

DEVELOPMENT OF TEST INSTRUMENTS TO MEASURE MATHEMATICAL COMMUNICATION SKILLS OF MADRASAH IBTIDAIYAH STUDENTS

Elisha Farmah¹, Novaliyosi², Hepsi Nindiasari³, Nurul Anriani⁴

^{1,2,3,4}Department of Mathematics Education, Faculty of Teacher Training and
Education, Universitas Sultan Ageng Tirtayasa, Indonesia

E-mail: elishafarmah@gmail.com¹
novaliyosi@untirta.ac.id²
hepsinindiasari@untirta.ac.id³
nurul.anriani@untirta.ac.id⁴

Received June 10, 2025; Received in revised form September 08, 2025; Accepted March 15, 2026

ABSTRACT

Mathematical communication skills are fundamental competencies that students must have, but there are no valid and reliable test instruments to measure the mathematical communication skills of Madrasah Ibtidaiyah students specifically. This research aims to develop valid, reliable, and practical mathematical communication ability test instruments for Madrasah Ibtidaiyah students. The research uses a Research and Development (R&D) approach with the Plomp model which includes the initial investigation, prototype phase, and assessment phase. The research subjects consisted of 39 students of class V MIN 1 Cilegon City who were selected using the purposive sampling technique. The instrument was developed based on the theoretical framework of NCTM with four mathematical communication indicators in KPK and FPB materials. The developed instrument met psychometric standards with very high construct validity (correlation coefficient $r > 0.80$ for all question items) and high reliability ($\alpha = 0.861$). The distribution of difficulty level showed one question item in the medium category (65.40%) and three questions in the very difficult category (15.90%-17.80%). The differentiating power showed that one question item was categorized as sufficient (0.341) while the other three items needed improvement. The instrument was successfully developed with adequate psychometric quality and can be used as an evaluation tool for mathematical communication skills at Madrasah Ibtidaiyah. Further research is suggested to expand the research sample and develop instruments for other mathematical materials.

Keywords: assessment; mathematical communication skills; test instruments

ABSTRACT

Kemampuan komunikasi matematika merupakan kompetensi mendasar yang harus dimiliki siswa, namun tidak ada instrumen tes yang valid dan handal yang tersedia untuk secara khusus mengukur kemampuan komunikasi matematika siswa Madrasah Ibtidaiyah. Penelitian ini bertujuan untuk mengembangkan instrumen uji kemampuan komunikasi matematika yang valid, handal, dan praktis bagi siswa Madrasah Ibtidaiyah. Penelitian menggunakan pendekatan Research and Development (R&D) dengan model Plomp yang terdiri dari penelitian pendahuluan, fase pembuatan prototipe, dan fase penilaian. Subjek penelitian terdiri dari 39 siswa kelas lima MIN 1 Cilegon City yang dipilih menggunakan teknik purposive sampling. Instrumen ini dikembangkan berdasarkan kerangka teoritis NCTM dengan empat indikator komunikasi matematis pada materi KPK dan GCD. Instrumen yang dikembangkan memenuhi standar psikometri dengan validitas konstruksi yang sangat tinggi (koefisien korelasi $r > 0,80$ untuk semua item) dan keandalan tinggi ($\alpha = 0,861$). Distribusi tingkat kesulitan menunjukkan satu item dalam kategori sedang (65,40%) dan tiga item dalam kategori sangat sulit (15,90%-17,80%). Kekuatan diskriminasi menunjukkan satu item dalam kategori yang memadai (0,341) sementara tiga lainnya membutuhkan perbaikan. Instrumen ini berhasil

dikembangkan dengan kualitas psikometrik yang memadai dan dapat digunakan sebagai alat evaluasi kemampuan komunikasi matematika di Madrasah Ibtidaiyah. Penelitian di masa depan disarankan untuk memperluas sampel penelitian dan mengembangkan instrumen untuk materi matematika lainnya.

Kata Kunci: kemampuan komunikasi matematika; penilaian; instrumen tes



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Introduction

Mathematical communication skills are a fundamental competency that every student must have in learning mathematics. These skills include the ability to express mathematical ideas orally and in writing, represent concepts in various visual forms such as diagrams and graphs, and use mathematical symbols appropriately and meaningfully. The importance of mathematical communication skills is not only limited to the context of classroom learning, but also has a strategic role in helping students develop a deep conceptual understanding and apply mathematical knowledge in everyday life. The formation of solid mathematical communication skills from primary education will provide a strong foundation for the development of more complex mathematical skills at the next level of education. Therefore, identifying and measuring students' mathematical communication skills is a crucial aspect in efforts to improve the quality of mathematics learning. However, the complexity in measuring mathematical communication skills requires appropriate measurement instruments and in accordance with the characteristics of students at each level of education. The existence of valid and reliable measurement instruments is the main prerequisite in conducting a comprehensive evaluation of students' mathematical communication skills. Improper instruments can produce biased and inaccurate data, thus hindering efforts to improve the overall quality of mathematics learning.

