Hipotenusa
Journal of Research Mathematics Education *p-ISSN: 2621-0630*
VOL.9 NO.I 2026 *e-ISSN: 2723-486X*

ANALYSIS OF INSTRUMENTS FOR ASSESSING THE MATHEMATICAL CRITICAL THINKING ABILITY ON THE PYTHAGOREAN THEOREM MATERIAL

Afdella Zuniawati[1,a], Desta Nadia Putri[1,b], Anis Mar'atus Sholehah[1,c] Shindi Purwaningtias[1,d], Sri Astuti Anggra Niwi Ningsih[1,e], Agung Putra Wijaya[1,f], Desrina Hardianti[1,g]
[1]Program Studi Pendidikan Matematika, FKIP, Universitas Lampung
Jl. Prof. Dr. Soemantri Brojonegoro No. 01, Gedong Meneng, Bandar Lampung
e-mail: [a]afdellazuniawati@gmail.com, [b]pdestanadia@gmail.com, [c]anismaratusssholehah@gmail.com, [d]xyindik@gmail.com, [e]anggraniwi57@gmail.com, [f]agung.wijaya@fkip.unila.ac.id, [g*]hardiantidesrina@fkip.unila.ac.id.

Abstract

Critical thinking skills are important to develop in students. This study was conducted to evaluate the quality of a test instrument in measuring the critical thinking skills of junior high school students on the Pythagorean Theorem material. The analysis was carried out through testing the validity, reliability, discriminating power, and difficulty level of the questions with the support of Microsoft Excel. The method used was descriptive quantitative with Grade VIII students selected through purposive sampling as research participants. The results of the study showed that the test instrument was 100% valid, had a reliability of 0.668 (moderate category), the discriminating power indicated questions 1 and 3 were in the fair category while question 2 was in the good category, and all questions had an easy level of difficulty.

Keywords: *Critical Thinking; Instrument Quality; Pythagorean Theorem.*

1. INTRODUCTION

The rapid expansion of globalization and technological advances require education systems to prepare individuals with the skills to compete in the 21st century. Some of the important competencies required include creativity, critical thinking skills, collaboration skills, effective communication, and independence in the learning process. Critical thinking is considered one of the key skills because it involves the ability to analyze, weigh, and draw logical conclusions in order to solve various problems effectively.

Previous studies have shown that critical thinking skills play a major role in decision making and mathematical problem solving. Some researchers view critical thinking as a complex cognitive process that involves the ability to identify problems, analyze information, and be aware of the issues at hand. There are six main indicators of critical thinking skills, namely focus, reasoning, inference, situation, clarity, and review, which are used as references in measuring the quality of students critical thinking [1], [2]

However, in reality students critical thinking skills are still low in practice. Many students face obstacles in understanding problems, managing time, and working on math problems that require complex thinking.[3] This situation is reinforced by the PISA report, which places Indonesia's math scores below the international average, indicating the importance of efforts to improve students critical thinking skills. In the Pythagorean Theorem material, critical thinking skills are essential because the solution process does not only depend on mastery of the concept but also requires the ability to connect prior knowledge. Initial observations show that students still face various obstacles, such as a lack of in depth understanding of the concept and a lack of sill in solving problems related to this material.

Based on these circumtances, this study aims to analyze the quality of assessment instruments so as to provide a comprehensive picture of the level of students critical thinking skills and to determine the extent to which they can apply critical thinking indicators in solving mathematical problems. The results of this study are expected to serve as a reference in the development of more effective learning strategies to improve students critical thinking skills.

2. METHODS

A descriptive quantitative approach was used to describe the results of the analysis of the quality of students critical thinking instruments on the Phythagorean Theorem material based on the numerical data obtained. This approach was chosen because it was able to provide a clear picture of the students condition and abilities based on test results and analysis of their answers. The research subjects consisted of 23 students in class VIII C of SMP Negeri 20 Bandar Lampung in the 2025/2026 academic year. The subjects were selected using purposive sampling, namely students who had studied the Phythagorean Theorem in mathematics [4], [5], [6].

