process should be carried out systematically, progressively, and continuously to gain insights into the learning progress of the students.

According to Dunn there are three main objectives of the assessment process in learning: diagnosing students learning difficulties, measuring improvement over time, and providing information that can be used by students to enhance their performance. Thus far, the learning assessment process conducted by teachers has been closely associated with the acquisition and collection of data on students' learning abilities, enabling a comparison of the abilities of different students'. Additionally, learning assessment can be used to gauge the teacher's success in the classroom teaching process. In practice, the measurement of specific competencies during the learning process can be accomplished through testing methods. The selection of the appropriate testing method is expected to effectively measure students abilities[2].

Russell & Airasian assert that assessment is the most critical component in education[3]. Assessment requires a tool to measure students' abilities, and one of the tools used is tests. To ensure that tests accurately measure the true abilities of students, a good test instrument is required. A good test instrument should meet three criteria: the content is aligned with the material to be tested (content validity), the test has a sound construct (construct validity), and it exhibits

---

[1] RK Hambleton, "Swaminathan. H., & Rogers, HJ (1991). Fundamentals of Item Response Theory" (n.d.).

[2] Lee Dunn et al., *The Student Assessment Handbook: New Directions in Traditional and Online Assessment* (Routledge, 2003).

[3] Sri Sumarni, "Designing Ict Competences-Integrated Assessment Instruments Of Practical Key Teaching Competences For English Language Education Study Program," *IJLECR (International Journal of Language Education and Cultural Review)* 5, no. 1 (2019): 47–55.

consistency (reliability). A test is considered reliable if, when used to measure the same or different participants several times, the results remain relatively consistent[4].

In reality, many test instruments still have unknown quality in their items, leading to superficial assessments and, consequently, an inaccurate measurement of students' true abilities. This situation can be attributed to teachers' limited proficiency in evaluating and constructing test instruments as a means of assessment. Evaluation here serves the purpose of determining whether or not learning objectives have been achieved and assessing the quality of the test items that have been prepared.

Item analysis helps enhance the quality of test items through revision or the elimination of ineffective items. Furthermore, it can be used as diagnostic information for students to determine their comprehension of the taught material. Retnawati mentions that item analysis in education can be performed through two approaches: the classical and modern approaches. Classical item analysis involves examining test items based on student responses to improve the quality of those items, using classical test theory[5]. On the other hand, modern item analysis involves examining test items using Winstep. Winsteps is a Windows-based software that aids in the computation of Rasch models, particularly in the fields of educational evaluation, attitude surveys, and scale analysis.[6]

In this study, test results in the form of multiple-choice answers were analyzed using Rasch model software. The Rasch model is a modern assessment theory that classifies item and person measurements on a distribution map[7]. It is part of item response theory[8]. The Rasch model is based on two principles. The first principle is that the ability of subjects, in this case, students, to respond to a question can be predicted using a set of factors referred to as "traits." Traits represent dimensions of individual abilities, which can include verbal, cognitive, and psychomotor abilities. The second principle describes the relationship between a subject's, in this case, a student's, ability to answer a particular question and their

---

[4] Heri Retnawati, *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, Dan Psikometrian)* (Parama publishing, 2016).

[5] Retnawati, *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, Dan Psikometrian).*

[6] John M Linacre, "Understanding Rasch Measurement: Estimation Methods for Rasch Measures," *Journal of outcome measurement* 3 (1999): 382–405.

[7] AR Rozeha, Z Azami, and M Mohd Saidfudin, "Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering," 2007.

[8] David Thissen, "Psychometric Engineering as Art," *Psychometrika* 66 (2001): 473–485.

ability to answer other questions, which can be depicted in an item characteristic curve[9].

In the Rasch model, test participants with high ability should have a higher probability of answering a question correctly compared to other students. Conversely, students with lower ability have a lower chance of answering correctly on questions with higher levels of difficulty[10]. In the Rasch model approach, aside from item analysis, attention is also paid to the aspects of response and their correlations[11].

