Suska Journal of Mathematics Education
(p-ISSN: 2477-4758|e-ISSN: 2540-9670)
Volume 11, Issue 2, 2025, page 109-120

Development of Test Instruments for Mathematical Problem-Solving Ability in Algebra Content

Aghniya Syakira Mufidah, Maimunah*, Kartini

Program Studi Pendidikan Matematika, Universitas Riau, Pekanbaru, Indonesia
e-mail: *mimunah@lecturer.ac.id

ABSTRACT. Mathematical problem-solving ability (MPSA) is an important competence that supports the development of higher-order thinking skills. However, research shows that students' MPSA remains low, partly due to the lack of valid and reliable test instruments. This study aims to develop a MPSA test instrument in the form of essay questions on the content of algebraic forms, equations, and linear inequalities in one variable for seventh-grade students at SMP Negeri 13 Pekanbaru. The method used is development research with the Tessmer model, covering the preliminary and formative evaluation stages. The preliminary results indicate that mathematics learning in schools remains dominated by routine, result-oriented questions, while MPSA practice remains limited. Curriculum analysis indicates that the content of phase D algebra requires solving contextual problems, so the instrument was designed to measure the ability to understand problems, plan strategies, solve them, and interpret answers. The design phase produced Prototype I, consisting of 16 essay questions based on MPSA indicators and everyday contexts. Formative evaluation included self-evaluation, expert review, one-to-one, small group, and field tests. Self-evaluation and expert review showed that the instrument was highly valid (average score: 86.20%), with revisions to language, wording, illustrations, and instructions. One-to-one and small group tests produced Prototype II, which was then refined into Prototype III. Field testing with 37 students showed that 15 of 16 items were valid and reliable (coefficient 0.91 for package I; 0.85 for package II) and had good difficulty and discrimination levels, while one item was removed. Thus, the developed instrument is valid, reliable, and feasible for measuring students' mathematical problem-solving ability.

Keywords: algebra; mathematical problem-solving ability; test instrument development; validity and reliability

ABSTRAK. Kemampuan pemecahan masalah matematis (KPMM) merupakan kompetensi penting yang mendukung pengembangan kemampuan berpikir tingkat tinggi. Namun, penelitian menunjukkan bahwa KPMM siswa masih rendah, salah satunya karena belum tersedianya instrumen tes yang valid dan reliabel. Penelitian ini bertujuan mengembangkan instrumen tes KPMM berbentuk uraian pada konten bentuk aljabar, persamaan, dan pertidaksamaan linear satu variabel untuk siswa kelas VII SMP Negeri 13 Pekanbaru. Metode yang digunakan adalah *development research* dengan model Tessmer, meliputi tahap *preliminary* dan *formative evaluation*. Hasil *preliminary* menunjukkan bahwa pembelajaran matematika di sekolah masih didominasi soal rutin dan berorientasi hasil, sementara latihan KPMM masih terbatas. Analisis kurikulum mengindikasikan bahwa konten aljabar fase D menuntut penyelesaian masalah kontekstual, sehingga instrumen dirancang untuk mengukur kemampuan memahami masalah, merencanakan strategi, menyelesaikan, dan menafsirkan jawaban. Tahap desain menghasilkan prototipe I berupa 16 soal uraian berdasarkan indikator KPMM dan konteks sehari-hari. *Formative evaluation* meliputi *self-evaluation, expert review, one-to-one, small group,* dan *field test*. *Self-evaluation* dan *expert review* menunjukkan instrumen sangat valid (skor rata-rata 86,20%) dengan revisi pada bahasa, redaksi, ilustrasi, dan petunjuk. Uji *one-to-one* dan *small group* menghasilkan prototipe II, yang kemudian disempurnakan menjadi prototipe III. *Field test* pada 37 siswa menunjukkan 15 dari 16 butir soal valid, reliabel (koefisien 0,91 paket I; 0,85 paket II), serta memiliki taraf kesukaran dan daya pembeda yang baik, sedangkan satu soal dihapus. Dengan demikian, instrumen yang dikembangkan terbukti valid, reliabel, dan layak digunakan untuk mengukur KPMM siswa.

DOI:http://dx.doi.org/10.24014/sjme.v11i2.38500

Kata kunci: aljabar; kemampuan pemecahan masalah matematis; pengembangan instrumen tes; validitas dan reliabilitas

INTRODUCTION

The ability to solve mathematical problems is a core competence that every student should develop during the mathematics learning process. This ability is essential because it sharpens critical, creative, and reflective thinking skills, enabling students to address a range of problems in academic and daily settings. According to Agustami, Aprida, & Pramita (2021), strengthening mathematical problem-solving skills is crucial so that students become accustomed to facing a variety of challenging learning situations. In line with this view, Sri Septiani & Nurhayati (2019) stated that problem-solving is a key element in the discipline of mathematics, as each step in the solution process demonstrates deep mathematical thinking.

