UJMER Volume 1 No 3 2026

Unnes Journal of Mathematics Education Research

http://journal.unnes.ac.id/sju/index.php/ujmer

Development of a Mathematical Reasoning Test for Prospective Mathematics Teachers

Sabrina Aprilia[1], Elisa Mutiara Fatmadewi[2], Ahmad Taufik Hidayat[3], Nafa Nur Rakhima[4], Indah Maulidya[5], Nuriana Rachmani Dewi[6]✉

Kampus UNNES Sekaran, Gunungpati, Kota Semarang, Jawa Tengah, Indonesia 50229

@nurianaramadan@mail.unnes.ac.id

Article Info

Article History:
Received 7 July 2026
Accepted 21 August 2026
Published 9 September 2026

Keywords:
Mathematical Reasoning, Prospective mathematics teachers, Test instrument.

Abstract

This study aimed to develop a mathematical reasoning test that is valid, reliable, possesses good discriminating power, and has an appropriate level of item difficulty for assessing the mathematical reasoning ability of prospective mathematics teachers. A descriptive quantitative method was employed to examine the characteristics of nine test items in terms of validity, reliability, discriminating power, and item difficulty. The participants were 21 fourth-year students from Semarang State University. The findings indicated that all test items met the validity criteria, with a reliability coefficient of 0.75, which is classified as high. In terms of discriminating power, six items were categorized as good, while three items were classified as satisfactory. Regarding item difficulty, seven items were categorized as easy and two as moderate. Overall, the developed mathematical reasoning test is suitable for assessing the mathematical reasoning ability of prospective mathematics teachers, as it satisfies the criteria of validity, reliability, and good discriminating power. This study is expected to serve as a reference for the development of mathematical reasoning assessment instruments in higher education. Nevertheless, further research is recommended to refine the distribution of item difficulty, thereby producing an instrument with a more balanced and varied range of difficulty levels.

✉Correspondence Address (author1):
E-mail (sabrinaaprilia@students.unnes.ac.id):

p-ISSN 2252-6455
e-ISSN 2502-4507

INTRODUCTION

Mathematical reasoning is one of the essential competencies that prospective mathematics teachers are expected to possess (Ulfah & Miatun, 2023). eachers need to develop their own mathematical reasoning knowledge before they can effectively foster students' mathematical reasoning, and therefore initial teacher education should provide opportunities for prospective teachers to develop mathematical reasoning from the beginning of their professional preparation (Rodrigues et al., 2021). According to NCTM (dalam Ramdan & Roesdiana, 2022), mathematical reasoning refers to the ability to think logically, serving as the foundation for understanding relationships among mathematical concepts, constructing arguments, and drawing valid conclusions. This competency is widely recognized as a central component of mathematical proficiency and meaningful mathematics learning (Jeannotte & Kieran, 2017). This ability encompasses four key indicators: making generalizations to predict solutions and solution processes, performing mathematical manipulations, using patterns and relationships to analyze mathematical situations, and drawing logical conclusions. Mastery of these competencies is crucial because prospective mathematics teachers are expected not only to obtain correct solutions but also to explain mathematical reasoning processes to students as an integral part of their professional competence (Rodrigues et al., 2021).

Current evidence indicates that the mathematical reasoning ability of university students remains underdeveloped. Dewi and Zahid (2020) reported that students encounter difficulties in providing explanations based on facts and relationships, predicting solutions and solution processes, and constructing mathematical proofs. These findings were reinforced by Miatun and Ulfah (2023), who found that 63% of prospective mathematics teachers demonstrated very low levels of mathematical reasoning ability. Furthermore, students' abilities in performing mathematical manipulations (37%), constructing mathematical proofs (43%), and drawing conclusions (36%) were also classified as very low. Similarly, Subarinah et al. (2026) found that prospective mathematics teachers frequently made errors in mathematical proof, particularly in algebraic manipulation, the application of proof concepts, and logical reasoning when constructing mathematical arguments. These persistent challenges highlight the need to strengthen mathematical reasoning during initial teacher education. In this regard, Oliveira and Henriques (2021) emphasized that prospective mathematics teachers should be better prepared to understand and design learning tasks that promote students' mathematical reasoning, while Martins et al. (2023) highlighted that fostering mathematical reasoning remains a central objective of teacher education programs. These findings indicate that the mathematical reasoning ability of prospective mathematics teachers requires greater attention, as it forms the foundation for understanding mathematical concepts, developing proofs, and effectively communicating mathematical ideas to students.