Various previous studies have examined the importance of mathematical communication skills in mathematics learning and identified various problems related to its measurement. Suhenda & Munandar (2023) identified that the mathematical communication skills of grade VIII students are still diverse, with high-ability students able to recognize and use mathematical terms and symbols even though they are not perfect, while low-ability students experience more significant difficulties in mastering mathematical communication indicators. The research indicates the need for measurement instruments that can accurately identify the level of ability. La'ia & Harefa (2021) shows a positive and significant relationship between mathematical problem-solving skills and students' mathematical communication skills, which indicates that the two abilities support each other in the mathematics learning process. These findings emphasize the importance of developing instruments that can measure mathematical communication capabilities holistically and integrate with other mathematical aspects. Moreover Erayani et al. (2022) It found that there was a positive correlation between confidence and mathematical communication skills on mathematics learning outcomes, suggesting that affective aspects also play a role in the development of mathematical communication skills. The results of this study imply

that the instrument for measuring mathematical communication ability must be able to accommodate the cognitive dimension while considering the factors that affect its performance. Hanisah & Noordyana (2022) reported that students of grade VIII MTs have not mastered all indicators of mathematical communication ability, especially in expressing mathematical ideas orally using the right mathematical language. Purnamasari & Afriansyah (2021) Identify that most students have difficulty explaining the idea of the problem orally or in writing, even though they are able to describe the problem in the form of diagrams. This condition indicates a gap between visual representation ability and verbal communication ability that requires comprehensive measurement instruments and sensitivity to various aspects of mathematical communication.

However, the results of empirical research reveal that the mathematical communication skills of students at various levels of education are still relatively low, and this problem is exacerbated by the limitations of available measurement instruments. Wahyuni (2022) develop mathematical communication ability test instruments for relational and functional materials with an adequate level of validity and reliability, but these instruments are limited to the learning context at the junior high school level and cannot be applied directly to different levels of education. These limitations suggest that the development of mathematical communication ability measurement instruments must take into account the specific characteristics of each level of education. Yuliani et al. (2022) developing strategy-based learning tools *REACT* to improve mathematical communication skills, however, the focus is on the development of learning tools rather than measurement instruments that can be used for evaluation. Dalimunthe et al. (2022) develop an interactive learning model based on *think pair share*, which demonstrated effectiveness in improving mathematical communication skills, but did not produce measurement instruments that could be widely used by teachers in daily learning practice. Masitoh & Aedi (2020) and Kusumo & Nindiasari (2024) has developed an instrument development model that has proven to be effective, but the context of its application has not specifically targeted the characteristics of learning in Madrasah Ibtidaiyah. This gap shows that previous studies have focused more on identifying the level of mathematical communication ability or the development of learning models, but no one has specifically developed measurement instruments that suit the needs of Madrasah Ibtidaiyah. In fact, proper measurement is a crucial first step in efforts to improve students' mathematical communication skills.

The characteristics of Madrasah Ibtidaiyah students have their own uniqueness that distinguishes them from other levels of education, so they require specific and appropriate measurement instruments. Madrasah Ibtidaiyah students are in the age range of 6-12 years old with a level of cognitive development that is still in the concrete operational stage according to Piaget's theory, where they can more easily understand mathematical concepts through concrete objects and visual representations. Their verbal communication skills are still in the developmental stage, with limited mathematical vocabulary and simple sentence structure, so the measurement instruments must accommodate these limitations by using language appropriate to their developmental level. The learning system at Madrasah Ibtidaiyah also integrates Islamic values in every subject, including mathematics, so

that the context of the questions in the measurement instrument should use familiar situations and in accordance with the cultural and religious background of the students. In addition, Madrasah Ibtidaiyah students have limited writing skills, especially in expressing complex ideas in writing, so measurement instruments need to provide alternative means of communication such as images, symbols, or other simple representations. The duration of students' concentration at this age is also relatively short, so the measurement instrument must be designed taking into account the practicality and time efficiency aspects. This fundamental difference suggests that mathematical communication ability measurement instruments developed for higher levels cannot be directly adopted without significant adaptation. The need for instruments specific to Madrasah Ibtidaiyah became even more urgent given the importance of establishing a strong foundation of mathematical communication skills from an early age. The right instrument will allow teachers to accurately identify students' ability levels and design learning strategies that suit individual and classical needs.

Based on the gap analysis, the main problem faced is the lack of a valid and reliable test instrument to measure the mathematical communication skills of Madrasah Ibtidaiyah students specifically. This condition has an impact on teachers' difficulties in conducting a comprehensive evaluation of students' mathematical communication skills, so that the process of identifying students' weaknesses and strengths in the aspect of mathematical communication becomes less optimal. The absence of appropriate measurement instruments also hinders teachers' efforts in designing learning strategies that are in accordance with the level of students' mathematical communication skills. A further impact is the difficulty in monitoring and evaluating the development of students' mathematical communication skills on an ongoing basis, which can ultimately hinder efforts to improve the quality of mathematics learning at the Madrasah Ibtidaiyah level as a whole. In addition, the absence of standard instruments that can be used consistently causes inconsistencies in the assessment of mathematical communication skills between teachers and between schools, making it difficult to compare and standardize learning quality. This problem also has an impact on the difficulty in identifying students who need special intervention in the development of their mathematical communication skills. The complexity of this problem demands a comprehensive solution through the development of measurement instruments that are specifically designed in accordance with the characteristics and needs of Madrasah Ibtidaiyah students. The instruments developed must meet the criteria of validity, reliability, and practicality in order to be used effectively in the context of daily learning at Madrasah Ibtidaiyah.