Item analysis in mathematics education is highly valuable for teachers. This is because mathematics is considered a challenging subject for many primary school students. [12] The analysis conducted can help teachers identify the difficulties students encounter in learning mathematics. Suwarsito & Sutomo mention that many primary school teachers have not performed item analysis. One reason for this is the excessive workload involved in manual analysis, and some teachers are concerned about potential leaks of the test items they create. These reasons contribute to the low quality of the questions provided to students[13].

An item can be considered good when it meets the criteria for item fit, taking into account the analysis results of Outfit MNSQ, Outfit ZSTD, and Pt-measure Corr. An item is also considered good when it exhibits appropriate difficulty levels in the item characteristic test. The aim of this research is to assess the quality of the Summative Mathematics Test Items for Grade V in SD Salafiyah Fityatul Huda, Pekalongan, for the academic year 2022/2023.

This research employs an evaluative methodology with a quantitative descriptive approach. This indicates that the study is conducted quantitatively not to either accept or reject hypotheses but rather to analyze the test items, item fit, and item difficulty.

---

[9] Hambleton, "Swaminathan. H., & Rogers, HJ (1991). Fundamentals of Item Response Theory."

[10] Bambang Sumintono and Wahyu Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan* (Trim komunikata, 2015).

[11] Difa Ardiyanti, "Aplikasi Model Rasch Pada Pengembangan Skala Efikasi Diri Dalam Pengambilan Keputusan Karir Siswa," *Jurnal Psikologi* 43, no. 3 (2016): 248–263.

[12] Nita Syahputri, "Rancang Bangun Media Pembelajaran Matematika Sekolah Dasar Kelas 1 Menggunakan Metode Demonstrasi," *Jurnal sistem informasi kaputama (JSIK)* 2, no. 1 (2018).

[13] Hindayati Mustafidah and Harjono Harjono, "Implementasi Program QUEST Untuk Menganalisis Butir Soal Bagi Guru-Guru SMP Muhammadiyah 2 Karanglewas," *JPPM (Jurnal Pengabdian dan Pemberdayaan Masyarakat)* 3, no. 2 (2019): 321–328.

RESULT AND DISCUSSION

Rasch Model

The Rasch Model Theory is a statistical approach used in the analysis of tests and assessments in various fields, especially in the context of measuring abilities, assessing item difficulty, and educational evaluation. This theory is named after Georg Rasch, a Danish statistician who first developed it in the 1960s.[14]

The core of the Rasch Model is to measure individual abilities in a specific domain (such as mathematics, language, or other aspects) and to measure the level of difficulty or ease of test items within that domain. The Rasch Model theory aims to examine the extent to which test items in an instrument or test are consistent with the principle that more capable respondents have a higher success rate in answering more difficult test items.

Key points in the Rasch Model theory:[15]

a. Respondent's Ability: The Rasch Model theory focuses on measuring the respondent's ability. This means not only looking at how many items are answered correctly but also measuring the extent of an individual's ability in a specific domain.
b. Item Difficulty Level: The Rasch Model measures the level of difficulty or ease of each test item. More challenging items have a greater chance of being answered correctly by more capable participants.
c. Probability Model: The Rasch Model uses a probability model that estimates how likely someone is to answer a test item correctly based on the comparison between their ability and the item's difficulty.
d. Invariance Property: This theory assumes that the measurement aspect being analyzed is invariant, meaning the measurement remains consistent regardless of variations in time, respondents, or measurement devices.
e. Logit Scale: The results of the Rasch Model analysis are measured on a logit scale, providing an understanding of the difference between an individual's ability and the difficulty level of test items.

The Rasch Model Theory is used in various educational evaluation contexts, the development of quality test instruments, the analysis of individual performance in various aspects, as well as research in the fields of psychometrics and evaluation. It helps identify the quality of test items and the ability levels of participants in an assessment domain.

