Developing four-tier diagnostic test to measure students' misconceptions on simple harmonic motion material

A F Janah, B N Mindyarto, and Ellianawati
Graduated Program of Universitas Negeri Semarang, Indonesia

*Corresponding author: ellianawati@mail.unnes.ac.id

Abstract. This research developed four-tier multiple choice test items to measure its validity, reliability, characteristics, and students' misconceptions on Simple Harmonic Motion material. This research was Research and Development (R & D) using 3D (Define, Design, and Develop) development model. The tenth graders of 1 Slawi Senior High School participated in tryout test and the eleventh graders of 1 Kramat Senior High School participated in the final field test. This research used documentation, interviews, questionnaires, and tests to collect the data. The developed test consisted of 30 four-level multiple choice items covering questions, the level of confidence in answering questions, reasons, and the level of confidence in giving reasons. This research yielded validity, reliability, discrimination power, difficulty level, and the percentage of misconceptions analyses. Validation results from 8 validators stated that the instrument was very feasible with the average percentage of 86.23%. The content validity result using Aiken coefficient got the average score of 0.80 with the valid category at the 0.05 significance level. The reliability of the developed test was 0.892. The difficulty level indicated 5 easy items, 22 medium questions, and 3 difficult questions. Students experienced misconceptions on Simple Harmonic Motion material of 58.63% in the medium category.

1. Introduction
The problems experienced by students in the learning process must be immediately recognized by the teacher, so that the teacher can determine the right steps to solve the problem. One of the problems students often face is misconception. Misconceptions can be interpreted as concepts that are not in line with the understanding or theory agreed upon by scientists in the scientific field [1]. Students are categorized as having misconceptions if they cannot provide a correct explanation regarding the concept of Physics in accordance with the theory agreed upon by scientists [2]. The causes of misconceptions experienced by students can come from the students themselves, which are related to the students' prior knowledge (preconception), cognitive developmental stages that are not in accordance with the concepts being studied, students' limited and wrong learning, students' ability to capture and understand the concepts being learned, and students' interest in learning the concepts being taught [3]. Misconceptions are not only seen from student factors but also influenced by several other things such as the teacher, the learning carried out by the teacher, even the teaching materials used by students are also factors causing the emergence of misconceptions [1].

In Physics learning, misconceptions are easy to find because the concept of Physics material is classified as abstract. The concept is an abstract framework of the characteristics of objects, events or phenomena, and situations, which are understood and facilitate communication between humans that involves thinking or cognitive processes [5]. Concept can be considered as a way to categorize and understand the world around and know the interpretation of something [5] [6]. Misconceptions
experienced by students arise based on daily experiences when students interact with their surroundings so that they can form initial knowledge in students [3]. However, the initial knowledge obtained by students can be right or wrong, this is because the sources of information are inaccurate and the experiences experienced by students are also different. The initial knowledge that students have has a very important role because it can affect the next level of education. This can be interpreted that the ability of students to learn depends on the conceptions obtained based on previous student experiences.

This experience will make students build their own theories in their minds which are not necessarily true. If the concept that is formed in the mind is not true it will be very difficult to correct, because accidentally the wrong concept has become a guideline or guideline. In learning Physics, a good understanding of concepts is closely related to learning achievement and scientific consistency. Learning achievement is a comprehensive indicator of learning outcomes, while scientific consistency is the basis for discipline in applying scientific principles so that it can be said that both are very important in the learning process [7] [8]. Based on the results of observations it is known that students of Senior High School of 1 Dukuhwaru have relatively low learning outcomes, this is due to misconceptions in some Physics material. One way to find out misconceptions in students is by using a diagnostic test. The use of diagnostic tests at the beginning or at the end of learning can help teachers find misconceptions experienced by students about the material that has been studied [9].

A good diagnostic test is a test that is able to provide an accurate condition of the misconceptions experienced by students based on the error information they make [10]. Good questions for the diagnostic test not only show that students do not understand certain parts of the material, but also show how students think in answering the questions given even though their answers are not correct. Four-tier multiple choice diagnostic test is a test that is used to accurately identify misconceptions experienced by students and ascertain the weaknesses and strengths of students in the concepts they receive [11]. The four-level multiple choice diagnostic test consists of: (1) multiple choice questions, (2) students' confidence level in choosing answers, (3) students' reasons for answering questions, and (4) students' confidence level in choosing reasons [12]. The advantages possessed by a four-level multiple choice diagnostic test are that the teacher can: (1) differentiate the level of confidence in the answers and the level of confidence in the reasons chosen by students so that they can explore deeper understanding of concepts, (2) diagnose misconceptions experienced by students more deeply, (3) determine the part of the material that requires deeper emphasis, and (4) plan more effective learning to reduce student misconceptions [13]. Therefore, the present research attempts to develop a four-tier diagnostic test to measure students’ misconceptions.

