THE DEVELOPMENT OF DIAGNOSTIC TEST IN WAVES AND SOUND FOR SENIOR HIGH SCHOOL

Winda Yola Yohana dan Sahyar

Department Of Physics Education Faculty Of Mathematic and Science State University Of Medan
windayolayohana@gmail.com, sahyarekonofisika@gmail.com

ABSTRACT

The objectives of the research was to determine the level of validity, reliability, difficulty level, distinguishing power and the effectiveness of the distracting index of the two-stage multiple-choice diagnostic tests developed in senior high school. The types of research is R&D with Borg and Gall development model. The participants of the research is all of student class XII of SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan with online. Participants were selected randomly from all class XII students when the test was conducted online. Students who became participants for the small class test were 10 people and for the large class test were 33 people. From the research results obtained the number of valid questions on the small class test test as many as 25 items and reliable. Then carried out a large class test, obtained the number of valid questions on the field test as many as 18 items and reliable questions.

Keywords: *Diagnostic Test, Validity, Reliability, Difficulty level,*

INTRODUCTION

Learning is a process that brings about a relatively permanent change in a student's behavior after a particular exercise or experience. By learning, it is expected that the results as student assessments and the achievement of the expected results at the completion of certain instructional activities are expected. These results included changes in students' retention of knowledge, skills, attitudes, and other post-instructional changes. Learning outcomes for students are usually facilitated by adopting appropriate learning strategies (Yildirim and Onder, 2018).

The teaching and learning process or teaching process is an activity to implement the curriculum of an educational institution in order to influence students to achieve the educational goals that have been set. To achieve these educational goals, a teacher as an educator has a very important role and influence for students in building student character at school and must be able to create an active, fun teaching and learning atmosphere and impress students in learning activities so that students can absorb and receive the knowledge they have learned. They get for their future life.

Physics learning is a teaching and learning process that studies natural events in everyday life. In studying physics, students' understanding of concepts is needed to get maximum learning outcomes. The concept that is embedded in students must be in accordance with the actual scientific concept of physics (Azizah. et al, 2015). So that students have an

understanding of Physics concepts in accordance with the literature and the agreement of experts. If students understand the wrong concept and continue to be left alone, students will experience not understanding the concept or misconception. Based on this, it is important for a teacher to be able to explore what concepts students bring before the learning process (Sheftyawan and Trapsila, 2018). One way to find out the learning weaknesses of students is with a diagnostic test. The use of diagnostic tests at the beginning and at the end of learning can help teachers find student learning weaknesses in the material being studied (Lin, 2004).

Diagnostic tests are useful to find out the learning difficulties faced by students, including misconceptions in understanding concepts. Where the results of the diagnostic test provide information about concepts that have not been understood by students and which have been understood by students (Djemari, 2004).

The two-tier diagnostic test is one of the most widely used to understand student understanding. This test can be defined as two parts. The first level is a content question which refers to students' understanding of the subject. The second is the level of reasoning of the first response (Chuenmanne and Kongsak, 2017). The test there are two tiers and each item in the first tier consists of multiple choice questions covering content knowledge. The second level of each item contains the reasons students choose to go to the first level. The second level includes a choice of reasons where one of them is the expected answer (Mutlu and Acar, 2015).

Researchers have made initial observations by giving questionnaires to class XII science students at SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan, where students do not know the two-level diagnostic test. Researchers also conducted interviews with physics teachers at SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan, that the teacher did not know the diagnostic test and had never seen it before. The objectives of the research was to determine the level of validity, reliability, difficulty level,

distinguishing power and the effectiveness of the distracting index of the two-stage multiple-choice diagnostic tests developed in senior high school.

RESEARCH METHODS

This research was conducted at SMA Negeri 1 Sijamapolang, Kab. Humbang HAsundutan T.P 2019/2020. The participants in this study were students of class XII IPA in the odd semester of T.P 2019/2020 as many as 10 students in the small class and 33 students in the large class.

