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.21-30, Edisi November 2024

Validity and Reliability Survey of Mathematical Reasoning Ability of Senior High School Students Using Statistical Applications

Dinda Triasti[1], Qathrunnada Farhanah[2], Sofia Rengganis Santosa[3], Ayu Faradillah[4*], Syafika Ulfah[5]

[1,2,3,4,5]Mathematics Education, 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;
Reasoning;
Statistics application

ABSTRACT

In Indonesia, students' reasoning ability is still very low, especially in mathematical reasoning ability. It was found that the data of mathematical reasoning ability test results of class X students where only 47.22% reached the minimum learning completeness. Reasoning ability also plays an important role in the process of thinking creatively, logically, and systematically so that students can conclude solutions from several known facts as well as proof of mathematical statements, find new ideas, overcome challenges in solving mathematical problems, and reduce dependence on memorization. This study aims to determine the validity and reliability in mathematical reasoning ability of high school students using two statistical applications, namely SPSS and Winstep. The research method used is a quantitative research method with a survey approach, which has a total of 252 respondents from class X high school students in Jakarta. The results of the validity test research using SPSS have 6 valid items so that they are in accordance with the criteria. While the results of the validity test using Winstep there are seven items that do not match the three, and there are twenty-nine items that only meet one requirement, so 154 items must be maintained because they are fulfilled. Meanwhile, the results of reliability calculations using SPSS and Winstep obtained the results (α = 0,74). By using two different statistical applications, namely SPSS and Winstep, it can be stated that the results of the mathematical reasoning ability test have been declared valid and reliable.

DOI: https://doi.org/10.30598/ppcst.2024.knmxxii.21-30

1. INTRODUCTION

Reasoning skills are the skills that students need to evaluate new situations, draw conclusions, express ideas, and make logical assumptions, while mathematical reasoning skills are an integral part of the mathematical thinking process that is central to the school curriculum [1]. In the learning of mathematics, the development of mathematical reasoning skills is crucial for the improvement of educational standards. The importance of the development of these skills is due to the focus of mathematics on the process of reasoning, and not just on the results of experiments or observations [2]. Mathematical reasoning skills enable students to engage in the creative process of developing innovative ideas, drawing conclusions, and proving statements. Furthermore, these skills facilitate the solution of a wide range of mathematical problems [3]. From the definition of reasoning and mathematical reasoning, mathematical thinking skills can be defined as the process of connecting various known facts to reach a conclusion. Therefore, it can be concluded that this is the essence of mathematical reasoning skills [4][5]. To conclude then, the capacity to reason mathematically can be defined as the ability to draw conclusions from accepted mathematical principles by attending to the relationships between these principles [6]. The capacity to reason mathematically, employing logic and a structured methodology to address mathematical problems, is the fundamental definition of mathematical reasoning.

The capacity to reason effectively facilitates students' ability to navigate a range of problems encountered in their everyday lives [7]. Consequently, the capacity to employ mathematical reasoning is a crucial element in ensuring that the educational objectives of mathematics are attained. Furthermore, mathematical reasoning ability plays an important role in the development of creative, logical, and systematic thinking. It enables students to derive solutions from a combination of known facts and mathematical proofs, to generate new ideas, overcome challenges in mathematical problem solving, and to reduce dependence on memorization [3][8][9].

With regard to the significance of reasoning ability, it is evident that students' mathematical reasoning ability remains suboptimal. The findings of a recent study indicate that [1], A total of 32 students in class XII were observed to be engaged in trigonometry problems. Of these, 6.25% demonstrated an understanding of the material, while 93.75% exhibited a lack of comprehension. The observations conducted at one of the high schools in Padang City indicate that students in grade X exhibit deficiencies in their mathematical reasoning abilities. This is evidenced by the initial test results, which indicate that only 47.22% of students achieved or surpassed the minimum learning completeness (MLC) benchmark. Consequently, the potential for students to develop their mathematical reasoning abilities is constrained [10]. In order to ascertain the extent of students' mathematical reasoning abilities, it is necessary to utilise an instrument designed to assess these abilities. Consequently, in order to ascertain the veracity of the results obtained from the mathematical reasoning ability instrument, it is necessary to conduct a process of validity and reliability testing.

