Proceeding Konferensi Nasional Matematika (KNM) XXII
Universitas Bengkulu, 15-16 Juli 2024
Dipublikasikan online pada:
PATTIMURA PROCEEDING: Conference of Science and Technology
e-ISSN: 2829-3770, volume 5 issue 1, pp.11-20, Edisi November 2024

Validity and Reliability of Mathematical Reasoning Ability Instruments in High Schools with Winsteps

Keysha Alifia Syahwa[1], Satria Kamal Gustia[2], Ayu Faradillah[3*], Syafika Ulfah[4]

[1,2,3,4] Faculty of Teacher Training and Education, Universitas Muhammadiyah Prof. DR. HAMKA Jalan Tanah Merdeka No. 20, RT.11/RW.2, Rambutan, Ciracas Sub-district, East Jakarta City, Special Capital Region of Jakarta 13830, Indonesia

Corresponding author's e-mail: **** ayufaradillah@uhamka.ac.id***

Keywords:
Validity;
Reliability;
Mathematical Reasoning Ability;
Winsteps

ABSTRACT

The mathematical reasoning ability in understanding mathematical material in high school students still has not touched perfect scores. A valid and reliable instrument is needed in measuring mathematical reasoning skills in students. Validity and reliability testing of the instrument is carried out to ensure whether an instrument can be used to provide accurate results in a measurement being performed. Therefore, this study aims to determine the level of validity and reliability of mathematical reasoning skills in students using Winsteps. The method used in this study is a quantitative survey method, with 252 respondents in high schools in Jakarta. This study uses 3 indicators that are a reference in measuring the mathematical reasoning ability possessed by students. The validity test on the instrument shows that all items on the instrument are valid because they meet the criteria of MNSQ, ZSTD, and Pt Mean Corr except for item 1 which has a ZSTD value of > 2.0 and item 3 has a ZSTD value of < -2.0. Meanwhile, the reliability test obtained an Alpha Cronbach calculation of 0.74 which means that the value of the reliability item is high so that the instrument is declared reliable for use. The DIF analysis using winsteps shows that all question items used in the instrument are valid because none of the question items contain bias referring to the DIF Analysis criteria, namely the DIF Contrast value > 0.5 and the probability value < 0.05 where all the value of the question items actually obtained a DIF Contrast value of < 0.5 and a probability value > 0.05.

DOI: https://doi.org/10.30598/ppcst.2024.knmxxii.11-20

1. INTRODUCTION

Reasoning ability is the ability to think at a higher level to draw the right conclusion based on several existing rules and evidence [1][2]. Mathematical reasoning skills are one of the parts of mathematical thinking skills that students need to understand mathematical materials and solve problems faced in daily life [3][4]. Therefore, learners must have the reasoning ability to understand concepts before solving math problems [5]. The importance of mathematical reasoning skills is in line with the vision of mathematics which is directed to provide opportunities for the development of reasoning skills, awareness of the usefulness of mathematics, fostering confidence, an objective and open attitude, as well as the ability to effectively overcome mathematical problems, building logical, analytical, and critical thinking patterns that support decisions in daily life, which is overall a crucial thing to achieve success in Learning process and preparation for future challenges [6][7][8][9]. Seeing the importance of mathematical reasoning skills, in fact education in Indonesia is currently not in line with the development of education and the development of students' abilities which causes a lack of emphasis on the development of these skills in students, resulting in difficulties in finding patterns, relationships, and properties in understanding mathematical materials and in facing various problems in daily life [7][8][9][10]. In addition, the problem with students' mathematical reasoning ability is the weak mathematical reasoning ability in students, which is shown from several survey reports that state that the average score of students' mathematical ability in Indonesia is still low, even the mathematical achievement of students in Indonesia is ranked 44th out of 49 countries [10][11]. This is also shown by the results that only about 20% of students are able to solve reasoning problems and other students are only able to solve problems at a low level of comprehension ability, which means that the score obtained is only 49% of the ideal score of 100 [12].

In this regard, to find out the level of mathematical reasoning ability, an instrument is needed. Test instruments that are commonly used in schools are important tools in research because they serve as a means to collect, check, and investigate data so as to ensure the correctness of research results and be able to measure students' abilities appropriately [13][14]. An instrument is said to be good if it has high validity and reliability, is reliable, is able to accurately measure what should be measured, and provides accurate information related to the ability of students to test their abilities [13][14][15]. Validity means the validity or truth which is defined as the level of accuracy and precision of the measuring tool in carrying out its measurement tasks so that it can measure the reliability and variables to be measured [16][17][18].After knowing the valid question items, a reliability test will be carried out to determine the accuracy of the questions [19][20]. Reliability means a trusted or test resistant that concerns the consistency of a measuring instrument so that it can measure the same variable over and over again on different subjects and at different times [13][16][18].

Testing the validity and reliability of an instrument in research is important to ensure that the instrument can be understood properly, measure exactly what should be measured, how well the instrument is used, and have a high level of confidence, so that the research results obtained are accurate and reliable [13][15][21][22][23]. By conducting validity and reliability tests, researchers can ensure that the instruments used are reliable and provide accurate results in the measurements taken.