In response to these challenges, numerous studies have attempted to improve students' mathematical reasoning ability. Dewi and Zahid (2020) demonstrated that the implementation of Web-Assisted Brain-Based Learning significantly enhanced students' mathematical reasoning ability compared with conventional instruction. In addition, a recent systematic literature review of 80 publications reported that research on reasoning and proof in mathematics education has predominantly focused on instructional strategies and interventions for improving students' reasoning and proof abilities, while several research gaps remain to be addressed (Lessing & Ogbonnaya, 2025). These findings suggest that previous studies have predominantly emphasized instructional interventions to enhance mathematical reasoning.

Alongside studies on improving mathematical reasoning ability, research on instrument development has also begun to emerge. Merona and Santi (2018) developed a mathematical reasoning assessment instrument for a Complex Functions course and produced an instrument that was valid, practical, and effective for use within that specific course. However, the instrument was designed exclusively for the Complex Functions course and was not intended as a general instrument for assessing the mathematical reasoning ability of prospective mathematics teachers. Although studies on mathematical reasoning and instrument development have been conducted, research specifically aimed at developing mathematical reasoning test instruments for prospective mathematics teachers remains limited. Most existing studies have concentrated on the effectiveness of instructional models, while instrument development research has generally been restricted to particular courses and has not comprehensively evaluated empirical instrument quality, including validity, reliability, discriminating power, and item difficulty. Consequently, there is still no assessment instrument specifically designed for prospective mathematics teachers that is supported by sufficient empirical evidence regarding the quality of each test item.

The quality of assessment outcomes depends largely on the quality of the instrument employed. A sound assessment instrument should accurately measure the intended construct, produce consistent results, distinguish individuals according to their ability levels, and include items with appropriate levels of difficulty. Therefore, instrument development should be accompanied by rigorous item analysis encompassing validity, reliability, discriminating power, and item difficulty to ensure that the resulting measurements are trustworthy and can serve as a sound basis for instructional decision-making (Arikunto, 2013).

Based on these considerations, this study offers a novel contribution by developing a mathematical reasoning test instrument specifically designed for prospective mathematics teachers and evaluating its empirical quality in terms of validity, reliability, discriminating power, and item difficulty. Unlike previous studies that primarily focused on improving mathematical reasoning through instructional interventions or developing instruments for specific courses, this study produces a more general assessment instrument for measuring the mathematical reasoning ability of prospective mathematics teachers. This study also contributes to the achievement of Sustainable Development Goal (SDG) 4 on Quality Education, which emphasizes the importance of preparing high-quality teachers through effective teacher education. Since prospective mathematics teachers are expected to possess strong mathematical reasoning to effectively foster students' reasoning (Rodrigues et al., 2021), developing a valid and reliable assessment instrument is essential to support this goal. Accordingly, this study aims to develop a mathematical reasoning test instrument that meets established psychometric standards and is suitable for accurately assessing the mathematical reasoning ability of prospective mathematics teachers.

METHOD

This study employed a descriptive quantitative approach with an evaluative analysis to examine the quality of a mathematical reasoning test instrument developed for prospective mathematics teachers. The evaluation focused on the psychometric characteristics of each test item, including validity, reliability, item difficulty, and discriminating power. The results of these analyses were used to determine the suitability of the instrument as a measure of mathematical reasoning ability.

The participants were fourth-year undergraduate students enrolled in the Mathematics Education Study Program who had completed the coursework relevant to the competencies assessed by the instrument. The sample was selected using a simple random sampling technique, ensuring that each student had an equal opportunity to be selected as a participant.