The main instrument in this study was a test of mathematical critical thinking skills on the Pythagorean Theorem, consisting of three essay questions, with a score range of 0-6 for each question. The data were analyzed using validity, reliability, discriminating power, and difficulty index tests with the help of Microsoft Excel. The research procedure consisted of several stages. The preparation stage included: (1) conducting observations; (2) developing a mathematical critical thinking competency grid; (3) developing an instrument consisting of three essay questions; and (4) validating the questions. The implementation stage included: (1) administering the critical thinking competency test to students; and (2) analyzing students answers. Furthermore, the data management stage includes: (1) analyzing quantitative and qualitative data; (2) describing the results of data processing; and (3) drawing conclusions based on the findings.

The criteria related to validity, reliability coefficient, discriminating power, and difficulty level for interpreting students mathematical critical thinking skills are as follows.

Table 2.1 Criteria for Test Validity

Test Value	Level of Corellation
$r_{calculated} < r_{table}$	Invalid

$r_{calculated} < r_{table}$	Valid

Source: [7]

Table 2.2 Criteria for Test Reliability Coefficient

Reliability Coefficient (r_{11})	Level of Corellation
$r_{11} < 0{,}20$	Very Low
$0{,}20 \leq r_{11} < 0{,}40$	Low
$0{,}40 \leq r_{11} < 0{,}70$	Moderate
$0{,}70 \leq r_{11} < 0{,}90$	High
$0{,}90 \leq r_{11} < 1{,}00$	Very High

Source: [8]

Table 2.3 Discrimination Index Criteria

Discrimination Index (DP)	Criteria
$0{,}71 \leq DP \leq 1{,}00$	Excellent
$0{,}41 \leq DP \leq 0{,}70$	Good
$0{,}21 \leq DP \leq 0{,}40$	Fair
$0{,}01 \leq DP \leq 0{,}20$	Poor
$-1{,}00 \leq DP \leq 0{,}00$	Very Poor

Sumber: [8]

Table 2.4 Interpretation of Difficulty Index

Difficulty Index (TK)	Interpretation
$TK = 0{,}00$	Very Difficult
$0{,}00 < TK \leq 0{,}30$	Difficult
$0{,}30 < TK \leq 0{,}70$	Moderate
$0{,}70 < TK \leq 1{,}00$	Easy
$TK = 1{,}00$	Very Easy

Source: [5], [8], [9]

3. RESULTS AND DISCUSSIONS

The results of this study are in the form of a test instrument designed to assess the critical thinking skills of eight grade students at junior high school. The instrument was developed to evaluate the quality of the measuring tools developed by researchers and is expected to serve as a reference for the world of education,

especially in mathematics. This is because the resulting instrument has undergone a series of quality tests, including validity, reliability, discriminating power, and difficulty level tests. Furthermore, this instrument is expected to be an alternative solution to the problem of low quality evaluation instruments that often occurs today. Instruments that have undergone a series of tests should be able to overcome the problems of low instrument feasibility and quality[8], [10], [11]. Many mathematics teachers rarely test the feasibility of instruments due to limitations in time, understanding, and resources. As a result, test instruments have the potential to be biased if they do not undergo an adequate testing process, making it unlikely that they will be able to accurately measure students abilities. Referring to the research results, the quality of critical thinking ability test instruments analyzed in terms of validity, reliability, discriminating power, and level of difficulty obtained the following results [12], [13], [14].

3.1 Question Item Validity

Validity testing is a very important part of analyzing test item because it serves to ensure that each question is truly capable of measuring the competencies that should be assessed [8]. If a test item is valid, it means that the item truly measures the intended competencies or learning indicators, in accordance with the predetermined learning objectives. Validity emphasizes the accuracy of the item function as a measuring tool.

The ability of a question to distinguish between students with high and low levels of mastery of the material is a characteristic of discriminating power, not merely validity. Questions with good discriminating power will results in higher scores for high ability students compared to low ability students. Conversely, if an item is invalid, it is not aligned with the learning indicators or competencies being measured. This invalidity causes the

assessment results to not accurately reflect students abilities, thereby potentially causing unfairness in the evaluation process.

The findings in Table 3.1 show that valid items are a positive indication that the questions have been carefully constructed, taking into account the alignment between learning objectives, content, and the required cognitive demands. Item validity is not only important in terms of technical measurement, but also plays a role in ensuring that the evaluation process truly reflects students level of understanding of the material taught [9]. With good validity, the learning process becomes more meaningful because teachers have the right tools to monitor student progress and provide appropriate feedback [17], [18], [19].