In conducting validity and reliability tests, tools in the form of software are needed, one of which can be used is Winsteps software. Winsteps is a Windows-based software designed specifically for the computing of Rasch models used in educational evaluations, attitude surveys, and scale analysis, including test score analysis to Measure Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), Point Measure Correlation (Pt Mean Corr), Item Reliability, and Alpha Cronbach [13][14][24]. This software helps in processing data to obtain relevant information, researchers can also identify invalid items in the questionnaire and obtain in depth information about the quality of the instruments used, as well as can analyze the question items, so that it can help in evaluating the validity, reliability, and difficulty level of the test instrument [25]. Winsteps' advantage is its ability to perform Rasch Model analysis, which is a statistical method used to measure the psychometric characteristics of measurement instruments, such as validity and reliability [21]. The Rasch model is considered to be able to predict lost data and generate standard error measurement values for the instruments used. This can improve the accuracy of calculations [26].

In recent years, there have been several studies that are relevant to this study. The first research is a study conducted by Silvia, Zulfah, and Lussy in 2023 [20] namely to test the validity level of the instrument

in the form of an essay that was tested on 28 respondents with the product moment correlation technique and the reliability of the instrument with the Cronbach's Alpha formula to measure the influence of the Ethno- Realistic Mathematics Education (RME) approach on mathematical reasoning skills in grade VII using SPSS software version 22. This study obtained the validity result value where the calculation value of all > problems is 0.374 with a significant level of 5% and the Cronbach's Alpha value is 0.729 which means that the instrument of this study has high validity and reliability, so that the instrument instrument used by the researcher can be relied on in measuring the influence of the Ethno-RME approach on mathematical reasoning ability with a good level of confidence. Then the second research conducted by Muntazhimah, Syifani, and Hikmatul [27] with the aim of examining the validity and reliability of mathematical resilience instruments using the Rasch model assisted by Winsteps software on students who obtained the results that the instrument was declared valid because there were 29 out of 32 valid items that had been used and met the criteria of "very reliable" because it produced an Alpha Cronbach nlai of 0.71, an item reliability value of 0.98, and a person reliability of 0.63, So that the instrument is valid and reliable to be used in researching mathematical resilience skills in students.

Another research is to test the validity and reliability of non-test instruments consisting of 10 questions in the form of anxiety questionnaires in mathematics learning for students at the secondary school level involving as many as 116 respondents using the Rasch Model on Winstep. This study obtained the results that there was only one item that was invalid because it had an Outfit value of MNSQ > 1.5, a Outfit ZSTD value of > +2.0, and a Pt Mean Corr value of > 0.85 with the results of the Person Fit test there were 11 (9%) people who were invalid in meeting the specified criteria. The results of the reliability calculation show that the person reliability is 0.71, the item reliability is 0.94, and Alpha Cronbach shows the number 0.76, so that this study produces a valid and reliable mathematical anxiety instrument [21]. The next research discussed the use of Rasch Model analysis on Winsteps was also carried out to test the validity of the EPUB3-based Mathematics E-Module conducted by Rahmi Ramadhani and Yulia Fitri [28] by using three stages, namely expert validation and validation of items and constructs using the Rasch Model which obtained the result of a valid and reliable number of question items at a percentage of 24% (below 50%) and still needs to be revised to the gender bias found so that there needs to be improvement, but one of the materials in the E-Module is already suiTable for use in statistics learning.

From several research findings relevant to the theme of mathematical reasoning ability and validity and reliability tests using Winsteps, researchers see that there has been no validity and reliability testing of Winsteps-assisted instruments for mathematical reasoning skills at the high school level. Therefore, this study has the purpose of determining the level of validity and reliability of the mathematical reasoning ability instrument of high school students using Winsteps software. The formulation of the problem in this study is (1) how the results of the validity of the mathematical reasoning ability instrument with Winsteps in high school students? (2) how the results of the reliability of the mathematical reasoning ability instrument with Winsteps in high school students? and (3) how to ensure that the mathematical reasoning test instrument does not contain bias against different demographic groups, so that each item of the test is valid and fair for all groups of respondents?

2. RESEARCH METHODS

2.1 Research Design and Participants

In this study, the researcher uses a quantitative research method using a survey approach. Quantitative research is a research method used to research a specific population or sample by collecting data using research instruments and involving statistical analysis of the numerical data collected [29][30][31]. The survey approach is an approach that in its research uses instruments that have been previously analyzed and are considered valid and reliable to measure variables and their indicators, with the aim of collecting data from the population without changing these variables, so that researchers can research the actual situation [31][32][33]. This approach was chosen because it allows for accurate and representative data collection from a sample of the population.

The subject selection technique was carried out by purposive sampling. Purposive sampling is a sampling technique based on certain considerations, where a sample is selected because it knows the expected information or has specific characteristics that are relevant to the research [34][35][36]. This technique was chosen because it allows researchers to get samples that are most suiTable for the purpose of the research, namely selecting samples that have gained an understanding of the material of the Three-Variable Linear Equation System in class X of high school. This study involved 252 students who were categorized based on the Table below.

