Al-Jabar: Jurnal Pendidikan Matematika
Volume 12, Number 2, 2021, Pages 399 – 412
http://ejournal.radenintan.ac.id/index.php/al-jabar/index

Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables(SLETV)

Adilla Desy Rizbudiani[1]*, Amat Jaedun[1], Abdul Rahim[1], Arief Nurrahman[1]

[1] Universitas Negeri Yogyakarta, Indonesia.
✉ adilladesy.2019@student.uny.ac.id

Article Information
Submitted Sep 24, 2021
Revised Nov 07, 2021
Accepted Nov 09, 2021

Keywords
System of Linear Equations in Two Variables; Item Response Theory; Rasch Model.

Abstract

A high-quality test has a balanced level of difficulty and can be completed by the respondent with their level of abilities. This study analyzed the test instrument used to measure students' mathematics abilities in the semester final exam on System of Linear Equations in Two-Variables. The purposive sampling technique was applied to select the respondent students (N=195). The test items were twenty multiple-choice questions. The researchers performed the data analysis using Rasch model Item Response Theory (IRT) approach with the QUEST program. The analysis revealed that the twenty items' validity matched the Rasch model with a range of INFIT MNSQ values between 0.89 – 1.17. Items on the final semester exam can be used based on the estimated OUTFIT t-value less than equal to 2.00. The OUTFIT t analysis obtained nineteen qualified items and one unqualified item.

INTRODUCTION

Various significant social changes occur along with the development of the fourth industrial revolution, especially in the educational transformation (Cotet et al., 2020; Johari et al., 2021; Slavinec et al., 2019). It is only natural that education can change individuals' way of thinking for the better (Purwandari et al., 2018). A better way of thinking will improve the quality of the individual's self-quality (Schunk & DiBenedetto, 2020). One way to develop self-quality is to establish a way of thinking through mathematics lessons (Noortsani, 2019). Concepts and principles can be constructed through mathematics learning through the internalization process to reshape ideas (Agnesti & Amelia, 2020; Suryani et al., 2020).

Mathematics determines the ability to understand concepts that have been obtained and mastered by students after following the learning process. There is an understanding of concepts, thinking skills, reasoning abilities, and mathematical concepts in posing problems (Rachmawati et al., 2019). Realizing good mathematics learning needs to involve several problems that must be solved (Syafitri et al., 2018). Mathematics is more appropriate by introducing context-based problems (contextual problems) (Achir et al., 2017). In reality, students in Indonesia do not understand and cannot apply mathematical models to existing contextual problems. They cannot interpret mathematical solutions into contextual problems and explain concepts and strategies to solve them (Nasution & Ahmad, 2018). In other words, Indonesian students possess a relatively poor ability to solve mathematical problems (Asih & Ramadhani, 2019; Suryani et al., 2020).

How to cite: Rizbudiani, A. D., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables (SLETV). *Al-Jabar: Jurnal Pendidikan Matematika*, *12*(2), 398-412.
e-ISSN: 2540-7562.
Published by: Mathematics Education Department, UIN Raden Intan Lampung.

The Program for International Student Assessment (PISA) survey published in March 2019 reveals that Indonesian students' reading, science, and math skills are ranked 74 out of 79 countries (Yulistianti & Megawati, 2019). The average PISA score obtained by Indonesian students in mathematics was 379. The average score was far from the minimum score of 489 (Panyahuti et al., 2020). Indonesia's PISA results were behind Mexico's, which reached level 2 in mathematics (Rojas-kramer et al., 2020). One of the mathematics materials is the System of Linear Equations in Two Variables (SLETV) (Murtiyasa & Al Karomah, 2020; Suryani et al., 2020). SLETV material is presented through stories so that students can apply the material learned in everyday life (Yusuf & Fitriani, 2020). Besides, the results of other studies show that mathematical communication skills through the System of Linear Equations in Two-Variables (SLETV) are still relatively low in junior high school (Bey & Asriani, 2013). The other research results support this statement that junior high school students in Cianjur were having difficulty solving the System of Linear Equations in Two-Variables (SLETV) questions (Yusuf & Fitriani, 2020).

