(JIML) JOURNAL OF INNOVATIVE MATHEMATICS LEARNING
Volume 9, No. 1, March 2026

P-ISSN 2621-4733
E-ISSN 2621-4741

https://dx.doi.org/10.22460/jiml.v9i1.p30071

# DEVELOPMENT OF DIAGNOSTIC TEST FOR LEARNING DIFFICULTIES IN GEOMETRY MATERIAL ON JUNIOR HIGH SCHOOL STUDENTS

**Diendha Chaniesya Achma Dany[1], M. Duskri[2]**
[1]Universitas Islam Negeri Ar-Raniry, Lorong Ibnu Sina No.2, Banda Aceh, Indonesia.
210205006@student.ar-raniry.ac.id
[2]Universitas Islam Negeri Ar-Raniry, Lorong Ibnu Sina No.2, Banda Aceh, Indonesia.
m.duskri@ar-raniry.ac.id

**ARTICLE INFO**

***Article history:***

Received Nov 24, 2025
Revised Nov 29, 2025
Accepted Dec 12, 2025

***Keywords:***

Diagnostic Tests
Learning Difficulties
Geometry
Junior High School

**ABSTRACT**

Geometry is one of the mathematics topics that often causes difficulties for junior high school students because it requires visual-spatial abilities and a structured understanding of concepts, leading to misconceptions and procedural errors that impact students' low learning achievement. This research aims to develop a diagnostic test to identify learning difficulties in seventh-grade students regarding geometry material. Researchers used the Research and Development method with the Tessmer model. This stage begins with the preliminary stage, which is conducted thru an analysis of materials and curriculum to design the initial 40 test items. At the self-evaluation stage, the questions are reviewed and refined before being validated. The expert review stage involved two validators, and the initial results showed that 35 questions were suitable and 5 needed revision. In the second stage, all questions were declared suitable. One-on-one testing with several students showed the questions were understandable without further revision, while small group testing with 60 students resulted in 18 questions being acceptable, 17 needing revision, and 5 being eliminated. The final results show 11 easy questions, 28 medium questions, and 1 difficult question, with a reliability coefficient of 0.826, which is considered very high. This diagnostic test is able to identify students' difficulties in understanding concepts, procedures, and misconceptions. Thus, the developed instrument is effective in mapping students' learning weaknesses and supporting more targeted lesson planning.



***Corresponding Author:***

Diendha Chaniesya Achma Dany,
Department of Mathematics Education,
Universitas Islam Negeri Ar-Raniry,
Lorong Ibnu Sina No.2, Banda Aceh, Indonesia
Email: 210205006@student.ar-raniry.ac.id

***How to Cite:***

Dany, D.C.A., & Duskri, M. (2026). Development of Diagnostic Test for Learning Difficulties in Geometry Material on Junior High School Students. *JIML*, 9(1), 110-121.

## INTRODUCTION

Geometry is a branch of mathematics that deals with points, lines, planes, space, and their relationships to each other. Geometry is also one of the branches of mathematics that has been introduced since early childhood education up to the university level. Beside education, geometry also plays an important role in other fields such as applied science, technology, and is one way to develop logical thinking. Another benefit that can be gained from studying geometry is that it can be applied in daily life. For example, to calculate the area of a rectangular piece of land, we can apply the principles of geometry. Thus, the expectation is for students to be able to understand all the concepts in the subject and material, particularly in the mathematics subject, specifically geometry (Sudihartinih & Purniati, 2020).

In reality, many students still have relatively low abilities, experiencing difficulty solving problems, particularly in geometry. This is due to a lack of understanding of geometry concepts, making it difficult for students to solve problems in geometry and leading to errors in solving geometry problems they don't understand. Based on interviews conducted by the researcher on February 10, 2025, with one of the 7th-grade mathematics teachers at MTsN 4 Aceh Besar, it was found that many students experienced difficulties when solving mathematical problems in geometry. These difficulties are related to students' inability to choose and use the correct formulas when solving problems involving the volume and surface area of both flat-sided and curved-sided three-dimensional shapes. Additionally, students have not yet grasped basic geometric concepts, such as the characteristics of three-dimensional shapes, making it difficult for them to identify the type of three-dimensional shape and determine the appropriate solution steps. This aligns with the research findings of Fiskha et al. (2022), whose study aimed to analyze the numeracy skills of 8th-grade students in Sungai Penuh city in solving AKM problems. In the content category of numeracy assessment, there are four aspects of AKM assessment: numbers, geometry, statistics, and algebra. Based on the analysis results, it can be seen that the average score on number problems is higher than other aspects, reaching 72.09, while the second highest average score is statistics at 59.62, followed by the algebra aspect with an average score of 51.27. The average score for the measurement and geometry aspects is in last place, at 43.59 (Fiskha et al., 2022). One way to identify students' difficulties in understanding material concepts, particularly in geometry, is to administer diagnostic tests (Ardila Yuliarisma et al., 2022).