Table 1. Demographic Information

Demographics	Description	Code	Amount
Class	X	A	252
Age	14-15	B	1
	15-16	C	217
	17 years and over	D	34
Domicile	East Jakarta	E	238
	Depok	F	12
	Bekasi	G	2
Gender	Female	H	127
	Male	I	125

The researcher disseminated the instrument directly and through Google Form to several schools in East Jakarta, Depok, and Bekasi. Based on Table 1, the researcher divided the respondents into several categories listed and adjusted to demographics consisting of class, age, domicile, and gender. The respondents consisted of class X students aged 14 and above with a total of 127 women and 125 men.

2.2 Instrument

Data collection in this study is done in the form of tests. A test is an instrument used to measure students' ability to understand the material and its learning outcomes through steps such as preparing questions, sharing questions, supervising work, checking results, evaluating, and analyzing test results [37], [38]. The test in this study is prepared based on indicators of mathematical reasoning ability consisting of six questions with each indicator consisting of two essay questions related to the material of the Three-Variable Linear Equation System. Indicators of mathematical reasoning ability in this study include making conjectures, performing calculations based on certain rules or formulas, and checking the validity of an argument [39]. The scoring guidelines used in the mathematical reasoning ability test are as follows.

Table 2. Guidelines for Scoring Mathematical Reasoning Ability Tests

Indicator	Description	Scoring			
		3	2	1	0
Raise Allegations	Students can design assumptions and suspect various possibilities that could be a solution to the problem given	Convey the alleged solution clearly, correctly and and relate it to the problem given	Presents alleged solution clearly and correctly, but is unable to relate it to the problem given	Presenting allegations about a problem but the allegations given are incorrect	There was no answer, even though there was only a show of lack of understanding
Perform calculations based on certain rules or formulas	Students can read the questions and then use the formula well	Using mathematical calculations and formulas correctly	Doing mathematical calculations but using formulas is not perfect	Performing mathematical calculations using formulas but not precisely	Cannot perform mathematical calculations using formulas
Checking the Validity of an Argument	Students present proof of the truth of a statement based on known mathematical results	Shows the validity of a statement along with strong evidence	Shows the validity of a statement but is accompanied by weak evidence, provide incomplete evidence, evidence does not match the conclusion	Showing the validity of a statement but not being able to show proof	There was no answer, even though there was only a show of lack of understanding

Data source: [39]

In Table 2, it is explained that the highest score is 3 which means that students can answer questions according to the right steps and in accordance with the indicators of mathematical reasoning skills, and the lowest score is 0 which means that students are not able to answer the given questions.

The data analysis technique used in this study is using the Rasch model on Winsteps software. Question item validation uses two types of measurements, namely (1) construct validation, and (2) content validation. Construct validation is a validity process that describes the construction of an instrument from the aspects of compilation, framework, language arrangement, and etc. [40]. Content validity is the extent to which the test measures what is intended to be measured with precision and precision and evaluates its validity and validity through analysis by experts to ensure the measurement includes adequate and representative items to uncover the concept being measured [17][23][41].

The instruments, which have been compiled based on the indicators and their descriptions in Table 2, have gone through the process of construct validation of concepts by Mathematics Education lecturers and mathematics teachers before being used in collecting data on mathematical reasoning skills in high schools with instrument components guided by validation sheets that include aspects of concept suitability, component construction, and language arrangement. The results of the validation are.

(a)	(b)
For example, x , y, and z are three integers which add up to give the result 75. The first number is five more than the sum of the other two numbers. The second number is equal to $\frac{1}{4}$ of the sum of the other two numbers. What is the value of x?	Three whole numbers are known which when added together the result is 75. The first number is five more than the sum of the other two numbers while the second number is equal to $\frac{1}{4}$ of the sum of the other two numbers. Which is the largest of the three numbers?

Figure 1. (a) Questions Before Validation by Experts; (b) Questions After Validation by Experts

Based on Figure 1, it can be seen that the questions have changed in the core question section which adjusts to the mathematical reasoning ability indicators used by the researcher and the reasoning part is changed so that students can make mathematical models according to their ability to make guesses to get appropriate results.

After the data is obtained from the respondents, the validity of the content is carried out. Content validity is the extent to which the test measures what is intended to be measured with precision and precision and evaluates its validity and validity through analysis by experts to ensure the measurement includes adequate and representative items to uncover the concept being measured [17][23][41]. The data obtained is grouped in Microsoft Excel according to the code contained in Table 1, then the data is processed using Winsteps. The next stage is to eliminate zero responses and identify outliers that are then excluded from the data.

3. RESULTS AND DISCUSSION

This study aims to measure the validity and reliability of the instrument through analysis using Winstep software. In this effort, a number of data are collected and analyzed to ensure that the instruments used are capable of producing accurate and consistent measurements. Quantitative research through Winsteps can provide comprehensive data on item quality, person quality, and even interactions between respondents and items at once [24][42].

