PERJ Vol. 7 No. 1 (2022), 35-42

PERJ
Physics Education Research Journal
ISSN 2685 - 6190 (p)
ISSN 2714 - 7746 (e)
Department of Physics Education
Faculty of Science and Technology
UIN Walisongo Semarang-Indonesia

Development of Three-Tier Diagnostic Test to Identify Physics Fluid Misconceptions in Class XI SMAN 2 Medan

Ribka Kezia* **and Motlan**

Universitas Negeri Medan, Medan, Indonesia

ARTICLEINFO

Article history:
Submitted : January 21st, 2025
Revised : Mei 17th, 2025
Accepted : June 28th, 2025

Keywords:
Three-tier diagnostic test; student misconceptions; fluid physics; physics education; instrument development.

ABSTRACT

This research aims to develop a three-tier diagnostic test instrument to identify misconceptions in fluid physics among grade XI students at SMAN 2 Medan. Misconceptions are a persistent barrier in learning physics and require specific instruments for accurate identification. The developed instrument includes multiple-choice questions (first tier), reasoning options (second tier), and confidence levels (third tier) to assess students' conceptual understanding comprehensively. Using the 4D (Define, Design, Develop, Disseminate) development model, the instrument underwent validation by experts, small-scale, and large-scale trials. The results indicated that the diagnostic test instrument is valid and reliable, effectively identifying students' misconceptions in fluid concepts such as hydrostatic pressure, Archimedes' principle, Pascal's law, and Bernoulli's principle. This instrument not only aids teachers in recognizing areas that need instructional emphasis but also contributes to improving learning strategies. The findings revealed a significant portion of students experienced misconceptions, highlighting the need for targeted pedagogical interventions. This study contributes to physics education by offering a practical diagnostic tool to improve conceptual learning and reduce misconceptions.

Introduction

Physics is a science that discusses a collection of systematically structured knowledge resulting from detailed observations and measurements (Silaban, et al., 2022). Secondary physics education plays an important role in building a strong understanding of science concepts among students. Given the important role of physics in human life, physical science should be properly understood and continuously developed. One of the most important things when studying physics is to have a correct understanding of the concepts. However, based on direct observation conducted at SMAN 2 Medan, there are still many students who often experience misconceptions and have different understandings in physics learning.

Misconception is an understanding of a concept that exists in one's mind that is not in accordance with the actual scientific concept. Misconceptions occur when someone understands information in an inaccurate or erroneous way, which can be caused by ignorance, incorrect assumptions, or improper interpretation of existing facts. Misconceptions are prone to occur and are very difficult to eliminate (Makhrus, 2021). Misconceptions in physics learning can hinder the development of students' overall understanding. Misconceptions are often caused by students' inability to connect physics concepts with their daily experiences, as well as difficulties in understanding mathematical representations of physical phenomena.

Based on the results of interviews with several physics teachers at the school, it is said that students also often experience misconceptions in a physics

*Correspondence email: ribkakezia14@gmail.com
doi: 10.21580/perj.2025.7.1. 27200

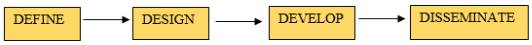
material taught, one of which is the concept of fluid. Therefore, it is important for teachers and researchers to develop evaluation tools that can identify misconceptions more specifically and accurately. The physics teacher at the school also revealed that the right test instrument to measure students' misconceptions is rarely used in learning evaluations due to the limited ability of teachers to make the right test instrument to identify these misconceptions. One way that can be used is to develop a three-tier diagnostic test instrument. Three-tier diagnostic test can be used to determine misconceptions or conceptual errors is a diagnostic test (Maryam, 2020).

This three-tier diagnostic test includes the first tier of each item in the test is a proportional statement and part of a concept map made in multiple-choice form. The second tier contains reasons to be chosen by learners that explain the answers in the first tier. The third tier contains the level of confidence of learners in answering the first tier. From the student's answer, it can be seen to what extent the student understands the correct concept because the diagnostic test questions developed can reveal student misconceptions which can be seen from the varied student answers in answering diagnostic test questions. The level of confidence in each answer and reason can measure differences in the level of knowledge of students so that it will help detect the level of misconceptions of students (Nazura et all, 2021). After the development of this diagnostic test instrument, it is hoped that it can help teachers in determining the parts of the material that need more emphasis during learning, and planning more good lessons to help reduce student misconceptions (Atsilah, 2024).