Thus, it is necessary to have a good quality test to determine mathematical understanding related to the System of Linear Equations in Two Variables. Assessment of learning outcomes is one way to determine students' achievement (Panyahuti et al., 2020). Each assessment is designed to measure and classify the test taker's performance in a specific domain. Depending on the scoring design, scores can be on an ordinal, interval, or ratio scale (Himelfarb, 2019). The assessment determines whether the learning objectives have been achieved and how students acquire the expected goals (Alfarisa & Purnama, 2019). Students' low learning outcomes do not always indicate poor competence and vice versa. The low learning outcomes also does not measure the meaninglessness of the learning. Thus, to get good learning evaluation results, it is necessary to investigate test quality as an indicator (Sainuddin, 2018).

Learning evaluation is a series of processes in improving the quality, performance, and productivity in carrying out the program or its objectives (Isnani et al., 2019; Mardapi, 2012). Item analysis can be analyzed using classical test theory and modern test theory. The weakness of classical test theory is the difficulty and the discriminating index of the items depending on the group of students working on them. The test analysis methods and techniques compare students' abilities by dividing several parts into specific groups. The concept of score reliability is defined from the term test. There is no fundamental theory to determine how participants do the tests relevant to their abilities, and standard error measurement (SEM) is applied to all test-takers (Alfarisa & Purnama, 2019; Hambleton et al., 1991).

There is a weakness in classical item analysis, so modern problem analysis is utilized. Modern problem analysis can be done with Item Response Theory. Item Response Theory contains a broad set of statistical models that calculate the probability of a learner with some or all of the abilities to answer individual items on a test correctly (Stewart et al., 2018). The Item Response Theory approach is a measurement approach developed to overcome the weaknesses of classical test theory (Ayub et al., 2020; Ramdani et al., 2020). One item response theory model is the Rasch model, which is developed from a simple logistic model (Rahayu et al., 2021).

Rasch model was developed to analyze categorical data (Himelfarb, 2019). It offers a new paradigm in longitudinal educational research. The Rasch model's analysis improves the accuracy and quality of tests and surveys, allowing various measuring instruments (Al Ali & Shehab, 2020). The Rasch model is a modern valuation theory that classifies the calculation of items and people in a distribution map (Azizah & Wahyuningsih, 2020). It is a simple model (Mardapi, 2012). Besides, (Sumintono & Widhiarso, 2015), the Rasch model meets the measurement principles, such as having a linear measurement with equal intervals, overcoming the problem of missing data, providing an accurate estimate, being able to detect an inappropriate model, and providing independent measurement instruments parameters (Pratama, 2020).

Research on the Rasch model has been widely discussed (Chan et al., 2021; Che Lah et al., 2021; Nielsen et al., 2021; Scoulas et al., 2021; Tseng & Wang, 2021). Previous findings show that the Rasch model can be used by teachers and can be applied in final semester exams in various subjects: Chemistry (Purnamasari & Kartowagiran, 2019), Economics (Alfarisa & Purnama, 2019), Morality (*Aqidah-Akhlak*)(Pratama, 2020), and Physics (Asriadi & Hadi, 2021; Palimbong et al., 2018). Several similar studies were found, such as investigating the quality of test instruments in measuring students' abilities in exams. However, there has been no research on the quality of the items investigated using the Rasch model approach with QUEST in mathematics subjects on System of Linear Equations in Two-Variables. This research aimed to determine the quality of the test instruments used to measure students' abilities in the second-semester final exam in one of the junior high schools in Yogyakarta using the Rasch model. The quality is measured based on several indicators, namely items that match the Rasch model, level of item difficulty, and item reliability.

METHODS

This quantitative descriptive research focuses on analyzing the final semester examination test instrument using the Rasch model approach. The sampling technique used was purposive sampling to select the research subjects. The subjects were eighth-grade Junior High School (SMP) students in Yogyakarta (N=195 students). There were 20 multiple choice questions on the System of Linear Equations in Two-Variable material. The items were composed of easy, medium, and hard difficulties. The correct answer was given a score of 1, and the incorrect answer was given a score of 0. Therefore, the data obtained was dichotomous.

The test results were analyzed using the Rasch Item Response Theory (IRT) Model with QUEST. From the output of the QUEST, several parameter items were obtained that fit the Rasch model. Also, Cronbach's Alpha value was obtained from the overall items reliability test. The overall fit of items in the QUEST program was based on the average value of the INFIT Mean of Square (INFIT MNSQ) and the standard deviation or the average value of the INFIT Mean INFIT t. The fit of each item in the QUEST was based on the size of the INFIT MNSQ value or the INFIT t value of the items. The importance of the Unweighted Mean Square (OUTFIT MNSQ) in the QUEST program and the Weighted Mean Square was expected to be 1, and the variance was 0.