3.1 Validity

The validity of the instrument was carried out to test whether the mathematical reasoning ability instrument of class X students could be used to measure these abilities. The analysis of the validity of the instrument in Winsteps is called the fit and misfit test on items that can be seen from the values of Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), Point Measure Correlation (Pt Mean Corr). The purpose of validity is to evaluate the suitability of the question items with the model used, which is known as item fit [43]. The purpose of conducting this validity test is to determine whether an instrument is suiTable for

use or not [44] . Item limits are declared fit to the model if they meet one or both of the conditions that can be seen in Table 3 of the item suitability section of the question. Item fit means that the item is normal for the measurement it should be, while if the item shows a mismatch, it indicates that the respondent has a misunderstanding of the item [45] . Here are the results of the fit item test.

Table 3. Item Validity Analysis Results

Entry Number	Item	MNSQ	ZSTD	Pt Mean Corr
1	P1	1.38	3.6	0.38
2	P2	0.95	-0.4	0.66
3	P3	0.71	-2.6	0.67
4	P4	1.08	0.9	0.75
5	P5	0.92	-0.5	0.59
6	P6	0.73	-1.4	0.52

It can be seen that in Table 3, the valid items are all items. This is indicated by the Outfit MNSQ value for all items meeting the criteria, namely 0.5 < MNSQ < 1.5 [46]. Meanwhile, the Outfit ZSTD values for items P1 and P3 are outside the criteria, namely -2.0 < ZSTD < +2.0 contained in Ngadi [46], with the value of the ZSTD Outfit item P1 which is 3.6 and item 3 which is -2.6. For the correlation value of items with a total score (Pt Measure Corr), all items have met the criteria, namely 0.4 < Pt Measure Corr < 0.85 [46].

In general, the instrument has good validity with most items meeting the criteria of Outfit MNSQ and Pt Measure Corr. However, there are two items (P1 and P3) that show deviations in the value of the ZSTD Outfit, which requires further attention. Researchers may need to revise or re-evaluate those items to ensure they match the expected model.

Furthermore, a check was carried out on the person. Person fit is used when the criteria for MNSQ and ZSTD are met even though the total score (Pt Measure Corr) is very low. The information for person fit in Rasch's analysis is based on unusual response patterns. It is possible that the respondent is not serious in answering the question items in the instrument so that unusual patterns are detected. The following are the results of the person fit test.

Table 4. Misfit Order of The Person

No.	Person	MNSQ	ZSTD	Pt Mean Corr
1.	004ACEI	1.85	1.4	-0.04
2.	009ACEI	1.97	1.6	-0.12
3.	011ACEH	0.31	-1.5	0.91
4.	013ACEI	2.44	1.9	0.25
5.	015ACEH	2.09	1.3	-0.44
6.	016ACEH	2.09	1.3	-0.44
7.	017ACEI	**0.11**	**-2.5**	**0.97**
8.	026ACEI	1.56	1.0	0.14
9.	032ACEI	**3.65**	**2.1**	**0.05**
10.	035ADEI	0.16	-0.9	0.87
11.	050ACEH	0.30	-1.0	0.86
12.	053ACEH	1.96	1.0	-0.15
13.	070ACEH	**5.79**	**3.4**	**-0.14**
14.	077ACEH	0.28	-1.3	0.86
15.	089ACEI	1.82	1.3	0.29
16.	094ACEH	3.32	1.7	-0.14
17.	112ACEH	**5.68**	**4.2**	**0.01**
18.	116ACEI	0.16	-0.9	0.87
19.	118ACEI	0.29	-1.3	0.93
20.	123ACEH	0.35	-1.3	0.89
21.	124ACEH	1.61	1.0	0.36

No.	Person	MNSQ	ZSTD	Pt Mean Corr
22.	173ACEH	0.16	-0.9	1.87
23.	174ACEH	2.82	1.4	-0.30
24.	176ACEH	2.32	1.9	0.04
25.	178ACEI	**4.05**	**2.8**	**-0.12**
26.	180ADEI	1.95	1.3	-0.06
27.	190ACEH	2.44	1.3	0.05
28.	203ACFH	**2.44**	**2.1**	**0.08**
29.	210ACEH	0.19	-1.9	0.95
30.	212ACEH	0.28	-1.6	0.93
31.	213ACEH	0.24	-1.9	0.85
32.	214ACEH	0.18	-1.9	0.95
33.	224ACFH	0.31	-1.5	0.91
34.	235ACEH	0.24	-1.9	0.85
35.	241ACEH	**0.11**	**-2.2**	**0.97**
36.	242ACEI	-0.24	-1.9	0.85

Table 4 shows the responses of items that do not conform to the rules in the Rasch Model. The person fit Table shows that only 190 (75.4%) of the 252 respondents read the data. The results of the analysis showed that there were 36 (14.29%) person misfits with the same conditions as in the item analysis [46]. There were 7 people whose criteria did not comply with the provisions for MNSQ, ZSTD, and Pt Mean Corr, while the other persons did not comply with the provisions for MNSQ and Pt Mean Corr. This can be caused by data errors, unserious responses, or difficulties in understanding the item or item of the question. All three indicators (making a conjecture, performing calculations, checking the validity of an argument) may require a review to ensure each item is valid and precise in measuring the capabilities in question.