Methods

This study uses a Research and Development (R&D) approach aimed at producing a valid and effective instrument in the form of a three-tier diagnostic test to identify misconceptions in the topic of fluids among Grade XI students. The development model applied is the 4D model proposed by Thiagarajan et al., which consists of four stages: Define, Design, Develop, and Disseminate as Figure 1. The model was chosen because it systematically guides the development of educational tools from the initial concept stage to dissemination. The research was conducted at SMAN 2 Medan during the even semester of the 2024/2025 academic year.

Figure 1.
Research procedure using the 4D R&D model



The collected data will be analyzed to find out the assessment and opinions of the developed product.

Content Validity Test

Content validity analysis was conducted by analyzing the results of content validation by experts using Content Validity Ratio (CVR).

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}} \quad (1)$$

Information :
ne = number of raters who agree;
N = Number of all validating raters

After identifying each sub question on the test instrument using CVR, CVI is used to calculate the overall number of sub questions, CVI

$$CVI = \frac{\sum CVR}{\sum CVI} \quad (2)$$

Table 1.
Categories of Content Validity Index

Range	Category
0 – 0.33	Unsuitable
0.34 – 0.67	Suitable
0.68 - 1	Very Suitable

Empirical Validity Test

The validity of a measuring instrument refers to the extent to which the instrument can achieve the desired goals or whether the instrument is truly appropriate in measuring what it should measure (Purba, 2011).

$$r_{xy} = \frac{N\Sigma X_i Y - (\Sigma X_i)(\Sigma Y)}{\sqrt{\{N\Sigma X_i^2 - (\Sigma X_i)^2 (N\Sigma Y^2 - (\Sigma Y)^2\}}} \quad (3)$$

Information :
r_{xy} = Correlation coefficient between all students' scores for question i and all scores for each student
ΣX_i = The total score of all students for question i
$\sum Y$ = The total of all grades for each student
$\Sigma X_i Y$ = The multiplication of the scores of all students for question i and the sum of all scores for each student

ΣX_i^2 = The sum of the squares of the scores of all students for question i

ΣY^2 = The sum of the squares of all grades for each student.

$(\Sigma X_i)^2$ = The total score of all students for question i is then squared.

N = The number of students

Reliability Test of Question Items

Reliability testing of question items is carried out after the wide-scale product testing stage. Reliability in this research is a measuring instrument that is used consistently to give the same results, so that the data is used as a data collection instrument, test reliability testing can be calculated using Croanbach's alpha. If a value is obtained from $r_{11} > r_{table}$ then it can be stated as a whole that the question items are reliable. The criteria for a realistic question item are as follows Table 2.

Table 2.
Reliability Criteria

Range	Category
$0.80 < r_{11} \leq 1.00$	Very high reliability
$0.60 < r_{11} \leq 0.79$	High reliability
$0.40 < r_{11} \leq 0.59$	Reliability is sufficient
$0.20 < r_{11} \leq 0.39$	Low reliability
$0.01 < r_{11} \leq 0.19$	Very low reliability
0.00	Not reliable

Source: Sugiyono (2015)

Result and Discussions

This section presents the results obtained from each stage of the research and discusses the findings in relation to the research objectives. The development process was carried out through several phases, including needs analysis, design, development, and evaluation. The findings are analyzed to determine the effectiveness and feasibility of the developed product based on both qualitative and quantitative data.

Define stage

The purpose of this stage is to establish and define the learning needs and analyze the objectives and limitations of the material.

The definition stage consists of :

a. Needs Analysis Results

The needs analysis was conducted to identify the core problems in learning. This process included interviews with three 11th-grade physics teachers from SMA Negeri 2 Medan. The analysis was further supported by a literature review involving books, journals, and references related to diagnostic test construction, particularly the three-tier diagnostic test and misconceptions in fluid physics. Additionally, a student characteristics analysis was performed using an initial observation questionnaire. Based on the interview and questionnaire data, it was revealed that students in class XI often experience misconceptions and have varying levels of understanding in physics. These findings were confirmed by class test results and student interest levels.

b. Concept Analysis Result

Concept analysis was carried out to review the basic competency standards and compile relevant concepts for developing the diagnostic test instrument.