The expected value of the Mean INFIT t was equal to 0 with a variance equal to 1. The INFIT MNSQ compatibility with the Rasch model ranged from 0.77 to 1.33. On the other hand, the OUTFIT t compatibility with the Rasch model ranged from OUTFIT t ≤ 2.00. The reliability of the test was calculated using the following Cronbach Alpha formula:

$$r = \left[\frac{k}{(k-1)}\right]\left[1 - \frac{\sum \sigma_b{}^2}{\sigma_t}\right]$$

Where:
r: Test reliability coefficient (Cronbach Alpha)
k: Number of questions
$\sum \sigma_b{}^2$: Total item variance
σ_t: Total variance

The level of difficulty of the Rasch model, which describes the relationship between the test items and the person (Bozdağ & Türkoğuz, 2021), can be stated as follows (Hambleton et al., 1991):

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

Where:
$P_i(\theta)$: The probability that a randomly chosen examinee with ability θ answers item *i* correctly
b_i: The item *I* difficulty parameter
e: A transcendental number whose value is 2.718
n: The number of items in the test
θ: The ability scale
D: Scaling factor whose value is 1.7 as a logistic model

RESULTS AND DISCUSSION

Final Semester Exam Test Construction Map

The test focused in this research was the mathematics final semester test on System of Linear Equations in Two Variables (SLETV). The eighth-graders studied this material in the second semester. The test contained twenty items, as displayed in Table 1.

Table 1. Problems' Specification

Basic competencies	Indicators	Cognitive Level	Items
3.5 Explain the system of linear equations of two variables and the solutions associated with contextual problems.	Students can determine the solution to the system of linear equations of two given variables.	C2	1,2
4.5 Solve related problems.	Students can determine the value of the system of linear equations of two variables if two are known equation forms.	C2	3,4

	Students are presented contextual problems about the system of linear equations of two variables and solve the problems.	C3	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
	Students are presented contextual problems about the system of linear equations of two variables and solve the problems.	C4	16,17,18, 19, 20

The test's blueprint followed the government's syllabus in the 2013 curriculum. Each indicator was adjusted to the cognitive level according to Bloom's taxonomy. In 21st-century learning, mastery of competencies is more critical in binding critical thinking skills, creativity, and finding solutions to problems. The test was designed to increase thinking ability and competence, especially in the System of Linear Equations in Two Variables (SLETV) System material.

Estimation of Item Validity

The test's validity using the QUEST program (Bashooir & Supahar, 2018) can be compared through the criteria in Table 2.

Table 2. The Criteria for INFIT MNSQ

INFIT MNSQ Value	Description
>1.33	Infit the model
0.77-1.33	Fit the model
<0.77	Infit the model

The results of the analysis of the INFIT MNSQ value in the QUEST program can be seen in Table 3:

Table 3. The Recapitulation of Item Validity

Item Estimates (Thresholds) In input Order
all on all (N = 195 L = 20 Probability Level= .50)

	ITEM NAME	SCORE	MAXSCR	THRSH 1	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1	item 1	151	195	-2.07 .18	.89	.82	-1.2	-1.4
2	item 2	17	195	1.53 .26	1.01	1.18	.1	.7
3	item 3	174	195	-2.95 .24	.95	.82	-.2	-.7
4	item 4	24	195	1.15 .22	1.02	.99	.2	.0
5	item 5	23	195	1.20 .23	.96	.97	-.2	-.1
6	item 6	16	195	1.59 .26	1.02	1.32	.2	1.1
7	item 7	24	195	1.15 .22	1.06	1.26	.4	1.2
8	item 8	62	195	-.04 .16	.99	1.01	-.1	.2
9	item 9	148	195	-1.99 .17	.92	.86	-.9	-1.1
10	item 10	56	195	.11 .16	1.17	1.39	2.4	3.0
11	item 11	116	195	-1.21 .15	.97	.98	-.7	-.2
12	item 12	74	195	-.31 .15	.94	.94	-1.3	-.6
13	item 13	153	195	-2.13 .18	.92	.87	-.8	-.9
14	item 14	29	195	.94 .20	1.03	1.11	.3	.6
15	item 15	26	195	1.06 .21	1.02	1.17	.2	.9
16	item 16	19	195	1.41 .25	1.05	1.48	.3	1.8
17	item 17	92	195	-.70 .15	1.01	1.04	.3	.5
18	item 18	24	195	1.15 .22	.94	.81	-.3	-.9
19	item 19	115	195	-1.18 .15	.91	.93	-2.0	-.7
20	item 20	21	195	1.30 .23	1.04	1.42	.3	1.7
Mean				.00	.99	1.07	-.1	.3
SD				1.47	.07	.21	.9	1.1

Table 3 provides information on item validity where all items fit or match the Rasch model with a range of INFIT MNSQ values between 0.89 – 1.17. The item fit map can be used to find out whether the item is suitable or not, as displayed in figure 1.