3.2 Reliability

The reliability of the instrument was carried out to see if this mathematical reasoning ability instrument could be used to reduce this ability in class X students anytime and anywhere. The reliability of the instrument was carried out to see if this instrument was reliable and could be used as a measure of the mathematical reasoning ability of class X students. One of the pieces of information that can be interpreted from the output of Summary Statistics is the value of measuring reliability, both from the respondent's side (Person Reliability) to see the consistency of the respondent's answers, as well as from the review of question items (Item Reliability) which shows the reliability of the question items, as well as the interaction between the respondents and the items in Alpha Cronbach [24], [26]. Here are the results of the Output Summary Statistic.

Table 5. Review of Output Summary Statistics

Statistics	Value
(KR-20)	0.74
Person Reliability	0.62
Item Reliability	0.99
Person Separation	1.29
Item Separation	9.89

Table 5 shows the values of Alpha Cronbach (KR-20), Person Reliability, Item Reliability, Person Separation, and Item Separation based on Rasch analysis in Winsteps. The results show that the Person Reliability value is 0.62 which means it is below 0.67 with the Weak category and the Person Separation value is 1.29 [46]. Item Reliability of 0.99 which belongs to the Special category because it is more than 0.94 with Item Separation 9.89 [46]. This research resulted in an Alpha Cronbach value of 0.74 which means it has a Good category according to Ngadi [46].

3.3 Differential Item Functioning (DIF)

Rasch modeling can detect biases that can be called DIF (differential item functioning) detection or grain functionality [45]. Furthermore, the Differential Item Functioning (DIF) analysis was carried out to identify the existence of bias in the research instrument in order to find out whether there is an instrument that tends to benefit one of the parties carried out after controlling the level of respondents' ability to answer a question item based on several existing groups [43][47].

Using Winsteps, researchers can identify items that exhibit bias and take the necessary steps to correct or replace those items, thereby improving the overall validity and fairness of the test. The question items on the instrument are said to be biased if there is one individual with certain characteristics who benefits more than an individual with other characteristics. A measurement is said to be valid when the instrument and question items do not contain bias. Here's the result of the DIF output on Winsteps.

Table 6. Result of Output DIF

No.	Item	DIF Measure	DIF Contrast	t	Prob
1	P1	-1.37	0.39	2.00	0.0750
2	P2	-0.35	-0.30	-1.47	0.2895
3	P3	0.22	-0.28	-1.22	0.4544
4	P4	-0.77	0.51	2.60	0.0705
5	P5	0.61	-0.40	-1.55	0.1571
6	P6	1.58	-0.82	-2.16	0.0096

Table 6 shows the DIF Measure, DIF Contrast, t-value, and probability values. The criteria for DIF analysis are a DIF Contrast value > 0.5 and a probability value < 0.05 [47]. It can be seen in Table 6 that there are no question items that contain bias. This is because each question item obtained a DIF Contrast value of < 0.5 which means that the difference in the function of the item between the analyzed groups is not significant and the probability value > 0.05 which indicates that the existing difference is not statistically significant. Therefore, it can be said that all the questions in this research instrument are proven to be valid.

4. CONCLUSIONS

Based on the results of the item and person fit tests, the mathematical reasoning ability instrument has several strengths and weaknesses in its validity. All items are valid based on the MNSQ Outfit and Pt Measure Corr criteria, indicating that the item generally matches the Rasch model and correlates well with the total score. The majority of respondents indicated a match with the Rasch model, supporting the overall validity of the instrument. Items P1 and P3 do not meet the ZSTD Outfit criteria, indicating that there are items that need review or revision to reduce non-conformities. A total of 36 (14.29%) respondents showed significant discrepancies, especially 7 persons who did not match the three main indicators (MNSQ, ZSTD, and Pt Mean Corr). This instrument can be considered valid as a whole because the majority of items meet the validity criteria specified by the Rasch model.

The reliability test of the mathematical reasoning ability instrument with Winsteps on students in high school shows that this instrument has excellent reliability of question items, with strong consistency in measuring mathematical reasoning ability and excellent ability to distinguish the difficulty level of the problem, as shown by the high value of Item Reliability and Item Separation. However, a low Person Separation score (< 2.0) indicates that this instrument is less effective in distinguishing learners' abilities, which may be caused by variability in respondents' answers due to different understandings of questions or variations in participants' performance. Although the Alpha Cronbach value shows adequate internal consistency, improvements are needed to improve the instrument's ability to measure the variation in mathematical reasoning ability among learners.

The DIF test can be concluded that there are no question items that contain bias. Each question item in this research instrument showed a DIF Contrast value < 0.5 and a probability value > 0.05. Therefore, no item functioned differently for the analyzed group, which means that this instrument proved to be valid and fair for all groups of respondents. Thus, these results provide confidence that the instruments used in

this study can be relied upon to measure mathematical reasoning ability without bias towards certain groups. This is important to ensure that the results of the study are accurate and fair, reflecting the mathematical reasoning abilities of the respondents objectively.