Diagnostic tests are used to determine students' strengths and weaknesses when learning something, so the results can be used as a basis for providing follow-up (Rosida & Susatyo, 2021). To achieve maximum test results, it is necessary to develop diagnostic tests. Diagnostic test development will be used to identify and determine the elements within a material that students find difficult to master, as well as to find the causes of the difficulties experienced by students. Based on the developed diagnostic test results, teachers can design more effective learning strategies and focus on students' errors in understanding the material that needs improvement. This can help teachers gain a deeper understanding of students' individual learning needs.

The development of diagnostic tests is becoming increasingly relevant in the implementation of the Merdeka Curriculum, which emphasizes differentiated learning, student-centeredness, a focus on essential materials, and contextual and meaningful learning. Good spatial ability in mathematics can improve student well-being and help students solve everyday life problems. Diagnostic assessment, as one form of evaluation in the Merdeka Curriculum, serves to identify students' abilities, strengths, and weaknesses so that teachers can design learning according to students' needs and characteristics (Cholilah, 2023). This assessment is important to determine students' basic abilities, especially in Latin reading and writing and arithmetic, which are prerequisites for mastering learning outcomes in the Merdeka Curriculum.

The development of diagnostic tests will continue to be developed considering the needs of the independent curriculum. Until now, the development of diagnostic tests will continue to be well developed and pursued. This aligns with several studies, such as Dian Mutmainna's research, which showed that two-tier multiple-choice diagnostic tests are effective in identifying students' understanding levels, with a higher percentage of students understanding the concepts compared to those who have misconceptions or do not understand the concepts. However, some misconceptions were still found in specific materials. This research uses the Research & Development (R&D) method with the Tessmer development model (formative evaluation), which includes four stages: (1) Preliminary: reference review, determination of location, and test subjects; (2) Self-evaluation: design of diagnostic test instruments; (3) Prototyping, including expert review (expert validation), one-to-one (testing on 3 non-subject students), and small groups (testing on small groups with student response questionnaires); and (4) Field test: field testing after instrument revision (Mutmainna et al., 2018).

As for the research conducted by Putri (2021), it resulted in a diagnostic test that functions to determine students' level of understanding and areas of difficulty on specific material, making it easier for teachers to design more effective and needs-based learning. This instrument also helps identify students' strengths and weaknesses so they can be improved. This research is a development study using the Plomp model, which includes three phases: (1) Preliminary research, namely field studies, literature studies, and findings analysis; (2) Prototype phase, namely designing grids, descriptions, assessment rubrics, instrument validation, and small group trials; and (3) Assessment phase, namely diagnostic test trials on a larger group (Oktaviana Putri, 2021). Based on a review of previous studies, this research has significant differences compared to the studies by Mutmainna et al. (2018) and Putri (2021). The research conducted by Mutmainna et all. (2018) focuses on the effectiveness of diagnostic instruments in identifying student understanding, while the research by Putri (2021) emphasizes the development of instruments to determine the overall level of student understanding and learning difficulties.

Meanwhile, this research specifically developed a two-tier multiple-choice diagnostic test on geometry material by utilizing the Tessmer development model thru more complete stages up to the field test stage. Additionally, this study not only measures students' final answers but also delves into the underlying reasons or reasoning behind each student's choice of answer, thus enabling a deeper identification of thinking processes, sources of error, and forms of learning difficulties. Thus, this research fills the gap by addressing the need for a diagnostic instrument in geometry that not only detects misconceptions but also traces their causes, and produces a learning difficulty map for students that teachers can use as a basis for improving learning to meet their needs. Based on the above explanation, the researcher aims to find a solution by scientifically examining the development of a diagnostic test for learning difficulties in geometry for seventh-grade junior high school students and presenting the stages in the diagnostic test development process in this study.