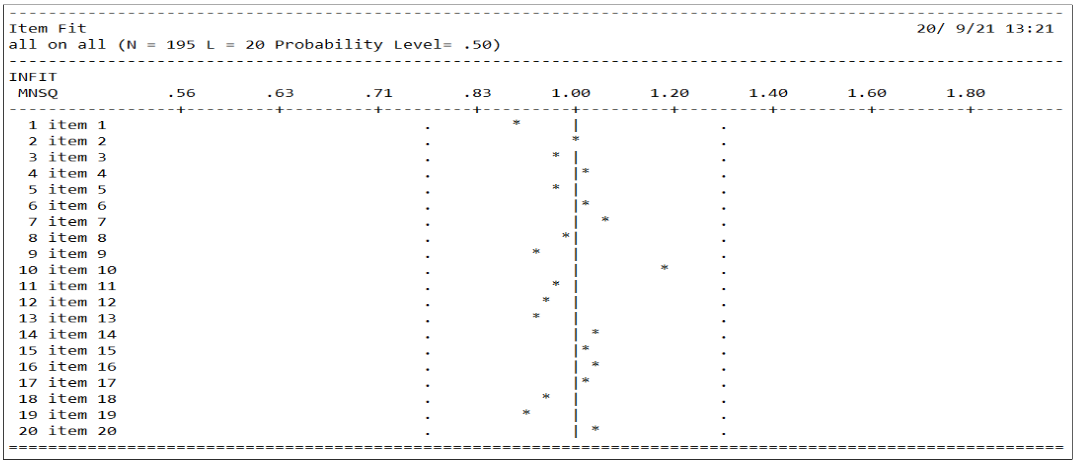

Figure 1. Fit Map of Rasch Model

When viewed from the fit map of the model above, all items are in the INFIT MNSQ value range of 0.77 – 1.30. The dots on the left show a value of 0.77, while the dots on the right show a value of 1.30.

Estimation of Difficulty Level

The results of threshold analysis can show the difficulty level of an item through the QUEST program. The criteria of the items' difficulty level are ranged from -2.0 to 2.0. If the range or distribution of items or test-takers is less than -2.0, the items are included in the easy category. Furthermore, if the range of items or test-takers is more than 2.0, the item is in the difficult category. For a more detailed view of the distribution of item difficulty levels, see the figure below.