The instrument consisted of nine open-ended test items. The test was developed based on indicators of mathematical reasoning ability, including making generalizations to predict solutions and solution processes, performing mathematical manipulations, using patterns and relationships to analyze mathematical situations, and drawing logical conclusions.

The study was conducted in three stages: the planning stage, the instrument development stage, and the pilot testing and instrument analysis stage.

During the planning stage, the researchers identified the intended learning outcomes, determined the mathematical content to be assessed, and formulated indicators of mathematical reasoning ability that served as the basis for constructing the test blueprint. This stage ensured that each test item was explicitly aligned with the competencies to be measured.

The second stage involved instrument development. At this stage, nine open-ended test items were constructed based on the established blueprint. The development process emphasized conceptual accuracy, alignment with the mathematical reasoning indicators, the use of clear and communicative language, and appropriate wording to ensure that the items were easily understood by participants without reducing the level of reasoning required.

The final stage consisted of pilot testing and instrument analysis. The developed instrument was administered to the participants to obtain empirical data. The collected data were subsequently analyzed using validity testing, reliability testing, item difficulty analysis, and discriminating power analysis. Based on the results of these analyses, each test item was evaluated to determine whether it should be retained, revised, or discarded. This process was intended to produce an assessment instrument with satisfactory measurement quality.

The research data were collected using a documentation technique, specifically through students' answer sheets after they completed the mathematical reasoning test. All data obtained from the pilot testing were processed and analyzed using Microsoft Excel.

Item validity was analyzed using the Pearson Product–Moment correlation coefficient. The validity of each test item was determined by comparing the calculated correlation coefficient r_{hitung} with the critical correlation coefficient r_{tabel}. A test item was considered valid if its r_{hitung} value exceeded the r_{tabel} value (Arifin, 2017).

The reliability of the instrument was evaluated using Cronbach's Alpha coefficient. This analysis was conducted to determine the internal consistency of the instrument in producing stable and consistent results when administered under relatively similar conditions (Arikunto, 2013). The resulting reliability coefficient was subsequently interpreted according to the criteria proposed by Hairun (2020).

Table 1. Reliability Criteria

Cronbach's Alpha	Reliability Criteria
$0,00 \leq r < 0,20$	Very Low
$0,20 \leq r < 0,40$	Low
$0,40 \leq r < 0,60$	Moderate
$0,60 \leq r < 0,80$	High
$0,80 \leq r < 1,00$	Very High

Source: Hairun, 2020.

Discriminating power (DP) analysis was conducted to determine the ability of each test item to distinguish between students with high and low levels of mathematical reasoning ability. The discriminating power index for each item was calculated using the following formula.

$$DP = \frac{Mean\ Score\ of\ the\ Upper\ Group - Mean\ Score\ of\ the\ Lower\ Group}{Maximum\ Item\ Score}$$

The criteria for interpreting the discriminating power index are presented in Table 2.

Table 2. Criteria for the Discriminating Power Index

Discriminating Power Index (DP)	Criteria
$DP \leq 0,00$	Very Poor
$0,00 < DP \leq 0,20$	Poor
$0,20 < DP \leq 0,40$	Fair
$0,40 < DP \leq 0,70$	Good
$0,70 < DP \leq 1,00$	Very Good

Source: Arikunto, 2013.

Item difficulty analysis was conducted to determine the level of difficulty of each test item for the participants. The item difficulty index was calculated using the following formula.

$$Item\ Difficult\ Index = \frac{Mean\ Item\ Score}{Mximum\ Item\ Score}$$

The criteria for interpreting the item difficulty in Table 3.

Table 3. Criteria for the Item Difficulty Index

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

Source: Arikunto, 2013.