2. Methods
This research was conducted in Senior High School of 1 Kramat. The subjects of small and wide-scale trials were conducted at SMA Negeri 1 Slawi. The number of students in the small-scale trial was 10 tenth graders of Natural Science Program 3 and the broad-scale trial was 36 tenth graders of Natural Science Program 1. This research was a Research and Development (R & D) research. The development model used was the 3-D model (an adaptation of the 4-D model). Figure 1 presents the major steps of 3-D development model.

![Figure 1. Major Steps of 3-D Development Model](image-url)
The purpose of this research was to develop a four-level multiple choice diagnostic test to reveal students' misconceptions on simple harmonic motion material. The initial activity in this research was to analyze the potential and problems of misconceptions on simple harmonic motion material. The data obtained were then analyzed and used as a reference to design a diagnostic test for studying the misconceptions of simple harmonic motion material.

Data collection methods consisted of documentation, interviews, questionnaires, and tests. Interviews were conducted with the teacher to determine the level of student understanding of the material of simple harmonic motion and the instruments developed. The data analysis consisted of validity, reliability, level of difficulty, discriminatory power, and analysis of student misconceptions. The validity score was determined by the Aiken coefficient which obtained from the judgments of 8 validators consisting of 5 lecturers and 3 physics teachers, while Cronbach Alpha formula was used to determine reliability. Analysis of student misconceptions uses the equation used by Caleon & Subramaniam (2010), to determine CDQ (Confidence Discrimination Quotient) using Equation 1.

\[
CDQ = \frac{CFC - CFW}{S}
\]  

Information:

- \(CFC\): the average of students’ certainty in answering correctly
- \(CFW\): the average of students’ certainty in answering incorrectly
- \(S\): the deviation standard of certainty level

3. Results and Discussion

3.1. The Validity of Four-Tier Diagnostic Test

The validity test is the initial stage before the instrument is tested on research subjects. The instrument developed was validated by 8 validators, consisting of 5 lecturers and 3 teachers. The percentage of feasibility were obtained from the experts’ judgements, while the scores for validity were obtained from Aiken coefficient. The results of the instrument feasibility are presented in Figure 2.

![Figure 2. The Percentage of the Validity tested by Validators](image)

The information obtained from Figure 1 is the average feasibility percentage of 8 validators: 86.23% with very valid criteria so that it can be used without revision. Based on these results it can be said that the instrument is suitable for use in evaluating Physics learning. However, the instrument is still being improved according to the validator's recommendation. Furthermore, the validation results were analyzed based on the Aiken coefficient to determine 3 aspects of the assessment, namely the aspects of material, construction, and language. In the material aspect there are 4 items, construction aspects 9 items, and language aspects 4 items. The results of the Aiken coefficient can be seen in Table 1.

| Table 1. Results of Aiken Coefficient Analysis |
|-----------------------------------------------|
| Aspect    | Aiken Coefficient | Criteria |
| Material  | 0.76              | Valid    |
| Construction | 0.85          | Valid    |
| Language  | 0.80              | Valid    |
Table 1 presents the Aiken value of the instruments that have been developed, from the table it can be seen that the lowest value is in the material aspect (0.76) compared to the construction and language aspects. The overall average score of Aiken is 0.80 with valid criteria. The content validity test can be used to ensure if the assessment instrument is relevant to the objectives [14].

3.2. Reliability of the Four-Tier Diagnostic Test
Reliability test is for showing the extent to which the results of a measurement can be trusted [15]. Reliability refers to the consistency or stability of the assessment results. The results of the instrument reliability analysis that have been developed can be seen in Table 2.