In implementing the Independent Curriculum, problem-solving skills are made a primary focus in mathematics learning. This includes the ability to understand problems, build mathematical frameworks or models, and interpret the results of solutions (Anggraena, Ginanto, Kesuma, & Setiyowati, 2025). This action aligns with the suggestions of the National Council of Teachers of Mathematics (NCTM), as described by Hafriani (2021), which state that mathematics learning should develop five core competencies: connecting concepts, reasoning, communicating, problem-solving, and representing ideas. Nevertheless, according to several international studies, such as TIMSS (Hadi & Novaliyosi, 2019) and PISA (OECD, 2023), Indonesian students' achievement in mathematical problem-solving remains relatively low. This situation indicates that problem-solving-oriented learning has not yet been optimally implemented in school settings.

In addition, the national assessment results show that many students still struggle to interpret information and analyze contextual problems (Yusuf & Ratnaningsih, 2022). Previous research by Risma & Isnarto (2019) and Soniawati (2022) showed that most middle school students completed only the initial steps of problem-solving according to Polya, while only a few reached the final stage. This situation is worsened by learning habits that still focus on routine textbook exercises rather than on contextual problems that challenge higher-order thinking skills (Fitri Rahmadhani, Heleni, & Yuanita, 2024).

Findings from the analysis of the Merdeka Curriculum mathematics textbooks for junior high school also show that around 81.3% of the questions presented still fall into the category of the mathematical world and are not yet contextual (Oktavia, Susanto, & Meifani, 2024). This means that students have not had enough exposure to questions that stimulate thinking skills for problem-solving. Based on observations in several junior high schools in Pekanbaru, many students still struggle to understand the meaning of questions, often make mistakes when using formulas, and feel less confident when faced with non-routine problems. The lack of contextual practice questions leaves students unaccustomed to applying mathematical concepts in real-world situations, even though this is crucial for improving mathematical problem-solving skills (Damayanti & Kartini, 2022).

The use of measurement tools, such as tests, is considered adequate for comprehensively assessing students' abilities to solve mathematical problems. According to Putra, Novaliyosi, Nindiasari, & Fathurrohman (2025), a well-designed test instrument not only assesses outcomes but also evaluates the thinking processes students engage in when working on problems. One type of instrument deemed suitable is the essay question, as it provides students with the opportunity to express their thoughts and solution strategies freely (Arikunto, 2021; Putri, Susiani, Wandani, & Putri, 2022). Therefore, developing valid, reliable essay tests with adequately high-quality items is a strategic step to support mathematics learning that emphasizes improving mathematical problem-solving skills.

The selection of algebra content in this study is based on its importance as a fundamental foundation in mathematics learning, serving as a basis for understanding advanced concepts.

Wibowo & Faizah (2021) stated that mastery of algebraic concepts helps students recognize patterns, understand relationships between objects, and build mathematical models to solve real-life problems. However, some studies have found that many students still struggle to understand and apply algebraic concepts in everyday situations (Endah, Kesumawati, & Andinasari, 2019). Therefore, the development of test instruments that effectively and comprehensively measure students' abilities in mathematical problem-solving grounded in algebraic content is needed.

Based on the explanation above, this study aims to develop a written test instrument on algebra content, specifically designed to measure students' mathematical problem-solving abilities in phase D. The resulting instrument is expected to meet the criteria of validity, reliability, and balanced levels of difficulty and discriminative power. In addition to serving as an accurate assessment tool, this instrument can be used by teachers to conduct both formative and summative assessments. This development is carried out because existing instruments generally still focus on multiple-choice formats, cover different material, or do not specifically assess problem-solving abilities in phase D. Therefore, this instrument is presented to fill the gap that previous research has not addressed and to provide a more precise tool for assessing students' mathematical problem-solving skills.

METHOD

This research falls under development research, which focuses on creating learning products that are valid, reliable, and practical to apply in real-world settings (Dewi & Syofiana, 2020). The product developed is an instrument for testing mathematical problem-solving ability (MPSA) in the form of essays on algebra material, intended for seventh-grade junior high school students.

The development model used adapts the Tessmer model, which comprises two main stages: preliminary and formative evaluation. The selection of this model is based on its advantages, including a systematic development process and empirical validity grounded in tiered testing stages ranging from expert assessment to field implementation. Thus, the final product produced is expected to have a thoroughly tested quality. Overall, the flow for developing the MPSA test instrument is shown in Figure 1.