Validity is the degree to which an instrument or measurement tool is suitable and accurate in estimating or representing the variables or concepts that are being investigated in a study. Validity is the extent to which an instrument or method of measurement accurately reflects the intended variables or concepts under investigation [11][12]. Validation is an essential element in the evaluation of the extent to which a measurement tool can represent the variable or concept being studied. The greater the accuracy of the measurement instrument in measuring data, the higher the instrument's validity. It is crucial to administer this validity test to ascertain that responses to questions do not result in data that differs from the intended definition of the variable [13]. If the conditions of validity are not fulfilled, the research results derived from the measurement instrument are rendered irrelevant and inaccurate in relation to the research or measurement objectives.

Furthermore, reliability is defined as a test that gauges the dependability and integrity of a measuring instrument. Reliability is a quantitative index test that assesses the extent to which a measuring instrument can be trusted or reliable [6][14]. In this context, reliability can be defined as the extent to which a measurement tool or test produces consistent and reliable results over time and in the context of different situations. A test is considered highly reliable if it consistently yields the same results, regardless of whether it is administered to the same subject on multiple occasions. This reliability can be assessed by measuring the consistency of results produced by a measurement tool when repeated within the same context [15]. In

the absence of reliability, the results obtained from measurement instruments may be inconsistent and inaccurate. Consequently, a reliable measurement tool must be both valid and reliable.

The importance of validity and reliability in a study was acknowledged, prompting the use of two statistical applications, namely SPSS (Statistical Product and Service Solutions) and Winsteps, to test for these qualities. In the context of data analysis, SPSS serves as a crucial tool for providing extensive access to data. SPSS is a software program that offers exceptional capabilities for statistical analysis, facilitating precise and accurate data analysis [16][17]. Winsteps is a Windows application that supports Rasch model computations. Its applications include educational evaluation, attitude surveys, and scale analysis. [18][19].

Winsteps and SPSS were selected as the statistical software for this study because they possess the necessary capabilities to achieve the research objectives [19]. The SPSS statistical software can be employed to read data in various formats or to permit the input of data directly via the SPSS Data Editor. Furthermore, SPSS facilitates the interpretation of the results. Winsteps is advantageous due to its ability to make predictions on missing data based on a systematic response pattern, which results in accurate measurement results [20].

Therefore, the selection of Winsteps and SPSS as statistical applications in this study is based on previous relevant studies, such as those conducted by [21], related to testing the instrument of mathematical communication skills on the material of relations and functions by calculating the validity and reliability using statistical applications, namely SPSS, resulting in validity of more than 0.3 by construct and meeting a high reliability of 0.740 so that the test instrument that has been designed to measure students' mathematical communication skills on the material of relations and functions can be used effectively. Furthermore, with reference to other relevant research conducted by [17], took the theme of testing students' critical thinking skills in teaching materials by calculating validity and reliability using two statistical applications, namely Excel and SPSS, which obtained valid and reliable results on teaching materials. Thus, it is feasible to use in the learning process to improve students' critical thinking skills.

Furthermore, other researchers [22] discussed mathematical reasoning ability on ethno-RME by calculating validity and reliability using SPSS which obtained high statistical testing results with a validity level of 0.374 and reliability of 0.729 thus, this instrument has proven that this research is considered relevant and accurate. Research conducted by [23] also used the SPSS application to test the validity and reliability which resulted in 10 valid statements and a high reliability score of 0.84 on the ethnomathematics problem solving ability instrument.

The research gap related to the validity and reliability of mathematical reasoning ability remains unresolved at the senior high school level. Some studies have been conducted at the junior high school and college levels. Moreover, previous studies have indicated that it is challenging to assess the validity and reliability of mathematical reasoning ability using two distinct systems, namely SPSS and Winsteps. Furthermore, the objective of this study is to assess the validity and reliability of mathematical reasoning ability in high school students by employing two statistical software programs, namely SPSS and Winsteps. The problem formulations of this study are (a) How are the results of the validity and reliability of mathematical reasoning ability using SPSS? and (b) How are the results of the validity and reliability of mathematical reasoning ability using Winsteps?

