Journal of Philosophy and Research in Islamic Mathematics Education
J-PRIMa
Vol. 1, No. 1, 2025, pp 63–73 e-ISSN:
https://doi.org/10.33474/
online at: https://riset.unisma.ac.id/index.php/j-prima/article/view/25137

Development of test instruments to measure students' mathematical problem-solving skills

Samsul Ma'arif[1*]

[1]SMK Modern Ngawi
*Correspondence: 22201072002@unisma.ac.id

Received: 17-09-2025; Revised: 27-10-2025; Accepted: 17-11-2025

Abstract. *This study aims to develop a test instrument that can measure the ability to solve mathematical problems, focusing on the development procedures and the quality of the resulting instrument. The research method used is Research Development (research and development/R&D) with a development model based on the type of formative evaluation Tessmer, which involves the preliminary stage, self evaluation, prototyping (expert review, one to one, small group), and field tests. The test subjects in this study were 36 students of Class XI TKR Vocational High School. The test results showed that the mathematical problem solving test instruments have high validity, assessed based on the Coefficient Validity Ratio (CVR) of 1, and meet the criteria with the value of the Coefficient Validity Index (CVI) of 1, which is in the very appropriate category (0.68-1.00). In addition, this test instrument is also considered reliable with a value of 0.732, which is categorized as high and is in the range of 0.60<r11<0.80. The difficulty level of the test questions is categorized as good, with 60% of the questions having a good difficulty level and an average difficulty level of 0.58, which is in the medium category (0.31-0.70). The distinguishing power of the test instrument was considered good, with 93.34% of the questions meeting the criteria for distinguishing power, and the average distinguishing power of 0.35, categorized as sufficient (Dp>0.2). Thus, this student math problem solving test instrument can be considered to be of good quality based on the results of the tests carried out.*

Keywords: *development of test instruments, problem solving, mathematics.*

How to cite: Ma'arif, S. (2025). Development of test instruments to measure students' mathematical problem-solving skills. *Journal of Philosophy and Research in Islamic Mathematics Education (J-PRIMa)*, *1*(1), 63–73. Retrieved from https://riset.unisma.ac.id/index.php/j-prima/article/view/25137

Introduction

The learning process involves two main aspects, namely the learning and teaching processes. Learning, as the core of education, is a process in which positive changes occur in a person's behavior and knowledge (Idrus et al., 2023). Meanwhile, teaching is providing guidance to individuals during the learning process. The success of achieving educational goals is highly dependent on the learning experiences of students. To measure the effectiveness of a learning process, it is important to know the learning outcomes obtained by students. Evaluation is a necessary means to understand student learning outcomes. Therefore, to assess the quality of learning, it is necessary to understand student learning outcomes.

Evaluation is an action or assessment process carried out to assess the extent of the success of an educational program, teaching, or training that has been implemented. The results of the evaluation can be a source of motivation for teachers and students, encouraging them to be more diligent in learning and improving their thinking skills. Therefore, to obtain accurate evaluation results, measurements and assessments are carried out (Arif et al., 2022). Evaluation results are influenced by students' thinking abilities, which in turn affect their logical intelligence and numeracy skills, thereby maximizing and improving learning evaluation results (Rumiati, 2023).

The term commonly used for measuring instruments in research is research instruments. Instruments can be defined as devices used to collect data. Many experts have established certain standards or instruments to meet measurement needs in the social sciences. If, in the context of social research, measuring instruments are used as standards for measurement, then data collection needs to include reporting of the results of quality tests by individuals who validate the instruments.

The role of research instruments is significant in achieving the objectives of a study. Often, the weight or quality of a study is assessed based on the quality of the instruments used. This is reasonable, considering that research instruments are the means used to collect data to solve research problems or achieve research objectives. If the data obtained is inaccurate or invalid, the decisions made will be inappropriate. Discussions about research instruments in the social and educational context require differentiation, especially when comparing qualitative and quantitative research approaches. Validation is important to assess the material presented, student achievement, and student learning outcomes (Amhar, 2023).