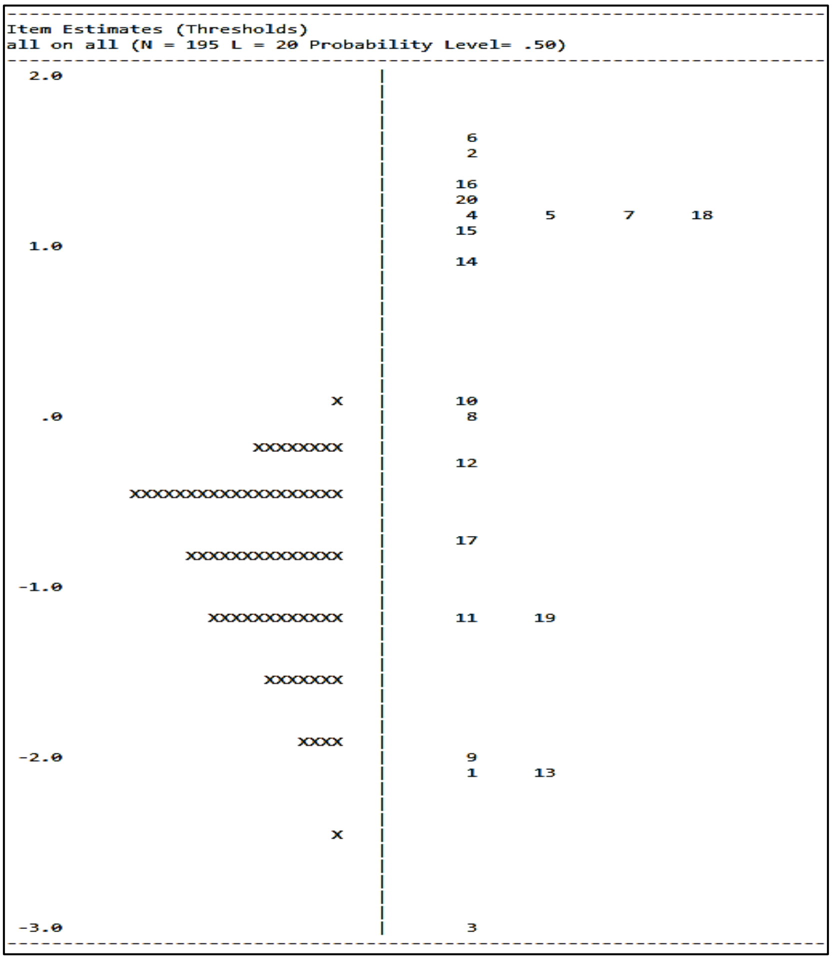

Figure 2. Distribution of Items' Difficulty Levels

Based on Figure 2, item number 6 is the most difficult item. When compared to the test-taker's ability, the possibility of the test-taker correctly answering item number 6 is low or impossible. Also, item number 3 is the easiest item according to the ability of the test-takers. The difficulty level of items through the QUEST program can also be seen from the threshold criteria displayed in Table 4.

Table 4. Difficulty Level Criteria

Threshold Value	Description
$b > 2$	Very difficult
$1 < b \leq 2$	Difficult
$-1 \leq b \leq 1$	Medium
$-2 \leq b < -1$	Easy
$b < -2$	Very easy

The following is a recapitulation of the difficulty level of each item in Table 5 and Table 6.

Table 5. The Recapitulation of the Difficulty Level of the Rasch Model Questions

Item	Threshold Value	Interpretation	Item	Threshold Value	Interpretation
1	-2.07	Very Easy	11	-1.21	Easy
2	1.53	Difficult	12	-.31	Medium
3	-2.95	Very Easy	13	-2.13	Very Easy
4	1.15	Difficult	14	0.94	Medium
5	1.20	Difficult	15	1.06	Difficult
6	1.59	Difficult	16	1.41	Difficult
7	1.15	Difficult	17	-0.70	Easy
8	-0.04	Medium	18	1.15	Difficult
9	-1.99	Easy	19	-1.18	Easy
10	0.11	Medium	20	1.30	Difficult

Table 6. Category Level of Difficulty Item

Category	Frequency	Percentage (%)
Very Easy	3	15
Easy	4	20
Medium	4	20
Difficult	9	45
Very Difficult	0	0
Total	20	100

The level of difficulty based on Table 6 illustrates that the very difficult category items are 0 items or 0%. The difficult category consists of nine items or 45%, the medium category consists of four items or 20%, the easy category consists of four items or 20%, and the very easy item category consists of three items or 15%. Therefore, the test-takers abilities are below the item difficulty levels shown by the small number of test participants who can answer very difficult or difficult items correctly. The test-takers abilities can be seen through the QUEST program's *Summary of Case Estimate* on the reliability of the estimate. The criteria are as follows: the high ability category has an estimated value of more than 1.00. The moderate ability category has an estimated value of -1.00 to 1.00. The low ability category has an estimated value of less than -1.00.

```
--------------------------------------------------------------------
Case Estimates
all on all (N = 195 L = 20 Probability Level= .50)
--------------------------------------------------------------------

Summary of case Estimates
=========================

Mean                                  -.85
SD                                     .56
SD (adjusted)                          .00
Reliability of estimate                .00

 Fit Statistics
===============

 Infit Mean Square           Outfit Mean Square

    Mean     1.00              Mean     1.07
    SD        .51              SD        .83

      Infit t                    Outfit t

    Mean     -.15              Mean     -.09
    SD       1.48              SD       1.27

   0 cases with zero scores
   0 cases with perfect scores
====================================================================
```

Figure 4. Estimation of Respondents' Abilities

Figure 4 reveals that the test-takers have moderate abilities, with a reliability estimate value of 0.00 or with a range of -1.00 to 1.00.

Estimation of Item Fit

The OUTFIT t value in the QUEST program can be used to determine the items' passing criteria. If the value of OUTFIT t is equal to or lower than 2.00, then the item passes. In contrast, if the value of OUTFIT t is equal to or higher than 2.00, the item fails.