This type of research is research and development (R & D) to develop a form of diagnostic test on wave and sound material. With a two-level diagnostic test design with a choice of reasons. This research is a development research that develops a diagnostic test instrument by adopting the Borg & Gall model, need analysis; planning; Developing Preliminary Form Of Product; Preliminary Field Testing; Revising Main Product; Main Field Testing; Operational Product Revision. The assessment tools developed in this study include the development of grids, diagnostic questions, and validation sheets. Data analysis techniques in this study are validity test, reliability test, difficulty level test, discriminating power test, and distractor effectiveness (Borg and Gall, 1983).

RESULT AND DISCUSSION

a. Research Results

Diagnostic tests on wave and sound material were tested on students of class XII IPA SMA Negeri 1 Sijamapolang, Kab. Hasundutan's relationship with the Borg & Gall development model. In this section, the researcher will describe the results of the diagnostic test research on the developed sound and wave material.

The steps are organized in the following order. The first step is need analysis. This stage is the stage in collecting references related to research, such as journals and books. Then an analysis of some of these references is carried

out so that it can be obtained that to determine the level of students' understanding of the concept of physics can be seen from the sources of errors that may be the cause of students' difficulties in understanding the concept of physics, indicators of students' ability to define concepts verbally and in writing, identify and make examples and not examples, use models, diagrams and symbols to represent a concept, change one form of representation to another, recognize various meanings and interpretations of concepts, identify the properties of a concept and recognize the conditions that define a concept and compare the concept.

At this stage, researchers conducted an analysis of the applicable curriculum. Based on the results of an interview with a physics teacher at SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan researchers obtained some information, including that the curriculum used was the 2013 curriculum that had been implemented in the school, the teacher did not know about the diagnostic test and had never seen it so that the teacher only used formative tests in measuring students' knowledge. The researcher chose material whose test development was adapted to the material contained in the 2013 curriculum for class XII IPA, namely the Wave and Sound material that had been taught by the teacher to students in the hope that the diagnostic test that would be given to students could identify the difficulties experienced by students so that the developed test refers to in the 2013 curriculum.

Then, the researcher made a planning. It includes formulating skills and expertise regarding the problem of the research, formulating objectives of each stage, and designing research steps and necessary feasibility study. The development of the diagnostic test was carried out to determine the level of validity, reliability, discriminating power, level of difficulty, distractor index and level of understanding of students' concepts. The developed diagnostic test is designed based on a grid, namely there are number of questions, physics concepts, question indicators, questions, types of concepts in questions and diagnosis of questions. The questions are formed in the form of two-level multiple-choice diagnostic test

questions. Where the question has a choice of answers and a choice of reasons that support the answer. The developed test consists of 30 questions covering the material of waves and sound.

The development of a diagnostic test is carried out in several stages starting with the preparation of a grid of questions. The initial stage is to design diagnostic test questions to identify students' knowledge of physics concepts. The test questions are designed based on the material that has been analyzed and also based on indicators of concept knowledge. The researcher also made test grids and answer keys as consideration for the validator to check the validity of the multiple-choice diagnostic test questions. The test grids are designed based on or referring to the indicators of each question. In addition, researchers also designed scoring guidelines that are used to facilitate researchers, teachers or other researchers in providing an assessment of the results of multiple-choice diagnostic tests to identify the level of understanding of students' physics concepts.

After the development of the test is complete, then the validity test is carried out to the expert to determine the feasibility level of the diagnostic test that has been designed. The validity of the instrument is done by providing a validation sheet of the test grid, test questions, test answer criteria, and answer criteria to the validator. In this validation stage, the validator assesses 20 aspects related to the designed test. The validator provides an opinion that includes the test criteria can be used without revision, the test can be used with few revisions, the test can be used with many revisions, and the test is not feasible to use. The results of expert validity (expert review) are then used as the basis for designing questions at a later stage to be tested on a limited scale. The trial was attended by class XII IPA students, which consisted of 10 students. The trial was conducted online for 45 minutes, where the questions were in pdf form and students completed them in handwriting and then photographed and sent to the physics teacher at the school. Students are asked to do 30 questions on a two-level multiple choice diagnostic test. Students are asked to work on a two-level multiple choice diagnostic test. Based

on the results of students' diagnostic tests, then the validity and reliability were calculated again using the validity and reliability formulas.

