24% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

104 Not Cited or Quoted 24%
Matches with neither in-text citation nor quotation marks

1 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

18% Internet sources
16% Publications
15% Submitted works (Student Papers)

Match Groups

- **104** Not Cited or Quoted 24%
  Matches with neither in-text citation nor quotation marks
- **1** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

18% Internet sources
16% Publications
15% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	jurnalbeta.ac.id 4%
2	Internet	ojs.fkip.ummetro.ac.id 2%
3	Publication	Sekar Purbarini Kawuryan, Banu Setyo Adi, Mujinem Mujinem, Evy Nur Rochmah,... 2%
4	Internet	journal-gehu.com 1%
5	Internet	eduresearch.web.id <1%
6	Internet	hipatiapress.com <1%
7	Internet	repository.unipa.ac.id <1%
8	Student papers	Universitas Pendidikan Indonesia on 2021-07-18 <1%
9	Internet	repository.upi.edu <1%
10	Internet	www.frontiersin.org <1%

11 Internet

acervodigital.ufpr.br <1%

12 Student papers

UIN Raden Intan Lampung on 2025-08-04 <1%

13 Internet

fyi.org.nz <1%

14 Internet

jurnal.umt.ac.id <1%

15 Publication

"Advances in Applications of Rasch Measurement in Science Education", Springer ... <1%

16 Internet

eudl.eu <1%

17 Internet

ebin.pub <1%

18 Student papers

Universitas Negeri Semarang on 2023-08-30 <1%

19 Internet

eprints.uny.ac.id <1%

20 Publication

Dian Rachmawati, Rizza Megasari, Sri Umi Mintarti Widjaja, Ro'ufah Inayati, Restu... <1%

21 Internet

ejournal.undiksha.ac.id <1%

22 Internet

www.atlantis-press.com <1%

23 Student papers

Institute of Graduate Studies, UiTM on 2013-03-26 <1%

24 Internet

afaaj.com <1%

25	Publication	"Pacific Rim Objective Measurement Symposium (PROMS) 2015 Conference Proce... <1%
26	Student papers	Universitas Islam Riau on 2025-09-05 <1%
27	Internet	jurnal.syntaxtransformation.co.id <1%
28	Internet	repository.nie.edu.sg <1%
29	Publication	Peng Yuanyuan, Bity Salwana Alias, Azlin Norhaini Mansor, Mohd Rashid Ab Hami... <1%
30	Student papers	Universitas Negeri Jakarta on 2025-07-16 <1%
31	Internet	cahaya-ic.com <1%
32	Internet	ejournal.upi.edu <1%
33	Publication	Indri Margaretha, Nizlel Huda, Khairul Anwar. "ANALYZE NUMERACY LITERACY SK... <1%
34	Publication	Maharani Kartikasari, Triyanto, Laila Fitriana, Farida Nurhasanah. "Using Interac... <1%
35	Student papers	Universitas Negeri Padang on 2020-03-13 <1%
36	Internet	aip.scitation.org <1%
37	Internet	ia801407.us.archive.org <1%
38	Internet	journal.stkipsingkawang.ac.id <1%

39 Internet
jurnaledukasia.org <1%

40 Internet
www.hindawi.com <1%

41 Internet
www.researchgate.net <1%

42 Internet
www.shs-conferences.org <1%

43 Publication
Mustafa Bekmezci, Lütfi Sürücü. "Determination of validity, reliability and sample... <1%

44 Student papers
Universitas Mataram on 2021-07-30 <1%

45 Student papers
University of the West Indies on 2020-04-04 <1%

46 Internet
jurnal.feb-umi.id <1%

47 Internet
jurnal.stkippgritulungagung.ac.id <1%

48 Internet
publisher.unimas.my <1%

49 Internet
www.medcraveonline.com <1%

50 Publication
Dwi Rismi Ocy, Wardani Rahayu, Makmuri Makmuri. "RASCH MODEL ANALYSIS: D... <1%

51 Student papers
University Tun Hussein Onn Malaysia on 2016-04-25 <1%

52 Publication
Y Romika, S Atun. "Chemistry student's virtual laboratory self-efficacy: A scale de... <1%

53 Publication

Yusriyyah Febriani Putri, Kadir Kadir, Ahmad Dimyati. "Analysis of Content Validit... <1%

54 Publication

Muhtarom Muhtarom. "Developing an instruments to measure prospective teach... <1%

55 Publication

Rahmi Ramadhani, Soeharto Soeharto, Fitria Arifiyanti, Rully Charitas Indra Prah... <1%

56 Publication

Shiau-Wei Chan, Chee-Kit Looi, Bambang Sumintono. "Assessing computational t... <1%

Journal of General Education and Humanities
Vol. 4, No. 4, November 2025, pp. 2309 – 2321, https://doi.org/10.58421/gehu.v4i4.736
ISSN 2963-7147

Rasch Model Analysis of Algebraic Numeracy Literacy Items Using Winsteps

Rifqi Hidayat[1], Turmudi[2], Jarnawi Afgani Dahlan[3], and Elah Nurlaelah[4]
[1,2,3,4]Universitas Pendidikan Indonesia, Bandung, Indonesia

Article Info

Article history:

Received 2025-09-27
Revised 2025-11-14
Accepted 2025-11-28

Keywords:

Algebra
Numeracy Literacy
Rasch Model
Winsteps.

ABSTRACT

The limited availability of valid and reliable numeracy literacy test instruments, particularly in the algebra domain, for junior high school students in preparation for the Minimum Competency Assessment exam makes this research important to conduct. This study aims to describe the quality of a numeracy literacy test instrument in the algebra domain for junior high school students. It is a quantitative study with a descriptive approach. This method is used to describe the quality of the numeracy literacy test instrument in the algebra domain, which was developed based on Rasch modeling. The research procedure comprises three main phases: item development, limited trial testing with 8th-grade students, and item quality analysis using the Rasch model analysis with the Winsteps software. The developed numeracy literacy test instrument consists of 12 essay questions. The instrument was then used to collect quantitative data from 34 8th-grade students selected as research respondents through purposive sampling. The results indicate that one item is categorized as invalid (requiring revision), while the remaining items are deemed valid. The reliability index is 0.80, which falls into the ideal category. The analysis of item difficulty levels reveals that one item is categorized as difficult, nine items as moderate, and two items as easy. Overall, the developed algebra numeracy literacy instrument meets the criteria of Rasch modelling, although one item requires revision. Teachers can use this instrument as an assessment tool to evaluate students' algebraic numeracy literacy skills in junior high school. This instrument can also serve as a model for developing numeracy literacy test instruments in other domains.

