

Development and Validation of Mathematical Critical Thinking Ability Instrument Using Rasch Measurement

Nur Afifah¹, Tasya Pebrianti², Ayu Faradillah^{3*}, Syafika Ulfah⁴
^{1, 2, 3, 4} Universitas Muhammadiyah Prof. DR. HAMKA
*) ayufaradillah@uhamka.ac.id

Abstract

This instrument has been confirmed to be suitable for measuring students' mathematical thinking skills. Its validity and reliability have been confirmed through verification, including Wright Map analysis conducted using Winstep software. This research is important to produce a valid, reliable, and Wright map instrument for mathematical critical thinking skills. A properly developed instrument will help measure students' abilities more accurately and objectively. This research utilized a quantitative methodology through the application of a survey approach. The data was obtained by applying a tool for measuring critical thinking skills in mathematics to 299 10th grade students at a secondary school in North Jakarta. This measurement instrument consists of three questions covering four indicators of critical thinking skills in mathematics. The measurement tool consists of three questions with four indicators of mathematical critical thinking skills. The validity test results show that all three questions are valid based on the three categories in the item fit order table. Additionally, in the reliability test using Cronbach's Alpha (KR-20), the results showed that the questions were reliable with a value of 0.75. The results of the Wright Map analysis showed two categories, namely indicators and gender. In the indicator category, students have difficulty in drawing conclusions from the results obtained, which are included in the inference indicator. Meanwhile, the gender category shows that female students have higher mathematical critical thinking skills than male students with a percentage value of 16.38%. Proven instruments can help teachers to accurately support and assess the development of students' cognitive abilities.

Keywords: validity, reliability, Wright map, Rasch, mathematical critical thinking skills.

Introduction

The purpose of learning mathematics is to equip students with the ability to handle real-life problems (Hamimi et al., 2020; Purnaningsih & Zulkarnaen, 2022). Mathematical critical thinking skills are one of the essential competencies that play a role in supporting the achievement of these goals, because it involves the process of analyzing and evaluating situations so that students are able to formulate problem-solving strategies independently (Purnaningsih & Zulkarnaen, 2022; Wilujeng & Sudihartinih, 2021). Developing mathematical critical thinking is crucial, as it enables students to create their own approaches and generate new ideas based on personal experience (Wahid et al., 2023). This skill helps students comprehend mathematical problems, evaluate them effectively, and make well-informed decisions in the problem-solving process (Kurniasih & Hakim, 2019).

The results of the current study show that the level of mathematical critical thinking skills in students has not reached the optimal category. This is largely due to their inability to process problems based on given data, making it difficult for them to identify appropriate

strategies to solve the same problem (Cahyani & Sritresna, 2023). One more factor is the absence of students' logical type of thinking in solving problems (Anggraini et al., 2022). What's more, too many students fail in learning basic math. They tend to learn by rote, tending to listen and memorize rather than actively participate. This in turn makes it difficult for them to solve new problems in the face of new challenges (Rosliani & Munandar, 2022; Wahidin & Romli, 2020). Students are frequently guided to focus solely on getting the correct answers, without being encouraged to think creatively or explore alternative ideas. This approach limits the development of their mathematical critical thinking abilities (Dewi et al., 2020). In many cases, the tools teachers use to assess students do not adequately consider the importance of validity and reliability. Therefore, it is essential to design instruments that are both appropriate and thoroughly tested (Rosdiani et al., 2021). Validity refers to the degree to which an instrument accurately and consistently measures what it is intended to assess, ensuring dependable outcomes (Amini, 2023; Dewi et al., 2020; Herlina & Ihsan, 2020). Alongside validity, reliability must also be achieved. Reliability refers to how consistently an instrument can measure students' mathematical critical thinking skills across different conditions, providing a dependable basis for developing evaluation methods and instructional strategies (Dewi et al., 2020; Juliani & Erita, 2023). Instruments that are both valid and reliable are essential for accurately assessing students' critical thinking in mathematics and can be used as a solid foundation for improving teaching practices (Nopita & Wijoyo, 2022; Puspasari et al., 2022).