## METHOD

This research is Research & Development (R&D) research using the Tessmer development model. In this study, the development procedure consists of two stages: preliminary and formative evaluation. The preliminary stage includes preparation and design. Formative evaluation includes self-evaluation, expert reviews, one-to-one, small group, and field test stages.

As for the Tessmer model development stages, as shown in the following diagram (Nuryenisa et al., 2022):

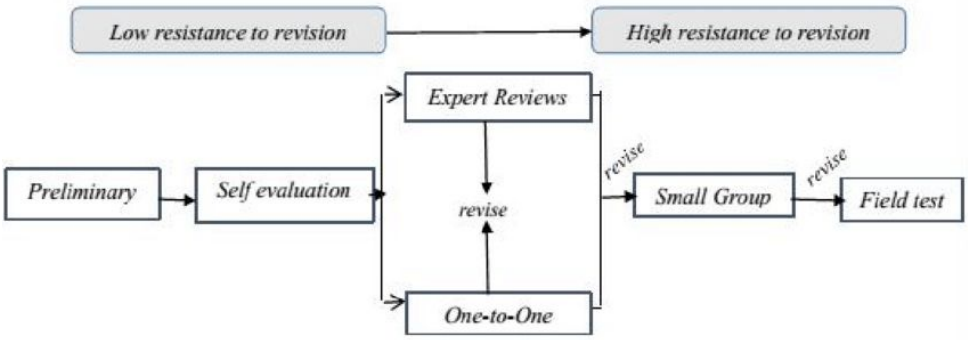


**Figure 1:** Flowchart of formative evaluation design Tessmer 1993 (in Nuryenisa et al.)

This research was conducted at MTsN 4 Aceh Besar with the research subjects being students from classes VIII-A and VIII-B during the first semester of the 2024/2025 academic year. The research instrument is a validation sheet used to assess the quality of the test items in terms of content, construction, and language. The data collection techniques used in this study are validation sheets and interviews. As for the data analysis techniques, they consist of:

Validity analysis: The validity of this test is used to measure how accurate or inaccurate the test questions are (Magdalena et al., 2021). A test can be considered valid if its content can measure a specific objective that aligns with the material listed in the curriculum. In this study, a validity test was conducted to assess the product developed and provided by the researcher. The formula for testing the validity of test items is as follows (Ma'aniyah & Mintohari, 2019):

$$P(\%) = \frac{total\ data\ collection\ score}{maximum\ score} x100\%$$

The content validity of the instrument was assessed by experts (validators) using the following rating scale (Novianti, 2015):

**Table 1.** Validation Criteria

| Scale | Quality | Description |
|---|---|---|
| 1 | Not Good | Not Feasible |
| 2 | Less Good | Less Feasible |
| 3 | Quite Good | Fairly Feasible |
| 4 | Good | Feasible |
| 5 | Very Good | Highly Feasible |

Reliability analysis: The reliability test is used to determine the consistency of the measurement tool if repeated (Kasanova & Sulistiyono, 2023). The formula for testing the reliability of test items is as follows (Ambarwati & Ismiyati, 2022):

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{s^2 - \sum pq}{s^2}\right)$$

Explanation:
$r_{11}$ = overall test reliability
$p$ = proportion of subjects who answered the item correctly
$q$ = proportion of subjects who answered the item incorrectly
$\sum pq$ = Sum of the product of $p$ and $q$
$n$ = number of items
$s$ = standard deviation of the test

The criteria for determining the high or low reliability of a test instrument can be seen in the table below (Basri et al., 2021):

**Table 2.** Interpretation of Product Reliability

| Interval r | Criteria |
|---|---|
| $0{,}00 \leq r < 0{,}20$ | Very Weak |
| $0{,}20 \leq r < 0{,}40$ | Weak |
| $0{,}40 \leq r < 0{,}60$ | Moderate |
| $0{,}60 \leq r < 0{,}80$ | Strong |
| $0{,}80 \leq r < 1{,}00$ | Very Strong |

Item discrimination index analysis: Item discrimination is the ability of an item to differentiate between high-achieving and low-achieving students (Loilatu et al., 2021). The formula for item discrimination is as follows (Bano et al., 2022):

$$DI = \frac{2(BA - BB)}{N}$$

Explanation:
$DI$ = Item dicrimination index
$BA$ = Number of correct answers in the upper group
$BB$ = Number of correct answers in the lower group
$N$ = Number of students who passed the test