Corresponding Author:

Rifqi Hidayat
Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia.
Email: rifqi.hidayat@upi.edu

1. INTRODUCTION

Numeracy literacy is one of the fundamental competencies that students must master to face the challenges of the 21st century, particularly in the context of the Minimum Competency Assessment, which emphasizes the ability to reason using mathematical

concepts to solve problems [1]. One important domain of numeracy literacy is algebra, since algebra serves as the foundation for mastering advanced mathematical concepts and is connected to various disciplines [2], [3]. However, several studies indicate that junior high school students in Indonesia continue to face challenges in grasping algebraic concepts and applying them in numeracy [4], [5], [6], [7].

One factor contributing to students' low numeracy achievement is the limited availability of valid and reliable test instruments, particularly in the algebra domain. Previous research indicates that the availability of numeracy literacy test instruments specifically for the algebra domain that are truly valid and reliable in Indonesia remains limited, with development just beginning. This means that the number of numeracy literacy test instruments that have been tested for validity and reliability in the algebra domain at the junior high school level in Indonesia is limited and still few [8]. Instruments that have not been thoroughly tested for validity and reliability may yield biased assessment data, thereby failing to accurately reflect students' abilities [9]. Therefore, the development of high-quality numeracy literacy test instruments is an urgent need, allowing teachers and researchers to conduct accurate learning evaluations.

The Rasch Model is one approach for analyzing the quality of test instruments. This approach enables researchers to evaluate validity, reliability, and item difficulty levels more objectively than classical analysis [4]. With the assistance of Winsteps software, the Rasch Model can visualize the distribution of student ability and item difficulty through a Wright Map, ensuring that the resulting instrument can be scientifically justified [10].

This study aims to describe the quality of a numeracy literacy test instrument in the algebra domain, specifically in essay format, for junior high school students, developed using Rasch modeling. The novelty of this research lies in the development of a Rasch-based algebra numeracy essay instrument specifically tested on eighth-grade students, making it the first instrument to utilize the Rasch approach to measure algebraic numeracy literacy at the junior high school level. The resulting instrument not only meets validity and reliability criteria but also provides information on varying levels of item difficulty, allowing for a deeper identification of students' strengths and weaknesses. The advantage of essays over multiple choice in measuring algebraic numeracy literacy, from the perspective of the Rasch Model, lies in their ability to assess students' understanding more holistically, reduce bias that may arise from limited answer options, and provide a more accurate depiction of students' analytical abilities and critical thinking skills in solving algebra problems.

2. METHOD

This study employs a quantitative research approach with a descriptive design. A quantitative rather than qualitative design was selected because the goal is to objectively and systematically measure and describe the quality of a numeracy-literacy test instrument in the algebra domain. Data are collected in numerical form or as scores that can be statistically analyzed [11]. The study focuses on developing a test instrument whose validity and reliability can be examined and that measures students' numeracy skills in a structured and measurable manner through psychometric analysis [12], [13], [14].

A quantitative approach was chosen because the test instrument developed is based on Rasch modeling, which requires statistical analysis to assess validity, reliability, and item difficulty levels [13], [15], [16]. Using this approach, researchers collect numerical data (e.g., students' test scores) and analyze them to determine how well the instrument measures students' numeracy literacy [12], [17].

In contrast, a qualitative approach focuses on an in-depth understanding of phenomena or contexts through techniques such as interviews, observations, or text analysis and does not produce numerical data suitable for statistical analysis [11]. Because this study centers on instrument development and statistical analysis of test results, the quantitative approach is the most appropriate to achieve these goals [11], [13].

Meanwhile, the descriptive approach is not intended to accept or reject hypotheses but rather to describe the situation under study [18]. Thus, the descriptive-quantitative method is used to describe the quality of the numeracy-literacy test instrument in the algebra domain for junior high school students, developed based on Rasch modeling [12], [16]. The research was conducted at a junior high school in Cirebon Regency, Indonesia. The population in this study consisted of 34 eighth-grade students selected by the teacher.

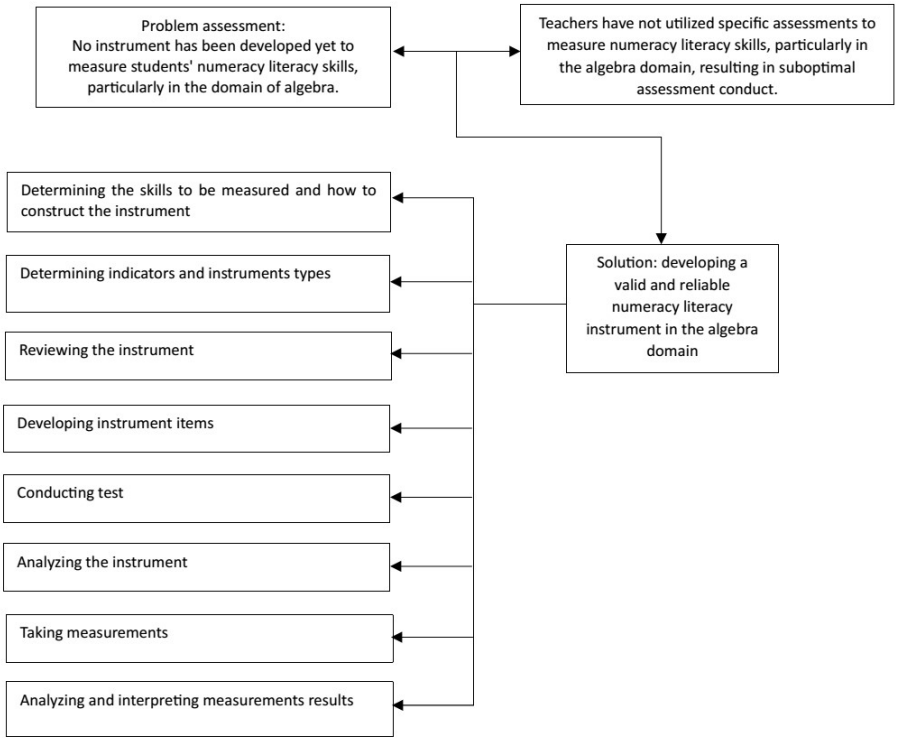


Figure 1. Research Procedure [19]

The research procedure consisted of three main stages: the development of the test instrument, a limited trial of the instrument, and an evaluation of its quality using Rasch model analysis with the assistance of Winsteps software, as illustrated in Figure 1. In the initial stage, instrument development was preceded by identifying the core student competencies in mathematics, specifically numeracy literacy. An essay-type format was selected for the instrument, as it enables students to convey their problem-solving strategies and demonstrate their understanding through mathematical language.