Previous studies have explored the development of instruments for assessing mathematical critical thinking. First, study by Faradillah & Adlina (2021) when using research results to validate the reliability of mathematical critical thinking tools, the mathematical critical thinking tools developed in this study were proven to be reliable and trustworthy. However, it is still necessary to consider eliminating differences in students' answers to ensure the optimisation of measurement quality. Second, study examining mathematical stability at the high school level (using winsteps 3.73) showed that three items were invalid, while the remaining six items only partially met the validity criteria (Saputri et al., 2023). Comparing these two studies shows a certain gap, the first study focused on critical thinking skills for prospective teachers by testing validity and reliability, while the second study also tested validity and reliability but did not assess students' mathematical critical thinking skills. Thus the previous studies have not evaluated mathematical critical thinking instruments comprehensively by integrating validity, reliability, and analysis using

Wright map simultaneously. The uniqueness of this study lies in its integrated approach, namely by using the Rasch model to assess all three aspects. Therefore, the purpose of this study is to develop and evaluate an instrument to measure students' mathematical critical thinking through validity, reliability, and Wright Map analysis.

Method

This research applies a quantitative approach by administering the same set of questions to all participants, aiming to compare responses across the entire sample (Faradillah & Septiana, 2022). To gather empirical data, the researcher conducted a survey using observation methods. The data collection involved the use of a test as the primary research instrument. Therefore, this study employs a survey-based methodology. The main objective is to illustrate the procedures used to validate and confirm the reliability of an instrument developed to measure mathematical critical thinking skills. The collected data will then be analyzed statistically to test the proposed hypothesis.

Table 1. Research Sample

Demography	Descripts	Code	Sum
Class	X	A	299
Gender	Male	G	125
	Female	F	174
School	SMK Hang Tuah 1 North Jakarta	B	79
	SMAN 110 North Jakarta	C	220

Table 1 presents data from a sample of grade X students selected from two senior high schools and vocational schools located in North Jakarta, consisting of 41.81% male and 58.19% female students. The study includes a diverse sample in terms of gender, school type, and grade level. The three assessment questions focus on the Pythagorean Theorem, with each question reflecting one of the following four aspects of critical thinking in mathematics: interpretation, inference, analysis, and evaluation. Experts in mathematics education, including university lecturers and high school teachers, validated the instrument, and their review confirmed that it met the required standards for use. Data analysis was performed using the Rasch model through the Winstep application. This study developed an assessment tool that aims to evaluate students' mathematical critical thinking, based on four main indicators and utilizing a Likert rating scale worth 0 to 4.

Table 2. Indicators of Mathematical Critical Thinking Skills

Indicator	Description	Score
Interpretation	Does not mention either the known facts or the question being asked.	0
	States the known and asked information, but both are incorrect.	1
	Provides only either the known information or the question, but not both.	2
	Correctly states the known information from the question, but the explanation is not fully complete.	3
	Clearly and accurately identifies both the known information and the question asked.	4
Analysis	Fails to construct a mathematical representation of the problem.	0
	Attempts to form a mathematical model, but it is incorrect.	1
	Accurately creates a mathematical model without adding an explanation.	2
	The model is correct, but the explanation contains some mistakes.	3
	Accurately builds a mathematical model with a clear and correct explanation.	4
Evaluation	Does not apply any strategy to solve the problem.	0
	Applies an incorrect or incomplete strategy to address the problem.	1
	Uses a correct but partial strategy, or solves the problem fully with an incorrect method.	2
	Applies a suitable strategy but makes errors in calculation or explanation.	3
	Effectively uses an appropriate strategy with accurate calculations and clear explanation.	4
Inference	No conclusion is provided.	0
	Draws an incorrect and unrelated conclusion.	1
	Conclusion is relevant to the context but factually incorrect.	2
	Conclusion is accurate and fits the context but lacks completeness.	3
	Conclusion is both accurate and complete, and fully aligned with the question's context.	4

Table 2 describes that each indicator is scored on a scale from 0 to 4, with the assessment tailored to the definition of each indicator. For the interpretation indicator, students earn a score of 4 if they can clearly understand and express the meaning of the

question by writing down both the known information and what is being asked. The analysis indicator involves students accurately forming mathematical models that effectively link the problem to relevant concepts. In the assessment aspect, learners are directed to use problem-solving approaches that are appropriate to the situation at hand.