2. RESEARCH METHODS

2.1 Design Research and Participants

This research uses quantitative research methods with a survey approach. Survey is a type of research that collects data on the characteristics, behavior, and opinions of a group of respondents who are considered as a population [24]. The survey method in quantitative research is a positivist approach in social science that aims to collect data from sources or research informants through empirically based observations and interviews. This approach uses instruments such as questionnaires, structured interviews, and documentation. The collected data is analyzed quantitatively or statistically to test the hypotheses that have

been formulated previously [24][25]. In this research, data is obtained through the use of specialized instruments to evaluate the variables under study in a structured and systematic way.

The subject selection technique used was purposive sampling. Purposive sampling is a non-random sampling technique in which the researcher selects a sample selectively based on specific characteristics that are in accordance with the research objectives and are expected to provide answers to research questions [26]. The researcher set specific criteria for this study, namely X grade high school students who have mathematical abilities, especially to students who have studied the material of the system of linear equations of three variables and are willing to take the mathematical reasoning ability test. Based on these criteria, it is able to make the basis for researchers in selecting samples that are in accordance with the research objectives, namely testing the validity and reliability of the mathematical reasoning ability test instrument and is considered to be representative of the population. To ensure that the results of the validity and reliability testing can be generalized, it is necessary to ensure that the sample size is representative.

Table 1. Demography Participant

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

Table 1 contains information about the demographics of the research participants. The demographics include several aspects such as class, age, domicile, and gender. The Table provides a clear picture of the demographic background of the participants involved in the study, helping to understand their distribution and characteristics.

2.2 Instrument

The instruments used in this study include the Test of your mathematical reasoning ability on the topic of the system of linear equations of three variables. The terms and language used in the test were adapted to the education relevant to the topic. The test structure consists of six essay questions on the system of linear equations of three variables. The test assesses three mathematical reasoning skills, namely making conjectures, performing calculations based on certain rules or formulas, and checking the validity of an argument [27]. The scoring guidelines for mathematical reasoning ability instruments for each question used in this study are adjusted to the indicators of each mathematical reasoning ability question.

Table 2. Indicator of Mathematical Reasoning Ability

Question Number	Indicator of Mathematical Reasoning Ability	Score	Criteria
1 and 2	Making conjectures	3	Present the solution hypothesis clearly and precisely and relate it to the problem being discussed.
		2	Presents solution ideas clearly and accurately, although not yet able to relate them to the problem being discussed.
		1	Presents an analysis of the problem, but the predictions are not accurate.
		0	No answer is given.
3 and 4	Perform calculations based on certain rules or formulas	3	Perform math calculations accurately and apply formulas appropriately.
		2	Performs mathematical calculations using formulas but not yet optimal.
		1	Perform mathematical calculations using formulas but

Question Number	Indicator of Mathematical Reasoning Ability	Score	Criteria
			inaccurately.
		0	Unable to perform math calculations using formulas.
5 and 6	Checking the validity of an argument	3	Affirming the truth of a statement with strong evidence.
		2	Asserts the truth of a statement, but uses evidence that is less strong, incomplete, or irrelevant to the conclusion.
		1	Asserts the truth of a statement but cannot provide adequate evidence.
		0	No answer given.

Source: [27]

Based on Table 2, it explains the indicators of mathematical reasoning ability of participants based on research by [27]. This Table categorizes mathematical reasoning ability into indicators, related questions, maximum achievable scores, and scoring criteria. In addition, it is used to evaluate students' mathematical reasoning ability through a series of questions with different scores, depending on the clarity and accuracy of their answers. This evaluation is important to understand the extent to which students can make conjectures, perform rule-based calculations, and check the correctness of arguments in a mathematical context.

This instrument went through a content validity test process. Content validity is an important basis for checking construct validity because it focuses on the components of the measurement instrument and their relevance to the research objectives [28]. Content validity provides information about the clarity of the items and the ability of the instrument to achieve the objectives through expert evaluation and recommendations. The validation process involved two mathematics education lecturers and one mathematics teacher, which produced good validation results and was suitable for mathematical reasoning ability research. The validation results showed a difference between the conditions before and after validation.

Before Validation	After Validation
Suppose x, y, and z are three integers whose sum has 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?	You know three integers that when added together equal 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. What is the value of these three numbers?