Table 1. Numeracy Literacy Indicators

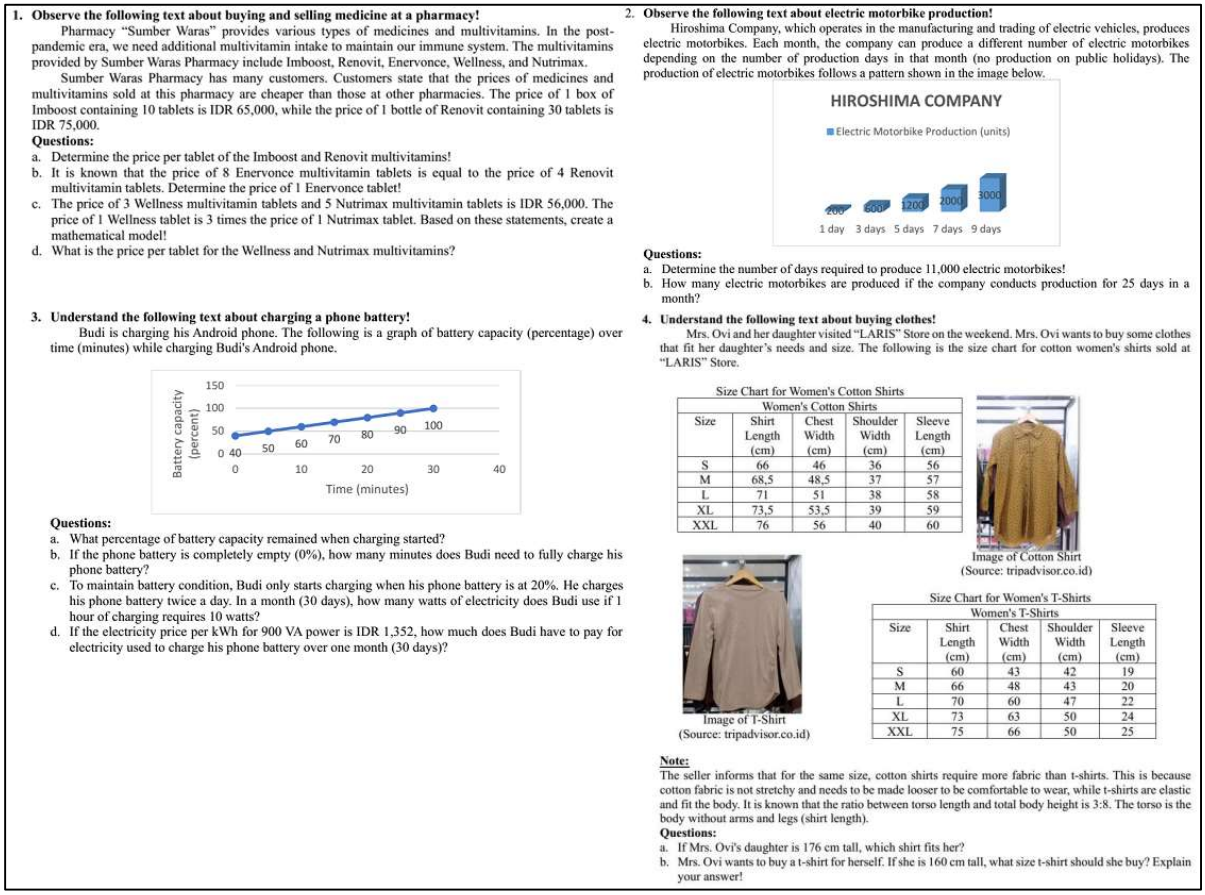
No	Numeracy Literacy Indicators
1	The ability to use various types of numbers or symbols related to basic mathematics in solving everyday problems.
2	The ability to analyze information in various forms (graphs, tables, charts, diagrams, etc).
3	The ability to interpret the results of problem analysis to make predictions and decisions.

Source: Kemendikbud [20]

The outline for the Algebra numeracy literacy test for grade 8 can be seen in the following table 2:

Table 2. The Outline for Algebra, Numeracy, and Literacy Test for Grade 8

Domain	Sub-Domain	Context	Competency	Cognitive Level	Numeracy Literacy Indicator	Question Indicator	Question Number
Algebra	Equations and Inequalities	Socio-Cultural	Solve linear equations and inequalities with one variable and systems of linear equations with two variables.	Applying (C3)	The ability to use various types of numbers or symbols related to basic mathematics in solving everyday problems.	Students use various types of numbers or symbols related to one-variable linear equations to solve a problem of buying and selling medicine/multivitamins at a pharmacy.	1a 1b
				Applying (C3)		Students use various types of numbers or symbols related to two-variable linear equations to solve a problem of buying and selling medicine/multivitamins at a pharmacy.	1c 1d
	Relations & Functions	Socio-Cultural	Generalize patterns of number sequences and object configurations.	Reasoning (C6)	The ability to interpret the results of problem analysis to predict and make decisions.	Students generalize number sequence patterns in problems related to electric motor production.	2a 2b
		Socio-Cultural	Present, analyze, and solve problems using relations, functions, and linear equations along with their graphs.	Reasoning (C4)	The ability to analyze information in various forms (graphs, tables, charts, diagrams, etc.)	Students analyze problems related to the percentage of battery charging on Android phones presented in the form of a graph.	3a 3b 3c 3d
	Ratio & Proportion	Socio-Cultural	Solve everyday problems related to rates of change.	Reasoning (C4)	The ability to analyze information in various forms (graphs, tables, charts, diagrams, etc.)	Students analyze problems related to the rate of change in price and discounts on clothing presented in the form of a table.	4a 4b



1. Observe the following text about buying and selling medicine at a pharmacy!

Pharmacy "Sumber Waras" provides various types of medicines and multivitamins. In the post-pandemic era, we need additional multivitamin intake to maintain our immune system. The multivitamins provided by Sumber Waras Pharmacy include Imboost, Renovit, Enervonce, Wellness, and Nutrimax.

Sumber Waras Pharmacy has many customers. Customers state that the prices of medicines and multivitamins sold at this pharmacy are cheaper than those at other pharmacies. The price of 1 box of Imboost containing 10 tablets is IDR 65,000, while the price of 1 bottle of Renovit containing 30 tablets is IDR 75,000.

Questions:

a. Determine the price per tablet of the Imboost and Renovit multivitamins!
b. It is known that the price of 8 Enervonce multivitamin tablets is equal to the price of 4 Renovit multivitamin tablets. Determine the price of 1 Enervonce tablet!
c. The price of 3 Wellness multivitamin tablets and 5 Nutrimax multivitamin tablets is IDR 56,000. The price of 1 Wellness tablet is 3 times the price of 1 Nutrimax tablet. Based on these statements, create a mathematical model!
d. What is the price per tablet for the Wellness and Nutrimax multivitamins?