Design Stage

After obtaining problems in the analysis stage, the design stage is then carried out. This design is carried out to design the instrument developed. Activities at this stage include.

a. Format Selection

This stage is the selection of the diagnostic test format. The format used is a diagnostic test with a three-tier multiple choice type with closed reasons.

b. Initial Design

This stage is the initial design of the prototype of the diagnostic test instrument. instrument prototype in the form of a draft grid of test questions, 12 test questions. Writing items on the three-tier multiple choice diagnostic test developed includes levels C2-C5 of Bloom's taxonomy (Understanding, Application, and Analysis), Answer Keys, Scoring Guidelines, and Guidelines for Interpretation of Results.

c. Grid of Three-Tier Multiple Choice Diagnostic Test Instrument

The components in the lattice of multiple choice diagnostic test instruments include submaterial, indicators, question indicators, cognitive levels and question numbers. Making indicators is adjusted to the learning teaching modules used with the independent

curriculum. These indicators are used to create a test grid for the multiple choice diagnostic questions developed. The grid of three-tier diagnostic test questions is made with the aim of placing each item into each indicator provided, knowing the category of items developed. The C2 level is the second level used to measure the realm of understanding. The C3 level is the third level used to measure the realm of applying and the C4 level is the fourth level used to measure the realm of analyzing and C5 level is used to measure the domain of evaluating.

d. Three-Tier Multiple Choice Diagnostic Test Question Instructions
Instructions for working on diagnostic test questions are instructions that explain some things that can be done and what not to do during the time of working on the questions. Instructions for working on questions contain procedures for working on questions, information about the number of questions, information on the length of time to work. Instructions for working on questions contain several points that are important and must be considered by students because they can help to understand the rules for working on Three-Tier Mutiple Choice Diagnostic Test questions properly and correctly.

e. Scoring Guidelines
The scoring guideline for the three-tier multiple choice diagnostic test has a use as a guide in scoring the answers to the tests that students have done, the answers to the reasons chosen, and the level of confidence in the reasons. The results of the acquisition of students' scores are then interpreted in the categories of understanding, not understanding, and misconceptions on each item.

Development Stage

During the development stage, both small-scale and large-scale trials were conducted to evaluate the effectiveness and practicality of the developed product. The small-scale trial involved a limited group of participants to identify any initial flaws, gather feedback, and make necessary improvements. Following this, a larger-scale trial was carried out involving a broader group of users to test the product in a more diverse and realistic setting. This process ensured that the product met the expected standards of quality, usability, and relevance before being finalized for broader implementation.

a. Instrument Validation by Experts
The three-tier multiple choice diagnostic test instrument that has been designed before being tested on students is first submitted to the validator for the validation process. Validation was carried out by 2 UNIMED physics department lecturers and 1 physics teacher of SMA Negeri 2 Medan. There were 12 questions validated by each validator. Instrument validation is carried out with the aim of knowing the validity of the test questions developed so that the test questions can be used. The questions were checked by the validator by giving the question grid, question sheet, and validation sheet to the validator.

There are three aspects assessed in the validation sheet, namely material, construction, and language aspects. Each indicator is given a score of 1 if it is in accordance with the assessment criteria and a score of 0 if it is not in accordance with the assessment criteria. Based on the results of the assessment by the three validators of the three-level multiple choice diagnostic test instrument, it can be concluded that the 10 items are suitable for testing while taking into account some suggestions and comments on improvements given by the validators and 2 other questions are declared invalid because they do not meet the validity.

The average validator assessment results show that there are 10 questions that can be used and there are questions that cannot be used. A total of 10 questions were declared valid and 2 questions were declared invalid to be tested.

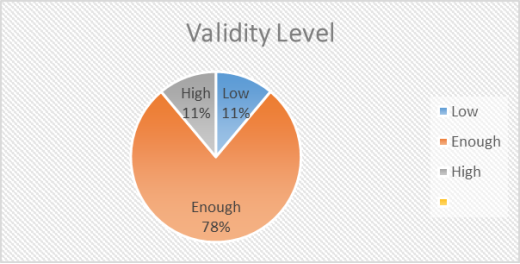
b. Item Validity Test
This trial was conducted twice where in the small scale test (first) involved 25 students of class XI Orbit and in the large scale test (second) involved 36 students of class XI Comet. This trial was conducted to determine the results of item analysis which included item validity, reliability, difficulty level, distinguishing power, and the effectiveness of the question instrument. In the trials carried out, it produced diverse student answers, there were the lowest scores and the highest scores. Furthermore, student answers are analyzed to find the validity of the items on each item. Item validity is sought through the Moment Product Correlation equation found in the appendix. Based on the results of the analysis conducted from 10

multiple choice questions, all questions are said to be valid.