Based on this background, this study aims to develop a mathematical communication ability test instrument specifically designed for Madrasah Ibtidaiyah students by adapting the theoretical framework of mathematical communication from *National Council of Teachers of Mathematics* (NCTM) which includes the ability to express mathematical ideas, use visual representations, and apply mathematical language. The developed instruments will integrate development approaches that have proven to be effective from research (Masitoh & Aedi, 2020) and (Kusumo & Nindiasari, 2024) with special adaptations for the

context of Madrasah Ibtidaiyah. This research aims to produce a mathematical communication ability test instrument that:

1. Valid based on expert assessment (*content validity*) by referring to indicators of mathematical communication skills that have been tested and adjusted to the characteristics of cognitive development of Madrasah Ibtidaiyah students at the concrete operational stage;
2. Have empirical validity (*construct validity*) through field trials involving Madrasah Ibtidaiyah students with diverse social, economic, and academic characteristics to ensure that the instrument can be used universally;
3. It has high reliability with a *Cronbach's Alpha* coefficient of at least 0.70 which allows the instrument to be used consistently and stably in various learning contexts at Madrasah Ibtidaiyah;
4. Practical and easy to use by Madrasah Ibtidaiyah teachers in daily learning evaluation activities without the need for complex special training, with clear usage guidelines and a simple but comprehensive scoring system.

The resulting instrument is expected to be an effective tool for teachers in evaluating and developing students' mathematical communication skills more precisely and in a more targeted manner, as well as facilitating the early identification of students who require special intervention in the aspect of mathematical communication. This research is also expected to make a theoretical contribution to the development of instruments for measuring mathematical communication skills, especially for basic education levels with learning characteristics based on Islamic values, as well as providing practical benefits for improving the quality of mathematics learning in Madrasah Ibtidaiyah as a whole. In addition, this instrument is expected to be a reference for further research in developing other mathematical ability measurement instruments that are in accordance with the characteristics of education in the madrasah environment. The results of this research are also expected to support the government's efforts to improve the quality of mathematics education in madrasas through the provision of valid, reliable, and practical evaluation instruments for widespread use by educators throughout Indonesia.

Research Methods

Research Design

This study uses the Research and Development (R&D) with a quantitative descriptive method that aims to produce products in the form of test instruments to measure the mathematical communication skills of Madrasah Ibtidaiyah students. The development model used adopts the Plomp model which consists of three main phases: initial investigation (Preliminary research), prototype phase (prototyping phase), and the assessment phase (Assessment phase) (Nieveen & Plomp, 2018). This model was chosen because it provides a systematic framework with continuous feedback (Feedback loop) which allows the instrument to be improved iteratively to achieve optimal quality. The constructivism paradigm is implemented through the preparation of questions that facilitate students in constructing mathematical

understanding through diverse representations and active communication of mathematical ideas.

Research Subject and Location

The research subjects consisted of 39 students of class V in MIN 1 Cilegon City, Banten Province, who were selected using the *purposive sampling* based on the criteria: have studied the material of Small Commons Multiples (KPK) and Large Common Factors (FPB), have adequate reading and writing skills, and are willing to participate in research. The number of subjects of 39 learners meets the minimum requirements for the psychometric analysis of the instrument, referring to the minimum recommendation of 30-50 subjects for a limited trial in the development of educational instruments (Scott, 2020). The selection of class V is based on the consideration that the learner at that level already has sufficient cognitive ability to express mathematical ideas in writing and orally.

Research Variables and Indicators

The main variable of the study is the mathematical communication ability of students which is operationalized through four indicators based on the National Council of Teachers of Mathematics (NCTM) standards: (1) the ability to organize and consolidate mathematical thinking through written communication (*organize and consolidate mathematical thinking through communication*); (2) the ability to communicate mathematical thinking coherently and clearly using mathematical representations (*communicate mathematical thinking coherently and clearly using mathematical representations*); (3) the ability to analyze and evaluate the mathematical strategies and thinking of others (*analyze and evaluate mathematical thinking and strategies of others*); and (4) the ability to use mathematical language to express mathematical ideas appropriately (*use the language of mathematics to express mathematical ideas precisely*) (Auliya et al., 2020). Each indicator was measured using an ordinal scale with a score range of 0-4 based on a holistic scoring rubric.

Instrument Development Procedure

The initial investigation stage includes needs analysis through literature studies and observation of mathematics learning to identify the characteristics of mathematical communication skills that are relevant to the context of madrasah ibtidaiyah. The preparation of the question grid is carried out based on the basic competencies of the Independent Curriculum for KPK and FPB materials by considering the revised Bloom taxonomy. The prototype stage begins with the construction of a description question item that refers to the grid, followed by expert validation using the Aiken's V formulation with the formula

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (1)$$

Where s is the validator score minus the lowest score, n is the number of validators, and c is the highest score that can be given. The validity of the content is declared valid if the value of $V \geq 0.60$ (Suhenda & Munandar, 2023). Limited trials were carried out on the research subjects to obtain empirical data. The assessment

stage includes the quality analysis of the question items including the level of difficulty using the formula

$$P = \frac{B}{JS} \quad (2)$$

Differentiating power with formula

$$D = \frac{BA}{JA} - \frac{BB}{JB} \quad (3)$$

And reliability using the Kuder-Richardson formula of 20 with the formula

$$r_{11} = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_{\text{item}}^2}{\sigma_{\text{total}}^2} \right) \quad (4)$$