3. Understand the following text about charging a phone battery!

Budi is charging his Android phone. The following is a graph of battery capacity (percentage) over time (minutes) while charging Budi's Android phone.

Questions:

a. What percentage of battery capacity remained when charging started?
b. If the phone battery is completely empty (0%), how many minutes does Budi need to fully charge his phone battery?
c. To maintain battery condition, Budi only starts charging when his phone battery is at 20%. He charges his phone battery twice a day. In a month (30 days), how many watts of electricity does Budi use if 1 hour of charging requires 10 watts?
d. If the electricity price per kWh for 900 VA power is IDR 1,352, how much does Budi have to pay for electricity used to charge his phone battery over one month (30 days)?

2. Observe the following text about electric motorbike production!

Hiroshima Company, which operates in the manufacturing and trading of electric vehicles, produces electric motorbikes. Each month, the company can produce a different number of electric motorbikes depending on the number of production days in that month (no production on public holidays). The production of electric motorbikes follows a pattern shown in the image below.

Questions:

a. Determine the number of days required to produce 11,000 electric motorbikes!
b. How many electric motorbikes are produced if the company conducts production for 25 days in a month?

4. Understand the following text about buying clothes!

Mrs. Ovi and her daughter visited "LARIS" Store on the weekend. Mrs. Ovi wants to buy some clothes that fit her daughter's needs and size. The following is the size chart for cotton women's shirts sold at "LARIS" Store.

Size Chart for Women's Cotton Shirts

Women's Cotton Shirts				
Size	Shirt Length (cm)	Chest Width (cm)	Shoulder Width (cm)	Sleeve Length (cm)
S	66	46	36	56
M	68,5	48,5	37	57
L	71	51	38	58
XL	73,5	53,5	39	59
XXL	76	56	40	60

Image of Cotton Shirt
(Source: tripadvisor.co.id)

Image of T-Shirt
(Source: tripadvisor.co.id)

Size Chart for Women's T-Shirts

Women's T-Shirts				
Size	Shirt Length (cm)	Chest Width (cm)	Shoulder Width (cm)	Sleeve Length (cm)
S	60	43	42	19
M	66	48	43	20
L	70	60	47	22
XL	73	63	50	24
XXL	75	66	50	25

Note:

The seller informs that for the same size, cotton shirts require more fabric than t-shirts. This is because cotton fabric is not stretchy and needs to be made looser to be comfortable to wear, while t-shirts are elastic and fit the body. It is known that the ratio between torso length and total body height is 3:8. The torso is the body without arms and legs (shirt length).

Questions:

a. If Mrs. Ovi's daughter is 176 cm tall, which shirt fits her?
b. Mrs. Ovi wants to buy a t-shirt for herself. If she is 160 cm tall, what size t-shirt should she buy? Explain your answer!

Figure 2. Algebraic Numeracy Literacy Test Instrument

The review stage concerns the process of selecting a mathematics topic. In this case, the algebra domain was selected based on Minimum Competency Assessment results from several junior high schools in Cirebon Regency (Indonesia), which indicated that the domain is in need of improvement. The number of test items was expected to align with the number of numeracy literacy indicators presented in Table 1. Each item measures one numeracy literacy indicator and one algebra competency based on the numeracy learning progression in the Minimum Competency Assessment. The complete outline of the Algebra numeracy literacy test for grade 8 in this study is presented in Table 2. The item development process involved expert judgment regarding face and content validity.

Face validity covers aspects such as language, symbols, and the potential for misinterpretation of sentences. Content validity, on the other hand, refers to the relevance of the items to the numeracy literacy indicators [21]. Content validity and face validity in this case were assessed by three experts: two mathematics education lecturers from Universitas Pendidikan Indonesia and one mathematics education lecturer from Universitas Swadaya Gunung Jati, as well as two practitioners, namely mathematics teachers from SMP Negeri 1 Weru and SMP Negeri 2 Sumber.

The next stage of the study involved pilot-testing the instrument with 34 students, who were asked to respond to each test item on the worksheet. The students' responses to each item were then scored using a holistic rubric with a scale of 0–4. This scale was

designed to describe indicators of ability and tier achievement criteria, as explained by Viyanti [22].

As the final and core stage of this study, a quantitative analysis of student responses was conducted using Rasch modeling to assess the quality of the developed numeracy literacy test instrument. An instrument is considered "of good quality" if it meets a number of evaluation criteria, including item fit, empirical validity, difficulty level, reliability, and the distribution of respondents' abilities [23]. The statistical calculations were performed using the Winsteps software.

In the application of the Rasch model, item fit is analyzed to determine the extent to which test items contribute to the measurement of the intended construct. Item fit is generally evaluated using fit statistics such as infit and outfit mean square (MNSQ) [24]. The sequence of items deemed appropriate is determined by examining the fit statistics for each item [25]. The quality of an item in the Rasch model is assessed according to specific criteria, as shown in Table 3 [25], [26], [27], [28].

Table 3. Criteria for Determining the Quality of an Item in the Rasch Model

No	Criteria	Value
1	Outfit MNSQ (Mean Square)	0,5 < outfit MNSQ < 1,5
2	Outfit ZSTD (Z-Standard)	−2,0 < outfit ZSTD < +2,0
3	Point Measure Correlation	0,4 < Pt Measure Corr < 0,85

Instrument items that do not meet all three criteria are categorized as "misfit" and must be replaced. Conversely, if an item fulfills at least two of the specified requirements, it can still be considered "fit" or suitable for use [25]. The item fit requirements based on the instrument quality assessment system are presented in Table 4 [19], [29].

Table 4. Outfit MNSQ criteria

Item model fit mean-square range extremes	Criteria
0,5 - 1,5	Good
1,5 - 2	Marginal
> 2,0	Poor

Table 4 indicates that an MNSQ value above 2.0 reflects a disturbance in the measurement process, implying that the item is not functioning properly within the model and may need to be revised or reevaluated. Values ranging from 1.5 to 2.0 suggest a moderate misfit that is not critically problematic; in certain contexts, such items can still be considered acceptable depending on the research objectives. An MNSQ value between 0.5 and 1.5 demonstrates that the item aligns well with the measurement model and contributes effectively to the assessment. Conversely, an MNSQ value below 0.5 indicates that the item provides limited measurement information, though it does not compromise the integrity of the measurement system.

Additionally, reliability is another important aspect that influences the quality of an item. Rasch reliability values, which range from zero to one and are often referred to as Cronbach's Alpha, describe the consistency of both respondents and items [30]. Reliability

is defined as the ability of an instrument to produce consistent results when applied repeatedly under the same conditions [31], [32]. Thus, reliability reflects both the quality of the instrument and the accuracy of the measurement results. In the Rasch model framework, reliability comprises both person and item reliability. Generally, an acceptable Cronbach's Alpha coefficient ranges between 0.70 and 0.99 [33]. The criteria for respondent and item reliability are presented in Table 5 [24], [26].

