



## THE DEVELOPMENT OF DIAGNOSTIC TEST WITH TESTLET MODEL TO DETECT DIFFICULT LEARNING STUDENTS IN TEMPERATURE AND ITS CHANGES MATERIALS

R. K. Safitri<sup>1</sup>, N. R. Dewi<sup>2</sup>✉

<sup>1</sup>Science Education Study Program, Faculty of Mathematics and Natural Sciences,  
Universitas Negeri Semarang

### Article Info

#### Article History:

Received January 2020

Approved May 2020

Published July 2020

#### Keywords:

Diagnostic test, testlet,  
learning difficulty

### Abstract

This Research aims to determine the characteristics of the testlet diagnostic model and the profile of students' learning difficulties on the material temperature and its changes. The research design using the 4D model reduced to 3D includes: define, design, and develop. The study was conducted at SMPN 1 Semarang with a total of 39 students. Data collection techniques through test techniques, questionnaires, interviews, and documentation. Data analysis techniques using descriptive-quantitative methods through the test of characteristics (validity, reliability, level of difficulty, distinguishing power), questionnaire analysis of teacher and student responses, and analysis of learning difficulties profiles. The results showed that the diagnostic test of the testlet model can be used to detect learners' learning difficulties in the temperature and its changes material that valid, reliable, relevant, representative, practical, discriminative, specific, and proportional characteristics. Profile of learning difficulties in the material temperature and its changes, namely, verbal ability, the highest and lowest of 64% and 31% in Competency Achievement Indicators 6 and 3; mastery of concepts, the highest and lowest of 68% and 31% in Competency Achievement Indicators 8 and 5; and the ability to count, the highest and lowest of 75% and 9% in Competency Achievement Indicators 5 and 1.

© 2020 Universitas Negeri Semarang  
p-ISSN 2252-6617  
e-ISSN 2502-6232

✉ Corresponding author:

**N. R. Dewi**

Science Education Study Program, Faculty of Mathematics and Natural Sciences,  
Universitas Negeri Semarang, Indonesia  
E-mail: noviratnadewi@mail.unnes.ac.id

## INTRODUCTION

Education assessment standards in Indonesia are regulated based on Permendikbud Number 23 of 2016. The Permendikbud explains that assessment is a process to collect and process information to measure the achievement of student learning outcomes. The process is carried out by educators in the learning process where interaction occurs between students, between students and educators and learning resources in a learning environment. Permendikbud Number 22 of 2016 concerning Basic and Secondary Education Process Standards explains that from the learning process an assessment process can be carried out which results can be used by educators to evaluate processes, learning progress and plan improvement programs that are better in accordance with the Education Assessment Standards.

Science subjects in junior high school are related to how to systematically find out about nature which is divided into four areas of study, namely, energy and its changes, living things and life processes, matter and nature, and the space earth. Students before learning about a concept, a conception has been formed by themselves at the beginning before the concept is studied so that the learning process is based on a constructivist approach (Rifa'I & Anni, 2016). Concepts that are formed by students themselves are often different from those brought by the teacher, causing students to often experience difficulty learning in understanding concepts.

The results of interviews with science teachers and direct observation in SMP Negeri 1 Semarang, it is known that the success of science learning is seen from the learning outcomes of students who exceed the minimum completeness value. The results of the analysis of the acquisition of Final Semester Assessment of VII grade students in science subjects showed that the classical learning completeness criteria were still low. The science concept does not only refer to rote memorization in the process of understanding but also mathematical abilities. The temperature and its changes material is one of material that is not only viewed from understanding concepts but also mathematical

abilities and converting units so that students often experience learning difficulties.

Learning difficulties are one of the factors that support the less than optimal level of achievement of learning outcomes. The causes of learning difficulties can be grouped into two, namely factors that come from inside the students (internal factors) and factors that come from the students' environment (external factors). Learning difficulties that occur on an ongoing basis can be detrimental to students to receive advanced material. Learning difficulties can be diagnosed by teacher as the first step in determining academic policy so that learning objectives can be achieved optimally.