Table 7. The Recapitulation of Item Fit

Item	Outfit t Value	Description	Item	Outfit t Value	Description
1	-1.4	Fit	11	-0.2	Fit
2	0.7	Fit	12	-0.6	Fit
3	-0.7	Fit	13	-0.9	Fit
4	0.0	Fit	14	0.6	Fit
5	-0.1	Fit	15	0.9	Fit
6	1.1	Fit	16	1.8	Fit
7	1.2	Fit	17	0.5	Fit
8	0.2	Fit	18	-0.9	Fit
9	-1.1	Fit	19	-0.7	Fit
10	3.0	Infit	20	1.7	Fit

Based on Table 7, there were 19 items passed, and there was 1 item that did not pass and could not be used. However, the easiest and the most difficult items should not be included in the test because very few test-takers can answer correctly. It should be noted that the item questions were very difficult. In this analysis, there were 45% difficult items. It will be better if the proportion of the difficult items is reduced to compensate for the test-takers. Moreover, the analysis showed that the test-takers belonged to the moderate ability category.

Estimation of Reliability

The reliability value of the Rasch model using the QUEST program is seen in the reliability of item estimate and case estimate. The reliability of the item estimate value was 0.98. This reliability is referred to as sample reliability in the Rasch model. The criteria for the reliability value of the Rasch model are as follows: less than 0.67 is weak, 0.67 to 0.80 is moderate, 0.81 to 0.90 is good, 0.91 to 0.94 is very good, and more than 0.94 is perfect.

The reliability of the item estimate value of 0.98 is related to the number of items that fit the model. The value of 0.98 is in the very good reliability category, so that it affects the items that fit the model. The higher the reliability, the more items fit with the model (Susdelina et al., 2018). These findings are supported by Asriadi and Hadi (Asriadi & Hadi, 2021), who claim that the Cronbach Alpha (KR20) measures the interaction between people and items as a whole. The person reliability value is 0.48, and the item reliability value is 0.88 in the calculation of the Rasch model for the physics final semester examination. Another finding shows that the instrument reliability value of 0.98 represents a comparative item which implies that the score for each question is consistent and stable (Zulkifli et al., 2018).

The reliability of case estimate value or the reliability of the test-takers of 0.00 is classified as weak. This value indicates that there is an inconsistency (Ardiyanti, 2016). The reliability of case estimate value or the reliability of the test-takers of 0.00 is classified as weak. This value indicates that the test-takers are inconsistent. It can also mean that the test-takers are careless in answering the questions, thus affecting the reliability value of the person or subject (Pratama, 2020).

The quality of the items in the System of Linear Equations in Two-Variables (SLETV) material is appropriate to be used in the final semester examination. However, it is not appropriate if the measurement results are used to make students' abilities (Primi et al., 2016).

CONCLUSIONS

Based on the final semester exam test analysis on the material of System of Linear Equations in Two Variables (SLETV), several characteristics of the test and test-takers can be described. The twenty items' validity fit or matched the Rasch model with a range of INFIT MNSQ values between 0.89 and 1.17. The items can be used based on the results of the estimated OUTFIT t value of 2.00. The OUTFIT t analysis obtained 19 items that passed and 1 item that did not pass. All items were analyzed with estimated difficulty levels. The item in the very difficult category was 0%. The items in the difficult category were nine items or 45%. The Items in the medium category were four items or 20%. The items in the easy category were four items or 20%. The items in the very easy category were three items or 15%. The reliability of the item estimate was 0.98 within the very good category, and the reliability of the case estimate was 0.00 within the weak category. The researchers suggest further research to calculate the quality of items using other models and approaches, such as 2PL or 3PL with different software applications.

AUTHOR CONTRIBUTIONS STATEMENT

Based on his contribution to this research, AJ prepared the research as a whole, developed the research instrument, and wrote the article. ADR contributes as an analytical instrument and writes down findings during the study. AN contributed as a documentary during the research and AR as an editor in writing articles.

REFERENCES

Achir, Y. S., Usodo, B., & Retiawan, R. (2017). Analisis kemampuan komunikasi matematis siswa dalam pemecahan masalah matematika pada materi sistem persamaan linear dua variabel (spldv) ditinjau dari gaya kognitif. *Paedagogia*, *20*(1), 78.

Agnesti, Y., & Amelia, R. (2020). Penerapan pendekatan kontekstual dalam menyelesaikan soal cerita pada materi perbandingan dan skala terhadap siswa SMP. *Mosharafa : Jurnal Pendidikan Matematika*, *9*(2), 347–358.