Table 2. Instrument Reliability Analysis Results

| $r_{table}$ | $r_{11}$    | Category       |
|------------|-------------|----------------|
| 0.329      | 0.892       | Very High      |

Reliability analysis is measured using the Alpha Cronbach formula. Based on the table above, it showed that the instrument that has been developed is reliable with a very high category, because $r_{11}$ is bigger than $r_{table}$.

3.3. Characteristics of Four-Tier Diagnostic Test Items
A good test questions must be valid and reliable. In addition, it should have a good level of difficulty and good discrimination power. The levels of difficulty and discrimination power are characteristics of the test items, including the four level multiple choice diagnostic test items. The analyses results of the difficulty level in the instrument showed there are 3 difficult questions, 5 easy questions, and 19 moderate questions. The difficult category questions are questions number 12, 15, and 19. These items are difficult because students are required to be able to understand the parts of each concept in depth and are required to be more thorough in mathematical calculations. The results of measuring the level of difficulty (TK) are presented in Table 3.

Table 3. Results of Instrument Difficulty Level

| Interval Value | Criteria | Question Number | Total Items | Percentage (%) |
|----------------|----------|-----------------|-------------|----------------|
| 0,00 ≤ TK ≤ 0,30 | Hard     | 12, 19, 30     | 3           | 10             |
| 0,31 ≤ TK ≤ 0,70 | Moderate | 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 | 22 | 73.33 |
| 0,71 ≤ TK ≤ 1,00 | Easy     | 1, 2, 17, 18, 23 | 5           | 16.67          |
| Total          |          |                 | 30          | 100            |

Questions on a test are good if they are not too difficult and not too easy. The try-out results show that the level of difficulty of the instrument is good, because there are only 3 difficult questions from a total of 30 questions. There rest is moderate questions. It means the moderate level item test is higher than the easy one. The moderate levels of item tests are important and should be exist so that the middle-low students can still answer the questions. The discrimination power analysis is used to find out whether the item can differentiate between students understand the concept and students do not understand the concept. This is in accordance with what was stated by [16], stating that a good test item must be able to distinguish students who really master the material from those who do not. The results of the power item discrimination (DP) are presented in Table 4.
Table 4. Instrument Discrimination Power Results

| Interval Value | Criteria                        | Question Number | Total Items | Percentage (%) |
|----------------|--------------------------------|-----------------|-------------|----------------|
| 0.40 < DP ≤ 1.00 | Accepted                       | 4,5,6,7,8,11,13,14,24,26,27 | 11           | 36.67          |
| 0.30 < DP ≤ 0.40 | Accepted with few correction   | 9,10,12,15,16,19,20 | 7            | 23.33          |
| 0.20 < DP ≤ 0.30 | Corrected                      | 1,2,3,17,18,21,22,23,25 | 9            | 30             |
| 0.00 < DP ≤ 0.20 | Not Used                       | 28,29,30        | 3            | 10             |

Total 30 100

Based on the results of the analysis in Table 4, it is found that 11 items were accepted, 7 items need a bit correction, 9 items were corrected, and 3 items were need to be eliminated. There are 3 unused items; they are item test number 28, 29, and 30. Those item tests cannot be used because the discrimination value is less than 0.20, so that those are belong to poor category if it is used for further research.

3.4. Student Misconceptions on Simple Harmonic Motion Material

Analysis of student misconceptions was carried out by determining the value of CDQ (Confidence Discrimination Quotient). A negative CDQ score indicates that students experience misconceptions but they did not know it. CDQ scores are determined based on the student's choice of answers, choice of reasons, or both. The results of the student misconceptions analysis are presented in Table 5.

Table 5. Results of Student Misconception Analysis

| CDQ   | Question Number                        |
|-------|----------------------------------------|
| CDQ < 0 | Answer 1, 6, 15, 23, 28 | Reason 2, 8, 15, 18, 22, 25, 29, 30 | Both 2, 3, 6, 7, 8, 10, 12, 14, 15, 18, 19, 22, 25, 26, 28, 29, 30 |
| Total  | 5 | 8 | 17 |

The results of the analysis in Table 5 show that students who experience misconceptions are as much as 16.67% in choosing answers; 26.67% in choosing reasons; and 56.66% in choosing both. Items with a negative CDQ value indicate that the average level of students’ confidence in answering incorrectly was greater than the average level of students’ confidence in answering correctly. This is due to the tendency of students to be unable to distinguish what is understood and what is not understood correctly, which causes misconceptions in students [2]. If it is not resolved, students have difficulty accepting new knowledge, because the wrong concept is deeply ingrained and they assume the concept they have understood is correct.