REFERENCES

[1] S. Suharti, S. Sulasteri, and H. Hairunnisa, "Analisis Kemampuan Penalaran dan Kemampuan Pemecahan Masalah Matematis Mahasiswa Pendidikan Matematika Ditinjau Dari Asal Sekolah," SJME (Supremum Journal of Mathematics Education), vol. 5, no. 1, pp. 11–21, Jan. 2021, doi: 10.35706/sjme.v5i1.4280.

[2] O. Romsih, Y. Yuhana, and H. Nindiasari, "Peningkatan Kemampuan Penalaran Matematis Siswa Melalui Problem Posing Ditinjau Dari Tahap Perkembangan Kognitif Siswa," SJME (Supremum Journal of Mathematics Education), vol. 3, no. 1, pp. 37–46, 2019, doi: http://dx.doi.org/10.35706/sjme.v3i1.1463.

[3] R. Pramujiyanti Khotimah and Hariyanti, "Kemampuan Penalaran Matematis Siswa dalam Menyelesaikan Soal Materi Bangun Ruang Sisi Datar Ditinjau dari Perbedaan Gender di Kelas VII SMP Negeri 1 Bendosari," Jurnal Pembelajaran Matematika Inovatif, vol. 5, no. 3, pp. 681–692, 2022, doi: 10.22460/jpmi.v5i3.681-692.

[4] Rohana and Y. L. Ningsih, "Peningkatan Kemampuan Penalaran Matematis Mahasiswa melalui Pembelajaran Reflektif Berbantuan Aplikasi Moodle," INDIKTIKA (Jurnal Inovasi Pendidikan Matematika), vol. 1, no. 2, pp. 134–143, 2019, doi: https://doi.org/10.31851/indiktika.v1i2.3034.

[5] Y. Venesia, A. Noornia, and T. Murdiyanto, "Upaya Meningkatkan Kemampuan Penalaran Matematis Siswa Menggunakan Pembelajaran Model Learning Cycle 7E (LC 7E) pada Pokok Bahasan Penyajian Data dan Peluang di Kelas X MIA 1 SMA Negeri 9 Jakarta," pp. 29–36, 2017.

[6] Y. Aprilianti and L. S. Zanthy, "Analisis Kemampuan Penalaran Matematik Siswa SMP pada Materi Segiempat dan Segitiga," Journal On Education, vol. 1, no. 2, pp. 524–532, 2019.

[7] D. K. Putri, J. Sulianto, and M. Azizah, "Kemampuan Penalaran Matematis Ditinjau dari Kemampuan Pemecahan Masalah," International Journal of Elementary Education, vol. 3, no. 3, pp. 351–357, 2019, [Online]. Available: https://ejournal.undiksha.ac.id/index.php/IJEE

[8] R. Aulya and J. P. Purwaningrum, "Pengaruh Model Pembelajaran PBL Berbantuan Alat Peragam dalam Peningkatan Kemampuan Penalaran Matematis," JURNAL MathEdu (Mathematic Education Journal), vol. 4, no. 3, pp. 401–406, 2021.

[9] Y. Ria, D. Risalah, and Sandie, "Kemampuan Penalaran Matematis Siswa dalam Menyelesaikan Soal Higher Order Thinking Skills (HOTS) pada Materi Teorema Phyytagoras Siswa Kelas VIII SMP Negeri 2 Moterado," Journal of Innovation Research and Knowledge, vol. 1, no. 5, pp. 767–772, 2021.

[10] P. A. Cahyani, "Kemampuan Penalaran Matematis Peserta Didik melalui Model Pembelajaran Reciprocal Teaching dengan berbantuan Alat Peraga 'Kartu Pintar' pada Materi Prisma dan Limas kelas VIII SMP Negeri 1 Pandaan Tahun Ajaran 2017/2018," JPM : Jurnal Pendidikan Matematika, vol. 5, no. 1, pp. 27–34, 2019.

[11] A. Iswanto and A. Faradillah, "Mosharafa: Jurnal Pendidikan Matematika Analysis of Mathematical Reasoning Ability in Trigonometry Materials Viewed from Students' Mathematical Resilience," Mosharafa : Jurnal Pendidikan Matematika, vol. 12, no. 3, 2023, [Online]. Available: http://journal.institutpendidikan.ac.id/index.php/mosharafa

[12] M. M. Amin, S. Prabawanto, and B. A. P. Martadiputra, "Peningkatan Kemampuan Penalaran Matematis Siswa dengan Metode Pembelajaran Project-Based Learning," AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 12, no. 2, p. 1873, Jun. 2023, doi: 10.24127/ajpm.v12i2.6550.

[13] E. Febrina Tarigan et al., "Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0," JIPK, vol. 16, no. 2, 2022, [Online]. Available: http://journal.unnes.ac.id/nju/index.php/JIPK

[14] Azizah and S. Wahyuningsih, "Penggunaan Model Rasch untuk Analisis Instrumen Tes pada Mata Kuliah Matematika Aktuaria," JUPITEK, vol. 3, no. 1, pp. 45–50, Jun. 2020, doi: 10.30598/jupitekvol3iss1ppx45-50.