Georg Rasch developed an analysis model from the Item Response Theory (IRT) in the 1960s, commonly known as the one-parameter logistic model (1PL). Benjamin Wright later popularized this mathematical model. Using raw data that

[14] Sumintono and Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*.

[15] Rozeha, Azami, and Mohd Saidfudin, "Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering."

was dichotomous (in the form of right and wrong) and indicated students' abilities, Rasch formulated it into a model that connects students and items[16].

For illustration, a student who can answer 80% of the questions correctly certainly has a better ability than another student who can only answer 65% of the questions correctly. This percentage data indicates that the raw data obtained is ordinal data, which represents rankings and is non-linear (Linacre, 1999). Since ordinal data doesn't have the same intervals, it needs to be transformed into ratio data for statistical analysis purposes. So, if someone scores 80%, their odds ratio would be 80:20 (meaning 80 correct scores compared to 20 wrong scores), which is essentially a more appropriate frequency/ratio comparison data for measurement purposes. Through this ratio data, Rasch developed a measurement model that determines the relationship between a student's ability (person ability) and an item's difficulty level (item difficulty) using logarithmic functions to produce measurements with the same interval. The result is a new unit called a logit (log odds unit), which represents a student's ability and the difficulty of items. Consequently, based on the logit values obtained, it can be concluded that a student's success in answering questions is highly dependent on their ability level and the difficulty of the questions[17].

Tem Fit Test (Model Fit)

The theory of fit statistics using the Rasch model is a fundamental aspect of Item Response Theory (IRT) analysis. It revolves around evaluating how well individual items (questions or tasks) fit the underlying model's expectations. In this context, the Rasch model provides a mathematical framework to link respondents' abilities and item difficulties, assuming that there is a one-to-one relationship between these abilities and difficulties.[18]

To assess the fit of individual items to the Rasch model, various fit statistics are employed. One common fit statistic is the Mean Square Residual (MNSQ), often divided into two types: Infit and Outfit.

Infit: This statistic examines the response pattern of a specific item in relation to the overall pattern of all items in the test. Infit statistics measure how well an item correlates with the latent trait being measured, and they should ideally be close to 1. Values less than 1 indicate overfitting (too high sensitivity to minor differences), while values greater than 1 indicate underfitting (insensitivity to variations).

---

[16] Sumintono and Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*.

[17] George Engelhard Jr, *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences* (Routledge, 2013).

[18] Engelhard Jr, *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*.

a. Outfit: Outfit statistics also measure the relationship between an item and the latent trait, but they are more sensitive to outliers. These statistics help identify items that may produce unexpected responses. Again, like Infit, Outfit values close to 1 are preferred. Values less than 1 imply an item is overly predictable, and values greater than 1 indicate an item is unpredictable.

The process involves examining these fit statistics and using their values to determine whether an item fits the Rasch model. Items that fit well contribute positively to the reliability and validity of the assessment, while poorly fitting items may need revision or removal.

The Rasch model and fit statistics are valuable tools for developing and refining assessment instruments, ensuring that the items accurately measure the latent trait being assessed. The goal is to create valid and reliable tests that provide meaningful insights into individuals' abilities or characteristics.

Item fit testing in this research is based on the Rasch model assumption. Item fit testing using the Rasch model assumption is done by assessing whether the item fits the model or not. This test is analyzed using the Winstep program. Criteria for an item to be considered fit in the Winstep program include having an Outfit MNSQ value between 0.5 and 1.5, an Outfit ZSTD value between -2 and 2, and a positive Pt-measure Corr value. If an item meets one of these three criteria, it is considered to be a fit item. The following are the results of the item fit analysis using the Winstep program[19].