Information about the learning process can be obtained through measurement activities, which are generally carried out using measuring tools or instruments. The most commonly used instrument in the context of measuring learning outcomes by teachers is tests. Tests are assessment methods designed and implemented for students at a specific time, place, and under specific conditions. As an assessment tool, tests consist of questions given to students in written format (written tests) or through actions (action tests). Generally, tests are used to evaluate and measure student learning achievements, particularly in terms of mastery of learning materials in accordance with predetermined objectives (Ahmad Asyraful Anam, Surahmat 2, 2021). Written tests and interviews are used to measure students' abilities and review how learning is carried out according to the students' circumstances. By using interview tests, it is also possible to find out how students represent their learning abilities (Sobirin, 2023).

A test that is considered effective as a measurement tool must meet several requirements, namely it must be valid, reliable, objective, practical, and economical (Arikunto, 2010). The main functions of tests include: 1) assessing students' ability levels, 2) measuring their growth and development, 3) diagnosing learning difficulties, 4) evaluating teaching outcomes, 5) determining learning achievements, and 6) providing encouragement for improving teaching and learning. Based on these characteristics, it can be concluded that tests play a significant central role because the quality of tests directly affects the accuracy of measuring student learning outcomes.

Mathematics is a field of study that emphasizes cognitive abilities. The cognitive domain involves mental activities or brain functions. All activities involving brain functions fall under the cognitive domain (Beginna Kurnia Marita, Zainal Abidin, 2019). The subject of mathematics covers all aspects of the cognitive domain, as students are expected to achieve various levels of ability in this subject, ranging from the ability to remember, understand, apply, analyze, to synthesize.

One very significant aspect of assessment in mathematics is problem-solving skills. The National Council of Teachers of Mathematics (NCTM, 1989) emphasizes the importance of problem-solving as an integral part of the mathematics curriculum, underlining that problem-solving should be the main focus in the design of the curriculum. Therefore, mathematical problem-solving skills are considered crucial and are made the main objective in mathematics learning. Problem solving is seen as a process that must be incorporated into all learning programs and provide a context in which mathematical concepts, principles, and skills are learned. Through mathematical problem-solving exercises, students are guided to understand certain concepts and principles of the material, as well as practice their skills in solving mathematical problems (Khusna et al., 2019).

Based on the PISA test results, Indonesia's position is still relatively low compared to other countries. The latest results from PISA 2015 show that Indonesian students' abilities are still at a low level, especially in the aspects of analyzing, reasoning, communicating effectively, and solving and interpreting problems in various contexts (Pulungan, 2014). PISA data ranks Indonesia 69th out of 76 countries participating in PISA 2015 in the PISA Mathematics category, with a score of 386 points. Therefore, efforts are needed to improve Indonesia's performance in mathematics, with a focus on improving students' mathematical problem-solving skills.

Based on mathematical problem-solving learning, it is hoped that students will experience situations where they use their existing knowledge and skills to solve new problems. By utilizing their prior knowledge, skills, and understanding, students are expected to be able to meet the demands of unusual situations. Therefore, the mathematics learning approach not only aims to improve students' ability to solve problems involving numerical calculations, but also aims to improve students' ability to solve contextual problems using mathematical concepts. Mathematical problem-solving questions are usually presented in the form of descriptive questions or story questions. Mathematical questions in descriptive form are considered more challenging for students than questions that only involve numbers or ordinary mathematical calculations (Rasid et al., 2021).

Problem solving is at the heart of mathematics. This statement is in line with the opinion of NCTM (1989), which asserts that problem solving is an integral part of mathematics learning and should not be separated from the learning process. The use of problem solving in the context of mathematics is not only to help students develop thinking skills, but also to help them hone basic skills in solving problems, especially in everyday situations. The thinking process in problem solving is considered an important aspect that requires special attention from educators, especially in helping students develop their ability to deal with problems.

Based on the previous explanation of mathematical problem solving as a skill in mathematics learning, the measurement of mathematical problem-solving skills can be done by analyzing each step applied in solving a problem. The evaluation of these abilities involves assessing students' abilities to understand problems by identifying their context, planning problem-solving strategies, implementing the planned strategies, and finally, double-checking by providing interpretations of solutions or conclusions. Problem solving with learning motivation affects student learning outcomes, by examining the components of learning that help students improve their learning outcomes (Widdah, 2023).