[15] S. Nuryanti, M. Masykuri, and E. Susilowati, "Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan," Jurnal Inovasi Pendidikan IPA, vol. 4, no. 2, pp. 224–233, Oct. 2018, doi: 10.21831/jipi.v4i2.21442.

[16] Sugiono, Noerdjanah, and A. Wahyu, "Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation," Jurnal Keterapian Fisik, vol. 5, pp. 1–61, 2020.

[17] H. Hendryadi, "Validasi Isi : Tahap Awal Pengembangan Kuesioner," Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT, vol. 2, no. 2, pp. 169–178, Jun. 2017, doi: 10.36226/jrmb.v2i2.47.

[18] A. Faradillah and S. Adlina, "Validity of critical thinking skills instrument on prospective Mathematics teachers," Jurnal Penelitian dan Evaluasi Pendidikan, vol. 25, no. 2, Dec. 2021, doi: 10.21831/pep.v25i2.40662.

[19] K. Hidayati and Z. E. K. Nisa', "Analisis Butir Soal Penilaian Akhir Semester Matematika," AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 12, no. 3, p. 3516, Sep. 2023, doi: 10.24127/ajpm.v12i3.7575.

[20] R. Oktaviyanthi and R. N. Agus, "Instrumen Evaluasi Kemampuan Penalaran Adaptif Matematis Mahasiswa," AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 9, no. 4, pp. 1123–1136, Dec. 2020, doi: 10.24127/ajpm.v9i4.3150.

[21] A. Faradillah and C. Septiana, "Mathematical Resilience : Validity and Reliability with Rasch Model and SPSS in Senior High School," AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 11, no. 4, p. 3545, Dec. 2022, doi: 10.24127/ajpm.v11i4.5204.

[22] A. Fauziyah, Z. A. Sakinah, Mariyanto, and D. E. Juansah, "Instrumen Tes dan Non Tes pada Penelitian," Pendas : Jurnal Ilmiah Pendidikan Dasar, vol. 08, Dec. 2023.