Preliminary products or revision trials use the data obtained in the fourth step. Revisions may be carried out more than once depending on the results of product trials. Revision is ready for better testing. From the limited scale test to 10 students with 30 questions, the students' scores were varied. Ranging from low to high scores. This is because students' conceptual knowledge is also different about the material being tested. Based on the results of students' diagnostic tests, then the validity and reliability were calculated again using the validity and reliability formulas. Based on student scores, it was found that there were 25 valid questions. Then the reliability test was carried out on the 30 items which obtained all reliable questions, meaning that the questions could be used repeatedly on the subject and under the same conditions. However, because in the validity test, there are only 25 items in the valid category, so the questions will be given in the field test.

This step is also called the main test where the revised educational product is tested on a larger scale to many parties. Data are usually collected using qualitative methods. Some products that need to be done in experimental research design to get the right data for the next step. The diagnostic test questions that have been tested on a limited scale, were tested on a large group, namely all class XII students of SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan which was held on the day of the trial was carried out on 33 students. The test activity is carried out for 90 minutes online, where the questions are in pdf form and students complete it by hand and then take a photo and send it to the physics teacher at the school. Students are asked to take a multiple-choice diagnostic test containing 25 questions. The results obtained from students' answers were analyzed to measure or determine the level of students' conceptual knowledge abilities. In addition, based on the results of the student's work, the level of validity, level of reliability, level of difficulty, discriminatory

power, distractor index, and level of students' conceptual knowledge of the developed test will be calculated.

Based on the results of field trials, it can be calculated the validity of the test on the Microsoft Excel program. The validity test was carried out based on the results of a field test involving students of class XII science at SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan with a total of 33 students, obtained questions with the status of valid items as many as 18 items and invalid questions as many as 7 items. The esults of the analysis of the diagnostic test validty shown in the Table 1.

Table 1. Results of the analysis of the diagnostic test validty

Item Status	Item Question	Total	Percentage (%)
Valid	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 15, 16, 17, 19, 21, 24	18	72
Invalid	11,14, 18, 20, 22, 23, 25	7	28

The test items can be said to be reliable if they have a reliability greater than 0.70. Based on the results of the field test, it can be calculated reliability test involving students of class XII science at SMA Negeri 1 Sijamapolang, Kab. Humbang Hasundutan with a total of 33 students. Based on the results of data analysis performed on the Microsoft Excel program, the reliability of the test obtained is 0.71. So based on this analysis, there is no revision of the test according to the reliability test and the diagnostic test is reliable or trustworthy to be used in identifying the level of knowledge of students' physics concepts. In other words, this test can be tested repeatedly to students.

The items on the test instrument can be said to be good if the test items have the smallest distinguishing power of -0.02. This shows that the items have sufficient discriminatory power. The following are the results of the analysis of the distinguishing power of a diagnostic test to identify students' knowledge of physics concepts. The results of distinguishing power analysis of diagnostic tests shown in the Table 2.

Table 2. Results of distinguishing power analysis of diagnostic tests

Question Item Status	Item Questions	Total	Percentage (%)
Question cannot be used	11, 14, 20, 22, 23, 25	6	24
Revised question	1, 2, 6, 8, 13, 17, 18	7	28
Questions can be used with minor revisions	5, 10, 16, 19	4	16
Question accepted	3, 4, 7, 9, 12, 15, 21, 24	8	32

The test items can be said to be good if the test items have a difficulty level at the interval of 0.3-0.7, this indicates that the items are not too difficult and not too easy. The following are the results of the analysis of the level of difficulty in the diagnostic test to identify students' knowledge of physics concepts. It can be seen that of the 25 items, all of the questions are in the status of middle items because it can be seen that as many as 25 items or 100% of the questions are in the status of these items and there are no items with the status of difficult or easy items.

The effectiveness of the distractor index is one of the characteristics in determining the quality of a question. The effectiveness of the distractor index in question is to find out the distractor on a question is functioning or not. The number of options on the developed diagnostic test is 100 answer options and 100 reason options from 25 questions. The 25 options are answer keys and 25 options are correct reasons so that it is known that there are 75 options for distracting answers and 75 options for reasoning.