Solving essay questions involves several specific steps. According to Polya (1971), there are four steps that can be taken in solving mathematical problems, namely understanding the problem, designing a solution plan, implementing the plan, and rechecking all the steps that have been taken. Therefore, to tackle essay questions, the first step is to understand the problem contained in the question. Next, students plan how to solve the problem based on the concepts relevant to the question. After that, students implement the plan and, in the final stage, they double-check to ensure the correctness of each step taken.

Generally, in every school, the use of multiple-choice questions or questions that require the use of formulas without prior analysis still dominates. These types of questions are less effective in evaluating students' mathematical problem-solving abilities. The goal of problem-solving learning is important for compiling questions that meet problem-solving criteria. According to the guidelines for developing problem-solving questions, problem-solving questions should require students to: (1) use various procedures, where students are expected to relate their previous experiences to the given problem to find a solution, (2) involve manipulation or operation of previously acquired knowledge, (3) understand mathematical concepts and terms, (4) note similarities, differences, and analogies, (5) identify critical aspects and select relevant procedures and data, (6) note irrelevant details, (7) visualize and interpret quantitative facts or information about the location and relationship between facts, (8) make generalizations from the examples given, and perform estimates and analyses.

After analyzing the questions compiled by teachers at SMK Modern Ngawi for grade XI, it can be concluded that these questions generally use formulas directly. However, when viewed from the criteria for mathematical problem solving described above, these questions do not fully meet the standard criteria for measuring mathematical problem-solving skills.

Based on this, it is necessary to develop a more appropriate test instrument to measure problem-solving skills. For this purpose, the use of essay questions is considered more relevant, as they can provide an overview of the extent to which students understand and master the problems being tested. For example, Evy Yosita Silva, Zulkardi, and Darmawijoyo collaborated to develop a test instrument to measure mathematical problem-solving skills, with the research title

"Development of PISA Model Mathematics Questions on Uncertainty Content to Measure the Mathematical Problem-Solving Skills of Junior High School Students." 's research produced test instruments or questions that had a positive impact on students' mathematical problem-solving skills.

Therefore, based on these considerations, the researcher was interested in conducting research entitled "Development of Test Instruments to Measure Students' Mathematical Problem-Solving Skills".

Method

This research adopts the Research & Development (R & D) or development research type. The development method applied in this research is the Tessmer development model, which is known for its formative evaluation model. Tessmer's development involves four stages, namely the preliminary stage, self-evaluation, the prototyping stage which includes expert reviews, one-to-one, and small groups, and the field test stage. The following are the development procedure steps carried out in this study.

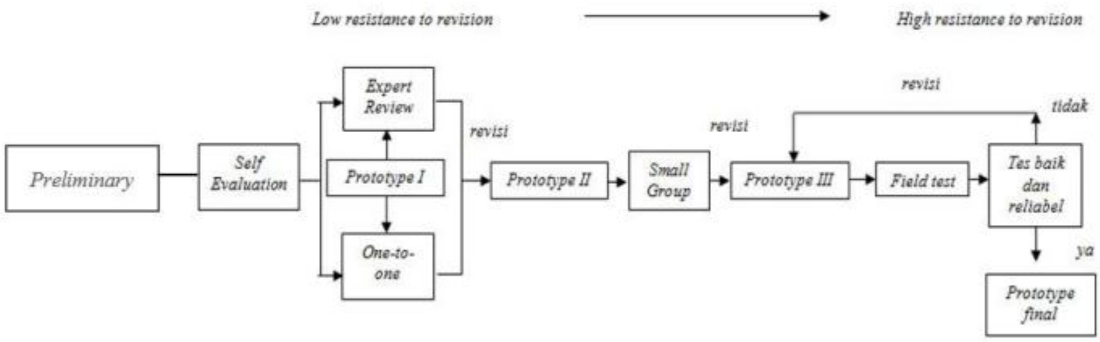

Figure 1. Tessmer's formative research model

The participants in this study were students in the 11th grade of TKR Modern Ngawi. The data collection approach involved the use of tests and questionnaires. The instruments used included test instruments, validation sheets, and questionnaires. The data analysis methods included validating the test instruments to measure problem-solving skills, analyzing student responses to the questionnaire, testing the reliability of the problem-solving skills test, analyzing the level of difficulty of the mathematical problem-solving skills test instruments, analyzing the discriminatory power of the mathematical problem-solving skills test instruments, and analyzing the results of the students' mathematical problem-solving skills data.