Results and Discussion

Validity is one of the most important aspects in the construction of measurement instruments. The adequacy of each item was assessed by applying the Rasch model. The results from table 3 show a summary of the validity test based on MNSQ, ZSTD, and CORR.

Table 3. Validity Results on Winsteps

Question	Outfit MNSQ	Outfit ZSTD	PT-Measure CORR
P1	1.30	3.0	0.49
P2	1.20	1.7	0.56
P3	1.13	1.0	0.56
P4	1.08	0.8	0.53
P5	0.96	-0.3	0.62
P6	1.10	0.9	0.42
P7	1.03	0.3	0.44
P8	0.97	-0.2	0.50
P9	0.86	-1.3	0.44
P10	0.78	-2.0	0.51
P11	0.71	-2.8	0.47
P12	0.66	-3.0	0.49

The validity of the questions is assessed based on the thresholds for MNSQ, ZSTD, and CORR values. A question is considered valid if at least two out of the three indicators meet the required standards (Yunika et al., 2021). As shown in Table 3, One of the ZSTD criteria was not accepted because it exceeded the maximum permissible value. However, since the other two indicators fall within the required boundaries, all items are still regarded as valid. Similar findings were reported in previous studies (Dwinata, 2019; Faradillah & Septiana, 2022), where despite one criterion not being met, the items were still validated due to two indicators fulfilling the requirements.

Following the above validity analysis, the second step is to check the reliability and dependability of the instrument and it is necessary to check it in the reliability test. This test is essential to validate the extent to which the instrument can provide stable and consistent results when used under the same conditions. Table 4 shows the Cronbach's Alpha value and the results obtained from the reliability test using the Rasch model.

Table 4. Reliability Results on Winsteps

Number	Statistic	Person	Item
1	(KR-20)	0.79	-
2	Reliability	0.75	0.99
3	Separation	1.74	8.51

Table 4 presents the reliability analysis of the instrument used in this research, utilizing the Cronbach's Alpha coefficient method. The resulting score of 0.75 suggests that the instrument can be considered reliable, as it surpasses the minimum standard value of 0.70. This outcome aligns with prior research, which suggests that a Cronbach's Alpha value greater than 0.70 demonstrates sufficient internal consistency in assessing the intended construct (Faradillah & Septiana, 2022; Febryana et al., 2024).

After confirming the instrument's validity and reliability, the next step involved analyzing how student abilities were distributed and how challenging each item was, using the Wright Map. The Wright Map, a tool used in Rasch analysis, displays both the respondents and the test items along a single logit scale. This allows for a comparison between student ability levels and item difficulty. The visualization of the Wright Map based on the measurement results is shown in picture 1 below.