Table 3.
Recapitulation of Question Item Validity in XI Orbits

Question Item	r_{table}	r	Criteria
1	0.396	0.455	Valid
2	0.396	0.494	Valid
3	0.396	0.695	Valid
4	0.396	0.456	Valid
5	0.396	0.398	Valid
6	0.396	0.519	Valid
7	0.396	0.503	Valid
8	0.396	0.644	Valid
9	0.396	0.544	Valid
10	0.396	0.450	Valid

Figure 3.
Percentage validity of the first trial



c. Item Reliability Test
Reliability analysis is carried out to determine the consistent level of the question. The equation used to determine the level of item reliability is Alfa Croanbach. To determine the reliability of the question using a significance level of 0.05. After the calculation, the results are categorized based on the reliability criteria.

The results of the reliability calculation on the small-scale test experiment in class XI Comets resulted in r11 of 0.846 with very high reliability criteria. So it can be concluded that the Three-Tier Multipe Choice Diagnostic Test questions used in this large-scale experiment are reliable. After conducting validity tests on small-scale and large-scale trials, it can be concluded that this three-level diagnostic test question has very high reliability criteria.

d. Test of Item Difficulty
The level of difficulty is seen from the ability or ability of students to answer each item. The level of difficulty of the questions is different, some are easy, moderate, and difficult. The data obtained in the trial were used to analyze the level of difficulty of the Three-Tier Multiple Choice Diagnostic Test items.

Based on the results of the small-scale trial that has been carried out, the results obtained for the test of the level of difficulty of the items on the three-level multiple choice diagnostic test have a difficulty level criteria of 2 criteria, namely 6 questions with medium criteria and 4 questions with difficult criteria.

e. Item Distinguishing Power Test
Furthermore, the analysis of the item differentiator test on the multiple choice three-level diagnostic test instrument. Differentiating power serves to distinguish groups of students who have mastered the material from students who have not mastered the material.

The results of the item discriminating test on small-scale tests can be concluded to produce 7 questions that fall into the category of acceptable questions and 3 questions fall into the category of acceptable questions but need to be corrected.

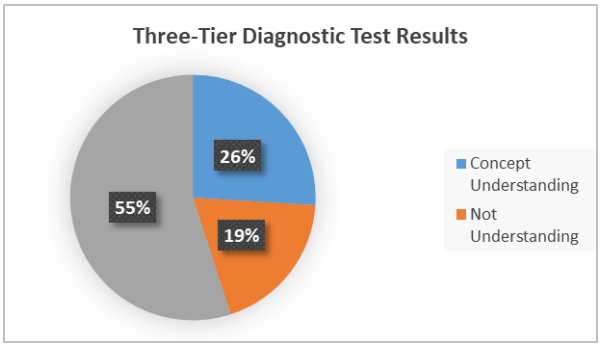
f. Distractor Functionality Test
Next, the distractor functionality test was analyzed. This analysis determines the effectiveness of the distractors used. A question has attractiveness if it can trick students who do not understand the concept and master a physics concept. Distractors are obtained by counting the number of participants who choose answers other than the answer key. An exception works well if it is at least 5% of the total number of participants who take the test. An exception that is selected by less than 5% of the total number of test takers does not have a good function because it does not have the attractiveness to be selected. The results of the analysis of question exemptions can be seen in the Appendix.

The instrument that was tested on this small scale consisted of 10 question items and each had 4 options consisting of 1 answer option and 3 distractor options. Then there are 30 distractor options on the test instrument.

The result analysis shows that the distractors on the usable or functioning answers are 76% and the distractors that are not functioning properly are 24%. **T**he distractors on the usable or functioning answers are 93% and the distractors that are not functioning properly are 7%.

g. Data Analysis
After the questions are tested to students, the students' answers will then be analyzed for the level of student misconceptions and specify the percentage of sub-material misconceptions in Fluid material. The analysis is carried out by giving an assessment according to the scoring guidelines in the Appendix. Furthermore, the results will be analyzed and grouped into three groups, namely Understanding the Concept (PK), Not understanding the concept (TPK), and Misconception (MK).

Figure 4.
Interpretation of Three-Tier Diagnostic Test Results



Based on Figure 4, it can be seen that the category in misconceptions occupies the highest percentage of 55%. The percentage of the category Does not understand the concept is 19% and the category understands the concept is 26%.