RESULTS AND DISCUSSIONS

The developed test instrument consists of nine open-ended items constructed based on the indicators of mathematical reasoning ability for prospective mathematics teachers. The instrument is designed to assess mathematical reasoning through four indicators: making generalizations to predict solutions and solution processes, performing mathematical manipulations, using patterns and relationships to analyze mathematical situations, and drawing logical conclusions. Items 1 and 2 were developed to assess students' ability to make generalizations by identifying patterns, formulating general rules, and predicting solutions based on the identified patterns. Items 3, 4, and 5 were designed to measure students' ability to perform mathematical manipulations through transforming mathematical expressions, proving mathematical concepts, and solving problems requiring algebraic and geometric manipulations. Furthermore, Items 6 and 7 were intended to assess the ability to use patterns and relationships in analyzing mathematical situations by connecting contextual information to appropriate mathematical representations. Finally, Items 8 and 9 were developed to evaluate students' ability to draw logical conclusions by analyzing the validity of arguments, constructing counterexamples, and identifying and analyzing errors in mathematical reasoning during problem solving.

Validity

The validity of the test items was evaluated by comparing the calculated correlation coefficient r_{hitung} with the critical correlation coefficient r_{tabel}. The results showed that all test items obtained $r_{hitung} > r_{tabel}$, indicating that all nine items were valid. These findings suggest that each test item appropriately represents the intended indicators of mathematical reasoning ability.

Table 4. Results of the Test Instrument Validity Analysis

Item	r_{hitung}	r_{tabel}	Decision
1	0,637	0,433	Valid
2	0,707	0,433	Valid
3	0,483	0,433	Valid
4	0,517	0,433	Valid
5	0,690	0,433	Valid
6	0,697	0,433	Valid
7	0,455	0,433	Valid
8	0,508	0,433	Valid
9	0,568	0,433	Valid

Source: Authors' research data, 2026.

The validity analysis indicated that all test items obtained $r_{hitung} > r_{tabel}$, , indicating that all nine items were valid. These findings suggest that each test item was capable of measuring mathematical reasoning ability in accordance with the intended indicators, namely making generalizations to predict solutions and solution processes, performing mathematical manipulations, using patterns and relationships to analyze mathematical situations, and drawing logical conclusions. Therefore, the developed instrument demonstrates satisfactory validity in measuring the construct of mathematical reasoning ability among prospective mathematics teachers. This finding is consistent with Arikunto (2021), who stated that validity refers to the extent to which an instrument measures the construct it is intended to measure.

Reliability

The reliability of the instrument was evaluated using Cronbach's alpha coefficient. The analysis yielded a reliability coefficient of 0.75, which falls within the high reliability category. This result indicates that the instrument has good internal consistency and can therefore be used as a reliable measure of the mathematical reasoning ability of prospective mathematics teachers.

Tabel 5. Results of the Test Instrument Reliability Analysis

Reliability Coefficienrt	Category
0,750	High

Source: Authors' research data, 2026.

The reliability analysis yielded a Cronbach's alpha coefficient of 0.75, which falls within the high reliability category. This result indicates that the instrument has good internal consistency and is capable of producing relatively consistent measurement results when administered to participants with similar characteristics. Therefore, the instrument can be considered a reliable measure of the mathematical reasoning ability of prospective mathematics teachers. This finding is consistent with Arikunto (2021), who stated that reliability refers to the degree to which an instrument produces consistent results when administered under relatively similar conditions..

Discriminating Power

Discriminating power analysis was conducted to determine the ability of each test item to distinguish between students with high and low levels of mathematical reasoning ability. The discriminating power index was calculated by subtracting the mean score of the lower group from the mean score of the upper group and then dividing the result by the maximum item score. The resulting index was subsequently interpreted according to the established criteria. The analysis showed that six test items were classified as good, while the remaining three items were categorized as fair.

Tabel 6. Result of the Discriminating Power Analysis

Item	Disriminating Power Index	Kategori
1	0,278	Fair
2	0,444	Good
3	0,389	Fair
4	0,500	Good
5	0,556	Good
6	0,500	Good
7	0,278	Fair
8	0,444	Good
9	0,500	Good

Source: Authors' research data, 2026.