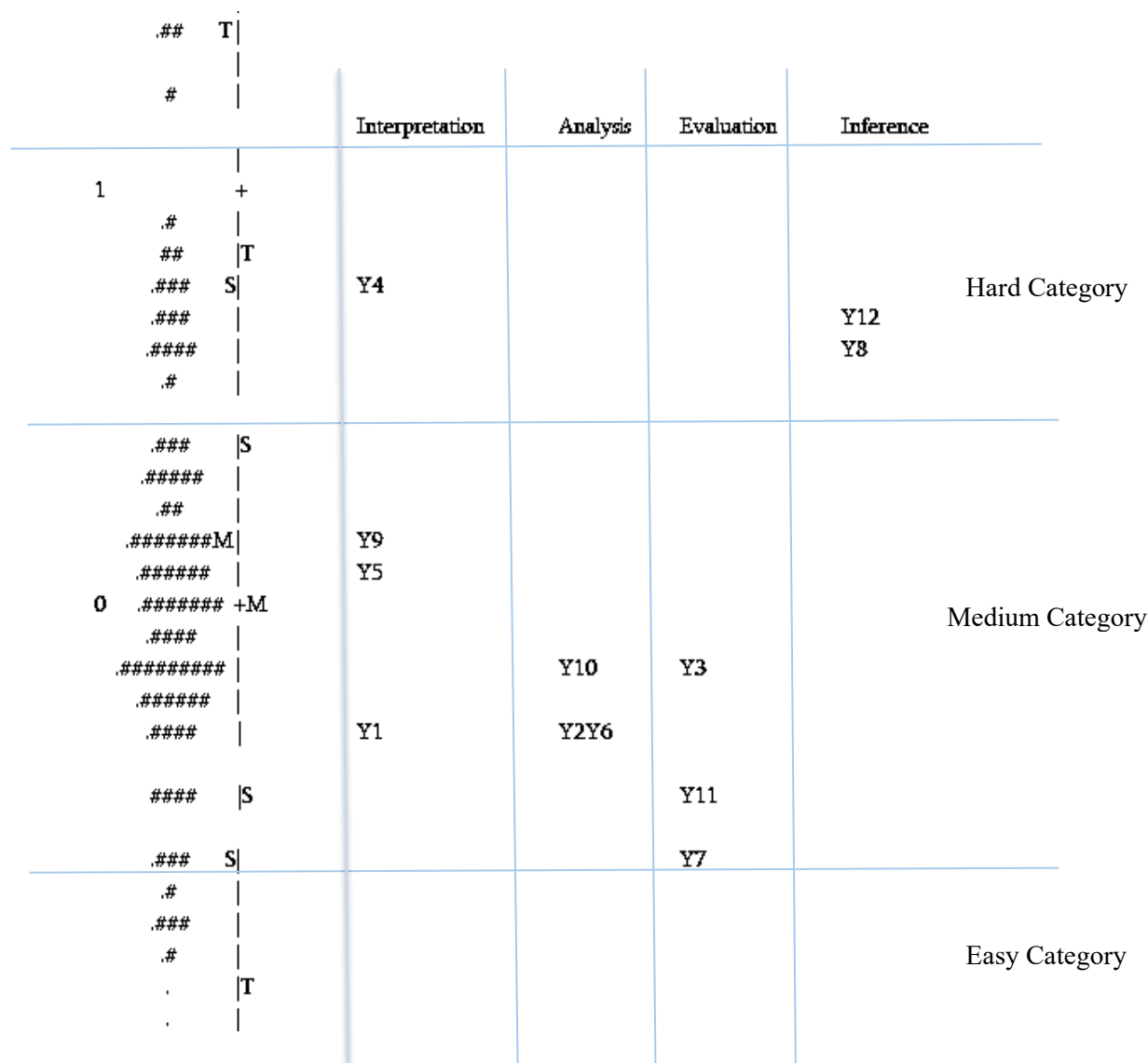


Figure 1. Mathematical critical thinking skills based on indicators

Items located at the lower part of the Wright Map suggest they are less difficult, meaning students with lower ability levels are more likely to answer them correctly (Fernanda & Wulan, 2024; Kristiyasari et al., 2022). This is supported by research by (Yuliana & Miatus, 2023), which states that students are more familiar with evaluative processes in learning, especially procedural-based ones. In contrast, questions on the inference and interpretation indicators occupy the highest position, indicating that students have difficulty in drawing conclusions and connecting information logically. This finding is in line with (Athifah & Khusna, 2022) who explained that inference ability is the most challenging aspect of critical thinking because it requires students to think reflectively and make decisions based on indirect information.