Table 5.
Percentage of each Category of Students' Level of Understanding of the Concept of Fluid

Question Items	Fluid Concept	Percentage (%)		
		Concept Understanding	Not Understanding	Misconceptions
1.2	Pressure in Fluids	48%	21%	32%
3.4	Archimedes Law	57%	7%	36%
5.6	Pascal's Law	48%	27%	26%
7.8	Fluid Flow	55%	21%	24%
9.10	Dynamic Fluid	67%	21%	12%

c. Disseminate
At this stage, the activity conducted was the distribution of the three-tier diagnostic test instrument to identify students' misconceptions in Fluid Material. This instrument is now ready to be used.The instrument was distributed to Physics teacher at SMA Negeri 2 Medan.

Conclusions

Based on the results of research on the Development of a Three-Tier Diagnostic Test Instrument to Measure the Misconceptions of Grade XI Students on Fluid Material, it can be concluded that the three-level multiple choice diagnostic test instrument developed has specific characteristics to measure misconceptions on fluid material through 10 items. This instrument has met good qualifications based on the assessment of three expert validators, with 10 items declared valid and 2 items invalid. The reliability test results showed a reliability value of 0.87 in the small-scale test and 0.84 in the large-scale test, indicating that this instrument can be used reliably to identify students' level of concept understanding. Based on the index of difficulty level, most of the questions fall into the category of medium difficulty level. The results of the pilot test showed that on a small scale, 33% of students understood the concept, 18% did not understand the concept, and 49% experienced misconceptions. Meanwhile, on a large scale, only 26% of students understood the concept, 19% did not understand the concept, and 55% experienced misconceptions. Thus, the three-tier diagnostic test instrument developed succeeded in identifying parts of the fluid concept that still contain misconceptions in students.

Acknowledgments

I would like to express my gratitude to my supervisor who has been willing to guide me and to SMAN 2 Medan who has been willing to accept me to conduct research at the school.

References

Aliyah, R, Sudibyo, E., Suyatno., & Wasis. (2022). The Profile of Misconceptions Using Three Tier Diagnostic Test on Dynamic Fluid. Journal of Physics. (2392) 012032.

Atsilah, B, M., Suhadi., Putri, K, J., Mabruroh, F., Sugiarti., & Pebralia J. (2024). Pengembangan Three Tier Test Multiple Choice Untuk Mengidentifikasi Miskonsepsi Pada Materi Hukum Newton. Jurnal Penelitian Pembelajaran Fisika.15(2).213-221

Didik, L. A., & Aulia, F. (2019). Analisa Tingkat Pemahaman dan Miskonsepsi pada Materi

Listrik Statis Mahasiswa Tadris Fisika Menggunakan Metode 3-Tier Multiple Choice Diagnostic. Phenomenon. 9(1). 99–112.

Entino, R, Hariyono, E., & Lestari, A, N. (2022). Analisis Miskonsepsi Peserta Didik Sekolah Menengah Atas pada materi Fisika. Journal of Science Education.6(1).177-182.

Makhrus, M., & Hidayatullah, Z. (2021). Detecting Student Misconceptions about Physics Using Three Tier Diagnostic Test with Analysis Certainty of Response Index. Jurnal Pendidikan Fisika dan Keilmuan.7(1).23-30.

Maryam, E. (2020). Identifikasi Miskonsepsi Menggunakan Three-Tier Diagnostik Test Berbasis Google Form Pada Pokok Bahasan Potensial Listrik. Silampari Jurnal Pendidikan Ilmu Fisika.2(2).149-162

Mesra, R. dkk. (2023)Research & Development dalam Pendidikan.Medan:Mifndi Mandiri Digital

Nazura, Saputri, F, D. &, Angraeni, L. (2021). Pengembangan Tes Diagnostik Three Tier Test pada Materi Pesawat Sederhana untuk Peserta Didik Kelas VIII SMP. Jurnal Pendidikan Sains dan Aplikasinya. 4(2).54-60

Saifullah,M,A.,Wartono., & Sugiyanto.(2021). Pengembangan instrumen diagnostik three-tier untuk mengidentifikasi miskonsepsi materi fluida statis pada siswa kelas X MIA. Jurnal MIPA dan Pembelajarannya.1(7).517-525

Silaban, Y. F. H., & Jumadi, J. (2022). Concept understanding profile of high school students on doppler effect and sound intensity levels. Momentum: Physics Education Journal, 6(1), 51–58.