Data Collection and Analysis Techniques

Data collection was carried out through three methods: expert validation using an assessment sheet containing aspects of the indicator's conformity with NCTM standards, language clarity, and accuracy of the context of the questions; the implementation of the test in controlled conditions with a duration of 90 minutes; and unstructured interviews to understand learners' difficulties. Data analysis was carried out in stages with interpretive criteria referring to Guilford (1956) for reliability: high ($r \geq 0.70$), adequate ($0.40 \leq r < 0.70$), and low ($r < 0.40$). The level of difficulty was categorized as easy ($P > 0.70$), medium ($0.30 \leq P \leq 0.70$), and difficult ($P < 0.30$), while the differentiating power was categorized as very good ($D \geq 0.40$), good ($0.30 \leq D < 0.40$), fair ($0.20 \leq D < 0.30$), and bad ($D < 0.20$) (Saragih et al., 2021). Data triangulation (*Data Triangulation*) is carried out by comparing the results of quantitative analysis and qualitative data to strengthen the validity of the findings and ensure that the quality of the developed instrument meets the required psychometric standards.

Results and Discussion

Research Results

Validity of Mathematical Communication Ability Test Instruments

The process of developing a mathematical communication ability test instrument for Madrasah Ibtidaiyah students has produced four descriptive questions that meet the standards of validity and reliability based on comprehensive statistical analysis. The validity of the instrument was analyzed through two main approaches: the validity of the content using expert assessment with the Aiken's V formula and the validity of the construct through the calculation of Pearson's *product moment* correlation in the data of the trial results was limited to 39 students of class V MIN 1 Cilegon City. The use of Pearson's *product moment correlation* in this study was based on the consideration that although the rubric score used an ordinal scale (0–4), the distribution of the test results showed a pattern close to the normal distribution with adequate sample sizes ($N = 39$). The selection of Pearson's correlation technique is also based on the nature of the interval of the total score which is the sum of the score of the item, so that it can be treated as continuous data for psychometric analysis. This is in line with the practice of developing instruments

in educational research that have been empirically validated. Based on Table 1, the results of the construct validity analysis show that the four question items have a very high correlation coefficient with consecutive correlation values of $r_1 = 0.820$, $r_2 = 0.864$, $r_3 = 0.859$, and $r_4 = 0.867$. The overall correlation value is above the table $r = 0.316$ at a significance level of 0.05 with $N = 39$, so that all question items are declared statistically valid. The data in the table shows a high consistency between each question item and the measured mathematical communication ability construct, indicating that the instrument is able to measure mathematical communication skills with optimal precision. The following are the results of the validity analysis of the question items presented in Table 1.

Table 1. Results of analysis of the validity of question items

Question Item	Correlation Coefficient (r_{xy})	r Table	Category	Status
X1	0,82	0,316	Very High	Valid
X2	0,864	0,316	Very High	Valid
X3	0,859	0,316	Very High	Valid
X4	0,867	0,316	Very High	Valid

Based on Table 1. validity categorization using correlation coefficient interpretation showed that the four question items were included in the category of very high validity, indicating that the developed instrument was able to measure students' mathematical communication skills with optimal accuracy. These findings are in line with research that shows the importance of construct validity in the development of mathematical ability evaluation instruments. The high validity of the instrument also supports the findings Yamtinah et al. (2021), which emphasizes the importance of developing valid mathematical communication instruments to accurately measure students' abilities. The success of the development of high-validity instruments in this study is in line with the findings Atikah et al. (2021), who develop *E-Module Elementary School Mathematics Based Approach Realistic Mathematics Education* with a validity rate of 93.8% for the material aspect and 93% for the design and language aspect. The consistency of high validity results in various types of mathematics learning instruments indicates that a systematic approach in the development of instruments, both in the form of conventional tests and digital media, can produce products that meet academic feasibility standards.

Test Instrument Reliability

The reliability analysis of the instrument using *Cronbach's Alpha* method yielded a reliability coefficient of $r_{11} = 0.861$ which is above the r table = 0.316. This value indicates that the instrument has a high level of internal consistency and can provide stable measurement results when used under the same conditions. The interpretation of the reliability coefficient shows that the instrument is included in the category of reliability with a high level of confidence to measure the mathematical communication ability of learners. Based on Table 2, the reliability calculation also considers item variance and total variance in the analysis. The item variance for each question item shows a proportional distribution, with the total number of item variances of 3,729 and the total variance of 10,518. The relatively

small ratio of item variance to total variance indicates that each question item contributes proportionally to the variability of the total score, indicating good internal consistency between the question items. The following instrument reliability analysis results are presented in Table 2.

Table 2. Instrument reliability analysis results

Analysis Components	Value
Variance Item X1	0,506
Variance Item X2	0,887
Variance Item X3	0,818
Item Variance X4	1,518
Total Item Variance	3,729
Total Variance	10,518
Reliability Coefficient (r_{11})	0,861
r Table	0,316
Conclusion	Reliable

Based on Table 2. high level of reliability in line with research Hotimah et al. (2021) which develops interactive multimedia to improve mathematical communication capabilities with effective results. The high reliability of the instrument ensures that the measurement of mathematical communication ability can be carried out consistently and reliably for the purposes of learning evaluation and further research. The study also showed that instruments with an Alpha Cronbach reliability of 0.708 have been considered reliable to measure mathematical communication skills, so the value of 0.861 in this study indicates a very adequate level of reliability.