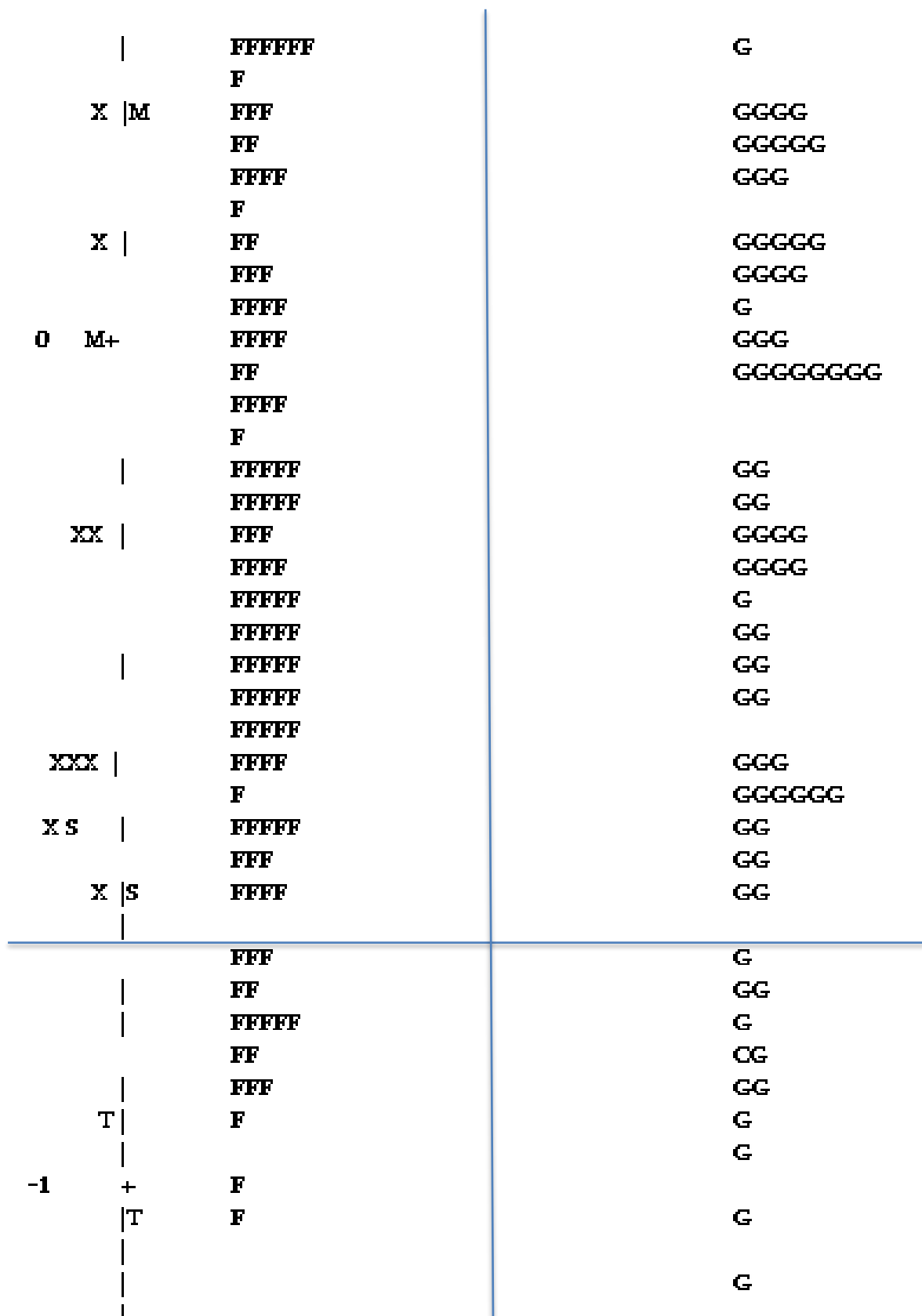


Figure 2. Wright map mathematical critical thinking skills based on gender

Picture 2 illustrates that female students generally perform better on questions categorized as difficult compared to male students. Those who demonstrate strong performance are the ones who correctly answer the greatest number of high-level items (Faradillah & Adlina, 2021). The data indicates that 16.38% of female students (49 out of 299) possess strong mathematical critical thinking skills, whereas only 11.03% of male students (33 out of 299) fall into this high-achievement category. This finding is supported by previous research, which also revealed that female students tend to excel in more challenging problems, while male students are more accurate in problems that are considered easier. This phenomenon shows a tendency for differences in cognitive styles between male and female students, as evidenced by the fact that female students have higher mathematical critical thinking skills than male students. According to (Yudiana et al., 2023) suggested that female students tend to have higher self-regulation skills, including in the aspects of planning, monitoring, and evaluating learning strategies. This has its own contribution in solving complex and non-routine problems as shown in the results of (Hilaliyah et al., 2019) who reported that female students are more thorough, reflective, and careful in making arguments and inferring information however, when compared to male students who although fast in solving problems, they lack deep reasoning and often neglect analysis.

Conclusion and Suggestion

Assessment tools are important for teachers to assess students' abilities, including tools to measure math critical thinking. Validity, reliability, and analysis based on the Wright Map were explored to test the quality of the instrument. The questionnaire used in this study consisted of three items based on four indicators of critical thinking in mathematics. The results showed that the three items were valid with appropriate MNSQ, ZSRD, and CORR values. In addition, the instrument turned out to be a reliable test with Cronbach's alpha of 0.75, which is higher than the lower limit of acceptable reliability, the instrument was also validated by Wright Map where the results with the conclusion indicator showed that students had difficulty in responding to the conclusion. Students' Mathematical Critical Thinking Ability Based on Gender In terms of gender, female students have higher mathematical critical thinking ability compared to male students with a mean difference of 16.38%. Thus, this instrument can be used as an assessment instrument to measure and map students' mathematical critical thinking skills more objectively and purposefully.

References

- Amini, R. P. (2023). Analisis Validasi dan Reliabilitas Instrumen Kemampuan Berpikir Kritis Sekolah Dasar. *Edukatika*, 01(01), 1–10.
- Dewi, R., Sariyasa, & Putrayasa, I. (2020). Pengembangan Instrumen Kemampuan Berpikir Kritis pada Pembelajaran IPA Siswa Kelas V SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, 10(2), 93–101. <https://repo.undiksha.ac.id/id/eprint/4298>