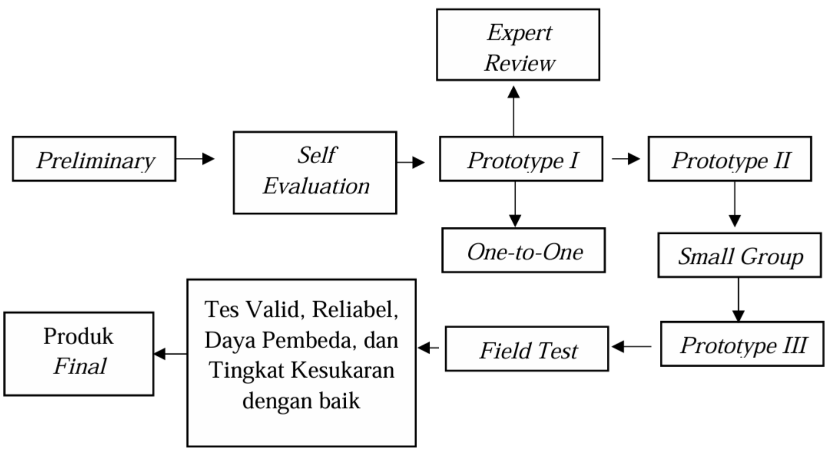

Figure 1. Development Flow of the MPSA Test Instrument Using the Tessmer Model

The preliminary stage of this study includes two main activities: analysis and design. The analysis stage aims to identify development needs, assess the feasibility of the product to be created, and determine the basic characteristics that the test instrument must have. Activities carried out at this stage include needs analysis, curriculum review, and student analysis to ensure that the developed test instrument aligns with the students' learning context. Based on the results of this analysis, the process continues to the design stage to prepare the initial product design.

The design stage focused on the preparation of the main components of the test instrument, which included the test blueprint, the development of test items for mathematical problem-solving abilities (MPSA), the creation of alternative solutions, the development of an assessment rubric aligned with the MPSA indicators, as well as the preparation of a validation sheet based on the essay test writing principles proposed by Ramadani & Handayani (2024). This initial design product then underwent an internal review process, including a self-evaluation stage conducted by the researcher and the supervising lecturer to ensure alignment between the research objectives, indicators, and the test instrument. The results of this stage produced Prototype I, which was then submitted for expert validation during the expert review stage.

At the expert review stage, Prototype I was validated by three validators based on language, construct, and material aspects. At the same time, a one-to-one stage was conducted with three seventh-grade students from SMP Negeri 13 Pekanbaru, with heterogeneous abilities, who had studied algebraic forms and one-variable linear equations and inequalities. Feedback from the validators and students was used to improve Prototype I, resulting in Prototype II, which will be tested during the small-group stage.

The small-group stage involved six seventh-grade students from class VII 2 at SMP Negeri 13 Pekanbaru, with varying abilities. At this stage, students were asked to complete the questions and provide feedback through direct interviews on the clarity of instructions, language, and the difficulty of the questions. The feedback obtained was used to improve the wording of the questions, clarify the illustrations, and adjust the context to make them easier to understand. The revisions resulted in prototype III, which was then tested more widely in the field test stage.

At the field test stage, Prototype III was administered to 37 seventh-grade students at SMP Negeri 13 Pekanbaru who had studied algebraic forms, equations, and linear inequalities in one variable. This trial aimed to obtain empirical data regarding the validity, reliability, difficulty level, and discriminative power of each test item.

Research data were collected through interviews, internal validation sheets, and essay tests. The research instruments included validation sheets and MPSA test sheets. Data analysis was conducted to assess internal validity, item validity (external validity), reliability, level of difficulty, and discriminating power.

Internal validity is established through expert assessments using a validation sheet that covers three main aspects: language, construct, and content. The average validity score of each question item is then interpreted based on validity categories, which are very valid for the interval $80<Va\leq100$, valid for the interval $60<Va\leq80$, reasonably valid for the interval $40<Va\leq60$, less valid for the interval $20<Va\leq40$, and invalid for the interval $0<Va\leq20$. These categories refer to the validity criteria proposed by Sa'adah, Rasmiwetti, & Linda (2019) and are used as a basis for determining the feasibility of each question item. Question items categorized as valid or very valid can proceed to the trial stage, while items categorized as less valid or invalid need to be revised before revalidation until they meet the valid or very valid criteria.

External validity was established through testing the instrument on students as research subjects. The trial data were analyzed in Microsoft Excel 2016 to empirically assess the quality of the MPSA test instruments by evaluating validity, reliability, difficulty level, and discriminative power. The instrument's reliability was calculated using Cronbach's Alpha with the same software. The reliability value is interpreted based on the reliability coefficient categories, namely very high in the interval $0.80<r_{11}\leq1.00$, high in the interval $0.60<r_{11}\leq0.80$, moderate in the interval $0.40<r_{11}\leq0.60$, low in the interval $0.20<r_{11}\leq0.40$, and unreliable if $r_{11}\leq0.20$. This category is used to determine the instrument's internal consistency before use in the next trial stage. An item is considered reliable if its reliability value falls into the high or very high category.