The criteria for item discrimination can be seen in the following interpretation table (Fitriati, 2016):

**Table 3.** Category of Discrimination Power Interpretation

| Index Range | Interpretation |
|---|---|
| $0{,}40 - 1{,}00$ | The item is well accepted |
| $0{,}30 - 0{,}39$ | The item can be accepted without revision |
| $0{,}20 - 0{,}29$ | The item still requires improvement |
| $-1{,}00 - 0{,}19$ | The item cannot be used or should be discarded |

Difficulty Level Analysis: Item difficulty analysis is used to determine how difficult the test items are (Nurhalimah et al., 2022). The formula for calculating item difficulty is as follows (Sanusi & Aziez, 2021):

$$P = \frac{B}{JS}$$

Explanation:
$P$ = Difficulty index
$B$ = Number of students who answered the item correctly
$JS$ = Total number of students

The criteria for item difficulty levels can be found in the following interpretation table (Bano et al., 2022):

**Table 4.** Category of Difficulty Level Interpretation

| Index Range | Interpretation |
|---|---|
| $0{,}1 - 0{,}29$ | Difficult; the test item is of low quality and should be revised |

| | |
|---|---|
| $0{,}30 - 0{,}70$ | Moderate; the test item is sufficiently good and can be used |
| $0{,}71 - 0{,}90$ | Easy; the test item is good but should be revised |
| $x < 0{,}1 atau\, x > 0{,}90$ | The test item should be discarded |

Distractor/Distractor Analysis: The effectiveness of a distractor is its ability to mislead less capable students into choosing the wrong answer option (Sofiani Putri Radja, et al., 2023). Distractor analysis is conducted to determine the extent to which the distractors function in the questions created by the researcher. A distractor or decoy has functioned if it has been chosen by 5% of all test participants. The formula for the effectiveness of the decoy is as follows (Sanusi & Aziez, 2021):

$$IP = \frac{P}{\frac{(N-B)}{n-1}}\ x\ 100\%$$

Explanation:
$IP$ = Distractor index
$P$ = Number of students selecting the distractor
$N$ = Total number of students taking the test
$B$ = Number of students with correct answers for each item
$n$ = Number of answer alternatives
1 = Constant value

## RESULTS AND DISCUSSION

### *Results*

The development of the diagnostic test in this study refers to the Tessmer development model, which includes several stages, namely: the preliminary stage, formative evaluation including: self-evaluation, expert review, one-to-one, small group, and field test.

### Preliminary

This research began with a literature review of relevant journals and books to obtain a strong theoretical foundation for designing research instruments, followed by a curriculum analysis of the Merdeka Curriculum. This analysis also includes a review of the 7th-grade mathematics textbook used as the primary learning resource, as well as a study of the school environment where the research was conducted, specifically MTsN 4 Aceh Besar. This analysis was carried out thru interviews with mathematics teachers to obtain information related to student characteristics. This stage aims to develop diagnostic tests for Geometry and Measurement material in Phase D, with the learning outcomes being that students can create nets of three-dimensional shapes (prisms, cylinders, pyramids, and cones) and construct three-dimensional shapes from their nets, as well as students being able to explain how to determine the surface area and volume of three-dimensional shapes (prisms, cylinders, spheres, pyramids, and cones) and solve related problems.

Researchers created a concept map by mapping the connections between subtopics in the geometry material for 7th grade. The concept map includes flat-sided solids (cubes, blocks, prisms, and pyramids) and curved-sided solids (cylinders, cones, and spheres), which are interconnected thru the concepts of surface area and volume of solids. Next, the researcher developed a test grid for the diagnostic test based on the competencies being tested, the scope of the material, the material itself, cognitive levels, question format, and question number. In developing the test grid, the cognitive levels used included C1 (remembering), C2 (understanding), and C3 (applying). The result of this stage is 40 multiple-choice questions with a column for reasoning, allowing students to explain the steps or information that support their chosen answers. This instrument design covers the subtopics of flat-sided and curved-sided solid figures. For flat-sided solid figures, the questions consist of 8 items about cubes, 7 items about rectangular prisms, 5 items about triangular prisms, and 6 items about pyramids. Meanwhile, for curved-sided solid figures, there are 6 items about cylinders, 5 items about cones, and 3 items about spheres.