Difficulty Level of Question Item

The difficulty level analysis using the formula $P = B/JS$ yields a varying difficulty index for each question item. Based on Table 3, the first question item (X1) has a difficulty index $P = 0.654$ which is included in the medium category, indicating that the question can be done well by most students but still provides adequate challenges. The second question item (X2) has a difficulty index $P = 0.159$ which belongs to the very difficult category, indicating that the question requires a high level of mathematical communication skills. The following analysis of the level of difficulty of question items are presented in Table 3.

Table 3. Results of analysis of the level of difficulty of question items

Question Item	True Amount	Difficulty Index (P)	Category
X1	102	0,654	Keep
X2	99	0,159	Very Difficult
X3	111	0,178	Very Difficult
X4	96	0,154	Very Difficult

Based on Table 3. the third (X3) and fourth (X4) question items have difficulty indices of $P = 0.178$ and $P = 0.154$ respectively, both of which fall into the very difficult category. The distribution of the difficulty level that tends to be difficult in the three question items indicates that the instrument is designed to measure mathematical communication skills at a high cognitive level, in accordance with the demands of developing students' critical thinking and analytical skills. The following students' answers based on indicators are presented in Table 4.

Table 4. Analysis of students' answers based on indicators

Indicators	Question Item	Response Characteristics	Success Rate
Organizing mathematical thinking	X1	The majority of students can use the familiar approach (multiples)	65,40%
Coherent communication with representation	X2	Difficulties in systematic visual representation	15,90%
Mathematical strategy analysis	X3	Difficulty interpreting abstract concepts	17,80%
Use of mathematical language	X4	Difficulty crafting logical and systematic arguments	15,40%

Based on Table 4. results of the analysis of the students' responses showed that question number 1 was included in the category of moderate difficulty because the majority of students were able to solve it correctly through an explicit and familiar approach, such as compiling multiples and determining the KPK. Students can recognize multiple patterns and infer results with basic logic, although explanations are still verbally limited. In contrast, questions 2, 3, and 4 are categorized as very difficult because they require high-level thinking skills, complex data representations, and deeper logical interpretation. The variation in difficulty levels found in this study is in line with findings that emphasize the importance of developing mathematical communication skills through questions of varying levels of difficulty in elementary school. Varying levels of difficulty allow the instrument to accommodate a broad spectrum of students' abilities, from basic abilities to advanced abilities, thus providing comprehensive diagnostic information about the learner's mathematical communication ability profile.

Differentiating Power of Question Items

The differentiating power analysis used the formula $D = (BA/JA) - (BB/JB)$ by dividing the study subjects into the upper group (top 27%) and the lower group (bottom 27%) based on the total score, referring to the quartile approach commonly used in psychometric analysis. Based on Table 5, the results of the analysis show a significant variation in differentiating power between the question items. The following differentiating power analysis of question items are presented in Table 5.

Table 5. Results of differentiating power analysis of question items

Question Item	Top Group Score (SA)	Lower Group Score (SB)	Differentiating Power (D)	Category
X1	3,273	1,909	0,341	Enough
X2	3,727	1,727	0,125	Drop
X3	3,818	1,909	0,119	Drop
X4	3,818	1,182	0,165	Drop

Based on Table 5. the first question item (X1) has a distinguishing power of $D = 0.341$ which is included in the sufficient category, showing adequate ability to distinguish between high-ability and low-ability students. This indicates that the first question can identify differences in mathematical communication skills between groups with good moderation. On the other hand, the second, third, and fourth question items each have a differentiating power of $D = 0.125$, $D = 0.119$, and $D = 0.165$, all three of which are included in the category that needs to be corrected (*drop*). The low differentiating power in the three question items is caused by several factors: first, the very high level of difficulty causes most students, both from the upper and lower groups, to experience the same difficulty in solving the problem; Second, the complexity of the scoring rubric that requires subjective assessment can cause variability in scoring; third, the characteristics of Madrasah Ibtidaiyah students who are still in the stage of developing written communication skills can affect the consistency of responses.

Score Distribution and Student Response Characteristics

Analysis of the score distribution showed a wide range of mathematical communication skills among learners. Based on the data obtained, the total score ranges from 4 to 16 with the maximum score that can be achieved is 16. The score distribution showed that there were 4 students who achieved the maximum score (S-4, S-10, S-14), indicating excellent mathematical communication skills. Meanwhile, 1 student (S-34) obtained the lowest score of 4, indicating mathematical communication skills that still need intensive development. The following distribution of total student scores are presented in Figure 1.

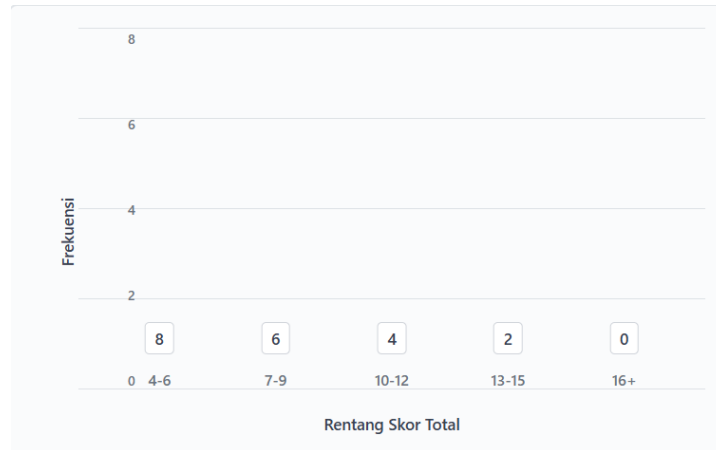


Figure 1. Distribution of total student scores

Based on Figure 1, the majority of students obtained scores in the range of 7-15 which showed mathematical communication skills at a moderate to good level. The relatively normal distribution of scores indicates that the instrument is able to accommodate a variety of students' abilities well, in line with research showing the importance of instruments that can measure mathematical communication skills at various levels. The following summary of response characteristics per question item are presented in Table 6.