Table 5. Person and item reliability

Person and Item Reliability	Criteria
> 0,94	Excellent
0,91 – 0,94	Very Good
0,81 – 0,90	Good
0,67 – 0,80	Fair
> 0,67	Poor

3. RESULTS AND DISCUSSION

3.1. Results

Figure 2 presents the test items developed in this study. The instrument consists of four essay questions and 12 sub-questions, designed to assess junior high school students' numeracy literacy skills in the algebra domain. The items cover the sub-domains of equations and inequalities, relations and functions, and ratios and proportions. Each item is contextualized within socio-cultural settings and targets cognitive levels in the categories of applying (C3) and reasoning (C4, C6). In addition, each item is equipped with indicators that cover the ability to apply different types of numbers or mathematical symbols in solving real-world problems, the skill to analyze information presented in various formats (such as graphs, tables, charts, or diagrams), and the competence to interpret analysis results for making predictions and informed decisions. All items were designed in alignment with the Minimum Competency Assessment framework. Based on evaluations by experts (mathematics lecturers) and practitioners (mathematics teachers), this test instrument meets the criteria for face and content validity.

Performance of Students' Algebraic Numeracy Literacy Skills

Based on the descriptive analysis of the pilot test results, the frequency distribution of algebraic numeracy literacy scores from 34 students is presented in Figure 3 below.

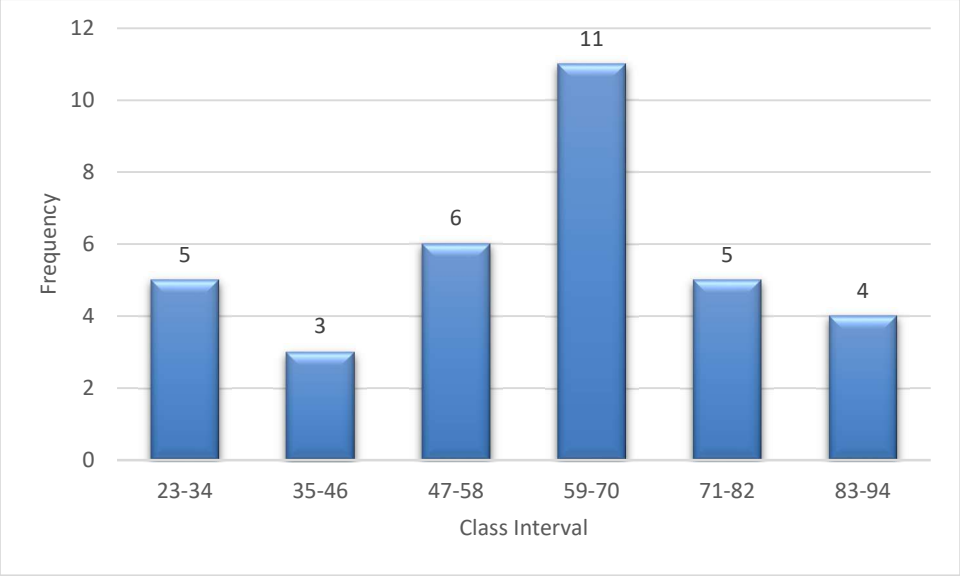


Figure 3. Frequency distribution of students' algebraic numeracy literacy scores

Figure 3 displays six class intervals, each with a class width of 12. A total of 32% of student scores fall within the range of 59–70, 15% within 71–82, and 12% within 83–94. The chart indicates that students' algebraic numeracy literacy skills are at a moderate level, as shown by more than 50% of students scoring above 60 on a scale of 100. These findings are not further explored, as they fall outside the scope of this study.

Item Fit Analysis and Empirical Validity

Figure 4 displays the output table generated by Winsteps, which presents the values of OUTFIT MNSQ, OUTFIT ZSTD, and OUTFIT PT MEASURE CORR. The twelve essay items are coded as 1a, 1b, 1c, 1d, 2a, 2b, 3a, 3b, 3c, 3d, 4a, and 4b in the "item" column, corresponding to question numbers 1 through 12 in the "ENTRY NUMBER" column. Meanwhile, the "TOTAL COUNT" column shows the number of participants in the study, which consisted of 34 students.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
11	39	34	1.01	.17	.57	-1.68	.55	-1.42	.73	.49	55.9	48.8	4a
3	48	34	.77	.15	1.19	.81	1.44	1.32	.12	.53	52.9	40.4	1c
12	54	34	.64	.15	1.04	.25	.96	-.05	.69	.55	38.2	36.1	4b
4	61	34	.49	.14	1.01	.14	.88	-.34	.58	.57	29.4	28.5	1d
10	61	34	.49	.14	1.32	1.43	1.14	.56	.59	.57	32.4	28.5	3d
6	69	34	.33	.14	.86	-.67	.75	-.83	.68	.59	38.2	33.6	2b
5	79	34	.14	.14	.91	-.40	.76	-.69	.62	.61	38.2	30.4	2a
9	89	34	-.06	.14	.89	-.43	.70	-.76	.65	.62	44.1	35.0	3c
7	108	34	-.51	.17	1.65	1.87	1.32	.68	.49	.62	52.9	59.1	3a
8	111	34	-.61	.18	.84	-.39	2.59	1.92	.60	.61	70.6	64.4	3b
2	123	34	-1.08	.22	.87	-.21	.98	.32	.51	.52	79.4	77.3	1b
1	131	34	-1.62	.32	1.44	.83	2.91	1.88	.15	.34	85.3	87.8	1a
MEAN	81.1	34.0	.00	.17	1.05	.13	1.25	.18			51.5	47.5	
P.SD	29.6	.0	.77	.05	.29	.94	.72	.99			17.7	19.3	

Figure 4. Output file for item analysis

As illustrated in Figure 4 (highlighted in red), the OUTFIT MNSQ values for items 4a, 3b, and 1a are x = 0.55, 2.59, and 2.91, respectively, indicating that these items are not fully consistent with the Rasch model assumptions. The same figure also shows that the OUTFIT ZSTD values for all items fall within the model's acceptable range. However, the OUTFIT PT MEASURE CORR values for items 1c and 1a are x = 0.12 and 0.15, which do not meet the criteria of the Rasch model.

Furthermore, Table 6 summarizes the OUTFIT MNSQ, OUTFIT ZSTD, and OUTFIT PT MEASURE CORR values along with their validity criteria. The analysis reveals that 11 items are empirically valid, while 1 item does not meet the validity standards. This invalid item requires revision to ensure the test instrument can effectively assess junior high school students' algebraic numeracy literacy skills.