Figure 1. Question before validation and after validation

Based on Figure 3 after validation, the researcher was asked to improve the sentence by the validator. Furthermore, the sentence before validation which is listed at the end of the sentence is asked to find the value of x only, while after validation it is asked to find the value of the three numbers.

3. RESULTS AND DISCUSSION

3.1 Validity Test Using SPSS

The results of the quantitative study calculations were obtained through the help of the Statistics application, SPSS. The results of the bivariate analysis on SPSS can provide an overview of the validity of the tool for testing mathematical resilience [29]. The efficacy of mathematical reasoning questions was gauged via a study involving 252 respondents, who employed the Corrected Item to Total Correlation method. This entailed correlating each item score with the total score [30].

In this validity test, the correlated item total correlation value is also called r count, with the validity decision criteria if r count> t Table is declared valid and if r count < r Table is declared invalid. In Table 4 it is explained that the r count of the relationship between item 1 and the total score is 0.566. R Table with

df (N 2) is 0.124. It can be seen that, r count> t table, it can be said that this has a relationship. Also seen from Sig. its >0.5. Then for item 2 – item 6, they are valid because they meet Sig. > 0,5.

Table 3. Validity Test Using SPSS

Item	Pearson Correlation
I1	0.566
I2	0.685
I3	0.746
I4	0.766
I5	0.660
I6	0.599

3.2 Reliability Test Using SPSS

A reliability test analysis can be conducted using the SPSS software. This can be accessed by selecting the menu option "Analyze" and then "Scale" and finally "Reliability Analysis." The resulting SPSS output will display the Cronbach's Alpha value, which is used to determine the reliability of the test. A value of Cronbach's Alpha > 0.60 is indicative of a reliable test. [30]. An instrument is considered reliable if its Alfa Cronbach reliability coefficient is greater than 0.70. If the reliability coefficient falls below 0.70, it is recommended that question items exhibiting low reliability values be revised or eliminated [31]. Table 5 indicates a Cronbach's Alpha value of 0.743 for the research instrument score, which exceeds the threshold of 0.50, indicating high reliability. This result indicates that the mathematical reasoning ability instrument employed in this study meets the established criteria for reliability [29]. Table 4 presents the results of the reliability analysis, which indicate that the resulting question item Cronbach's Alpha value is 0.743. These results indicate that the Cronbach's Alpha value is greater than 0.7 and less than 0.9, thereby establishing the reliability of the question items [32].

Table 4. Reliability Test on SPSS

Cronbach's Alpha	N of Items
0.743	6

3.3 Validity Test Using Winsteps

3.3.1 Item Fit

In the validity analysis, it is carried out on question items and respondents with the aim of knowing and assessing the accuracy and relevance of the instrument in measuring the desired variables, ensuring that the results obtained reflect the actual reality. If an item meets the criteria for Outfit Mean Square Values (MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PTMEACORR), or perhaps two of the three criteria, then the item can be considered suitable. Any item that does not comply with the MNSQ, outfit ZSTD, or PTMEA-CORR criteria must be repaired or refurbished to ensure the quality and conformity of the item [29].

Table 5. Fit Item Validity Test Results

Item	Outfit MNSQ	Outfit ZSTD	PT-MEASURE CORR
P1	1.38	3.6	0.63
P2	0.95	-0.4	0.66
P3	0.71	-2.6	0.67
P4	1.08	0.9	0.75
P5	0.92	-0.5	0.59
P6	0.73	-1.4	0.52

Based on Table 5 of the validity test results on items, it is found that there are 2 questions or items that do not meet the Z-Standard Outfit (ZSTD) requirements, namely on question items P1 and P3 with

values of 3.6 and -2.6, respectively. The calculation results are outside the limits of -2.0 to 2.0 [33]. Meanwhile, items P4, P5, P2, and P6 are valid because they have met the three requirements of Winsteps. However, the other testing criteria have met the requirements in Winsteps; validity is met if at least each item has met 2 requirements. From the test results, it was found that overall, 6 items have met at least 2 requirements so that from these results it can be concluded that all question items are valid or meet the validity test.