Al Ali, R. M. A., & Shehab, R. T. (2020). Psychometric properties of social perception of mathematics: Rasch Model Analysis. *International Education Studies*, *13*(12), 102.

Alfarisa, F., & Purnama, D. N. (2019). Analisis butir soal ulangan akhir semester mata pelajaran ekonomi SMA menggunakan RASCH model. *Jurnal Pendidikan Ekonomi Undiksha*, *11*(2), 366-374.

Ardiyanti, D. (2016). Aplikasi model rasch pada pengembangan skala efikasi diri dalam pengambilan keputusan karir siswa . *Jurnal Psikologi*, *43*(3), 248–263.

Asih, N., & Ramadhani, S. (2019). Peningkatan kemampuan pemecahan masalah matematis dan kemandirian belajar siswa menggunakan model pembelajaran means end analysis. *Jurnal Pendidikan Matematika*, *8*(3), 435-446

Asriadi, M., & Hadi, S. (2021). Implementation of item response theory at final exam test in physics learning: Rasch model study. *Proceedings of the 6th International Seminar on Science Education (ISSE 2020)*, *54*(1), 336–342.

Ayub, M. R. S. S. N., Istiyono, E., Munadi, S., Permadi, C., Pattiserlihun, A., & Sudjito, D. N. (2020). Analisa penilaian soal fisika menggunakan model rasch dengan program r. *Jurnal Sains dan Edukasi Sains*, *3*(2), 46–52.

Azizah, & Wahyuningsih, S. (2020). Penggunaan model rasch untuk analisis instrumen tes pada mata kuliah matematika aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, *3*(1), 45–50.

Bashooir, K., & Supahar. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 168–181.

Bey, A., & Asriani. (2013). Penerapan pembelajaran problem solving untuk meningkatkan aktivitas dan hasil belajar matematika pada materi SPLDV. *Jurnal Pendidikan Matematika*, *4*(2), 224–239.

Bozdağ, H. C., & Türkoğuz, S. (2021). A rasch model analysis of primary school students' conceptual understanding levels of the concept of light. *IOJPE: International Online Journal of Primary Education*, *10*(1), 160–179.

Chan, S. W., Looi, C. K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: A rasch model measurement analysis. *Journal of Computers in Education*, *8*(2), 213–236.

Che Lah, N. H., Tasir, Z., & Jumaat, N. F. (2021). Applying alternative method to evaluate online problem-solving skill inventory (OPSI) using Rasch model analysis. *Educational Studies*, *00*(00), 1–23.

Cotet, G. B., Carutasu, N. L., & Chiscop, F. (2020). Industry 4.0 diagnosis from an imillennial educational perspective. *Education Science*, *10*.

Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals of item response theory library of congress cataloging-in-publication data*. SAGE.

Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, *33*(2), 151–163.

Isnani, I., Utami, W. B., Susongko, P., & Lestiani, H. T. (2019). Estimation of college students' ability on real analysis course using Rasch model. *Research and Evaluation in Education*, *5*(2), 95–102.

Johari, A. B., Wahat, N. W. A., & Zaremohzzabieh, Z. (2021). Innovative work behavior among teachers in Malaysia: The effects of teamwork, principal support, and humor. *Asian Journal of University Education*, *17*(2), 72–84.

Mardapi, D. (2012). *Pengukuran, penilaian & evaluasi pendidikan*. Nuha Medika.

Murtiyasa, B., & Al Karomah, I. I. (2020). The impact of learning strategy of problem solving and discovery towards learning outcomes reviewed from students learning motivation. *Universal Journal of Educational Research*, *8*(9), 4105–4112.

Nasution, D. P., & Ahmad, M. (2018). Penerapan pembelajaran matematika realistik untuk meningkatkan kemampuan komunikasi matematis siswa. *Mosharafa: Jurnal Pendidikan Matematika*, *7*(3), 389–400.

Nielsen, T., Martínez-García, I., & Alastor, E. (2021). Critical thinking of psychology students: A within- and cross-cultural study using rasch models. *Scandinavian Journal of Psychology*, *62*(3), 426–435.

Noortsani, I. (2019). Meningkatkan kualitas proses pembelajaran melalui penerapan model pembelajaran. *Simpul Juara*, *1*(1).