Results and Discussion

RESULTS

1. **Development Results**

a. Preliminary Stage

This process began with collecting several references related to the research, focusing on developing test instruments that could measure mathematical problem-solving skills. The next step involved determining the location and participants for the research trial. The research trial was conducted at SMK Modern Ngawi, while the trial participants consisted of 11th grade TKR students.

b. Self-Evaluation Stage

The self-evaluation stage is divided into two parts, namely the analysis stage and the design stage. The analysis stage consists of three aspects, namely curriculum analysis, student analysis, and material analysis.

1) Curriculum Analysis

Based on interviews with mathematics teachers at SMK Modern Ngawi, it can be concluded that the Merdeka curriculum has been implemented since its introduction.

2) Student Analysis

The focus of the student analysis was on 11th grade TKR students who were the subjects of the trial, as they had received the relevant material. There were 40 students in the class. Interviews with mathematics teachers revealed variations in ability among 11th grade TKR students at SMK Modern Ngawi, with some having low, medium, and high abilities.

3) Material Analysis

The focus of student analysis activities is on 11th grade TKR students who are the subjects of the trial, as they have received the relevant material. There are 40 students in this class. Interviews with mathematics teachers reveal variations in ability among 11th grade TKR students at SMK Modern Ngawi, with some having low, medium, and high abilities.

c. Prototyping Stage (Validation, Evaluation, and Revision)

1) Expert Review

The instrument validation process was carried out by providing the test grid validation sheet, test questions, and test answer keys to the validators. The validators consisted of two mathematics lecturers from Mathematics Education, namely Validator 1 (expert 1) and Validator 2 (expert 2), as well as one mathematics teacher from SMK Modern Ngawi, namely Validator 3 (expert 3). The validators provided their opinions on the prototype, which could be whether it could be used without revision, whether some components of the questions needed revision, or whether all components needed revision.

2) One to one

In addition to undergoing validation by experts, the test instruments for mathematical problem solving were also tested directly on several students at SMK Modern Ngawi. The three students involved in the trial were not included in the research subject group and included students with low, medium, and high ability levels. After that, the students were asked to fill out a student response questionnaire, the purpose of which was to evaluate the readability of the questions.

3) Small Group

The results of the revisions and comments provided through the expert review process and one-to-one trials formed the basis for designing questions in the next stage, which resulted in prototype 2. The prototype was then tested on a small group of non-research subject students consisting of six people. This group consisted of two high-ability students, two medium-ability students, and two low-ability students. The students were asked to answer the problem-solving questions and then fill out a prepared response questionnaire.

4) Field Test

The prototype that had undergone the validation and revision process (prototype 2) was tested on the research subjects, namely students of class XI TKR Modern Ngawi. Although the class was supposed to consist of 40 students, during the test, several students were absent, so the total number of students who became test subjects was 36. The students were asked to complete a mathematical problem-solving ability test consisting of 15 essay questions. At the beginning of the test, the researcher distributed the questions and answer sheets to the students.

2. Analysis of Mathematical Problem-Solving Ability Test Results

The results of the analysis of the students' mathematical problem-solving ability test results are shown below.

Table 1. Analysis of Students' Mathematical Problem-Solving Ability Test Results

Student Scores	Frequency	Percentage	Category
80	2	5.56	Very Good
60 - 79	14	38.89	Good
40–59	17	47.22	Fair
20 - 39	3	8.33	Insufficient
0 - 19	0	0	Very low
Number of Subjects	36	100	
Average Score	58		Fair

Based on the analysis of data on the mathematical problem-solving abilities of 36 test subjects, it can be concluded that 2 students (5.56%) belong to the group with excellent mathematical problem-solving abilities. A total of 14 students (38.89%) were classified as having good mathematical problem-solving skills, while 17 students (47.22%) were classified as having adequate mathematical problem-solving skills. Meanwhile, 3 students (8.33%) were classified as having poor mathematical problem-solving skills. It is important to note that no students were classified as having very poor mathematical problem-solving skills.