Table 3.1 Results of the Essay Item Validity Test

Item No.	r Table	Correlation Coefficient	Remarks
1	0.413	0.727	Valid
2	0.413	0.891	Valid
3	0.413	0.707	Valid

Source: Author's data analysis

Based on the validity test using *Product Moment* correlation, the $r_{calculated}$ value was obtained for each essay question, which was thes compared with the r_{table} value = 0.413 at a significance level of 5% with 23 participants. For essay questions, question number 1 had an $r_{calculated}$ value 0.727, question number 2 had an $r_{calculated}$ value of 0.891, and question number 3 had an $r_{calculated}$ value of 0,707. All three $r_{calculated}$ value are greater than r_{table}, so all three are declared valid, meaning that the questions are able to measure the abilities that should be measured.

3.2 Item Reliability

The purpose of reliability testing is to determine wheter an assessment instrument is capable of producing stable and

consistent data when applied in similar situations [10]. A commonly used technique is Cronbach's Alpha coefficient, where a value close to 1 indicates a high level of reliability.

Table 3.2 Results of the Essay Test Reliability Analysis

Cronbach 's Alpha	N of Items	Criteria	Remarks
0.668	3	Moderate	Moderately Reliable

Sources: Author's data analysis

Based on the results of the reliability test, the test instruments used is this study are classified as moderately reliable, ranging from $0.40 \leq r_{11} \leq 0.70$. After analyzing the data from the mathematical critical thinking test, a Cronbach's Alpha value of 0.668 was obtained, which is considered moderate. This means that the test instrument meets the reliability criteria but needs to be revised. In other words, the instrument is consistent enough in measuring students abilities, and students answers to the items are relatively stable, but there are still items that may be less than ideal or need revision. Therefore, this instrument must be revised.

3.3 Item Discrimination Index

The discriminating power of a test item serves to determine whether the item is capable of distinguishing between students in the upper and lower groups [11]. The results of processing the discriminating power data using Microsoft Excel are presented below.

Table 3.3 Results of the Essay Item Discrimination Test

Item No.	Corrected Item-Total Correlation	Remarks
1	0.36	Fair
2	0.42	Good
3	0.31	Fair

Source: Author's data analysis

Based on the results of the discriminating power test above, questions 1 and 3 (DP = 0.36 dan 0.31– Fair) were able to distinguish between high and low ability groups of students, but were not yet optimal. Minor revisions are needed to improve their discriminating quality. Question number 2 (DP = 0.42 – Good). This question is very effective in distinguishing between the upper and lower groups of students. The quality of the item is good and suitable for use.

3.4 Level of Difficulty of Questions

The level of difficulty of a question indicates the extent to which students are able to answer the question correctly, and is calculated using the formula P = $\frac{\text{Mean}}{\text{Skor Maksimum}}$.

Table 3.4 Results of the Item Difficulty Index Analysis for Essay Questions

Item No.	Difficulty Index	Remarks
1	0.78	Easy
2	0.75	Easy
3	0.75	Easy

Source: Author's data analysis

Based on the results of the difficulty level calculation for each essay question, the difficulty level index (0.78), (0.75) dan (0.75). All three questions have a low level of difficulty (easy for students to answer), indicating that most students were able to answer them well.

4. CONCLUSIONS AND RECOMMENDATION

The results show that all items in the test instrument are Hasil 100% valid because $r_{hitung} > r_{tabel}$. This instrument also has a reliability of 0.668 for all items, which is categorized as moderate. The limitation of this study lies in the quality of the items, where the discrimination power of item 1 is 0.36 and item 3 is 0.31, so

both items are classified as criteria items, while item 2 has a discrimination power of 0.42, so it is classified as a good criteria item. In addition, all items have a relatively easy level of difficulty, so the instruments ability to distinguish variations in students critical thinking skills in depth is still limited.

Based on the results of the study, the mathematical critical thinking test instrument on the Pythagorean Theorem material has met the validity criteria and has moderate reliability, but some items still need to be revised because of their sufficient discriminatory power and overly easy level of difficulty. Therefore, the quality of the evaluation instrument needs to be considered through testing its validity, reliability, discriminating power, and level of difficulty in order to optimally measure students critical thinking skills [21], [22], [23].