[23] V. C. Fatwa, A. Septian, and S. Inayah, “Kemampuan Literasi Matematis Siswa melalui Model Pembelajaran Problem Based Instruction,” Mosharafa: Jurnal Pendidikan Matematika , vol. 8, no. 3, 2019, [Online]. Available: http://journal.institutpendidikan.ac.id/index.php/mosharafa
[24] Muntazhimah, S. Putri, and H. Khusna, “Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika,” JKPM : Jurnal Kajian Pendidikan Matematika, vol. 6, pp. 65–74, 2020, [Online]. Available: http://journal.lppmunindra.ac.id/index.php/jkpm/
[25] A. Anggraini and M. Muntazhimah, “Pengembangan Instrumen Kemampuan Berpikir Reflektif Matematis Siswa Madrasah Aliyah,” AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 10, no. 4, p. 2465, Dec. 2021, doi: 10.24127/ajpm.v10i4.4223.
[26] D. R. Ocy, W. Rahayu, and M. Makmuri, “Rasch Model Analysis : Development of HOTS Based Mathematical Abstraction Ability Instrument According to Riau Islands Culture,” AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 12, no. 4, pp. 3542–3560, Dec. 2023, doi: 10.24127/ajpm.v12i4.7613.
[27] T. N. Efendi, K. Kartini, and R. D. Anggraini, “Pengembangan Instrumen Tes Kemampuan Penalaran Matematis pada Materi Barisan dan Deret Kelas XI SMA/MA,” Jurnal Cendekia : Jurnal Pendidikan Matematika, vol. 8, no. 1, pp. 811–826, Mar. 2024, doi: 10.31004/cendekia.v8i1.2650.
[28] R. Ramadhani and Y. Fitri, “Validitas E-Modul Matematika Berbasis EPUB3 Menggunakan Analisis Rasch Model,” Jurnal Gantang, vol. 5, no. 2, pp. 95–111, Sep. 2020, doi: 10.31629/jg.v5i2.2535.
[29] I. Arimurti, E. S. Praja, and F. Muhtarulloh, “Desain Modul Berbasis Model Discovery Learning untuk Kemampuan Pemahaman Matematis Siswa,” Mosharafa: Jurnal Pendidikan Matematika , vol. 8, no. 3, 2019, [Online]. Available: http://journal.institutpendidikan.ac.id/index.php/mosharafa
[30] A. Putri, D. Sumardani, W. Rahayu, M. N. Hajizah, and A. Rahman, “Kemampuan Literasi Matematika menggunakan Bar Model pada Materi Aljabar,” AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 9, no. 2, Jun. 2020, doi: 10.24127/ajpm.v9i2.2744.
[31] I. Zulkarnain, “Pengaruh Kemampuan Awal terhadap Prestasi Belajar Matematika Siswa,” Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara, vol. 11, no. 2, pp. 88–94, 2020, doi: https://doi.org/10.37640/jip.v11i2.94.
[32] Maidina, “Penelitian Survey,” ALACRITY : Journal of Education, vol. 1, no. 2, pp. 20–29, 2021.
[33] P. S. Balkist, D. Dasari, and P. Fitriasari, “Analisis Pengalaman Pengembangan Diri Guru Matematika terhadap Pembelajaran yang Berdiferensiasi dan Mendorong Berpikir Kritis,” AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 12, no. 1, p. 1297, Mar. 2023, doi: 10.24127/ajpm.v12i1.6829.
[34] F. Chan, A. R. Kurniawan, S. Kalila, F. Amalia, D. Apriliani, and S. V. Herdana, “The Impact of Bullying on the Confidence of Elementary School Student,” Jurnal Pendas Mahakam, vol. 4, no. 2, pp. 152–157, 2019.
[35] D. N. Muna and E. A. Afriansyah, “Peningkatan Kemampuan Pemahaman Matematis Siswa melalui Pembelajaran Kooperatif Teknik Kancing Gemerencing dan Number Head Together,” Mosharafa: Jurnal Pendidikan Matematika , vol. 5, no. 2, 2016, [Online]. Available: http://e-mosharafa.org/Jurnal"
[36] W. Lestari, T. A. Kusmayadi, and F. Nurhasanah, “Kemampuan Pemecahan Masalah Matematika Ditinjau dari Perbedaan Gender,” AKSIOMA: Jurnal Program Studi Pendidikan Matematika, vol. 10, no. 2, p. 1141, Jul. 2021, doi: 10.24127/ajpm.v10i2.3661.
[37] Yenni and R. Sukmawati, “Analisis Kemampuan Representasi Matematis Mahasiswa Berdasarkan Motivasi Belajar,” Mosharafa: Jurnal Pendidikan Matematika , vol. 9, no. 2, 2020, [Online]. Available: http://journal.institutpendidikan.ac.id/index.php/mosharafa
[38] D. N. Antika and E. Hastuti, “Pengaruh Pembelajaran Inkuiri Melalui Media Lingkungan Sebagai Sumber Belajar di Sekolah Dasar,” JTL : Journal of Teaching and Learning Research, vol. 2, no. 1, pp. 17–22, 2020, doi: 10.24256/jtlr.v2i1.1333.
[39] W. N. Rohmah, A. Septian, and S. Inayah, “Analisis Kemampuan Penalaran Matematis pada Materi Bangun Ruang Ditinjau dari Gaya Kognitif Siswa SMP,” PRISMA, vol. 9, no. 2, pp. 179–191, 2020, doi: https://doi.org/10.35194/jp.v9i2.1043.
[40] N. R. Nengsih, E. Yusmaita, and F. Gazali, “Evaluasi validitas konten dan konstruk bahan ajar asam basa berbasis REACT,” EKJ : EduKimia, vol. 1, no. 1, pp. 1–10, 2019.
[41] H. Ihsan, “Validitas Isi Alat Ukur Penelitian Konsep dan Panduan Penilaiannya,” PEDAGOGIA : Jurnal Ilmu Pendidikan, pp. 266–273, 2015.
[42] I. Nurjanah and F. Alyani, “Kecemasan Matematika Siswa Sekolah Menengah pada Pembelajaran Matematika dalam Jaringan,” Jurnal Elemen, vol. 7, no. 2, pp. 407–424, Jul. 2021, doi: 10.29408/jel.v7i2.3522.
[43] C. L. Zibar Parisu, W. Ekadayanti, L. Sisi, A. Juwairiyah, and Kasmawati, “Analisis Butir Soal Pengetahuan Dasar Matematika Menggunakan Pendekatan Rasch,” SCIENCE TECH : Jurnal Ilmu Pengetahuan dan Teknologi, vol. 10, no. 1, pp. 36–45, 2024, doi: 10.30738/st.
[44] L. Sari, Hadiyanto, D. Arif, and F. F, “Validitas LKPD berbasis Model Project Based Learning Pembelajaran Tematik di Kelas V Sekolah Dasar,” Jurnal Cakrawala Pendas, vol. 8, no. 4, pp. 1358–1370, 2022, doi: 10.31949/jcp.v8i2.3215.
[45] O. : Lilla, S. Fakultas, I. Tarbiyah, D. Keguruan, U. Sunan, and K. Yogyakarta, “Analisis Item Soal dengan menggunakan Rasch Model sebagai Ukuran Kualitas Madrasah Ibtidaiyah pada Mata Pelajaran IPA,” PIONIR: JURNAL PENDIDIKAN, vol. 12, no. 2, pp. 1–12, 2023.
[46] Ngadi, “Analisis Model Rasch untuk Mengukur Kompetensi Pengetahuan Siswa SMKN 1 Kalianget pada Mata Pelajaran Perawatan Sistem Kelistrikan Sepeda Motor,” Jurnal Pendidikan Vokasi Otomotif, vol. 6, no. 1, pp. 1–20, 2023.
[47] H. Mulyono, S. Kusuma Ningsih, and B. Sholeha Raufi, “Validating the Academic Writing Creativity and Self-Efficacy Scale: A Rasch Model Analysis,” IRJE |Indonesian Research Journal in Education| |Vol, doi: 10.22437/irje.