Table 6. Summary of response characteristics per question item

Question Item	Dominant Response	Main Difficulties	Success Percentage
X1	Simple algorithmic strategy	Limited verbal explanations	65,40%
X2	Partial representation	Visual systematization	15,90%
X3	Weak conceptual understanding	Abstract relationship analysis	17,80%
X4	Unstructured arguments	Use of mathematical language	15,40%

Based on Table 6, the response pattern showed that the first question item (X1) tended to be answered well by most learners, while the second to fourth question items showed a greater variation in responses. Analysis of response patterns per question item showed unique characteristics for each indicator of mathematical communication ability, confirming research findings on the complexity of mathematical communication skills that cover various cognitive aspects.

Discussion

Key Research Findings

The results showed that the mathematical communication ability test instrument developed met the psychometric criteria with very high construct validity ($r > 0.80$) and high reliability ($\alpha = 0.861$). These findings confirm that the

mathematical communication skills of Madrasah Ibtidaiyah students can be measured objectively and consistently using systematically designed instruments. The varying distribution of difficulty levels (1 medium item, 3 very difficult items) reflects the hierarchical complexity of mathematical communication skills, where the ability to organize mathematical thinking is easier to master than the ability to use mathematical language precisely.

Factors Causing Research Results

The high validity and reliability of the instrument is due to several key factors: first, the use of NCTM's tested theoretical framework provides a solid conceptual foundation in the operationalization of indicators; second, a systematic expert validation process ensures the suitability of question items with the characteristics of cognitive development of Madrasah Ibtidaiyah students; third, the context of questions that integrate Islamic values increases *engagement* and students' understanding of problem situations. The low differentiating power in the three questions is due to cognitive complexity that is too high for the level of Madrasah Ibtidaiyah, causing most students to experience similar difficulties regardless of their basic abilities. The development factor of students' limited writing ability also contributes to the inconsistency of responses, especially in indicators that demand complex verbal elaboration.

Advantages and Disadvantages of Research

The advantages of this research include: the use of the *Research and Development* systematic with the Plomp model, the integration of Islamic contexts in accordance with the characteristics of Madrasah Ibtidaiyah, and the use of multiple validation approaches (validity of content and constructs) that strengthen the quality of the instrument. The resulting instruments also have high practicality with clear scoring rubrics and can be implemented by teachers without complex training. The study's shortcomings include: a limited sample size in one school which can limit generalizability, a focus on one mathematical topic (KPK and FPB) which limits the scope of measurement, and low differentiation in some questions that require revision before operational implementation. The limited duration of the test (90 minutes) can also affect a student's performance, especially given the limited concentration characteristics of elementary school age (Ginting et al., 2021).

Comparison with Previous Research

The results of this study are in line with the findings Nurdiana & Asmah (2021) which identifies variations in the mathematical communication skills of grade VIII students, but this study makes a specific contribution to the Madrasah Ibtidaiyah level that has not been explored in depth. Consistency with research La'ia & Harefa (2021) confirmed by the high correlation between question items that measure different aspects of mathematical communication, supporting the argument that mathematical communication ability is an integrated multidimensional construct. These findings also support the results of the study Erayani et al. (2022) about the importance of affective aspects in mathematical communication skills, which is reflected in the variation in students' responses to the context of familiar questions. However, the results of this study differ from the

findings Purnamasari & Afriansyah (2021) who reported better visual representation skills than verbal communication, where in this study both aspects showed a comparable level of difficulty. Compared to research Wahyuni (2022) who developed instruments for junior high school, this study provides specific adaptations for the characteristics of Madrasah Ibtidaiyah with the use of simpler language, familiar contexts, and scoring rubrics that are tailored to the writing ability of elementary level students. This is in line with research that emphasizes the importance of adapting instruments according to the level of education (Ats-Tsauri et al., 2021; Karsim et al., 2023).

Theoretical and Practical Implications

Theoretically, this study contributes to the development of a theory of measuring mathematical communication skills at the primary education level by confirming that the NCTM framework can be adapted to specific cultural and religious contexts without compromising the validity of the construct. This study also enriched the literature on the development of psychometric instruments for student populations with specific linguistic and cultural characteristics. Practically, the developed instruments provide objective tools for Madrasah Ibtidaiyah teachers in evaluating students' mathematical communication skills, facilitate the early identification of students who need special interventions, and support the development of differentiating learning strategies. The implementation of this instrument can improve the quality of *Assessment* Mathematics learning in madrasahs through the use of standardized and empirical evidence-based evaluation criteria. This research also provides policy implications in the development of mathematics learning evaluation standards in the madrasah environment, supporting the Ministry of Religion's efforts to improve the quality of mathematical education through the provision of instruments *Assessment* that are valid, reliable, and in accordance with the characteristics of madrasah students. This contribution is in line with the direction of the independent curriculum policy which emphasizes the importance of *Assessment* authentic and contextual (Arina & Nuraeni, 2022).