The level of difficulty is analyzed to determine the extent to which students can solve a test item. The difficulty level categories used in this study refer to the general criteria for item analysis: difficult if $0.00<TK\leq0.30$, moderate if $0.30<TK\leq0.70$, and easy if $0.70<TK<1.00$. A question is considered to have a good level of difficulty if it falls into the moderate or difficult category.

Meanwhile, the discriminating power was analyzed to determine the test item's ability to differentiate between high- and low-ability students. The categories of discriminating power in this study refer to (Hamzah, 2014), namely: very good for the interval 70<DP≤100; good for the interval 0.40<DP≤0.70; sufficient for the interval 0.20<DP≤0.40; poor for the interval 0.00<DP≤0.20; and very poor if DP≤0.00. A test item is considered suitable for use if it falls into the sufficient, good, or excellent categories. Meanwhile, items classified as very poor or poor are excluded from the MPSA test instrument.

Through these stages, it is expected that the developed test items meet the criteria of validity, reliability, proportionate difficulty, and optimal discriminatory power, so they can be used to accurately, comprehensively, and consistently measure students' mathematical problem-solving abilities.

RESULTS AND DISCUSSION

The development of a test instrument for mathematical problem-solving ability (MPSA) in the content of algebraic forms, linear equations, and inequalities with one variable is carried out through several systematic stages, following the Tessmer model, as explained below.

Preliminary Stage

The preliminary stage is the initial step in developing the MPSA test instrument, focusing on identifying field needs, conducting curriculum analysis, and developing the initial product design. This process consists of two main activities, namely analysis and design.

Analisis Stage

The analysis stage aims to obtain empirical and theoretical foundations for the development of the MPSA test instrument. The needs analysis was conducted through literature reviews and interviews with mathematics teachers at three schools. The interview results indicated that mathematics learning assessment in schools remains oriented toward final results rather than students' thought processes. The questions used are generally routine and taken from textbooks, so they are not yet capable of training higher-order thinking skills. Teachers also reported that limited references and a lack of contextual question examples are significant obstacles in developing mathematical problem-solving abilities (Fitri Rahmadhani dkk., 2024; Putra dkk., 2025).

Curriculum analysis shows that the Merdeka Curriculum phase D emphasizes students' ability to use variables to solve linear equations and inequalities, as well as to understand the basic concepts of algebraic forms in real-world contexts (Anggraena dkk., 2025). The review of learning outcomes indicates that seventh-grade students are expected to be able to solve problems involving algebraic forms and apply variables to model and solve single-variable linear equations and inequalities. These findings serve as the basis for determining the scope of the material, the types of context, and the cognitive demands for each test item. Thus, the test instruments developed remain relevant to the curriculum and support more targeted measurement of MPSA.

In addition to needs analysis and curriculum, the development of test instruments also takes into account the characteristics of the students who are the target of the test implementation. Seventh-grade students are generally aged 13–15 years and have entered the formal operational stage according to Piaget's developmental theory. At this stage, students begin to think abstractly, recognize patterns, and use symbolic representations to model mathematical situations. Therefore, designing instruments in the form of essay questions that require higher-order thinking skills is appropriate for students' cognitive capacity at this level. The analysis of the characteristics of seventh-grade students at SMP Negeri 13 Pekanbaru shows differences in academic ability, with high-, medium-, and low-ability groups. A mathematics teacher at SMP Negeri 13 Pekanbaru revealed that students' MPSA (mathematical problem-solving ability) has never been specifically mapped into the learning, so the development of this MPSA test instrument is expected to provide a comprehensive picture of variations in students' abilities. The initial trial was conducted with seventh-grade students in the even semester who had studied algebraic forms, linear equations, and inequalities in one variable.

Curriculum analysis shows that the Merdeka Curriculum phase D emphasizes the importance of mastering algebraic concepts to solve contextual problems by using variables and mathematical models (Anggraena dkk., 2025). Therefore, the developed instrument must be able to measure students' critical and strategic thinking abilities in authentic contexts. Meanwhile, the analysis of the characteristics of seventh-grade students at SMP Negeri 13 Pekanbaru shows differences in academic ability levels (high, medium, low) and difficulties in understanding non-routine problems. The initial trial was conducted on seventh-grade students in the even semester who had studied algebraic forms, linear equations, and inequalities in one variable.

Desain Phase

The design phase focuses on creating a preliminary draft of the instrument based on the results of the previously conducted needs analysis, curriculum, and student characteristics. At this stage, Prototype I is produced, which includes essential components such as a test grid, essay test items, answer alternatives, assessment rubrics, and validation sheets. The instrument is developed as an essay test consisting of 16 questions, divided into two sets of 8 questions each.

The formulation of each question is adjusted to four indicators of mathematical problem-solving ability (MPSA), namely: (1) understanding the problem, (2) planning the problem-solving process, (3) solving the problem, and (4) interpreting the obtained answers. Each question is designed for everyday life situations so that students can demonstrate their mathematical thinking skills more realistically and contextually.