**Self evaluation**

The self-evaluation stage, where the researcher directly reviews and evaluates the 40 questions that have been prepared. In this process, the researcher reexamines the alignment between learning objectives, cognitive levels (C1-remembering, C2-understanding, and C3-applying), and the types of questions used. Additionally, the researchers also examined sentence clarity, term accuracy, the correctness of mathematical concepts, and the coherence between the stimulus and the question in each item. This evaluation is conducted to review and improve each test item so that there are no deficiencies or technical errors. Thru this evaluation, the questions are refined before being given to experts for further validation.

**Expert review**

At the expert review stage, the validation process will be carried out by two validators: one mathematics education lecturer and one teacher who teaches mathematics in 7th grade middle school. The selection of validators was made based on certain considerations, specifically teachers with experience teaching mathematics at the middle school level and mathematics education lecturers who teach courses on evaluation system development or have expertise and experience in designing and developing mathematics questions. This validation aims to assess the quality of the test items, considering content, construction, and language aspects. The suggestions and input from the validators served as the main reference for revising the instrument to better align with the expected standards. In the first-stage assessment, 35 questions were deemed suitable, and 5 questions needed revision. Revisions were made because there were slight inconsistencies in the question indicators. The following is a summary of the first-stage item validation in Table 5.

**Table 5.** Recapitulation of the first-stage item validation

| Item | Category | Follow-up Action |
|---|---|---|
| 1, 2. 3. 4, 5, 6, 7,8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 38, 39 | Valid | Used |
| 14, 20, 25, 30, 40 | Moderately Valid | Revised (contains a score of 3) |

After revising the test items, the next step is the second and final stage of this validation process. In this second stage, all items were deemed suitable for use without requiring revision.

**One-to-one**

The one-to-one product trial (individual trial) was conducted after the researcher revised the diagnostic test items based on the input received from the validators. At this stage, the researcher involved 4 students as test subjects. Before implementation, the researcher first explained the purpose and objectives of developing the diagnostic test to the students, then asked them to answer the questions while observing the readability and clarity of the test items.

After completing the test, the students were asked to provide comments and suggestions regarding the test items they had answered. Based on the results of the one-to-one trial, no writing errors or multiple interpretations were found that could cause students to misunderstand the content of the questions. This indicates that the test item has good readability and can be clearly understood by students.

**Small group**

The small group stage (small group trial) is the final stage in this research. This trial involves 60 students from two classes. The students' work results were then analyzed using ITEMAN software to test the item characteristics, including reliability, difficulty level (DL), discrimination index (DI), and distractor analysis. The recapitulation of the test item evaluation results with the criteria of accepted, revised, and rejected, covering difficulty level, discrimination index, and distractor effectiveness for each test item, is presented in the following table.

**Table 6.** Recapitulation of Test Items with Accepted Criteria

| Item | Difficulty Level | Discrimination Index | Distractor Functionality |
|---|---|---|---|
| 4, 5, 8, 9, 12, 15, 17, 21, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36 | $0{,}30 - 0{,}70$ | $0{,}30 - 1{,}00$ | A, B, C, D≥ 5% |

The table above shows 18 test items that were accepted as diagnostic test instruments because they met the quality criteria based on three main aspects: difficulty level, discrimination power, and distractor effectiveness. The accepted items include numbers 4, 5, 8, 9, 12, 15, 17, 21, 25, 26, 27, 28, 30, 31, 32, 33, 35, and 36.

The difficulty level (DL) values ranged from 0.40 to 0.70, indicating that the questions were of moderate difficulty and were able to measure students' abilities proportionally. The discrimination index (DI) values ranged from 0.310 to 0.617, which signifies the ability to effectively differentiate between high- and low-achieving students. Additionally, all distractors functioned effectively because they were chosen by over 5% of the learners. Thus, the 18 items have characteristics that are valid, reliable, and suitable for identifying students' learning difficulties.