DISCUSSION

The development of a test instrument aimed at evaluating students' mathematical problem-solving abilities has gone through a series of stages in the Tessmer model. This process began with a preliminary stage, followed by self-evaluation, and culminated in the prototyping stage, which involved expert review, one-to-one, small group, and field testing with the aim of producing a high-quality test instrument. The final product of this process is a test instrument consisting of 15 questions designed to measure students' mathematical problem-solving abilities. Before the development process began, test instrument quality criteria were established as guidelines for assessing the success of the final product.

The preliminary stage is the first step in the development process, where researchers make preparations by looking for references related to the research to be carried out. At this stage, researchers also determine the school or research location and the research schedule.

The third stage in development is the prototyping stage, which consists of expert review, one-to-one, and small group. Validators or experts validate the mathematical problem-solving test instrument. The validation results are then tested through one-to-one. The results from the validators and one-to-one are used to revise prototype I into prototype II. Next, prototype II is tested in a small group trial. The results of the small group are used to make revisions, resulting in prototype III. Prototype III is tested in the final stage, namely the field test, which is carried out in class XI TKR SMK Modern Ngawi involving 36 students. Based on the results of the students' work, the reliability, difficulty level, and discriminating power of the test instrument can be determined.

The quality of the test instrument for measuring mathematical problem solving can be evaluated based on predetermined criteria. The results of the test validation showed a CVR value of 1 for each item and a CVI value of 1, indicating the validity of the test instrument. According to the criteria, a mathematical problem-solving test instrument is considered valid if it has a CVR ≥ 0.99 and a CVI≥ 0.34. Furthermore, the results of the one-to-one trial showed adequate question readability, with an average positive response from students of 75% and an average negative response from students of 25%. At the small group stage, the results of the response questionnaire showed an average positive response from students of 77.08% and an average negative response from students of 22.92%. In accordance with the criteria, the results of the student response questionnaire met the established standards, indicating that the readability of the questions was good. Therefore, the prototype II obtained could proceed to the field test stage.

Based on the results of the test instrument reliability analysis using the SPSS version 20 program, a value of 0.732 was obtained, which can be interpreted as high. This indicates that the problem-solving test instrument has a high level of reliability, so it can be relied upon to measure students' mathematical problem-solving abilities. Furthermore, the level of difficulty of the instrument is considered good if it is in the range of 0.31-0.70 with a moderate category.

Based on the results of the difficulty level analysis in the development of the problem-solving ability test instrument, it was found that 26.67% of the questions had an easy difficulty level. A total of 60% of the questions had a moderate level of difficulty, while 13.33% of the questions were in the difficult category. Therefore, the developed test instrument was considered to be of fairly good quality, given that most of the questions were in the moderate category (60%). Based on the average difficulty level of the questions of 0.58, which is in the moderate category, the overall difficulty level of the problem-solving test instrument was considered to be good.

The discriminating power of the test instrument was measured based on the criteria for good questions in the range of $0.20 < Dp \leq 1.0$. The analysis results showed that 26.67% of the questions had good discriminating power, while 66.67% of the questions had sufficient discriminating power, and 6.66% of the questions had poor discriminating power. In line with the established criteria, the analysis results show that 93.34% of the questions meet the discrimination criteria, which includes the adequate and good categories.

The average discrimination power obtained was 0.35 in the adequate category. These results indicate that the test instrument developed is quite good at distinguishing between students with high abilities and students with low abilities.

Based on the mathematics context, it can be observed that the number of students who can be classified in the category of having a very good level of mathematical problem-solving ability is 2 students (5.56%). Meanwhile, there are 14 students (38.89%) who have a good level of mathematical problem-solving ability. A total of 17 students (47.22%) showed mathematical problem-solving abilities that could be categorized as fairly good. On the other hand, there were 3 students (8.33%) who had mathematical problem-solving abilities in the poor category.