The assessment rubric for the MPSA test instrument was developed based on a modification of guidelines from Purnamasari & Setiawan (2019) with different score ranges for each indicator, namely a score of 0–3 for the problem-understanding indicator, a score of 0–4 for the problem-solving planning and problem-solving indicators, and a score of 0–2 for the answer re-interpretation indicator. Using different score ranges adjusts the level of cognitive demand for each indicator. Meanwhile, a validation sheet was developed to assess three main aspects: content, construction, and language, which serve as a reference for experts in evaluating the quality of the test instrument. The results of the design stage produced a complete, systematic test instrument, ready to be further tested in the formative evaluation stage to assess the validity, reliability, difficulty level, and discriminatory power of each test item.

The assessment rubric was prepared by referring to the guidelines from Purnamasari & Setiawan (2019), using a score scale of 0 to 4 for each indicator. Meanwhile, the validation sheet was developed to assess three main aspects, namely content, construction, and language, which would later be used as a guide for experts to evaluate the quality of the test instruments. The result of this design stage is a complete and systematic draft of the test instrument, ready to be further tested in the next stage to assess the validity, reliability, difficulty level, and discriminative power of each test item.

Formative Evaluation Stage

Self Evaluation dan Expert Review

The self-evaluation stage is carried out by the researcher, together with the supervising lecturer, to assess the suitability of the indicators, the clarity of each question's wording, and the completeness of the assessment rubric. At this stage, 16 test items for MPSA were developed, divided into two sets of 8 questions each. Each question is reviewed to ensure that its content and cognitive demands address the four MPSA indicators: understanding the problem, planning problem-solving, solving the problem, and reinterpreting the obtained answers.

For example, one question concerns buying traditional cakes at a souvenir shop. Students are asked to identify the price relationship between two types of cakes, plan an appropriate mathematical model using a one-variable linear Equation, solve the Equation to determine the total cost, and then reinterpret the answer to fit the given purchasing situation. Questions like this demonstrate that each item involves understanding the context, modeling the situation mathematically, and solving the problem systematically. Once all items meet these criteria, prototype I is deemed suitable to proceed to the expert review and one-to-one stages.

Next, at the expert review stage, three Mathematics Education lecturers with doctoral qualifications were involved to validate the instrument. Based on the experts' assessments, the MPSA test instruments developed were highly valid, with an overall average score of 86.20%. The validation process covered three main aspects, namely content, construction, and language. The detailed validation results for each aspect are presented in Table 1.

Table 1. Results of the Validity Calculation of MPSA Test Instruments Package I and II

	Aspect		
	Material	**Construction**	**Languange**
Total Score	73,24	68,25	65,34
Mean	4,58	4,27	4,08
Persentage	91,60	85,40	81,60
Category	Highly Valid	Highly Valid	Highly Valid
Total Validity Percentage		86,20	Highly Valid

The assessment results indicate that the material aspect received the highest score, 91.60%, indicating that the content is aligned with the learning objectives and MPSA indicators and is relevant to the SMP/MTs level. The validity results for the material aspect indicate that the developed questions align with the question indicators, covering topics such as algebraic forms, linear equations, and inequalities in one variable. The content is relevant to the SMP/MTs level, learning objectives, and MPSA indicators. The questions are also considered free of SARA and PPPK elements, and they have apparent question limitations and answer scopes. These findings are consistent with Ramadhan, Siroj, & Afgani (2024), who stated that good material validity is achieved when each item represents the overall competence being measured.

The expert review provided several improvement recommendations. Validator I highlighted language aspects, including the use of foreign terms, punctuation, and consistency of wording. Validator II suggested that the questions should not be too direct and recommended adding supporting illustrations. Validator III emphasized the need to clarify the instructions and make some questions more contextual. Based on these suggestions, revisions were made to prototype I by improving the wording, adding illustrations to the stimulus, and refining the answer alternatives, resulting in prototype II, which is more communicative and aligned with the characteristics of junior high school students.

Validator I emphasized improvements in language aspects, such as consistency in writing foreign terms in italics, proper punctuation, and sentence structure to avoid ambiguity. Validator II provided feedback to clarify overly direct questions, giving students room to think when solving mathematical problems. The validators also suggested adjustments to illustration images, adding alternative solutions, and maintaining consistency in the use of terms such as "age" in both questions and answers. Meanwhile, Validator III assessed that the instructions were clear but needed to be made more specific for easier student understanding, and recommended that some questions be made more contextual by adding supporting illustrations and a variety of question forms. Based on the suggestions of the three validators, the researcher revised Prototype II by improving the wording of the questions, adding illustrations to the stimuli, consistently adjusting terminology, clarifying the instructions, and completing alternative solutions so that the test instrument became more communicative, valid, and suitable for the characteristics of junior high school students.