The discriminating power analysis showed that Items 2, 4, 5, 6, 8, and 9 were classified as good, whereas Items 1, 3, and 7 were categorized as fair. These results indicate that all test items were able to distinguish between prospective mathematics teachers with high and low levels of mathematical reasoning ability. For the indicator of making generalizations to predict solutions and solution processes, one item demonstrated good discriminating power (Item 2), whereas the other was classified as fair (Item 1). For the indicator of performing mathematical manipulations, two of the three items (Items 4 and 5) were categorized as good, while one item (Item 3) was classified as fair. Similarly, for the indicator of using patterns and relationships to analyze mathematical situations, one item (Item 6) demonstrated good discriminating power, whereas the other (Item 7) was categorized as fair. Both items assessing the indicator of drawing logical conclusions (Items 8 and 9) demonstrated good discriminating power. Overall, these findings indicate that most test items possess good discriminating power in accordance with the mathematical reasoning indicators being assessed. Although the remaining items were classified as fair, they are still suitable for use and may be further refined to enhance their ability to distinguish between different levels of mathematical reasoning among prospective mathematics teachers.

Item Difficulty

Item difficulty analysis was conducted to determine the level of difficulty of each test item in the instrument. The item difficulty index was calculated by comparing the mean score obtained by the participants on each test item with the maximum possible score for that item. The resulting index was then interpreted according to the established item difficulty criteria. The analysis showed that Items 1, 2, 4, 5, 6, 8, and 9 were classified as easy, whereas Items 3 and 7 were categorized as moderate.

Tabel 7. Result of the Item Difficulty Analysis

Butir Soal	Skor Kesukaran	Kategori
1	0,889	Easy

2	0,762	Easy
3	0,667	Moderate
4	0,778	Easy
5	0,794	Easy
6	0,810	Easy
7	0,698	Moderate
8	0,730	Easy
9	0,714	Easy

Source: Authors' research data, 2026.

The item difficulty analysis indicated that the instrument consisted of seven items classified as easy (Items 1, 2, 4, 5, 6, 8, and 9) and two items classified as moderate (Items 3 and 7). These findings suggest that most of the test items could be completed relatively easily by the participants, whereas the remaining two items provided a moderate level of difficulty, allowing the instrument to assess mathematical reasoning ability across a broader range of proficiency levels. According to Arikunto (2021), a high-quality instrument should ideally contain a balanced distribution of item difficulty levels, with a greater proportion of moderately difficult items, as they provide more informative evidence of participants' abilities. Therefore, several items currently classified as easy should be revised by enriching the problem contexts, increasing the complexity of the required reasoning processes, or refining the wording of the items without altering the intended mathematical reasoning indicators. Such revisions are expected to produce a more balanced distribution of item difficulty, thereby improving the instrument's effectiveness and accuracy in assessing the mathematical reasoning ability of prospective mathematics teachers.

The analysis results indicate that the developed mathematical reasoning test instrument meets the criteria of a high-quality assessment instrument. All test items were found to be valid, the instrument demonstrated high reliability, the discriminating power of the items ranged from fair to good, and the item difficulty analysis showed that most items were classified as easy, with several items categorized as moderate. These findings indicate that the instrument is suitable for assessing the mathematical reasoning ability of prospective mathematics teachers. Nevertheless, several items that remain classified as easy should be refined to achieve a more balanced distribution of item difficulty. Such improvements are expected to enable the instrument to provide more informative and accurate measurements of mathematical reasoning ability. Beyond its psychometric quality, the developed instrument also contributes to the achievement of Sustainable Development Goal (SDG) 4 on Quality Education. A valid and reliable mathematical reasoning assessment enables teacher education institutions to identify prospective mathematics teachers' competencies more accurately and use the results to improve teacher preparation. In this way, the instrument supports efforts to prepare high-quality mathematics teachers, which is an essential component of quality education.