**Table 7.** Recapitulation of Test Items with Revised Criteria

| Item | Criteria | Category |
|---|---|---|
| 1, 6, 7, 19, 23, 39, 40 | Difficulty Level $0{,}1 - 0{,}29$ and $0{,}71 - 0{,}90$ | Revised |
| 7, 13, 16, 18, 20, 22, 29, 40 | Discrimination Index $0{,}20 - 0{,}29$ | Revised |
| 1, 6, 7, 11, 13, 34, 38 | Distractor Functionality A, B, C, D ≤ 5% | Revised |

The table above shows a summary of 17 test items that need revision because they do not fully meet the instrument's quality standards. Revisions are needed based on the analysis of difficulty level, discrimination power, and distractor effectiveness. From the difficulty level (DL) aspect, items 1, 6, 7, 19, 23, 39, and 40 have values above 0.70, falling into the easy to very easy category, thus being less effective in optimally differentiating students' abilities. From the discrimination index (DI) aspect, items 7, 13, 16, 18, 20, 22, 29, and 40 have values below 0.20, indicating low discriminatory power. Meanwhile, from the perspective of distractor functionality, items 1, 6, 7, 11, 13, 34, and 38 have distractors that are not functioning effectively because they were chosen by less than 5% of the students. Generally, these 17 questions require revisions to their wording, answer choices, and answer keys to improve the quality of the instrument and make it suitable for measuring students' abilities.

**Table 8.** Recapitulation of Test Items with Rejected Criteria

| Item | Difficulty Level | Discrimination Index | Distractor Functionality |
|---|---|---|---|
| 2 | 0,917 | 0,133 | C≥ 5%;A, B, D≤ 5% |
| 3 | 0,933 | 0,135 | D≥ 5%;A, B, C≤ 5% |
| 10 | 0,900 | 0,159 | A, B≥ 5%; C, D≤ 5% |
| 14 | 0,933 | 0,290 | B, C≥ 5%; A, D≤ 5% |
| 24 | 0,217 | 0,118 | A, B, C, D≥ 5% |

This table presents five items that were rejected due to significant quality issues based on the results of the ITEMAN analysis. Most items were rejected due to two main weaknesses: difficulty level (DL) being in the too easy category (approaching a value of 1.00) and discrimination index (DI) being low and unable to optimally differentiate student abilities. This condition is evident in items 2, 3, 10, and 14, which have difficulty levels (DL) above 0.90 and discrimination indices (DI) below 0.30. Meanwhile, item 24 shows the opposite characteristics, being too difficult (DL = 0.217) and having a very low discrimination index. Additionally, almost all rejected items also had issues with distractor functionality, where one or more answer alternatives were not effective (selected by ≤ 5% of the participants). The combination of an unfavorable difficulty level, low discrimination power, and non-functional distractors led to the items being deemed unsuitable for use.

The diagnostic tests developed in this study proved to be of good and reliable quality. The analysis results using ITEMAN software show that this instrument has a reliability level of 0.826, which falls into the high category. This reliability value indicates that the instrument is consistent in measuring students' abilities, so the results obtained can be trusted. Additionally, the test items developed also demonstrate an adequate level of difficulty and discrimination power. This means the question can accurately differentiate between students who have a good understanding and those who are still struggling. Most of the test items do not require further revision as they have met the eligibility criteria as an evaluation instrument, both in terms of validity, reliability, and the effectiveness of the distractors.

***Discussions***

This research is relevant to the findings of the research conducted by Mohammad Syaifuddin et al., which also showed that the developed diagnostic test had high reliability, moderate difficulty level, and good item discrimination (Syaifuddin et al., 2022). Additionally, the findings of this study align with those of Aini Nurawwaluliza et al., which showed that diagnostic tests are effective in identifying students' misconceptions about flat-sided solid figures, with misconceptions reaching 60%, conceptual misunderstanding 30%, and conceptual understanding only 10%. These findings confirm that diagnostic instruments are capable of revealing students' conceptual errors more deeply than conventional tests (Nurawwaluliza et al., 2021). In line with this, Fattya Rosyada and Nurul Ikhsan Karimah found that using two-tier diagnostic tests was also more effective in mapping students' understanding profiles in detail, with only 29.40% of students truly understanding the concepts, while the rest experienced various forms of misconceptions and lack of understanding. These results reinforce the findings of this study that a tiered diagnostic instrument with good validity and reliability is capable of providing a comprehensive picture of students' conceptual understanding

(Rosyada & Karimah, 2025). Another research by Wahyu Hartono showed that diagnostic test instruments developed in a study must have high validity and reliability levels, as well as good item discrimination, in accordance with the established indices. Thus, diagnostic instruments that meet good validity, reliability, and discrimination criteria can be effective tools for comprehensively analyzing and mapping students' understanding profiles (Hartono et al., 2023). Similar findings were also shown by Mukhlis Hidayat et al., who stated that valid and reliable diagnostic instruments can be used as a basis for designing more targeted learning follow-up (Hidayat & Zubaidah, 2023). The consistency of the results of this study with previous research strengthens the belief that instrument development procedures carried out systematically, thru expert validation stages, trials, and revisions based on needs analysis, are capable of producing evaluation tools that are valid, reliable, and effective.