3.3.2 Person Fit

Screening using Winsteps provides accurate respondent data. Winsteps is able to identify appropriate or accurate respondents based on unusual response patterns. For example, the model can detect students who are not serious in answering questions. It is possible that students with lower grades can unexpectedly answer correctly, as well as students with higher grades [29]. This means that respondents who have a personal fit will only successfully answer questions if the difficulty level of the question is lower than their level of competence. The following are the test results using person fit.

Table 6. Misfit Order of Person Fit

No.	Person	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR
1.	070ACEH	5.79	3.4	-0.14
2.	112ACEH	5.68	4.2	0.01
3.	178ACEI	4.05	2.8	-0.12
4.	032ACEI	3.65	2.1	0.05
5.	203ACFH	2.44	2.1	0.08
6.	017ACEI	0.11	-2.2	0.97
7.	241ACEH	0.11	-2.5	0.97
8.	094ACEH	**3.32**	1.7	**-0.14**
9.	174ACEH	**2.82**	1.4	**-0.30**
10.	013ACEI	**2.44**	1.9	**0.63**
11.	190ACEH	**2.44**	1.3	**0.05**
12.	176ACEH	**2.32**	1.9	**0.04**
13.	015ACEH	**2.09**	1.3	**-0.44**
14.	016ACEH	**2.09**	1.3	**-0.44**
15.	009ACEI	**1.97**	1.6	**-0.12**
16.	053ACEH	**1.96**	1.0	**-0.15**
17.	180ADEI	**1.95**	1.3	**-0.06**
18.	089ACEI	**1.82**	1.3	**0.29**
19.	124ACEH	**1.61**	1.0	**0.36**
20.	004ACEI	**1.85**	1.4	**-0.04**
21.	026ACEI	**1.56**	1.0	**0.14**
22.	173ACEH	**0.16**	-0.9	**0.87**
23.	116ACEI	**0.16**	-0.9	**0.87**
24.	035ADEI	**0.16**	-0.9	**0.87**
25.	214ACEH	**0.18**	-1.9	**0.95**
26.	210ACEH	**0.19**	-1.9	**0.95**
27.	242ACEI	**0.24**	-1.9	**0.85**
28.	235ACEH	**0.24**	-1.9	**0.85**
29.	213ACEH	**0.24**	-1.9	**0.85**
30.	212ACEH	**0.28**	-1.6	**0.93**
31.	077ACEH	**0.28**	-1.3	**0.93**
32.	118ACEI	**0.29**	-1.3	**0.86**
33.	224ACEH	**0.31**	-1.5	**0.91**
34.	011ACEH	**0.31**	-1.5	**0.91**
35.	050ACEH	**0.30**	-1.0	**0.86**
36.	123ACEH	**0.35**	-1.3	**0.89**

Table 6 shows the calculation results of oufit MNSQ, outfit ZSTD and PTMEA-CORR for non-conforming requests. Table 6 shows significant evidence that the calculation results do not meet the set requirements. It was found that seven (3,11%) persons did not meet three conditions which are invalid. In addition, 29 (12,89%) persons only fulfills one condition, while the other 154 persons should be retained because they are in accordance with the MNSQ outfit, ZSRD outfit, and PTMEA-CORR.

3.4 Reliability Test Using Winsteps

The reliability test aims to evaluate the consistency and stability of the measurement results when the instrument is used under the same conditions repeatedly. If the value of the Spearman-Brown formula reaches 0.70 or more, it indicates that the scale can be considered reliable. Otherwise, the value is in the range between 0 to 1 [34]. Nonetheless, there is a significant correlation between the reliability coefficient and the number of items in the scale. Here are the results of the Winsteps reliability test noted below.

Table 7. Reliability Test Results Using Winsteps

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

Table 7 indicates that the results of Cronbach's Alpha (KR-20) have a Cronbach's alpha coefficient of 0.74, which is indicative of a high level of reliability. A reliability value below 0.67 is indicative of low reliability. Consequently, the results of the reliability test yielded a reliability value of 0.62, indicating a weak reliability. Moreover, the reliability of items with a value of 0.99 is included in the very good category. The person separation value is 1.29, while the item separation value is 9.89. A separation value that is high indicates that the instrument is of good value [29]. The analysis conducted with Winsteps demonstrated that the instrument exhibited satisfactory reliability, substantial consistency, and an elevated Cronbach's alpha value (KR-20). A strong correlation between the various item statements was observed, thereby providing support for the high level of reliability of this instrument.