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Analisis Butir Soal Ujian Matematika Kelas VII Semester Ganjil Tahun Pelajaran 2017/2018. *Mathema: Jurnal Pendidikan Matematika*, 2(1), 57. <https://doi.org/10.33365/jm.v2i1.459>
- Herlina, M., & Ihsan, I. R. (2020). Penelitian Pendahuluan mengenai LKPD Model PBL terkait Kemampuan Berpikir Matematis. *Mathema: Jurnal Pendidikan Matematika*, 2(2), 46. <https://doi.org/10.33365/jm.v2i2.733>
- Purnaningsih, I., & Zulkarnaen, R. (2022). Identifikasi Faktor Penyebab Kemampuan Berpikir Kritis Matematis Pada Siswa Kelas Viii. *Teorema: Teori Dan Riset Matematika*, 7(2), 291. <https://doi.org/10.25157/teorema.v7i2.7185>
- Amini, R. P. (2023). Analisis Validasi dan Reliabilitas Instrumen Kemampuan Berpikir Kritis Sekolah Dasar. *Edukatika*, 01(01), 1–10.
- Dewi, R., Sariyasa, & Putrayasa, I. (2020). Pengembangan Instrumen Kemampuan Berpikir Kritis pada Pembelajaran IPA Siswa Kelas V SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, 10(2), 93–101. <https://repo.undiksha.ac.id/id/eprint/4298>
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Analisis Butir Soal Ujian Matematika Kelas VII Semester Ganjil Tahun Pelajaran 2017/2018. *Mathema: Jurnal Pendidikan Matematika*, 2(1), 57. <https://doi.org/10.33365/jm.v2i1.459>
- Herlina, M., & Ihsan, I. R. (2020). Penelitian Pendahuluan mengenai LKPD Model PBL terkait Kemampuan Berpikir Matematis. *Mathema: Jurnal Pendidikan Matematika*, 2(2), 46. <https://doi.org/10.33365/jm.v2i2.733>
- Purnaningsih, I., & Zulkarnaen, R. (2022). Identifikasi Faktor Penyebab Kemampuan Berpikir Kritis Matematis Pada Siswa Kelas Viii. *Teorema: Teori Dan Riset Matematika*, 7(2), 291. <https://doi.org/10.25157/teorema.v7i2.7185>
- ini, R. P. (2023). Analisis Validasi dan Reliabilitas Instrumen Kemampuan Berpikir Kritis Sekolah Dasar. *Edukatika*, 01(01), 1–10.
- Anggraini, N. P., Siagian, T. A., & Agustinsa, R. (2022). Analisis Kemampuan Berpikir Kritis Matematis Siswa Dalam Menyelesaikan Soal Berbasis Akm. *ALGORITMA: Journal of Mathematics Education*, 4(1), 58–78. <https://doi.org/10.15408/ajme.v4i1.25325>
- Cahyani, N. D., & Sritresna, T. (2023). Kemampuan penalaran matematis siswa dalam menyelesaikan soal cerita. *Jurnal Inovasi Pembelajaran Matematika: PowerMathEdu*, 2(1), 103–112. <https://doi.org/10.31980/powermathedu.v2i1.2720>
- Dewi, N. P., Rahmi, Y. L., Alberida, H., & Darussyamsu, R. (2020). Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi ten-tang Materi Hereditas untuk Peserta Didik SMA/MA. *Jurnal Eksakta Pendidikan (Jep)*, 4(2), 138. <https://doi.org/10.24036/jep/vol4-iss2/512>
- Dewi, R., Sariyasa, & Putrayasa, I. (2020). Pengembangan Instrumen Kemampuan Berpikir Kritis pada Pembelajaran IPA Siswa Kelas V SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, 10(2), 93–101. <https://repo.undiksha.ac.id/id/eprint/4298>
- Dwinata, A. (2019). Analisis Kemampuan Pemecahan Masalah Menggunakan Pemodelan RASCH pada Materi Permutasi dan Kombinasi. *PRISMA, Prosiding Seminar Nasional Matematika*, 2, 124–131. <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28899>
-