Table 1. Item Fit of the Context Assessment Instrument

Outfit MNSQ	Outfit ZSTD	PT-Measure Corr	Item
1.01	.1	.36	Item_01
1.12	.5	.10	Item_02
.67	-.8	.60	Item_06
.85	-.3	.35	Item_07
1.12	-1.8	.27	Item_09
.65	.4	.61	Item_11
1.74	-.9	-.24	Item_04
.72	1.4	.44	Item_08
1.42	-.5	-.10	Item_05
.45	.8	.64	Item_13
2.32	-.6	-.25	Item_03
.36	1.9	.61	Item_10
.45	.3	.53	Item_12
1.47	-1.7	-.08	Item_14
MINIMUM MEASURE		.00	Item_15

[19] Sumintono and Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*.

The table above is the result of item fit or model fit analysis for the Context Assessment instrument. All the items of the Context Assessment instrument have been found to be a good fit with the model. This is evidenced by the fact that all the items meet the criteria for fit in the Winstep program, which include an Outfit MNSQ value between 0.5 and 1.5, an Outfit ZSTD value between -2 and 2, and a positive Pt-measure Corr value. Therefore, the Context Assessment instrument is considered to meet the criteria for model fit based on this analysis[20].

Item Characteristics

The characteristics of test items can help in measuring and understanding test items, including how well they reflect the level of difficulty and relevance appropriate to the learning objectives.[21]

Here are some characteristics of mathematics test items:

a. Difficulty of Test Items: Each test item has a different level of difficulty. This is related to how much students need to understand the material and the skills required to answer them. Test items should vary in difficulty, ranging from easy to difficult, to effectively measure the range of students' abilities.
b. Consistency: Test items should be consistent in measuring what they are supposed to measure. They should be relevant to the material taught and the learning objectives. Consistency is important to ensure that assessment results reflect actual abilities.
c. Precision: Test items should be designed accurately to avoid misleading or confusing students. They should be clear and easy to understand.
d. Objectivity: Test items should be objective, meaning that correct and incorrect answers can be clearly identified. This minimizes subjective interpretation.
e. Varied Formats: 5th-grade mathematics tests should include various formats of test items, including multiple-choice, short-answer, and essay questions. This helps measure various aspects of students' mathematical understanding.
f. Alignment with Curriculum Standards: Test items should be aligned with the applicable curriculum standards. They should measure the abilities and knowledge that should be taught in the curriculum.
g. Appropriate Cognitive Levels: Test items should stimulate higher-level thinking in students and measure their understanding of mathematical concepts, not just mechanical comprehension.

---

[20] Sumintono and Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*.

[21] Mustafidah and Harjono, "Implementasi Program QUEST Untuk Menganalisis Butir Soal Bagi Guru-Guru SMP Muhammadiyah 2 Karanglewas."

h. Reproducible Results: Test items should yield consistent results. If a student answers correctly, they should be able to answer correctly every time they are tested with similar items.[22]

After conducting the item fit analysis or model fit, the next step is to assess the characteristics of the items in the developed instrument. The characteristics of the items at this stage involve examining the level of item difficulty, as the instrument developed for this research uses Rasch Modeling. This is because Rasch Modeling only measures the item's difficulty level. According to the Rasch model, an item is considered good if the difficulty index is greater than -2.0 or less than 2.0[23]. The analysis of item characteristics at this stage is performed using the Winstep program, where the difficulty level is observed in the measure. The following are the results of the analysis of item characteristics for each instrument.

The items analyzed in the Context Assessment instrument are a total of 15 items. The output of the item characteristics analysis can be seen in the following table.