Table 6. Empirical validity of the item

Item	OUTFIT MNSQ value	OUTFIT ZSTD value	OUTFIT PT MEASURE CORR value	Category
1a	2.91	1.38	0.15	Invalid
1b	0.98	0.32	0.51	Valid
1c	1.44	1.32	0.12	Valid
1d	0.88	-0.34	0.58	Valid
2a	0.76	-0.69	0.62	Valid
2b	0.75	-0.83	0.68	Valid
3a	1.32	0.68	0.49	Valid
3b	2.59	1.92	0.60	Valid
3c	0.70	-0.76	0.65	Valid
3d	1.14	0.56	0.59	Valid
4a	0.55	-1.42	0.73	Valid
4b	0.96	-0.05	0.69	Valid

Item Reliability Analysis

Figure 5 displays the reliability analysis for both respondents (persons) and items. The KR-20 coefficient (highlighted in red) for respondent reliability is r = 0.80, which places student reliability in the fair category. Therefore, the test items can be considered suitable for assessing junior high school students' algebraic numeracy literacy.

```
SUMMARY OF 34 MEASURED Person
-------------------------------------------------------------------------------
|           TOTAL                        MODEL          INFIT        OUTFIT    |
|           SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD    MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       28.6      12.0         .53     .27      1.01    .01    1.22    .17 |
|   SEM       1.6         .0        .11     .01       .08    .19     .32    .18 |
| P.SD        9.0         .0        .66     .04       .49   1.11    1.84   1.04 |
| S.SD        9.2         .0        .67     .04       .50   1.13    1.87   1.05 |
| MAX.       45.0      12.0        1.80     .40      2.08   2.06    9.90   4.49 |
| MIN.       11.0      12.0        -.89     .25       .27  -2.04     .22  -1.39 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .30 TRUE SD     .59  SEPARATION  1.94  Person RELIABILITY   .79 |
|MODEL RMSE    .28 TRUE SD     .60  SEPARATION  2.16  Person RELIABILITY   .82 |
| S.E. OF Person MEAN = .11                                                   |
-------------------------------------------------------------------------------
LACKING RESPONSES:       1 Person
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .80 SEM = 4.08
STANDARDIZED (50 ITEM) RELIABILITY = .95
```

Figure 5. Output person reliability

Wright Map Analysis (Person-Item Map)

The Wright map, also known as a person–item map, illustrates the alignment between the distribution of student abilities and the difficulty levels of the test items [25]. Based on the pilot test results, the distribution produced by Winsteps is shown in Figure 6.

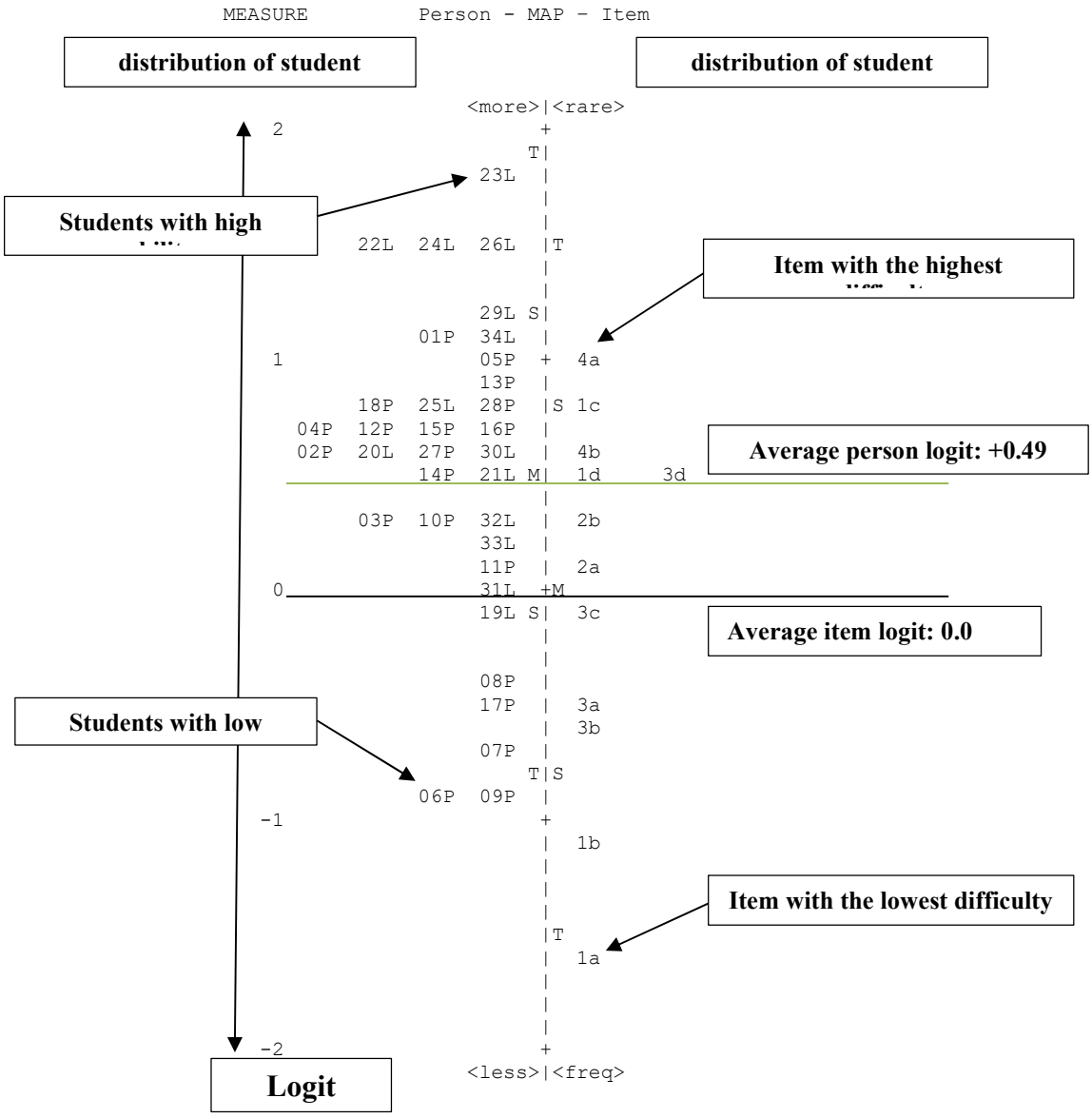


Figure 6. Wright map analysis

Figure 6 shows the distributions of student abilities (left) and item difficulty levels (right). Several points can be analyzed from the map, as follows:

First, on the upper left side of the map, there are eight students (23L, 22L, 24L, 26L, 29L, 01P, 34L, and 05P) whose ability levels are higher than all item difficulty levels provided. This means that all eight students would obtain the maximum possible score. Among them, student 23L has the highest ability (+1.88 logit). There are also four students (13P, 18P, 25L, and 28P) who successfully solved ten essay items but failed to correctly answer item 4a, which had a higher difficulty level (+1.01 logit). On the lower left side of

the map, there are two students with low ability levels (06P and 09P), who correctly answered only items 1b (-1.08 logit) and 1a (-1.62 logit). This indicates that these two students require special attention from the teacher, as the material taught has not been well understood when solving exam questions. In other words, a special approach is needed beyond classical teaching methods.