CONCLUSION

This study successfully developed a mathematical reasoning test instrument for prospective mathematics teachers consisting of nine open-ended items based on four indicators of mathematical reasoning: making generalizations to predict solutions and solution processes, performing mathematical manipulations, using patterns and relationships to analyze mathematical situations, and drawing logical conclusions. The results indicated that all test items met the validity criteria, with calculated correlation coefficients exceeding the critical correlation coefficient $r_{hitung} > r_{tabel}$. Furthermore, the instrument demonstrated high reliability, with a Cronbach's alpha coefficient of 0, 75. Discriminating power analysis showed that all test items were classified as fair to good, while item difficulty analysis indicated that most items were categorized as easy, with only two items classified as moderate. These findings demonstrate that the developed instrument satisfies the psychometric criteria of validity, reliability, discriminating power, and item difficulty, making it suitable for assessing the mathematical reasoning ability of prospective mathematics teachers. However, the item difficulty analysis also revealed that most items were classified as easy. Therefore, future research is recommended to refine several test items by increasing the complexity of the reasoning processes and problem contexts while maintaining the intended indicators. In addition, further validation using larger and more diverse samples is needed to provide stronger empirical evidence of the instrument's quality and enhance its generalizability across mathematics teacher education programs.

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation to Semarang State University, particularly the Department of Mathematics Education, for their valuable support and cooperation in facilitating the successful completion of this study.

REFERENCES

Arifin, Zaenal. (2017). “Kriteria Instrumen dalam Suatu Penelitian”. Jurnal THEOREMS (The Original Research of Mathematics, Vol 2(1), pp: 28-36.

Arikunto, S. (2013). Dasar-dasar evaluasi pendidikan (Edisi revisi). Jakarta: Bumi Aksara.

Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan* (Edisi 3). Jakarta: Bumi Aksara.

Dewi, N. R., & Zahid, M. Z. (2020, June). Enhancing mathematical reasoning: The effects of web-assisted brain-based learning. In *Journal of Physics: Conference Series* (Vol. 1567, No. 4, p. 042091). IOP Publishing.

Hairun, Y. (2020). Evaluasi dan Penilaian Dalam Pembelajaran. Yogyakarta: Deepublish.

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in mathematics*, *96*(1), 1-16.

Lessing, M. J., & Ogbonnaya, U. I. (2025). Reasoning and proof in mathematics education: a systematic literature review. *Discover Education*.

Martins, M., Ponte, J. P. D., & Mata-Pereira, J. (2023). Learning to promote students’ mathematical reasoning: Lesson study contributions in initial teacher education. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*.

Merona, S. P., & Santi, E. E. (2018). Pengembangan Instrumen Asesmen Penalaran Matematis pada Matakuliah Fungsi Kompleks. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, *4*(2), 113-122.

Miatun, A., & Ulfah, S. (2023). Kemampuan Penalaran Matematis Mahasiswa Calon Guru Matematika. *Teorema: Teori dan Riset Matematika*, *8*(2), 281-294.

Oliveira, H., & Henriques, A. (2021). Preservice Mathematics Teachers' Knowledge about the Potential of Tasks to Promote Students' Mathematical Reasoning. *International Journal of Research in Education and Science*, *7*(4), 1300-1319.

Ramdan, M. G. A., & Roesdiana, L. (2022). Analisis kemampuan penalaran matematis siswa SMP pada materi teorema phytagoras. *Jurnal Educatio FKIP UNMA*, *8*(1), 386-395.

Rodrigues, M., Brunheira, L., & Serrazina, L. (2021). A framework for prospective primary teachers’ knowledge of mathematical reasoning processes. *International Journal of Educational Research*, *107*, 101750.

Subarinah, S., Prayitno, S., Triutami, T. W., Salsabila, N. H., & Dasusmi, K. J. (2026). Analisis Kemampuan Pembuktian Matematis Masalah Bilangan pada Mahasiswa Calon Guru Matematika. *Mandalika Mathematics and Educations Journal*, *8*(1), 20-31.