Based on the above explanation, it can be concluded that the test instrument that has been developed meets the validity and reliability criteria standards. Although there are several questions that have a less than optimal level of difficulty, overall the test instrument still meets the difficulty level criteria. Similarly, in terms of discriminating power, although there was one question with poor discriminating power, overall the test instrument still met the discriminating power criteria standards. Therefore, the development of a test instrument to measure problem-solving skills has reached the final prototype stage.

Conclusion

The process of developing a test instrument by applying the Tessmer model followed several steps, including the preliminary stage, the formative evaluation stage, the prototyping stage (including expert review, one-to-one, and small group), and the field test stage. The resulting test instrument shows quality that meets the validity and reliability criteria standards and demonstrates an optimal level of difficulty and discrimination. Therefore, this test instrument can be considered to meet the criteria for a quality test.

References

Ahmad Asyraful Anam, Surahmat 2, S. S. F. 3. (2021). *Mathematical Problem-Solving Ability Based on Learning Style Ahmad Asyraful Anam 1, Surahmat 2, Surya Sari Faradiba 3*. *16*(12), 28–37.

Arif, N., Yuanita, P., & Maimunah. (2022). Development of a Test Instrument for Mathematical Problem-Solving Ability Based on SOLO Taxonomy in Sequences and Series Material. *Cendekia Journal: Journal of Mathematics Education*, *06* (02), 2318–2335. https://j-cup.org/index.php/cendekia/article/view/1498

Beginna Kurnia Marita, Zainal Abidin, I. N. (2019). Mathematics Education Study Program, FKIP, Islamic University of Malang. *Journal of Research, Education, and Learning*, *14*(30), 102–110.

Idrus, T., Pomalato, S. W., Ismail, S., & Mohidin, A. D. (2023). Development of Instruments for Assessing Students' Mathematical Problem-Solving Skills in Senior High School. *Jambura Journal of Mathematics Education*, *4* (1), 63–74. https://doi.org/10.34312/jmathedu.v4i1.16404

Khusna, A. N., Sunismi, & Isbadar Nursit. (2019). Mathematical Problem-Solving Skills through the Whole Brain Teaching (WBT) Method as Viewed from David Kolb's Learning Styles in Quadrilateral Material. *Journal of Mathematics Education*, *14* (9

Pulungan, D. A. (2014). Development of Pisa Model Mathematics Literacy Test Instruments. *Journal of Educational Research and Evaluation*, *3* (2), 75–78. http://journal.unnes.ac.id/sju/index.php/jere

Rasid, M., Surahmat, & Faradiba, S. S. (2021). Analysis of Mathematical Problem-Solving Ability in Terms of Self-Efficacy in Two-Variable Linear Equation Systems in Grade VIII of SMP Negeri 04 Satu Atap Mukok, Sanggau Regency. *Journal of Research, Education, and Learning*, *16* (32), 76–86. http://riset.unisma.ac.id/index.php/jp3/article/view/15558

Lina Rumiati, Surahmat, Sunismi & Sikky El Walida . (2023) The Effect of Logical-Mathematical Intelligence on the Numeracy Literacy Skills of Junior High School Students Assisted by E-Modules, *Journal of Mathematics Education* 7(2), 2023, 10-22. http://journal.stkip-andi-matappa.ac.id/index.php/histogram/index

Juliat Hadi Sobirin, Yayan Eryk Setiawan & Surya Sari Faradiba. (2023) Mathematical Representations Based on Field-Dependent and Field-Independent Cognitive Styles, *Journal of Mathematics Education* p-ISSN 2528-2468; e–ISSN 2528-2026, https://doi.org/10.31327/jme.v8i2.1981.

Yusuf Amhar, Zainal Abidin & Surya Sari Faradiba. (2023) Analysis of Success Rates: Results Validation of Tangkas Matematika 14 Level, *Journal of Mathematics Education* 7(2), 2023, 133-141, http://journal.stkip-andi-matappa.ac.id/index.php/

Hilmiyatul Widdah & Yayan Eryk Setiawan. (2023) Analysis of Mathematical Problem-Solving Ability in Terms of Students' Motivation to Learn Mathematics. *Journal of Mathematics Education Study* Program, Volume 12, No. 2, 2023, 2641-2650. https://doi.org/10.24127/ajpm.v12i2.7113