In addition, they tend to apply the wrong concept rather than the concept they just received. This is in accordance with [17], students who experience misconceptions will have difficulty accepting concept changes. In this case, the teacher has an important role in identifying the understanding of concepts of each student. However, in reality the teacher has problems identifying students ‘conceptual understanding and difficulty interpreting students’ conceptual understanding [18]. The grouping of students’ conceptual understanding on each criterion is presented in Figure 2.

Figure 3. Student Concept Understanding Grouping
Students who understanding the concept is in the low category, 51.01%; moderate is 30.43%; and a high is 18.56%. These results indicate that students who understand the concept of Simple Harmonic Motion are only 18.56%, so it can be said that most students have a low level of understanding the concept of Simple Harmonic Motion material. Furthermore, students who do not understand the concept belong to high category, 1.28%; moderate is as much as 53.10%; and low is as much as 45.62%. Students' lack of understanding of the concept is seen from the level of student confidence in answering the question. The results of the analysis of misconceptions experienced by students were divided into three categories, they are: the low category (24.76%); moderate category (58.63%); and a high category (16.61%). These results indicate all students who experience misconceptions as much as 16.61% in the material of Simple Harmonic Motion. It will be worse if among the students believe the same wrong concept. This condition needs to be fixed immediately because widespread misconceptions will cause problems for students when receiving new knowledge.

4. Conclusion

The instrument developed in this study was a four-tier diagnostic test to measure students' misconceptions on Simple Harmonic Motion material. The test questions consisted of four levels, they are: (1) the questions and the answer key, (2) the confidence level of answering questions, (3) reasons, and (4) the level of confidence in giving reason. The results of the instrument feasibility level of the 8 validators were 86.23% with very feasible criteria. The average value of the Aiken coefficient was 0.80 with valid criteria. The Cronbach Alpha value of the reliability test was 0.892 with the very good category. The test questions consisted of 5 easy items, 22 moderate questions, and 3 difficult questions. The results of the discrimination power test showed that 3 items were invalid because the correlation value was below 0.20. Based on the four-tier diagnostic test, it can be said that most of the students experienced misconceptions on the Simple Harmonic Motion material (58.63%) in the moderate category. Students tend to experience misconceptions in choosing answers and giving reasons (both), this is shown in the results of the percentage of misconception analysis which is 56.66%. Future researches are open to administer remedial on Simple Harmonic Motion material and are encouraged to develop the diagnostic test to identify students’ misconceptions on other material.

References
[1] Suparno P 2013 *Gramedia Widiasarana*
[2] Fariyani Q, Rusilowati A and Sugianto 2015 *J. Innov. Sci. Edu.* 4 42
[3] Yuliati Y 2017 *J. Bio. Edu.* 2 52
[4] Negoro R A and Karina V 2019 *J. Res. Dev. Phys* 5 70
[5] Chi M T H 2009 *Int. Handb. Res. Concept. Change* 90
[6] Streveler R A 2008 *J. Eng. Edu.* 97 279
[7] Krishnayanti R, Feranie S and Padri I M 2015 *J. Res. Dev. Phys* 1 25
[8] Permata A, Fenire S and Karim S 2015 *J. Res. Dev. Phys* 1 45
[9] Lin S 2004 *J. Sci. and Math. Edu.* 2 175
[10] Wahyuningsih T, Raharjo T and Masithoh D F 2013 *J. Phys. Edu.* 1 111
[11] Zaleha Z, Samsudin A and Nugraha M G 2017 *J. Phys. Edu. and Scho.* 3 36
[12] Caleon I S and Subramaniam R 2010 *J. Sci. Edu.* 32 939
[13] Rusilowati A 2015 *Proceeding SNFPF* 6 2
[14] Hendryadi 2014 *Teor. Pers. Pap.* 1 1
[15] Matondang Z 2009 *J. Tabularasa PPs Unimed* 6 87
[16] Nugraeni D, Jamzuri and Suwarto 2013 *J. Phys. Edu.* 1 12
[17] Kaur G 2013 *J. Educ. Confab* 2 79
[18] Hafizah D, Haris V and Eliwatis 2014 *J. Edu. MIPA* 1 100