The results of the data analysis, conducted using statistical applications such as SPSS and Winsteps, yielded slightly different outcomes. The results of the validity analysis conducted with the SPSS software indicate that the six items included in the mathematical reasoning ability questionnaire are indeed valid. In contrast, the Winsteps calculations indicate that two items do not align with the ZSTD outfit. However, the six items that comprise the calculation results meet at least two conditions. Consequently, the results of the item fit validity tests are deemed valid. In contrast, the results of the person fit calculations indicate that 29 individuals exhibited misfit and 154 exhibited fit according to the third requirement of Winsteps. This indicates that the results of validity calculations using Winsteps are more accurate and detailed in determining invalid items and persons. Moreover, there was no significant difference in the reliability calculation results between SPSS and Winsteps (α=0.74).

The strengths and weaknesses of this study are reflected in the subsequent discussion and explanation of the findings, which are set out above. One of the advantages of this study is the use of two statistical programs, Winsteps and SPSS, to conduct validity and reliability analysis. This study offers an in-depth comparison of the results of validity and reliability analysis by utilizing a variety of applications. However, a limitation of this study is that it focuses only on grade X high school students. Consequently, the instrument must undergo a further validation and reliability process should it be employed by other researchers to assess mathematical reasoning abilities.

4. CONCLUSIONS

A calculation of the validity and reliability of the mathematical reasoning ability test using Winsteps and SPSS revealed the presence of six valid items in the validity test results calculated using SPSS, as evidenced by the Pearson correlation value. Subsequently, the results of the validity test calculation on Winsteps according to item fit were found to be six valid items. However, according to person fit, 29 items only met one condition, while 154 items were declared suitable because they met all three calculation conditions. Meanwhile, the reliability of the instrument was calculated using both statistical applications, resulting in a very high category, namely (0.74). This study compares the calculation results of using Winsteps and SPSS software in calculating the validity and reliability of mathematical reasoning ability instruments. It is also recommended that the results of calculations in other studies that use statistical applications other than Winsteps and SPSS be compared. In addition to employing disparate applications, research can assess the outcomes of validity and reliability calculations on other instruments that may influence mathematical learning.

REFERENCES

[1] S. Agustiani, I. P. Dian, R. Fitriani, A. Faradillah, and F. Alyani, "Kemampuan penalaran siswa SMA dalam menyelesaikan soal matematika berdasarkan gender," J. GeoGebra Indones., vol. 1, no. 2, pp. 83–97, 2021,. Available: https://journal.geogebra.id/index.php/IGJ/article/view/14

[2] F. A. Agusti, T. Herman, and A. Zafirah, "Kemampuan penalaran imitatif dan kreatif matematis siswa SMP pada materi persamaan garis lurus," AKSIOMA J. Progr. Stud. Pendidik. Mat., vol. 12, no. 1, p. 1447, 2023, doi: 10.24127/ajpm.v12i1.7042.

[3] A. Octaviyunas and A. Ekayanti, "Pengaruh model pembelajaran giving question getting answer dan think pair share terhadap kemampuan penalaran matematika siswa kelas vii," Mosharafa J. Pendidik. Mat., vol. 8, no. 2, pp. 341–352, 2019, doi: 10.31980/mosharafa.v8i2.453.

[4] F. G. Khainingsih, M. Maimunah, and Y. Roza, "Analisis kemampuan penalaran matematis siswa smp dalam menyelesaikan soal open-ended pada materi teorema pythagoras," J. Kependidikan J. Has. Penelit. dan Kaji. Kepustakaan di Bid. Pendidikan, Pengajaran dan Pembelajaran, vol. 6, no. 2, p. 266, 2020, doi: 10.33394/jk.v6i2.2566.