Based on the results of the analysis using the Microsoft Excel program, it was obtained that of the 25 items by analyzing each of the number of quality answer options and the number of quality reasons, it was obtained that 37 answer options and 45 reason options were of very good quality, 28 answer options and 17 options. reasons with good quality, 9 answer options and 7 reasons with poor quality, 1 answer option and 6 reasons with poor quality, and 0 answer options and 0 reason options with poor quality.

b. Final Product Discussion

The development of diagnostic tests that have been carried out to identify students' conceptual knowledge regarding the material of Waves and Sound has gone through a series of phases of developing the R and D model using the Borg and Gall method starting from the preparation and data collection stages, planning, developing a preliminary form of the product, limited scale testing, revision based on the results of the limited-scale test, broad-scale test, revision based on the broad-scale test. The questions are in the form of a diagnostic test that is given to an expert review of 30 questions, then given to a small group with 25 valid questions which are then used for field tests.

The results of this study are in accordance with previous research related to the development of a diagnostic test instrument in the form of a description to identify concept understanding by Sriyanti and Mania (2019), which shows that the test instrument developed as a whole is of good quality in terms of validity, reliability, discriminatory power and level of difficulty. The developed test has a content validity index (CVI) of 1.00 and a high category reliability of 0.75. Compared to results obtained at the expert assessment stage, limited scale test, and field trials, the resulting test instrument reached the predetermined criteria, namely the status of the items that met the valid and reliable criteria. Diagnostic tests are generally declared valid with a percentage of 75%, test reliability is generally declared reliable because based on an analysis conducted using Microsoft Excel. Based on these results, the instrument can be used to measure students' ability to understand the concept of Waves and Sound. From the data in Table 1 states that 18 questions use words that are easy to understand and the questions presented are clear. As for the other seven questions, there are words that have multiple meanings, questions that are not clear and some words that are not relevant and must be replaced. With the instrument's level of

validity, the reliability, level of difficulty and discrepancy can also be measured.

The reliability value of the test 0.71 is obtained with a high interpretation. This means that the diagnostic test is reliable or can be used to identify the level of understanding of students' physics concepts. Because the test questions used are reliable, by doing the test twice to students, the test can be used for the next test.

The distinguishing power of the test based on the criteria of a good question is in the range of $0.3 < P < 0.7$ with good category. Based on the results of the discriminatory data analysis that has been carried out, it was found that, 32% of the questions were accepted, 16% of the questions could be used with a little revision, 28% of the questions were revised, and 24% of the questions could not be used. Based on these results, it means that the developed test is quite capable of distinguishing high-ability students and low-ability students. From the data in the Table. 2 there are 6 questions that cannot be used. Because the questions are too easy both from the questions and the language in the questions, so there is no difference between students who understand the concept and those who don't. There are also 7 questions that must be revised and 4 other questions with minor revisions. These questions can distinguish students who understand the concept and those who don't by slightly revising a few words. And there are 8 questions received, where the questions distinguish students who understand the concept and do not understand.

Followed by Maulidiansyah's research (2018), with the results of the study being a low level of difficulty, the researcher took questions with a moderate to easy level of difficulty.The level of difficulty of the test is seen from the index of each item of the question. Based on the data analysis that has been carried out, the average level of difficulty of the questions is 0.3-0.7 with middle item status. So that the overall difficulty level of the diagnostic test is low with a percentage of 100% with middle item status. Because, the level of difficulty affects the validation of the test. More items with high difficulty, resulted in most students getting poor

results. The level of difficulty on the test does not stimulate students to solve the problem, because the question attaches several questions that are the same.

Based on the results of the analysis of the effectiveness of the options that have been carried out, it can be concluded that the distractor (detractor) is sufficient to deceive the answers of the test takes, this can be seen from the answers of each participant who answered the diagnostic test. The effectiveness of distractors in these questions has greatly made students feel deceived in choosing the correct answer. Because the sentences in both the answer choices and the reason choices describe similarities. So that students assume the distractor's choice is the correct answer.