Further research is recommended to develop instruments with more varied and numerous questions and to apply them to different educational materials or levels with broader subjects, so that more accurate and comprehensive measurement results can be obtained [24].

ACKNOWLEDGEMENTS

The author would like to express his deepest gratitude and appreciation to all those who have provided support in the preparation of this research. Special thanks are given to the supervisors and reviewers for their guidance, input, and constructive criticism provided during the research process. The author also thanks the school, teachers, and students who participated as research subjects and assisted in the data collection process. Last but not least, the author thanks his family, colleagues, and all others who cannot be mentioned one by one for their

support, prayers, and assistance, which were invaluable in ensuring the successful completion of this research.

REFENRENCE

[1] S. Anggraini, W. Setyaningrum, and H. Retnawati, "How to improve critical thinking skills and spatial reasoning with augmented reality in mathematics learning?," in *Journal of Physics: Conference Series*, R. H., K. null, and M.-L. O.A., Eds., Graduate Program of Yogyakarta State University, Jl. Colombo No. 1, Karangmalang, Sleman, DI Yogyakarta, Indonesia: Institute of Physics Publishing, 2020. doi: 10.1088/1742-6596/1581/1/012066.

[2] R. De Amicis, M. Riggio, A. Shahbaz Badr, J. Fick, C. A. Sanchez, and E. A. Prather, "Cross-reality environments in smart buildings to advance STEM cyberlearning," *Int. J. Interact. Des. Manuf.*, vol. 13, no. 1, pp. 331–348, 2019, doi: 10.1007/s12008-019-00546-x.

[3] T. Wahyudi, Noerhasmalina, D. Desmayanasari, and F. Lestari, "Collaborative Learning on Mathematical Critical Thinking Skills," *Hipotenusa J. Res. Math. Educ.*, vol. 5, no. 1, pp. 32–45, 2022.

[4] M. N. Arofah and M. A. Noordyana, "Kemampuan Pemecahan Masalah Matematis Ditinjau dari Kemandirian Belajar Siswa pada Materi Lingkaran di Kelurahan Muarasanding," vol. 1, no. September, pp. 421–434, 2021.

[5] D. Desmayanasari, D. Hardianti, F. Lestari, T. Agnesa, and P. S. Balkist, "Designing mathematical creative thinking skill instrument," *YOUNG Sch. Symp. Sci. Math. Educ. Environ.*, vol. 2595, no. 1, p. 060008, 2023, doi: 10.1063/5.0123718.

[6] F. Lestari; *et al.*, "Mathematical communication in blended learning," *AIP Conf. Proc.*, vol. 3142, no. 1, pp. 020064-1–0200647, 2025, doi: https://doi.org/10.1063/5.0262599.

[7] S. Arikunto, "Metode penelitian kualitatif," *Jakarta Bumi Aksara*, 2006.

[8] M. Mardiana, F. G. Putra, and C. V. Anghel, "Students' Mathematical Communication Instruments: A Gender Perspective in Mathematics Learning," *Hipotenusa J. Res. Math. Educ.*, vol. 8, no. 2, pp. 217–228, 2025.

[9] D. Hardianti *et al.*, "Analysis of mathematical communication instruments on the material of flat side space," *5Th Int. Conf. Innov. Des. Anal. Dev. Pract. Aerosp. Automot. Eng. I-Dad'22*, vol. 2766, no. October 2021, p. 020045, 2023, doi:

10.1063/5.0141522.

[10] W. Arbeni, A. Windiani, D. Sariyani Br Sihotang, N. Anggraini, S. Wulandari, and A. Nugroho, "Test Reliability Analysis in Educational Evaluation: A Quantitative Approach to Consistency and Validity," *J. Nas. Holist. Sci.*, vol. 5, no. 1, pp. 59–64, 2025.

[11] S. A. Septiyaningrum, D. Hardianti, and F. Lestari, "Analisis kemampuan literasi matematis terhadap kemampuan menyelesaikan soal cerita matematika SMP," *J. Ilm. Mat. Realis.*, vol. 5, no. 2, 2024, doi: https://doi.org/10.33365/ji-mr.v5i2.5214.