Diagnostic tests as a form of test are used to identify problems or learning difficulties of students. Arikunto (2013) explains that a diagnostic test is a test that can be given to detect the weaknesses of students so that teachers can provide assistance appropriately. A teacher can find out the level of understanding of the concept or the absorption of students' learning of the material that has been given through a diagnostic test. Diagnostic tests have been developed and showed the results of the development instrument have been successfully used to capture the learning difficulties of junior high school students in some science learning materials.

The form of diagnostic test is generally developed in the form of multiple choice (multiple choice). The advantages of using multiple choice tests that can be used on a wide scale, can measure various levels of ability, reach broad material, easy in scoring and efficient in the use of time. Multiple choice tests also have the disadvantage, namely the possibility of the testee to guess the answer in a gambling way. Simkin & Kuechler (2005) conclude that multiple choice test forms give students the opportunity to guess the correct answer is greater than in the constructed response form. Several studies have tried to develop forms of testing that can be used to overcome the shortcomings of multiple-choice and constructed response form. Slepko and Shiell (2014) have revealed that the testlet can be

used as a combination of multiple choice form and constructed response form.

The testlet model was made by considering the efficiency of multiple choice and constructed response form. The items made are interrelated in providing information on supporting questions. Shiell and Slepko (2015) explained that the tiered items in the testlet were considered as an integrated assessment group. The correct answer to each question can provide information to students as a whole or in part as an initial result before continuing the next question. The first supporting question on the testlet will be the basis for the next supporting questions so that if the first supporting question is answered incorrectly, then the student will not be able to correctly answer the next question. Muna *et al* (2017) have developed a testlet model of diagnostic test instrument to detect students' learning difficulties on the buffering subject. The learning difficulties detected in the study were seen from the results of the achievement of students' indicators.

Based on the description described earlier, there are research gaps that have yet to be further investigated. There is no research that has developed a diagnostic test with testlet model on the temperature and its changes material for junior high school students that function to detect learning difficulties. From the explanation, the researcher chose the title "The Development of Diagnostic Test with Testlet Model to Detect Learning Difficulties in the Temperature and Its Changes Material".

## METHODS

This research was conducted as a Research and Development (R&D) research that is a research method used to produce certain products and test the effectiveness of these products (Sugiyono, 2017). The research design adopted the research model proposed by Thiagarajan (1974), namely the 4D model including: (1) define, (2) design, (3) develop, and (4) dissemination. The implementation of this research development is limited to the develop stage. This research was conducted at SMPN 1 Semarang in the even semester of the 2019/2020 school year. The sample used in this study were 7

students of class VIII A in the product trial and 32 students of class VII C in the use trial.

Data collection techniques used in this study were tests, questionnaires (validation from assessment experts and material experts as well as teacher and student responses), interviews, and documentation. Data analysis techniques used in this study were, test the characteristics (validity, reliability, level of difficulty, distinguishing power), analysis of questionnaire responses of teachers and students' responses, and analysis of learning difficulties profiles of verbal abilities, mastery of concepts, and ability to count.

## RESULTS AND DISCUSSION

### Characteristics of Test Diagnostics with Testlet Model

Diagnostic test with testlet model in temperature and its changes material are assessment instruments that are expected to be able to detect students' learning difficulties. As a measurement in evaluation, a test must meet certain requirements that can provide accurate data. Arifin (2012: 69) explained that the characteristics of good instruments, namely: (1) valid, (2) reliable, (3) relevant, (4) representative, (5) practical, (6) discriminatory, (7) specific, and (8) proportional.

#### (1) Valid

Validity is very important to reveal the quality of an instrument so that it provides accurate data in accordance with its function (Arifin, 2012). The validity test is carried out by 6 expert validators or raters each on the assessment and material experts. Validity is carried out by experts who have different ratings, comments and suggestions to improve the developed instrument. The validated diagnostic test with testlet model consists of 14 main items which are translated into 70 integrated testlet items. Validity testing is carried out before the diagnostic tests with testlet model are tested on product trials (small scale). The results of the validation of the diagnostic test with testlet model by the assessment and the material expert were then analyzed using the Aiken V formula.