One-to-One

The one-to-one stage was conducted concurrently with an expert review to assess the instrument's readability and clarity. This test involved three seventh-grade students from SMP Negeri 13 Pekanbaru, with high, medium, and low abilities, as determined by their teacher's learning outcome data. After the students completed two sets of questions, a direct interview was conducted to obtain suggestions and comments regarding the question sentences, difficulty level, and clarity of instructions. Based on the interview results, some question sentences were considered too long and difficult to understand, particularly items numbers 2 and 8 in set 1 and number 7 in set 2. Additionally, students had difficulty creating mathematical models for part (b) of the questions because they were not yet accustomed to question formats that require step-by-step reasoning. One student also suggested adding illustrations to specific questions to clarify the context. Following up on this, the researcher revised the wording of the sentences, added visual illustrations to question 4, and clarified the instructions for solving the problems. These revisions resulted in Prototype II, which is easier for students to read and understand and is suitable for testing during the small-group stage.

Small Group

The small group stage tested the comprehensibility of the instrument on six seventh-grade students, consisting of two high-ability, two average, and two low-ability students, using prototype II. The students worked on the questions for 80 minutes per set and were interviewed afterward. The results showed that the students were beginning to understand the questions and develop appropriate problem-solving strategies, although there were still errors in calculations and interpretations. Based on students' feedback, several questions were revised to clarify their wording and make them more transparent. This revision produced prototype III, which is ready for field testing.

Field Test

The field test stage was conducted to examine the instrument's empirical validity using Prototype III with 37 seventh-grade students at SMP Negeri 13 Pekanbaru. The trial lasted 80 minutes per set of questions. Analysis of the MPSA test results was used to assess the validity, reliability, difficulty level, and item discrimination of the questions in the two test sets. Detailed results are presented in Table 2.

From the validity test results using the Pearson Product-Moment, it was found that all the test items were externally valid, with t-values exceeding the t-table value (t table = ta (df = 37-2) = 2.040) at the 95% significance level. The reliability test using Cronbach's Alpha yielded values of 0.91 for package I and 0.85 for package II, both in the very high range, indicating that the instruments are consistent and reliable. These findings align with Sugiono, Noerdjanah, & Wahyu (2020), who explained that an instrument is considered reliable if it yields stable data across time for the same respondents.

Analysis of the difficulty level shows that the questions in package I consist of four moderate-level questions, two easy questions, and two difficult questions, while those in package II consist of three moderate, three easy, and two difficult questions. This composition indicates a balance in difficulty levels and suggests high-quality items. This aligns with Fatimah & Alfath (2019), who stated that good questions are neither too easy nor too hard. Another opinion is that of Saputra & Yuhana (2024), who stated that questions with very low or very high difficulty levels provide functional variation in assessing a wider range of student abilities.

The results of the discriminatory power test showed that eight questions were categorized as good, seven as adequate, and one as poor. The item with low discriminatory power, question 2 in package II, was then removed to avoid reducing the instrument's overall quality. This aligns with Anderha & Maskar (2021), who stated that an item has good discriminatory power if the average discrimination index is above 0.20. Therefore, based on this expert opinion, one question had low discriminatory power, so the researcher removed question 2 from package II.

Table 2. Final Results of the MPSA Test Instrument Sets I and II

No	t (Validity Criteria)	Level of Difficulty	Differentiating Factor
Package I			
1	11,59 (Valid)	0,71 (Easy)	0,4 (Enough)
2	8,37 (Valid)	0,68 (Moderate)	0,42 (Good)
3	7,82 (Valid)	0,78 (Easy)	0,26 (Enough)
4	6,88 (Valid)	0,69 (Moderate)	0,34 (Enough)
5	7,17 (Valid)	0,63 (Moderate)	0,38 (Enough)
6	8,57 (Valid)	0,51 (Moderate)	0,58 (Good)
7	7,86 (Valid)	0,37 (Difficult)	0,55 (Good)
8	8,07 (Valid)	0,22 (Difficult)	0,42 (Good)
Package II			
1	5,73 (Valid)	0,75 (Easy)	0,22 (Enough)
2	3,28 (Valid)	0,70 (Moderate)	0,16 (Bad)
3	5,84 (Valid)	0,83 (Easy)	0,23 (Enough)
4	5,67 (Valid)	0,76 (Easy)	0,39 (Enough)
5	8,56 (Valid)	0,72 (Moderate)	0,45 (Good)
6	8,59 (Valid)	0,63 (Moderate)	0,58 (Good)
7	8,15 (Valid)	0,58 (Difficult)	0,53 (Good)
8	5,49 (Valid)	0,55 (Difficult)	0,58 (Good)

Overall, of the 16 developed items (8 questions per set), 15 were deemed valid and reliable, with difficulty levels and discriminative power considered good. These results indicate that the MPSA test instrument meets the criteria for validity, reliability, appropriate difficulty level, and optimal discriminative ability, making it suitable for accurately and consistently measuring students' mathematical problem-solving skills.