[5] S. Agustiani, I. P. Dian, R. Fitriani, A. Faradillah, and F. Alyani, "Kemampuan penalaran siswa sma dalam menyelesaikan soal matematika berdasarkan gender," J. GeoGebra Indones., vol. 1, no. 2, pp. 83–97, 2021, doi: 10.30598/snpmunpatti.2021.pp35-43

[6] S. F. Akuba, D. Purnamasari, and R. Firdaus, "Pengaruh kemampuan penalaran, efikasi diri dan kemampuan memecahkan masalah terhadap penguasaan konsep matematika," JNPM (Jurnal Nas. Pendidik. Mat., vol. 4, no. 1, p. 44, 2020, doi: 10.33603/jnpm.v4i1.2827.

[7] Hariyanti and R. P. Khotimah, "Kemampuan penalaran matematis siswa dalam menyelesaikan soal materi bangun ruang sisi datar ditinjau dari perbedaan gender di kelas viii smp negeri 1 bendosari," J. Pembelajaran Mat. Inov., vol. 5, no. 3, pp. 681 692, 2022, doi: 10.22460/jpmi.v5i3.681-692.

[8] R. Fedistia and E. Musdi, "Efektivitas perangkat pembelajaran berbasis flipped classroom untuk meningkatkan kemampuan penalaran matematis peserta didik," J. Didakt. Mat., vol. 7, no. 1, pp. 45–59, 2020, doi: 10.24815/jdm.v7i1.14371.

[9] A. Iswanto, C. V. Putri, P. A. Rizkia, A. Faradillah, and W. Hadi, "Analisis kemampuan penalaran siswa sekolah menengah atas dalam menyelesaikan matematika berdasarkan gender dan geogebra," J. GeoGebra Indones., vol. 2, no. 1, pp. 37–51, 2022, Available: https://journal.geogebra.id/index.php/IGJ/article/view/21

[10] Dinda Kurnia Putri, Joko Sulianto, and Mira Azizah, "Kemampuan penalaran matematis ditinjau dari kemampuan pemecahan masalah," Int. J. Elem. Educ., vol. 3, no. 3, pp. 351–357, 2019, doi: 10.23887/ijee.v3i3.19497

[11] R. Fadli, S. Hidayati, M. Cholifah, R. A. Siroj, and M. W. Afgani, "Validitas dan reliabilitas pada penelitian motivasi belajar pendidikan agama islam menggunakan product moment," JIIP - J. Ilm. Ilmu Pendidik., vol. 6, no. 3, pp. 1734–1739, 2023, doi: 10.54371/jiip.v6i3.1419.

[12] F. F. Ida and A. Musyarofah, "Validitas dan reliabilitas dalam analisis butir soal," Al-Mu'Arrib J. Arab. Educ., vol. 1, no. 1, pp. 34–44, 2021, doi: 10.32923/al-muarrib.v1i1.2100.

[13] N. M. Janna and Herianto, "Konsep uji validitas dan reliabilitas dengan menggunakan SPSS," J. Darul Dakwah Wal-Irsyad, vol.3, no. 18210047, pp. 1–12, 2021, doi: 10.31219/osf.io/v9j52.

[14] L. Amanda, F. Yanuar, and D. Devianto, "Uji validitas dan reliabilitas tingkat partisipasi politik masyarakat Kota Padang," J. Mat. UNAND, vol. 8, no. 1, p. 179, 2019, doi: 10.25077/jmu.8.1.179-188.2019.

[15] Z. Nuraeni, "Pelatihan software SPSS untuk menghitung validitas, reliabilitas, dan analisis butir soal bagi mahasiswa calon guru di palembang," J. Anugerah, vol. 3, no. 1, pp. 15–23, 2021, doi: 10.31629/anugerah.v3i1.3383.

[16] M. Handayani et al., "Sosialisasi dan pengenalan aplikasi pengolahan data SPSS pada mahasiswa administrasi kesehatan fakultas ilmu keolahragaan dan kesehatan," JIPM J. Inf. Pengabdi. Masy., vol. 1, no. 2, pp. 25–31, 2023, doi: 10.47861/jipm nalanda.v1i2.170

[17] D. M. Kartikaningrum and Muhtarom, "Uji validitas dan reliabilitas bahan ajar menggunakan formula aiken's v dan spss.22 terhadap kemampuan berpikir kritis siswa," Pendas J. Ilm. Pendidik. Dasar, vol. 9, no. 1, pp. 879–885, 2024.