- Faradillah, A., & Adlina, S. (2021). Validity of critical thinking skills instrument on prospective Mathematics teachers. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 25(2). <https://doi.org/10.21831/pep.v25i2.40662>
- Faradillah, A., & Septiana, C. (2022). Mathematical Resilience: Validity and Reliability With Rasch Model and Spss in Senior High School. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 11(4), 3545. <https://doi.org/10.24127/ajpm.v11i4.5204>
- Febryana, U. I., Rahmania, F., Faradillah, A., & Jusra, H. (2024). *Evaluation of Validity and Reliability Using Rasch Model and SPSS on Students ' Mathematical Creative Thinking Ability*. 5(1), 1–10.
- Fernanda, J. W., & Wulan, E. R. (2024). Analisis Polychotomous Rasch Model untuk Kalibrasi Soal Analisis Real pada Program Studi Tadris Matematika. *Jurnal Universitas Muhammadiyah Metro*, 5(1), 157–166. <https://scholar.ummetro.ac.id/index.php/emteka/article/download/4177/2405/>
- Hilaliyah, H., Agustin, Y., Setiawati, S., Hapsari, S. N., Rangka, I. B., & Ratodi, M. (2019). Wright-Map to investigate the actual abilities on math test of elementary students. *Journal of Physics: Conference Series*, 1318(1). <https://doi.org/10.1088/1742-6596/1318/1/012067>
- Juliani, R. P., & Erita, S. (2023). Analisis Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Kritis dalam Konteks Sekolah Menengah. *JEID: Journal of Educational Integration and Development*, 3(3), 169–179. <https://doi.org/10.55868/jeid.v3i3.313>
- Kristiyasari, M. L., Asmaningrum, H. P., & Hanif, R. F. (2022). Wright-Map to Analyze Students' Abilities on Chemical Bond Test. *SHS Web of Conferences*, 149, 01050. <https://doi.org/10.1051/shsconf/202214901050>
- Kurniasih, R., & Hakim, D. L. (2019). Berpikir kritis siswa dalam materi segiempat. *Prosiding Sesiomadika*, 2(1e), 1135–1145. <https://journal.unsika.ac.id/index.php/sesiomadika/article/view/2911>, diakses Kamis 4 Maret 2021 pukul 16:52:02 WIB
- Nopita, E. N. S., & Wijoyo, Y. (2022). Validitas Dan Reliabilitas Instrumen Kuesioner Dan Video Edukasi Perkembangan Fitofarmaka Di Indonesia. *Jurnal Farmasi Dan Kesehatan Indonesia*, 2(1), 43–56. <https://doi.org/10.61179/jfki.v2i1.336>
- Purnaningsih, I., & Zulkarnaen, R. (2022). Identifikasi Faktor Penyebab Kemampuan Berpikir Kritis Matematis Pada Siswa Kelas Viii. *Teorema: Teori Dan Riset Matematika*, 7(2), 291. <https://doi.org/10.25157/teorema.v7i2.7185>
- Puspasari, H., Puspita, W., Farmasi Yarsi Pontianak, A., & Barat, K. (2022). Uji Validitas dan Reliabilitas Instrumen Penelitian Tingkat Pengetahuan dan Sikap Mahasiswa terhadap Pemilihan Suplemen Kesehatan dalam Menghadapi Covid-19 Validity Test and Reliability Instrument Research Level Knowledge and Attitude of Students Towards . *Jurnal Kesehatan*, 13(1), 65–71.
- Rosdiani, L., Munawar, B., & Dewi, R. (2021). Pelatihan Membuat Game Edukasi Wordwall Untuk Guru Di Kelurahan Karaton. *Jurnal Abdimas Bina Bangsa*, 2(2), 247–255. <https://doi.org/10.46306/jabb.v2i2.138>
- Roslani, V. D., & Munandar, D. R. (2022). Analisis Kemampuan Berpikir Kritis Matematis Siswa Kelas VII Pada Materi Pecahan. *Jurnal Educatio FKIP UNMA*, 8(2), 401–409. <https://doi.org/10.31949/educatio.v8i2.1968>
- Saputri, H. A., Zuhijrah, Larasati, N. J., & Shaleh. (2023). Analisis Instrumen Assesmen : Validitas, Reliabilitas, Tingkat Kesukaran, dan Daya Beda Butir Soal. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09(05), 2986–2995.
- Wahid, R., Eliza, R., & Susanto, A. (2023). Kemampuan Berpikir Kritis Matematis