Palimbong, J., Mujasan, & Allo, A. Y. T. (2018). Item analysis using rasch model in semester final exam evaluation study subject in physics class X TKJ SMK Negeri 2 Manokwari. *Kasuari: Physics Education Journal*, *1*(1), 48–51.

Panyahuti, Krismadinata, Jalinus, N., Rahmat, R., & Ambiyar, A. (2020). Analisis pemetaan kemampuan numerik siswa smk model teori respon butir. *INVOTEK: Jurnal Inovasi Vokasional Dan Teknologi*, *20*(3), 11–22.

Pratama, D. (2020). Analisis kualitas tes buatan guru melalui pendekatan item response theory (IRT) model rasch. *Tarbawy : Jurnal Pendidikan Islam*, *7*(1), 61–70.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, *29*(5), 453–469.

Purnamasari, U. D., & Kartowagiran, B. (2019). Application rasch model using R program in analyze the characteristics of chemical items. *Jurnal Inovasi Pendidikan IPA*, *5*(2), 147–157.

Purwandari, A. S., Astuti, M. D., & Yuliani, A. (2018). Evaluasi kemampuan komunikasi matematis siswa smp pada materi sistem persamaan linear dua variabel. *IndoMath: Indonesia Mathematics Education*, *1*(1), 55.

Rachmawati, Y. I., Sugandi, E., & Prayitno, L. L. (2019). Senior high school student's ability in posing system of linear equations in two variables problems. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, *4*(1), 57–65.

Rahayu, W., Putra, M. D. K., Rahmawati, Y., Hayat, B., & Koul, R. B. (2021). Validating an indonesian version of the what is happening in this class? (WIHIC) questionnaire using a multidimensional rasch model. *International Journal of Instruction*, *14*(2), 919–934.

Ramdani, R., Hanurawan, F., Ramli, M., Lasan, B. B., & Afdal, A. (2020). Development and validation of Indonesian academic resilience scale using rasch models. *International Journal of Instruction*, *14*(1), 105–120.

Rojas-kramer, C. A., Moreno-García, E., & Venegas-Martínez, F. (2020). Anxiety Determinants towards mathematics in mexican high school students. *European Journal of Contemporary Education*, *9*(4), 866–877.

Sainuddin, S. (2018). Analisis karakteristik butir tes matematika berdasarkan teori modern (teori respon butir). *Jurnal Penelitian Matematika Dan Pendidikan Matematika*, *1*(1), 1–12.

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, *60*, 101832.

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library and Information Science Research*, *43*(1), 101071.

Slavinec, M., Aberšek, B., Gačević, D., & Flogie, A. (2019). Monddisciplinarity in science versus transdiciplinarity in STEM education. *Journal of Baltic Science*, *18*(3), 435–449.

Stewart, J., Zabriskie, C., Devore, S., & Stewart, G. (2018). Multidimensional item response theory and the force concept inventory. *Physical Review Physics Education Research*, *14*(1), 10137.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim Komunikata.

Suryani, M., Jufri, L. H., & Putri, T. A. (2020). Analisis kemampuan pemecahan masalah siswa berdasarkan kemampuan awal. *Musharafa: Jurnal Pendidikan Matematika*, *9*(1), 119–130. 1

Susdelina, Perdana, S. A., & Febrian. (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan rasch model. *Jurnal Kiprah*, *6*(1), 41–48.

Syafitri, Q., Mujib, Anwar, C., Netriwati, & Wawan. (2018). The mathematics learning media uses geogebra on the basic material of linear equations. *Al Jabar: Jurnal Pendidikan Matematika*, *I*, 9–18.

Tseng, M. C., & Wang, W. C. (2021). The q-matrix anchored mixture rasch model. *Frontiers in Psychology*, *12*(March), 1–9.

Yulistianti, H. D., & Megawati, E. (2019). Analisis instrumen tes higher order thinking. *Jurnal Pendidikan Matematika*, *13*(1), 41–54.

Yusuf, A., & Fitriani, N. (2020). Analisis kesalahan siswa smp dalam menyelesaikan soal persamaan linear dua variabel di SMPN 1 Campaka Mulya-Cianjur. *Jurnal Pembelajaran Matematika Inovatif*, *3*(1), 59–68.

Zulkifli, F., Abidin, R. Z., Razi, N. F. M., Mohammad, N. H., Ahmad, R., & Azmi, A. Z. (2018). Evaluating quality and reliability of final exam questions for probability and statistics course using rasch model. *International Journal of Engineering and Technology(UAE)*, *7*(4), 32–36.