Second, on the right side of the Wright map, the twelve items display a wide variability of difficulty levels, ranging from item 4a as the most difficult to item 1a as the easiest. This is a positive finding, as the items provide helpful information about the abilities being tested in the students. Conversely, it would be problematic if all items were clustered at the same ability level, for example, all in the upper right (too difficult) or lower right (too easy). Based on student ability levels, item 4a is categorized as difficult, items 1a and 1b as easy, and the remaining items as having moderate difficulty.

3.2. Discussion

The results of this study show that the developed algebraic numeracy literacy test instrument meets the quality criteria based on Rasch Model analysis. Eleven out of twelve items were found to fit the model assumptions, while one item (1a) was identified as a misfit and thus requires revision. This finding is consistent with Sumintono and Widhiarso [34], who emphasized that the Rasch model can detect item misfit while providing detailed information on instrument validity. With a reliability index of 0.80, the instrument can be categorized as having "good" quality for use as a tool to measure junior high school students' algebraic thinking ability.

The analysis of item difficulty levels also shows a proportional distribution, with item 4a being the most difficult, items 1a and 1b the easiest, and the remaining nine items classified as moderate. This variation in difficulty levels is important for identifying the full range of students' abilities. These results align with William and Boone [29], who demonstrated that the Rasch Model can produce a more accurate mapping of student abilities by linking item difficulty to individual ability profiles. Furthermore, this finding reinforces the studies of Ambarwati [35], Rusyid [36], and Sheptian et al. [37], which confirmed that instruments with a balanced variation of difficulty levels are more beneficial for diagnostic assessments of numeracy literacy.

Practically, this instrument is useful for teachers in evaluating students' understanding of algebraic concepts, particularly within the context of the Minimum Competency Assessment, which emphasizes numeracy literacy as one of the key competencies of the 21st century. However, this study has limitations, including a relatively small number of items and a small sample size, which limit the generalizability of the results. Therefore, future research is recommended to expand the number of test items and involve more diverse participants, so that the instrument's quality can be more thoroughly tested and its applications broadened.

4. CONCLUSION

The development of a test instrument to measure students' algebraic numeracy literacy is highly relevant in the educational context. In this study, 12 essay items were designed and evaluated using Rasch model analysis, which confirmed that 11 items (1b, 1c,

1d, 2a, 2b, 3a, 3b, 3c, 3d, 4a, and 4b) fit the model assumptions, while item 1a required revision. Despite this misfit, the overall instrument demonstrated validity and reliability, with a reliability index of r = 0.80, categorized as "good." Item difficulty analysis showed that item 4a was the most challenging, items 1a and 1b were the easiest, and the remaining nine items were of moderate difficulty. These findings highlight the practicality of the instrument for assessing students' algebraic reasoning skills. Nonetheless, the study is limited by the small number of items and respondents, underscoring the need for future research with larger, more diverse samples.

REFERENCES

[1] Pusat Asesmen dan Pembelajaran, *AKM dan implikasinya pada pembelajaran*. Jakarta: Kemdikbud, 2020.

[2] J. Cai and E. Knuth, Eds., *Early Algebraization*. in Advances in Mathematics Education. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-17735-4.

[3] C. Kieran, "Learning and teaching algebra at the middle school through college levels," *Second Handbook of Research on Mathematics Teaching and Learning*, pp. 707–762, Jan. 2007.

[4] A. C. Valenzeo and E. Sumiaty, "Learning Obstacles Siswa SMP dalam Menyelesaikan Operasi Bentuk Aljabar," *Jurnal Pendidikan Matematika (JUDIKA EDUCATION)*, vol. 7, no. 1, pp. 37–48, Jun. 2024, doi: 10.31539/judika.v7i1.8675.

[5] N. Syamsuddin, S. S. Manik, and K. Safrina, "Pengembangan E-Modul Literasi Numerasi Aljabar Siswa Sekolah Menengah Pertama," *Jurnal Ilmiah Pendidikan Matematika Al Qalasadi*, vol. 8, no. 1, pp. 39–52, Jun. 2024, doi: 10.32505/qalasadi.v8i1.8230.

[6] R. Qolbiyah and C. K. Sari, "Analisis Literasi Matematis Siswa SMP Dalam Menyelesaikan Soal Aljabar Berorientasi AKM Numerasi Ditinjau Dari Gaya Kognitif," *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, vol. 10, no. 2, pp. 467–480, Sep. 2024, doi: 10.29100/jp2m.v10i2.6122.

[7] F. H. Fitri and K. N. S. Effendi, "Numeration ability of junior high school students in solving numeration problems algebra material," *Union: Jurnal Ilmiah Pendidikan Matematika*, vol. 11, no. 1, pp. 103–110, Mar. 2023, doi: 10.30738/union.v11i1.13395.

[8] N. Aprilia, Y. Setiani, and C. A. Hadi FS, "Pengembangan Instrumen Tes Numerasi Pada Asesmen Kompetensi Minimum Yang Bernilai Budaya Lokal," *Jurnal Educatio FKIP UNMA*, vol. 9, no. 2, pp. 850–857, Jun. 2023, doi: 10.31949/educatio.v9i2.4824.

[9] B. Sumintono and W. Widhiarso, *plikasi Pemodelan Rasch Pada Assesmen Pendidikan*. Cimahi: Trim Komunikata, 2015.

[10] C. Han, "William J. Boone, John R. Staver and Melissa S. Yale. Rasch Analysis in the Human Sciences," *Journal of Research Design and Statistics in Linguistics and Communication Science*, vol. 5, no. 1–2, pp. 208–211, Dec. 2019, doi: 10.1558/jrds.37535.

[11] J. W. Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*, 4th ed., no. c. Los Angeles: SAGE, 2014.

[12] R. T. Wardani, E. Nuraeni, and S. Diana, "Rasch model analysis of essay questions to measure literacy and numeracy skills in plant and animal bioprocess topics based on AKM," *REID (Research and Evaluation in Education)*, vol. 11, no. 1, pp. 29–44, Jul. 2025, doi: 10.21831/reid.v11i1.85614.