[18] Azizah and S. Wahyuningsih, “Penggunaan model rasch untuk analisis instrumen the use of rasch model for analyzing test,” J U P I T E K J. Pendidik. Mat., vol. 3, no. 1, pp. 45–50, 2020, doi; 10.30598/jupitekvol3iss1ppx45-50.
[19] E. F. Tarigan, S. Nilmarito, K. Islamiyah, A. Darmana, and R. D. Suyanti, “Analisis instrumen tes menggunakan rasch model dan software SPSS 22.0,” J. Inov. Pendidik. Kim., vol. 16, no. 2, pp. 92–96, 2022, doi: 10.15294/jipk.v16i2.30530.
[20] L. Tanujaya and A. Yudiarso, “Uji validitas dan reliabilitas multicultural attitude scale questionnaire (masque) versi indonesia dengan model rasch,” Humanit. (Jurnal Psikologi), vol. 7, no. 2, pp. 139–150, 2023, doi: 10.28932/humanitas.v7i2.7420.
[21] V. Wahyuni, “Validity and reliability of mathematical communication ability test instruments relation and function materials,” Sustain. J. Kaji. Mutu Pendidik., vol. 5, no. 1, pp. 89–99, 2022, doi: 10.32923/kjmp.v5i1.2232.
[22] S. Andani, Zulfah, and L. Midani Rizki, “Validitas Instrumen penelitian: pengaruh pendekatan ethno-rme terhadap kemampuan penalaran matematis peserta didik kelas vii di SMPN 2 Bangkinang Kota,” J. Pengabdi. Masy. dan Ris. Pendidik., vol. 2, no. 2, pp. 07–11, 2023, doi: 10.31004/jerkin.v2i2.190.
[23] R. Qomariyah, M. Zainudi, and I. I. T. Rohmah, “Validitas dan reliabilitas instrumen tes kemampuan pemecahan masalah berbasis etnomatematika,” IKIP PGRI Bojonegara, vol. 386, no. 46, pp. 481–484, 2023.
[24] A. Tohri, A. Rasyad, S. Sulaiman, and U. Rosyidah, “Indeks toleransi antarumat beragama di Kabupaten Lombok Timur,” J. Ilmu Sos. dan Hum., vol. 10, no. 3, p. 563, 2021, doi: 10.23887/jish-undiksha.v10i3.38822.
[25] G. R. Oldham and D. Jimeno-Ingrum, “Pengaruh job rotation dan job enlargement terhadap kepuasan kerja (suatu studi pada badan pengelolaan keuangan daerah kabupaten ciamis),” Wiley Encycl. Manag., vol. 2, pp. 1–1, 2015, doi: 10.1002/9781118785317.weom110177.
[26] I. Lenaini, “Teknik pengambilan sampel purposive dan snowball sampling,” Hist. J. Kajian, Penelit. Pengemb. Pendidik. Sej., vol. 6, no. 1, pp. 33–39, 2021, doi: 10.31764/historis.v6i1.4075.
[27] W. N. Rohmah, A. Septian, and S. Inayah, “Analisis kemampuan penalaran matematis materi bangun ruang ditinjau gaya kognitif siswa menengah pertama,” Prisma, vol. 9, no. 2, p. 179, 2020, doi: 10.35194/jp.v9i2.1043.
[28] E. Yuliati, S. P. Sari, and M. Metty, “Implementasi uji validitas isi: pengembangan kuesioner perubahan pola konsumsi minyak goreng pada rumah tangga,” Ilmu Gizi Indones., vol. 7, no. 1, p. 31, 2023, doi: 10.35842/ilgi.v7i1.412.
[29] A. Faradillah and C. Septiana, “Mathematical resilience: validity and reliability with rasch model and spss in senior high school,” AKSIOMA J. Progr. Stud. Pendidik. Mat., vol. 11, no. 4, p. 3545, 2022, doi: 10.24127/ajpm.v11i4.5204.
[30] R. Slamet and S. Wahyuningsih, “Validitas dan reliabilitas terhadap instrumen kepuasan kerja,” Aliansi J. Manaj. dan Bisnis, vol. 17, no. 2, pp. 51–58, 2022, doi: 10.46975/aliansi.v17i2.428.