The advantage of this research lies in the empirical evidence that the developed instrument has undergone comprehensive trials up to the field test stage, unlike previous studies Ayasa, Yunita, & Juwita (2023) that only reached the initial validation stage. Implicatively, the results of this study can serve as a model for the development of similar assessment instruments for other educational levels or subjects, and at the same time, provide a reference for teachers in designing evaluations that focus on students' critical and creative thinking skills.

However, this study has limitations because it did not include contextual indicators on the validation sheet, so some questions still need adjustments to be more relevant to real-life situations. Therefore, it is recommended that future research incorporate these aspects to make the MPSA instrument development more comprehensive, applicable, and aligned with the principles of contextual learning.

CONCLUSION

This study developed an instrument for testing mathematical problem-solving ability (MPSA) in the form of essays on algebraic forms, equations, and one-variable linear inequalities, intended for seventh-grade junior high school students. The development process was carried out by applying the Tessmer model, which includes systematic preliminary and formative evaluation stages. Based on expert validation and field trials, the developed MPSA test instrument met the criteria for validity, reliability, and proportionality in terms of difficulty level and discriminative power. Expert validation yielded an average score of 86.20%, indicating very high validity. The reliability test results using Cronbach's Alpha showed values of 0.91 for package I and 0.85 for package II, both considered very high. Empirical analysis also showed that most of the test items had good levels of difficulty and discriminative power. Thus, the developed MPSA test instrument is suitable as a measurement tool to comprehensively and objectively assess the mathematical problem-solving abilities of middle school students. This instrument can help teachers evaluate students' thinking processes, not just the final answers. This study also opens opportunities for further development, such as refining the contextuality indicators to make the questions more realistic in everyday life situations. Additionally, this instrument can serve as a reference in developing similar measurement tools on other mathematics topics to strengthen students' conceptual understanding and higher-order thinking skills.

REFERENCES

Agustami, A., Aprida, V., & Pramita, A. (2021). Analisi Kemampuan Pemecahan Masalah Mtematis Siswa dalam Meneyelesaikan Soal Materi Lingkaran. *JPPM: Jurnal Prodi Pendidikan Matematika*, *3*(1), 224–231.

Anderha, R. R., & Maskar, S. (2021). Pengaruh Kemampuan Numerasi dalam Menyelesaikan Masalah Matematika terhadap Prestasi Belajar Pendidikan Matematika. *Jurnal Ilmiah Matematika Realistik*, *2*(1), 1–10. https://doi.org/10.33365/ji-mr.v2i1.774

Anggraena, Y., Ginanto, D. E., Kesuma, A. T., & Setiyowati, D. (2025). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, Jenjang Pendidikan Menengah Edisi Revisi Tahun 2025*. Jakarta: Badan Standar, Kurikulum, dan Asesmen Pendidikan (BSKAP).

Arikunto, S. (2021). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT. Bumi Aksara.

Ayasa, R. N., Yunita, A., & Juwita, R. (2023). Pengembangan Instrumen Tes Pemecahan Masalah Matematika pada Materi SPLDV. *JOEL: Journal of Educational and Language Research*, *3*(5), 207–214.

Damayanti, N., & Kartini. (2022). Analisis Kemampuan Pemecahan Masalah Matematis Siswa SMA pada Materi Barisan dan Deret Geometri. *Mosharafa: Jurnal Pendidikan Matematika*, *11*(1), 107–118. https://doi.org/10.31980/mosharafa.v11i1.691

Dewi, A. P., & Syofiana, M. (2020). Pengembangan Soal Kemampuan Penalaran Matematis Siswa Kelas VII SMP. *Gammath : Jurnal Ilmiah Program Studi Pendidikan Matematika*, *5*(2). https://doi.org/10.32528/gammath.v5i2.5359

Endah, D. R. J., Kesumawati, N., & Andinasari, A. (2019). Kemampuan Pemecahan Masalah Matematis Berdasarkan Self Efficacy Siswa Melalui Logan Avenue Problem Solving-Heuristic. *JNPM: Jurnal Nasional Pendidikan Matematika*, *3*(2), 207. https://doi.org/10.33603/jnpm.v3i2.2331

Fatimah, L. U., & Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. *Al-Manar: Jurnal Komunikasi dan Pendidikan Islam*, *8*(2), 37–64. https://doi.org/10.36668/jal.v8i2.115

Fitri Rahmadhani, S., Heleni, S., & Yuanita, P. (2024). Implementasi Model Problem Based Learning untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Peserta Didik. *Sigma: Jurnal Pendidikan Matematika*, *16*(2), 544–556. https://doi.org/10.26618/sigma.v16i2.16348

Hadi, S., & Novaliyosi, N. (2019). TIMSS Indonesia (Trends in International Mathematics and Science Study). *Prosiding Seminar Nasional & Call For Papers*, 562–569.