## CONCLUSION

This research successfully developed a diagnostic test for Geometry material in Phase D using the Tessmer model thru the preliminary, self-evaluation, expert review, one-to-one, and small group stages. The resulting instrument consists of 40 multiple-choice questions with a column for reasons, which has been proven valid, reliable ($\alpha = 0.826$), and meets the criteria for test quality. Item analysis showed that 11 questions were in the easy category with a difficulty range of 0.717–0.933, 28 questions were in the moderate category with a difficulty range of 0.367–0.700, and 1 question was in the difficult category with a difficulty level of 0.217. This confirms that the instrument has a proportional distribution of difficulty levels, making it effective in measuring the variation in students' abilities.

This research provides a scientific contribution in the form of a diagnostic instrument that is not only reliable and valid but also able to identifying students' misconceptions and learning difficulties in more detail. This instrument enriches the study of learning evaluation because it is designed not only to measure learning outcomes but also to explore students' thinking processes thru the reasons they provide.

The results of this study have practical implications that can be applied in mathematics learning in schools, particularly for identifying students' understanding and learning difficulties in geometry material. However, this research is still limited to the small group stage and has not yet reached field testing. Therefore, further research is recommended involving a larger sample size, diverse school contexts, and testing effectiveness over the course of education. Thus, this research has the potential to be further developed into a standard evaluation tool that can be used continuously in education.

## REFERENCES

Ambarwati, Y. F., & Ismiyati, I. (2022). Analisis butir soal pilihan ganda ulangan akhir semester genap mata pelajaran kearsipan. *Measurement In Educational Research (Meter)*, *1*(2), 64. https://doi.org/10.33292/meter.v1i2.144

Ardila Yuliarisma, S., Masfingatin, T., & Andari, T. (2022). *Analisis Pencapaian Siswa Dalam Asesmen Kompetensi Minimum: Suatu Studi Pendahuluan*. *7*(2), 568.

Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, *8*(1), 145. https://doi.org/10.32884/ideas.v8i1.660

Basri, K., Baidowi, B., Junaidi, J., & Turmuzi, M. (2021). Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP. *Griya Journal of Mathematics Education and Application*, *1*(4), 682–694. https://doi.org/10.29303/griya.v1i4.107

Cholilah, M. (2023). Profil Kemampuan Spasial Siswa Smp Pada Pembelajaran Matematika Yang Berkaitan Dengan Geometri Dalam Implementasi Kurikulum Merdeka. *SCIENCE : Jurnal Inovasi Pendidikan Matematika Dan IPA*, *3*(3), 178–188.

Fiskha, S., Patri, D., & Heswari, S. (2022). Analisis Kemampuan Numerasi Siswa Kelas Viii Smp Se-Kota. *Jurnal Muara Pendidikan*, *7*(2), 235.

Fitriati. (2016). Menentukan Hubungan Antara Gender, Sikap Matematika dan Prestasi belajar Matematika: Uji Validitas dan Reliabilitas Instrumen. *Numeracy: Jurnal Ilmiah Pendidikan Matematika*, *III*, 11–22.

Hartono, W., Hadi, S., Rosnawati, R., & Retnawati, H. (2023). Exploration of Diagnostic Testing Instruments: Validity, Reliability, and Item Characteristics. *Pegem Journal of Education and Instruction*, *13*(3), 386–394. https://doi.org/10.47750/pegegog.13.03.39

Hidayat, M., & Zubaidah, T. (2023). *Students ' Errors in Solving Plane Geometry Problems Using E-Learning-based Diagnostic Tests*. *4185*, 93–110. https://doi.org/10.24815/jdm.v10i1.31938

Kasanova, R., & Sulistiyono, R. (2023). Evaluasi Butir Soal Pilihan Ganda Penilaian Tengah Semester dalam Pembelajaran Tematik untuk Kelas V di SDN Gladak Anyar 4 Pamekasan. *Journal on Education*, *6*(1), 5820–5834. https://jonedu.org/index.php/joe/article/view/3773

Loilatu, S. H., Mukadar, S., Kasmawati, K., & Hentihu, V. R. (2021). Strategi Belajar Mengajar Dengan Menerapkan Metode Eksperimen Untuk Meningkatkan Prestasi Belajar IPA Di SD Alhilaal Samalagi. *Edu Cendikia: Jurnal Ilmiah Kependidikan*, *1*(2), 65–73. https://doi.org/10.47709/educendikia.v1i2.1036

Ma'aniyah, S., & Mintohari. (2019). Pengembangan Media Kartu Gambar Berbasis Make a Match Dalam Pemahaman Konsep Materi Gaya Sekolah Dasar. *Jpgsd*, *7*(2), 2749–2759.