Research Limitations and Development Suggestions

This research has several limitations that need to be considered. First, the relatively small sample size (39 students) from one school limits the generalizability of the findings (Puspitaningtyas et al., 2025). Second, instruments are limited to the material of Small Commons and Large Commons Factors so it needs to be developed for other mathematical materials. Third, the limited duration of the test has not allowed for an in-depth evaluation of the students' mathematical communication process. Fourth, the potential for rubric assessment bias due to the evaluator's subjectivity in interpreting student responses. Fifth, the influence of teachers' backgrounds and learning environments has not been comprehensively explored. Based on these limitations, further research is recommended to expand the sample and research location, develop multi-material instruments, extend the duration of the test, develop more objective rubrics, and explore external factors that affect students' mathematical communication skills (Andini & Marlina, 2021; Rapsanjani & Sritresna, 2021).

Conclusions and Suggestions

Research on the development of mathematical communication ability test instruments for Madrasah Ibtidaiyah students succeeded in producing instruments that meet psychometric standards with very high construct validity (correlation coefficient $r > 0.80$ for all question items) and high reliability ($\alpha = 0.861$). The instrument developed consists of four questions based on the NCTM theoretical framework with an Islamic context, including indicators of organizing mathematical thought, coherent communication with representations, analysis of mathematical strategies, and the use of mathematical language. The distribution of difficulty level showed one question in the medium category (65.40%) and three questions in the very difficult category (15.90%-17.80%), reflecting the hierarchy complexity of mathematical communication skills. The differentiating power of the instrument showed that one question item was categorized as sufficient (0.341) while the other three items required improvement (0.119-0.165). The main findings confirm that the mathematical communication skills of Madrasah Ibtidaiyah students can be measured objectively and consistently, with the ability to organize mathematical thinking easier to master than the ability to communicate verbal and complex visual representations. This instrument makes a significant contribution as a valid, reliable, and contextual evaluation tool to support mathematics learning in the madrasah environment.

The next study is suggested to expand the scope of the research sample to multiple madrasas with variations in geographical and sociodemographic characteristics to increase the generalizability of the findings. The development of instruments needs to be extended to other mathematical materials such as geometry, measurement, and statistics by maintaining an Islamic context that has proven to be effective. Improving the differentiating power of question items is very difficult can be done through modifying the level of cognitive complexity and refining the scoring rubric that is more objective with specific response interpretation guidelines. The implementation of instruments in learning practice requires training of madrasah teachers on rubric scoring techniques and interpretation of assessment results to support data-based pedagogical decision-making. Longitudinal research is recommended to explore the development of mathematical communication skills of madrasah students throughout the basic education level, as well as the investigation of external factors such as learning methods, family environment, and teacher characteristics that affect mathematical communication skills. The development of digital versions of instruments with automatic scoring systems can also be a priority for future research to improve the efficiency and objectivity of mathematical communication skills assessment in the era of educational digitalization.

Reference

- Andini, S. F., & Marlina, R. (2021). Analysis of Mathematical Communication Skills of Junior High School Students in Solving Problems in Set Materials. *Calculation Formula*, 4(2), 343–354. <https://doi.org/10.22460/jpmi.v4i2.343-354>
- Arina, J., & Nuraeni, R. (2022). Mathematical Communication Skills of Vocational School Class X Students at Nurul Huda Boarding School. *Plusminus: Journal of Mathematics Education*, 2(2), 315–324. <https://doi.org/10.31980/plusminus.v2i2.1107>