Hafriani, H. (2021). Mengembangkan Kemampuan Dasar Matematika Siswa Berdasarkan NCTM Melalui Tugas Terstruktur dengan Menggunakan ICT (Developing The Basic Abilities of Mathematics Students Based on NCTM Through Structured Tasks Using ICT). *Jurnal*

Ilmiah Didaktika: Media Ilmiah Pendidikan dan Pengajaran, *22*(1), 63–80. https://doi.org/10.22373/jid.v22i1.7974

Hamzah, A. (2014). *Evaluasi Pembelajaran Matematika*. Depok: PT Raja Grafindo Persada.

OECD. (2023). PISA 2022 Assessment and Analytical Framework. In *PISA*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

Oktavia, T. A., Susanto, H. P., & Meifani, N. I. (2024). *Analisi Soal Latihan Buku Teks Matematika Kelas VII Kurikulum Merdeka ditinjau dari Literasi Matematika*. STKIP PGRI Pacitan.

Purnamasari, I., & Setiawan, W. (2019). Kemampuan Pemecahan Masalah Matematis Siswa SMP pada Materi SPLDV Ditinjau dari Kemampuan Awal Matematika (KAM). *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, *3*(2), 207–215. https://doi.org/10.31331/medivesveteran.v3i2.771

Putra, D. D., Novaliyosi, N., Nindiasari, H., & Fathurrohman, M. (2025). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Matematis Siswa Sekolah Menengah Kejuruan. *Edukasiana: Jurnal Inovasi Pendidikan*, *4*(3), 967–983. https://doi.org/10.56916/ejip.v4i3.1610

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, *4*(2), 139–148.

Ramadani, E. N., & Handayani, D. F. (2024). Instrumen Penilaian Hasil Pembelajaran Kognitif Pada Tes Objektif. *Jupendis: Jurnal Pendidikan dan Ilmu Sosial*, *2*(4), 86–96. https://doi.org/10.54066/jupendis.v2i4.2159

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975. https://doi.org/10.31004/joe.v6i2.4885

Risma, A., & Isnarto, H. (2019). Kemampuan Pemecahan Masalah Matematika dalam Menyelesaikan Soal Cerita Berdasarkan Langkah Polya. *Prosiding Seminar Nasional Pascasarjana*, *2*(1), 297–303.

Sa'adah, S. I., Rasmiwetti, R., & Linda, R. (2019). Pengembangan Soal HOTS dengan Wondershare Quiz Creator sebagai Media Display pada Materi Stoikiometri Kelas X. *JTK: Jurnal Tadris Kimiya*, *4*(2), 177–188. https://doi.org/10.15575/jtk.v4i2.5469

Saputra, R., & Yuhana, Y. (2024). Pengembangan Instrumen Tes Matematika Berbasis Higher Order Thinking Skills untuk Siswa SMA. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *9*(1), 34–46. https://doi.org/10.31004/cendekia.v9i1.3313

Septiani, E. S., & Nurhayati, E. (2019). Analisis Kemampuan Pemecahan Masalah Matematis ditinjau dari Adversity Quotient (AQ) Peserta Didik melalui Model Problem Based Learning (PBL). *Prosiding Seminar Nasional & Call For Papers*, 168–175.

Soniawati, S. (2022). Analisis Kemampuan Pemecahan Masalah Siswa Kelas VII SMP Negeri 4 Cibinong Materi Bentuk Aljabar dengan Problem Based Learning. *JPMI: Jurnal Pembelajaran Matematika Inovatif*, *5*(5), 1341–1350. https://doi.org/10.22460/jpmi.v5i5.10566

Sugiono, S., Noerdjanah, N., & Wahyu, A. (2020). Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation. *Jurnal Keterapian Fisik*, *5*(1), 55–61. https://doi.org/10.37341/jkf.v5i1.167

Wibowo, T. E., & Faizah, S. (2021). Pengembangan Soal tes untuk Mengukur Kemampuan Pemecahan Masalah Siswa pada Materi Bentuk Aljabar. *Alifmatika: Jurnal Pendidikan dan Pembelajaran Matematika*, *3*(2), 145–158. https://doi.org/10.35316/alifmatika.2021.v3i2.145-158

Yusuf, R. M. M., & Ratnaningsih, N. (2022). Analisis Kesalahan Numerasi Peserta Didik dalam Menyelesaikan Soal Assesmen Kompetensi Minimum. *Jurnal Paedagogy: Jurnal Penelitian dan Pengembangan Pendidikan*, *9*(1), 24–33. https://doi.org/10.33394/jp.v9i1.4507