Table 2. Output of the item characteristics analysis for the Context Assessment instrument

Entry Number	Total Score	Total Count	Measure	Model S.E
1	13	22	1.35	.46
2	16	22	.66	.51
6	17	22	.39	.54
7	17	22	.39	.54
9	17	22	.39	.54
11	17	22	.39	.54
4	18	22	.08	.58
8	18	22	.08	.58
5	19	22	-.29	.65
13	19	22	-.29	.65
3	20	22	-.78	.76
10	20	22	-.78	.76
12	20	22	-.78	.76
14	20	22	-.78	.76
15	22	22	-2.79	1.82

[22] Hindayati Mustafidah and Harjono Harjono, "Implementasi Program QUEST Untuk Menganalisis Butir Soal Bagi Guru-Guru SMP Muhammadiyah 2 Karanglewas," *JPPM (Jurnal Pengabdian dan Pemberdayaan Masyarakat)* 3, no. 2 (2019): 321–328.

[23] Hambleton, "Swaminathan. H., & Rogers, HJ (1991). Fundamentals of Item Response Theory."

The items' difficulty levels in the table are arranged in ascending order, starting from the highest difficulty level to the lowest. Item characteristics analysis in this stage examines the difficulty level of the items. The difficulty level of the items is assessed based on the results in the "measure" column. According to the table above, it shows that the item with the highest difficulty level is item 1 with a difficulty index of 1.35, whereas the item with the lowest difficulty level is item 15 with a difficulty index of -2.79. However, the results indicate that only 14 items in the Context Assessment instrument meet the criteria for good difficulty levels, while one item in the Context Assessment instrument does not meet the criteria for good difficulty level as the difficulty index exceeds 2.0.

CONCLUSION

From the above discussion, it can be concluded that the primary school assessment questions in the subject of Mathematics have items of good quality since they meet at least one of the three predetermined criteria. On the other hand, in terms of the difficulty level of the questions, there is one question that is relatively easy. This is because it was found that all students answered it correctly.

REFERENCES

Ardiyanti, Difa. “Aplikasi Model Rasch Pada Pengembangan Skala Efikasi Diri Dalam Pengambilan Keputusan Karir Siswa.” *Jurnal Psikologi* 43, no. 3 (2016): 248–263.

Dunn, Lee, Chris Morgan, Meg O’Reilly, and Sharon Parry. *The Student Assessment Handbook: New Directions in Traditional and Online Assessment*. Routledge, 2003.

Engelhard Jr, George. *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*. Routledge, 2013.

Hambleton, RK. “Swaminathan. H., & Rogers, HJ (1991). Fundamentals of Item Response Theory” (n.d.).

Linacre, John M. “Understanding Rasch Measurement: Estimation Methods for Rasch Measures.” *Journal of outcome measurement* 3 (1999): 382–405.

Mustafidah, Hindayati, and Harjono Harjono. “Implementasi Program QUEST Untuk Menganalisis Butir Soal Bagi Guru-Guru SMP Muhammadiyah 2 Karanglewas.” *JPPM (Jurnal Pengabdian dan Pemberdayaan Masyarakat)* 3, no. 2 (2019): 321–328.

———. “Implementasi Program QUEST Untuk Menganalisis Butir Soal Bagi Guru-Guru SMP Muhammadiyah 2 Karanglewas.” *JPPM (Jurnal Pengabdian dan Pemberdayaan Masyarakat)* 3, no. 2 (2019): 321–328.

Retnawati, Heri. *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, Dan Psikometrian)*. Parama publishing, 2016.

Rozeha, AR, Z Azami, and M Mohd Saidfudin. "Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering," 2007.

Sumarni, Sri. "Designing Ict Competences-Integrated Assessment Instruments Of Practical Key Teaching Competences For English Language Education Study Program." *IJLECR (International Journal of Language Education and Cultural Review)* 5, no. 1 (2019): 47–55.

Sumintono, Bambang, and Wahyu Widhiarso. *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Trim komunikata, 2015.

Syahputri, Nita. "Rancang Bangun Media Pembelajaran Matematika Sekolah Dasar Kelas 1 Menggunakan Metode Demonstrasi." *Jurnal sistem informasi kaputama (JSIK)* 2, no. 1 (2018).

Thissen, David. "Psychometric Engineering as Art." *Psychometrika* 66 (2001): 473–485.