The test results show that the developed diagnostic test is able to identify the level of students' conceptual understanding, it can be seen from how many students are identified as misconceptions, do not understand the concept or understand the concept. So from the results of this diagnostic test, it can be seen in which sub-materials students experience the most misconceptions through analysis of the level of understanding of the concept.

The implementation of this test development has many weaknesses and obstacles faced by researchers so that the implementation of this test development is not fully achieved. These obstacles include making question grids that are less effective and seem rushed and do not ensure that students understand the material or not because of online learning. The implementation of yes online is less than optimal. Where researchers only monitor students completing test questions online and efficiently from time to time. Researchers also do not know the conditions and circumstances of students in working on the questions.

CONCLUSIONS AND SUGGESTIONS

Based on the results of research that has been carried out for the development of a conceptual knowledge diagnostic test on Waves and Sound material in high school, a conclusion is obtained the validity level of the diagnostic

test developed for Waves and Sound material has met the good qualifications. The reliability level of the diagnostic test developed for Waves and Sound material has met the good qualifications with a reliability value in the high reliability category. The distinguishing power of the diagnostic test developed for Waves and Sound material has met the good qualifications. The level of difficulty of the diagnostic tests developed for wave and sound material has met the bad qualifications. The distinguishing power of the diagnostic test developed for Waves and Sound material has met the good qualifications. The level of effectiveness of the distracting index of diagnostic tests developed for Waves and Sound material has met that low answer option and low reason option. The diagnostic test developed to determine the level of understanding of students' physics concepts, it was found that the average number of students who were included in the concept understanding category.

Based on the results of the study and the conclusions above, then as a follow-up to this research, it is suggested the following:

1. For further researchers, who will carry out the development of diagnostic tests, they should ensure that students understand the material to be tested, so that the grid of questions made can be used effectively.

2. Designing the best possible test environment to ensure a good and conducive test execution.

REFERENCES

Azizah, R., Yuliati. L., dan Latifah. E., 2015, The Physic Problem Solving Difficulties on High School Student, Jurnal Penelitian Fisika dan Aplikasinya, (JPFA), 5(2), 44-50.

Borg, W.R., and Gall, M.D., 1983, Educational research: An introduction. New York: Longman.

Chuenmanee, C., and Thathong. T, 2017, The Development of Two-Tier Diagnostic Test for Evaluating Primary Students' Understanding on Plant Life, International Journal of Advanced Scientific Research and Management, 2(8), 79-85.

Djemari, M., 2004, Penyusunan Tes Hasil Belajar, Yogyakarta : Program Pascasarjana UNY.

Lin, S. W., 2004, Development and Application of a Two-tier Diagnostic Test for High School Students' Understanding of Flowering Plant Growth and Development. International Journal of Science and Mathematic Education, 2(4), 175-199.

Maudiansyah, D., 2018, Pengembangan Tes Diagnostik Menggunakan Aplikasi Google Form Materi Momentum dan Impuls untuk Siswa Sma, Jurnal Ilmiah Pendidikan Fisika Al-BiRuNi, 7(2), 207-217

Mutlu, A., and Burcin, A., 2015, Development of a Two-tier Diagnostic Test to Assess Undergraduates' Understanding of Some Chemistry Concepts, Procedia-Social and Behavioral Sciences, 174(7), 629-635.

Sheftyawan, W., dan Prihandono. T., 2018, Identifikasi Miskonsepsi Siswa Menggunakan Four-Tier Diagnostic Test Pada Materi Optic Geometri, jurnal pembelajaran fisika 7(4), 147-153.

Sriyanti, A., and Sitti, M., 2019, Pengembangan Instrumen Tes Diagnostik Berbentuk Uraian untuk Mengidentifikasi Pemahaman Konsep Matematika Wajib Siswa Man 1 Makassar, Jurnal Pendidikan Matematika, 2(1), 57-69.

Yildirim, H. I., and Sensoy, O, 2018, Effect of Science Teaching Enriched with Technological Practices on Attitudes of Secondary School 7th Grade Students towards Science Course, Universal Journal of Educational Research, 6(5), 947-959.