- Atikah, N., Gistituati, N., Fitria, Y., & Syarifuddin, H. (2021). The validity of the Elementary School Mathematics E-Module is Based on the Realistic Mathematics Education (RME) approach. *Journal of Basicedu*, 5(6), 6103–6109. <https://doi.org/10.31004/basicedu.v5i6.1799>
- Ats-Tsauri, M. S., Cholily, Y. M., Azmi, R. D., & Kusgiarohmah, P. A. (2021). Mathematical Communication Ability-Based Relationships and Functions Module. *JNPM (National Journal of Mathematics Education)*, 5(1), 109. <https://doi.org/10.33603/jnpm.v5i1.3569>
- Auliya, N. H., Andriani, H., Fardani, R. A., Ustiawaty, J., Utami, E. F., Sukmana, D. J., Istiqomah, R. R., & Abadi, H. (2020). *Qualitative & Quantitative Research Methods*. CV. Science Library. <https://books.google.co.id/books?id=qijKEAAAQBAJ>
- Dalimunthe, S. A. S., Mulyono, M., & Syahputra, E. (2022). Development of Think Pair Share-Based Interactive Learning Model to Improve Students' Mathematical Communication Skills. *Journal of Scholars: Journal of Mathematics Education*, 6(1), 735–747. <https://doi.org/10.31004/cendekia.v6i1.1229>
- Erayani, F. N., Sridana, N., Arjudin, A., & Baidowi, B. (2022). The Relationship between Confidence and Mathematical Communication Skills with Mathematics Learning Outcomes. *Scientific Journal of the Education Profession*, 7(3c), 1875–1884. <https://doi.org/10.29303/jipp.v7i3c.845>
- Ginting, P., Hasnah, Y., Hasibuan, S. H., & Batubara, I. H. (2021). Evaluating Cognitive Level of Final Semester Examination Questions Based on Bloom's Revised Taxonomy. *AL-ISHLAH: Journal of Education*, 13(1), 186–195. <https://doi.org/10.35445/alishlah.v13i1.385>
- Hanisah, H., & Noordiyana, M. A. (2022). Students' Mathematical Communication Skills in Data Presentation Materials in Bojong Village. *Plusminus: Journal of Mathematics Education*, 2(1), 131–140. <https://doi.org/10.31980/plusminus.v2i1.1588>
- Hotimah, H., Ermiana, I., & Rosyidah, A. N. K. (2021). Development of Macromedia Flash-based interactive multimedia to improve mathematical communication skills. *Educational Progress*, 2(1), 7–12. <https://doi.org/10.29303/prospek.v2i1.57>
- Karsim, Ratnaningsih, N., & Rahayu, D. V. (2023). Development of Realistic Mathematics Education Based Flipbook Media on Students' Mathematical Communication Skills Development of Realistic Mathematics Education Based Flipbook Media on Students' Mathematical. *Journal of Master of Mathematics Education*, 5(1), 8–19. <https://doi.org/10.30598/jumadikavol5iss1year2023page8-19>
- Kusumo, B. J., & Nindiasari, H. (2024). Development of Numeracy Literacy Instruments at the Reasoning Level in a Personal Context at the Junior High School Level. *Pendas : Scientific Journal of Basic Education*, 9 (September), 648–657. <https://journal.unpas.ac.id/index.php/pendas/article/view/18495/7929>
- La'ia, H. T., & Harefa, D. (2021). The Relationship between Mathematical Problem-Solving Ability and Students' Mathematical Communication Skills. *Journal of Nonformal Education*, 7(2), 463. <https://doi.org/10.37905/aksara.7.2.463-474.2021>
- Masitoh, L. F., & Aedi, W. G. (2020). Development of Higher Order Thinking Skills (HOTS) Mathematics Assessment Instruments in Junior High School Grade VII. *Journal of Scholars: Journal of Mathematics Education*, 4(2), 886–897. <https://doi.org/10.31004/cendekia.v4i2.328>
- Nieveen, N., & Plomp, T. (2018). *Curricular and Implementation Challenges in Introducing Twenty-First Century Skills in Education* (pp. 259–276). https://doi.org/10.1007/978-3-319-65368-6_15
- Nurdiana, R., & Asmah, S. N. (2021). Development of Mathematical Representation Skills to Improve Students' Number Sense through Open Ended Based Questions. *Educational: Journal of Educational Sciences*, 3(3), 738–748.

- <https://doi.org/10.31004/edukatif.v3i3.426>
Purnamasari, A., & Afriansyah, E. A. (2021). Mathematical Communication Skills of Junior High School Students on the Topic of Data Presentation in Islamic Boarding Schools. *Plusminus: Journal of Mathematics Education*, 1(2), 207–222. <https://doi.org/10.31980/plusminus.v1i2.1257>
- Puspitaningtyas, D., Suryanti, Suprpto, N., & Julianto. (2025). Systematic Literature Review: Analysis of STEAM-Based Science Learning on Increasing Student Creativity. *Journal of Education*, 14(2), 2209–2220. <https://doi.org/10.58230/27454312.2259>
- Rapsanjani, D. M., & Sritresna, T. (2021). Mathematical Communication Skills Reviewed from Student Self-Efficacy. *Plusminus: Journal of Mathematics Education*, 1(3), 469–480. <https://doi.org/10.31980/plusminus.v1i3.954>
- Saragih, H. S., Saragih, R. M. B., & Matondang, K. (2021). Improving the Quality of Distance Learning (PJJ) Based on FOCQIPOTU. *Journal of Scholars: Journal of Mathematics Education*, 6(1), 94–106. <https://doi.org/10.31004/cendekia.v6i1.1018>
- Sugiyono. (2020). *Quantitative, qualitative, R&D research methods* (Issue January).
- Suhenda, L. L. A., & Munandar, D. R. (2023). Students' mathematical communication skills in mathematics learning. *Journal of Educatio FKIP UNMA*, 9(2), 1100–1107. <https://doi.org/10.31949/educatio.v9i2.5049>
- Wahyuni, V. (2022). Validity and Reliability of Communication Ability Test Instruments Mathematics Material, Relationships and Functions. *Sustainable Journal of Education Quality Studies*, 5(1), 89–99. <https://doi.org/10.32923/kjmp.v5i1.2232>
- Yamtinah, the great son of Ngurah Marhaeni's wife, and I Wayan Lasmawan. (2021). Development of Instruments for Essay Writing Skills and Creative Thinking Skills in Narrative Essay Materials for Grade IV Elementary School Students. *PENDASI: Journal of Indonesian Basic Education*, 5(1), 94–104. https://doi.org/10.23887/jurnal_pendas.v5i1.262
- Yuliani, E. N., Arnawa, I. M., Musdi, E., & Hidayat, A. (2022). Development of mathematics learning tools based on the REACT strategy to improve mathematical communication skills. *AXIOM: Journal of the Mathematics Education Study Program*, 11(1), 407. <https://doi.org/10.24127/ajpm.v11i1.4340>