- Menggunakan Pendekatan Metaphorical Thinking. *Journal Cerdas Mahasiswa*, 5(1), 12–22.
- Wahidin, D., & Romli, L. A. M. (2020). Students critical thinking development in the national sciences and mathematics competition in Indonesia: A descriptive study. *Jurnal Pendidikan IPA Indonesia*, 9(1), 106–115. <https://doi.org/10.15294/jpii.v9i1.22240>
- Wilujeng, S., & Sudihartini, E. (2021). Kemampuan Berpikir Kritis Matematis Siswa Smp Ditinjau Dari Gaya Belajar Siswa. *JPMI: Jurnal Pendidikan Matematika Indonesia*, 6(2), 53–63.
- Yudiana, W., Triwahyuni, A., & Susanto, H. (2023). Multidimensional Rasch Analysis of Gender Differences in Tes Intelegensi Kolektif Indonesia–Tinggi (TIKI-T). *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 12(1), 1–16. <https://doi.org/10.15408/jp3i.v12i1.20417>
- Yunika, V. R., Rohmah, A. N., Istiqomah, S. N., & Faradillah, A. (2021). Validitas dan reliabilitas anxiety questioner dalam pembelajaran matematika dengan menggunakan rasch models. *Seminar Nasional Pendidikan Matematika Universitas Pattimura*, 2, 161–169.