[13] M. J. Peeters and J. M. Augustine, "Using Rasch measurement for instrument rating scale refinement," *Curr Pharm Teach Learn*, vol. 15, no. 1, pp. 110–118, Jan. 2023, doi: 10.1016/j.cptl.2023.02.015.

[14] F. Rasyid, E. Istiyono, and C. W. Gunawan, "Developing and analyzing items of a physics conceptual understanding test on wave topics for high school students using the Rasch Model," *REID (Research and Evaluation in Education)*, vol. 11, no. 1, pp. 1–16, Jun. 2025, doi: 10.21831/reid.v11i1.75575.

[15] J. Haruta, R. Nakajima, and T. Monkawa, "Development of a validated assessment tool for medical students using simulated patients: an 8-year panel survey," *BMC Med Educ*, vol. 24, no. 1, p. 399, Apr. 2024, doi: 10.1186/s12909-024-05386-2.

[16] G. Engelhard and J. Wang, *Rasch Models for Solving Measurement Problems: Invariant Measurement in the Social Sciences*. 2455 Teller Road, Thousand Oaks California 91320: Sage, 2021. doi: 10.4135/9781071878675.

[17] F. Rasyid, E. Istiyono, and C. W. Gunawan, "Developing and analyzing items of a physics conceptual understanding test on wave topics for high school students using the Rasch Model," *REID (Research and Evaluation in Education)*, vol. 11, no. 1, pp. 1–16, Jun. 2025, doi: 10.21831/reid.v11i1.75575.

[18] S. McCombes, "Descriptive Research | Definition, Types, Methods & Examples." Accessed: Nov. 25, 2025. [Online]. Available: https://www.scribbr.com/methodology/descriptive-research/
[19] J. Linacre, "Predicting responses from Rasch measures," *J Appl Meas*, vol. 11, pp. 1–10, Jan. 2010.
[20] Kemendikbud, *Materi Pendukung Literasi Numerasi*. Jakarta: Kementerian Pendidikan dan Kebudayaan, 2017.
[21] S. Azwar, *Reliabilitas dan Validitas*. Pustaka Pelajar, 2012.
[22] V. Viyanti, U. Rosidin, and R. E. Shintya, "Collaborative and Problem Solving Instruments in Project-Based Physics Learning," *Indonesian Journal of Science and Mathematics Education*, vol. 5, no. 1, pp. 96–108, Mar. 2022, doi: 10.24042/ijsme.v5i1.11636.
[23] P. E. Larasati, Supahar, and D. R. A. Yunanta, "Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material," *J Phys Conf Ser*, vol. 1440, no. 1, p. 012020, Jan. 2020, doi: 10.1088/1742-6596/1440/1/012020.
[24] J. M. Linacre, *A User's Guide to WINSTEPS® Rasch-Model Computer Programs: Program Manual 4.4.6*. Chicago: Mesa-Press, 2019.
[25] B. Sumintono, "Rasch Model Measurements as Tools in Assesment for Learning," in *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*, Paris, France: Atlantis Press, 2018. doi: 10.2991/icei-17.2018.11.
[26] C. J. Perera, B. Sumintono, and N. Jiang, "The Psychometric Validation Of The Principal Practices Questionnaire Based On Item Response Theory," *International Online Journal of Educational Leadership*, vol. 2, no. 1, pp. 21–38, Feb. 2018, doi: 10.22452/iojel.vol2no1.3.
[27] S.-W. Chan, C.-K. Looi, and B. Sumintono, "Assessing computational thinking abilities among Singapore secondary students: a Rasch model measurement analysis," *Journal of Computers in Education*, vol. 8, no. 2, pp. 213–236, Jun. 2021, doi: 10.1007/s40692-020-00177-2.
[28] M. Maryati, Z. K. Prasetyo, I. Wilujeng, and B. Sumintono, "Measuring Teachers' Pedagogical Content Knowledge Using Many-Facet Rasch Model," *Jurnal Cakrawala Pendidikan*, pp. 452–464, Oct. 2019, doi: 10.21831/cp.v38i3.26598.
[29] J. R. S. and M. S. Yale. William J. Boone, "Rasch Analysis in the Human Sciences," *Journal of Research Design and Statistics in Linguistics and Communication Science*, vol. 5, no. 1–2, pp. 208–211, Dec. 2019, doi: 10.1558/jrds.37535.
[30] P. W. Duncan, R. K. Bode, S. Min Lai, and S. Perera, "Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale11No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organ," *Arch Phys Med Rehabil*, vol. 84, no. 7, pp. 950–963, Jul. 2003, doi: 10.1016/S0003-9993(03)00035-2.
[31] J. A. Oluwatayo, "Validity and Reliability Issues in Educational Research," *Journal of Educational and Social Research*, vol. 2, no. 2 SE-Articles, p. 391, May 2012, [Online]. Available: https://www.richtmann.org/journal/index.php/jesr/article/view/11851
[32] L. SÜRÜCÜ and A. MASLAKÇI, "Validity and Reliability in Quantitative Research," *Business & Management Studies: An International Journal*, vol. 8, no. 3, pp. 2694–2726, Sep. 2020, doi: 10.15295/bmij.v8i3.1540.
[33] M. M. Mohamad, N. L. Sulaiman, L. C. Sern, and K. M. Salleh, "Measuring the Validity and Reliability of Research Instruments," *Procedia Soc Behav Sci*, vol. 204, pp. 164–171, Aug. 2015, doi: 10.1016/j.sbspro.2015.08.129.
[34] B. Sumintono and W. Widhiarso, *Aplikasi Pemodelan Rasch Pada Assesmen Pendidikan*. Cimahi: Trim Komunikata, 2015.
[35] D. Ambarwati and S. Maarif, "Rasch Model Analysis: Development of Students' Numeracy Literacy Interest Instruments in Mathematics Learning," *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, vol. 11, no. 1, pp. 41–52, Mar. 2025, doi: 10.29100/jp2m.v11i1.7279.
[36] H. K. Rusyid, D. Suryadi, T. Herman, M. Adnan, A. Lutfi, and A. Mukhibin, "Rasch modeling approach to measure the quality of algebraic thinking questions for junior high school students," *Beta: Jurnal Tadris Matematika*, vol. 17, no. 1, pp. 55–71, May 2024, doi: 10.20414/betajtm.v17i1.652.
[37] R. Sheptian, R. Arthur, and R. Wardhani, "Psychometric Analysis of Mathematics Readiness Assessment Tool for Elementary School Students: A Rasch Model Approach," *LEARNING : Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran*, vol. 4, no. 4, pp. 1213–1220, Dec. 2024, doi: 10.51878/learning.v4i4.3725.