[12] R. Diani, H. Herliantari, I. Irwandani, A. Saregar, and R. Umam, "Search, Solve, Create, and Share (SSCS) Learning Model: The Impact on the Students' Creative Problem-Solving Ability on the Concept of Substance Pressure," *J. Penelit. Fis. dan Apl.*, vol. 9, no. 1, p. 65, 2019, doi: 10.26740/jpfa.v9n1.p65-77.

[13] M. Yasin, S. Huda, F. G. Putra, M. Syazali, R. Umam, and S. Widyawati, "IMPROVE Learning Model and Learning Independence: Influence and Interaction on Mathematics Problem-Solving Abilities in Islamic Boarding School," *J. Phys. Conf. Ser.*, vol. 1467, no. 1, 2020, doi: 10.1088/1742-6596/1467/1/012003.

[14] M. Munifah, A. N. Romadhona, I. Ridhona, R. Ramadhani, R. Umam, and H. S. Tortop, "How to Manage Numerical Abilities in Algebra Material?," *Al-Jabar J. Pendidik. Mat.*, vol. 10, no. 2, pp. 223–232, 2019.

[15] T. M. Siregar, N. Agustina, F. B. Siahaan, L. L. Sinaga, I. Prabawa, and R. Z. Sagala, "Analisis Validitas dan Reliabilitas Butir Soal Pilihan Ganda Dalam Mata Pelajaran Matematika," *J. Educ. Learn. Eval.*, vol. 1, no. 2, pp. 725–730, 2024.

[16] Sudaryono, U. Rahardja, Q. Aini, Y. Isma Graha, and N. Lutfiani, "Validity of Test Instruments," *J. Phys. Conf. Ser.*, vol. 1364, no. 1, 2019, doi: 10.1088/1742-6596/1364/1/012050.

[17] S. Huda *et al.*, "Islamic education in supply chain system by prioritizing manners as a success factor of millennial generation on socializing," *Int. J. Supply Chain Manag.*, vol. 9, no. 2, pp. 853–863, 2020.

[18] S. Sumarni *et al.*, "Behavior in early childhood (2-3) years: A case study on the use of gadgets in social environments," *Int. J. Innov. Creat. Chang.*, vol. 8, no. 8, pp. 384–404, 2019.

[19] A. Saregar *et al.*, "International Journal of Educational Methodology Natural Disaster Education in School : A Bibliometric Analysis with a Detailed Future Insight Overview," *Int. J. Educ. Methodol.*, vol. 8, no. 4, pp. 743–757,

2022, doi: 10.12973/ijem.8.4.743.

[20] K. Basri, B. Baidowi, J. Junaidi, and M. Turmuzi, "Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP," *Griya J. Math. Educ. Appl.*, vol. 1, no. 4, pp. 682–694, 2021, doi: 10.29303/griya.v1i4.107.

[21] R. Prastowo, S. Huda, R. Umam, K. Jermsittiparsert, and A. Eko, "Academic Achievement and Conceptual Understanding of Electrodynamics : Applications Geoelectric Using Cooperative Learning Model," *J. Ilm. Pendidik. Fis. Al-Biruni*, vol. 8, no. 2, pp. 165–175, 2019, doi: 10.24042/jipfalbiruni.v0i0.4614.

[22] M. U. Ridwanulloh, S. Huda, and R. Umam, "Innovative Leadership Management: The Pattern of School Quality Development at SMP Muhammadiyah 2 Inovasi Malang," *Progres. J. Pemikir. dan Pendidik. Islam*, vol. 11, no. 01, pp. 25–42, 2022, doi: 10.22219/progresiva.v11i01.20742.

[23] S. Huda, Suherman, Komarudin, M. Syazali, and R. Umam, "The Effectiveness of Al-Qurun Teaching Model (ATM) Viewed from Gender Differences : The Impact on Mathematical Problem-Solving Ability The Effectiveness of Al-Qurun Teaching Model (ATM) Viewed from Gender Differences : The Impact on Mathematical Probl," *J. Phys. Conf. Ser.*, vol. 1467, p. 012001, 2020, doi: 10.1088/1742-6596/1467/1/012001.

[24] S. Huda, L. Anggraini, R. Saputri, M. Syazali, and R. Umam, "Learning model to improve the ability to understand mathematical concepts," *Prisma*, vol. 173, no. 2, pp. 173–181, 2019, [Online]. Available: https://jurnal.unsur.ac.id/prisma