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG : Jurnal Pendidikan Dan Sains*, *3*(2), 198–214. https://ejournal.stitpn.ac.id/index.php/bintang

Mutmainna, D., Mania, S., & Sriyanti, A. (2018). Pengembangan Instrumen Tes Diagnostik Pilihan Ganda Dua Tingkat Untuk Mengidentifikasi Pemahaman Konsep Matematika. *MaPan*, *6*(1), 56–69. https://doi.org/10.24252/mapan.2018v6n1a6

Novianti, D. A. (2015). Pengembangan Modul Akuntansi Aset Tetap Berbasis Pendekatan Saintifik Sebagai Pendukung Implemetasi K-13 Di SMKN 2 Buduran. *Jurnal Pendidikan Akuntansi*, *3*, 2. https://ejournal.unesa.ac.id/index.php/jpak/article/view/13193

Nurawwaluliza, A., Zubaidah, T., & Hidayat, M. (2021). *Analisis Miskonsepsi Siswa Kelas VIII MTsN 1 Bireuen Menggunakan Three Tier Test pada Materi Bangun Ruang Sisi Datar*. *6*(3), 273–281.

Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda Pas. *Natural Science Education Research*, *4*(3), 249–257. https://doi.org/10.21107/nser.v4i3.8682

Nuryenisa, N., Karim, K., & Rahmita, N. (2022). Pengembangan Soal Matematika Model Pisa Level 4 Berbasis Etnomatematika Budaya Banjar Untuk Mendukung Pembelajaran Matematika Smp. *Jurmadikta*, *2*(2), 61–72. https://doi.org/10.20527/jurmadikta.v2i2.1376

Oktaviana Putri, A. (2021). Pengembangan Tes Diagnostik Berbentuk Uraian Berdasarkan Ranah Kognitif untuk Mengetahui Letak Kesulitan Belajar Siswa SMP pada Materi Bangun Ruang. *Jurnal Pusaka*, *10*(1), 40–54. https://doi.org/10.35897/ps.v10i1.597

Rosida, I., & Susatyo, E. B. (2021). Analisis Pemahaman Konsep Pada Pembelajaran Larutan Penyangga Model Discovery Learning Menggunakan Tes CBT. *Chemistry in Education*, *10*(2), 62–69. http://journal.unnes.ac.id/sju/index.php/chemined%0AANALISIS

Rosyada, F., & Karimah, N. I. (2025). *Analysis of Students ' Misconceptions On The Characteristics of Flat Shapes Using Two-Tier Multiple-Choice Diagnostic Test in Grade V Elementary School*. *4*(1), 61–77.

Sanusi, R. N. A., & Aziez, F. (2021). Analisis Butir Soal Tes Objektif dan Subjektif untuk Keterampilan Membaca Pemahaman pada Kelas VII SMP N 3 Kalibagor. *Metafora: Jurnal Pembelajaran Bahasa Dan Sastra*, *8*(1), 99. https://doi.org/10.30595/mtf.v8i1.8501

Sofiani Putri Radja, et al. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda Dan Efektivitas Pengecoh Di Sman 1 Pandawai. *Jurnal Edusavana*, *1*(1), 30–41.

Sudihartinih, E., & Purniati, T. (2020). Students' Mistakes and Misconceptions on the Subject of Conics. *International Journal of Education*, *12*(2), 92–100. https://doi.org/10.17509/ije.v12i2.19130

Syaifuddin, M., Darmayanti, R., & Rizki, N. (2022). Development Of A Two-Tier Multiple Choise (TTMC) Diagnostic Test For Geometry Materials To Identify Miscopcentions Of Middle School Students. *Jurnal Silogisme*, *7*(2), 66–76. http://journal.umpo.ac.id/index.php/silogisme