The results of the validity analysis of the diagnostic test with testlet model resulted in a valid instrument. The criteria used to state an item is said to be valid for the number of assessors of 6 people with a significant level of  $\alpha = 5\%$  is 0.78 (Aiken, 1985). Each integrated testlet item is said to be valid because  $V_{\text{count}} > V_{\text{tabel}}$  Aiken. The diagnostic test with testlet model whose validity has been tested is suitable for use in the field after being revised according to suggestions and improvements from the assessment and materials experts.

The diagnostic test with testlet model that valid and revised was then selected by the researcher. Selection of questions is done by eliminating the main questions number 6, 8, and 11. Elimination of the main questions is done by considering the time of the test. Arikunto (2013: 104) explained that the preparation of a test consisting of many items must be proportional to the length of time the test was given in order to be meaningful. The diagnostic test with testlet model is rearranged and ready for testing to students in offline and online through the address <https://bit.ly/tesdiagnostik-testlet> so that further analysis can be done.

## (2) Reliability

The test instrument is said to have high reliability if it provides relatively fixed results when used by a group of students on different occasions with the same test instrument (Arikunto, 2013). The reliability of the diagnostic test with testlet model was obtained from the analysis using the Cronbach Alpha formula with the help of the Microsoft Excel 2010 application. The reliability test on the diagnostic test with testlet model resulted in  $r_{11}$  of 0.70. Rusilowati (2014) states that the reliability of the instrument is high if the reliability coefficient ( $r_{11}$ ) is  $0.6 \leq (r_{11}) < 0.8$ . The reliability coefficient only shows the level of the coefficient, so to complete the calculation, a comparison is made with the product moment table  $r_{\text{tabel}}$ . The amount of  $r_{11}$  is compared to the  $r_{\text{tabel}}$  with 32 testees, namely 0.35. This comparison results in the  $r_{11}$  value obtained which is higher than the  $r_{\text{tabel}}$  so that the diagnostic test with testlet model is declared reliable in the

high category. This shows that the diagnostic test with testlet model that has been developed can be trusted and gives the same results if it is used on different occasions such as different subjects, places, or conditions.

## (3) Relevant

The diagnostic test with testlet model to detect learning difficulties of students is said to be relevant because the preparation is in accordance with the competencies both competency standards, basic competencies, and indicators of competency achievement. The preparation is done in accordance with the domain of learning outcomes of the measured knowledge. Analysis of the results of the teacher's questionnaire responses showed an average achievement in percent of 100% with a very good category. The average achievement of student questionnaire responses obtained in percent is 83% with a very good category.

Making of assessment instruments according to competence will allow alignment between the form of the instrument, the material and the appropriate time allocation with the depth of the material. Rachmawati (2018) states that understanding Basic Competency (KD) is important to support the learning and assessment process carried out by the teacher. The arrangement of instruments based on KI and KD is then used as an evaluation tool to detect difficulties in understanding the cognitive aspects of students

## (4) Representative

The diagnostic test with testlet model is compiled represents all material in the learning process so that it is said to be representative. The results of the analysis showed the average achievement of teacher questionnaire responses in percent by 100% with a very good category. The teacher assesses the diagnostic test with testlet model that represents all temperatures and its changes sub-material, questions containing specific uses, selections, or operations and can measure concepts, formulas, and laws on temperature and its changes material that students have learned.

Analysis of the results of the questionnaire responses of students showed an average achievement in percent of 71% with good categories. The results of the students' responses have achievements with different categories from the teacher because students feel less accepting of the material explanation at school but do not seek additional materials independently at home. This affects students in giving responses because they feel they have not gotten the material fully so they do not understand the concepts, formulas, and laws on the temperature and its changes material.

Analysis of the results of validation by material experts also supports representative characteristics. Quality of material within indicators of accuracy, correctness and appearance of material. Analysis of the results of the validation by the material experts showed a valid test diagnostic with testlet model on each item. Obtaining material quality from the results of material expert validation was also carried out in the research of Roza & Bulan (2019) which has developed a diagnostic test of misconceptions in the form of a three-tier test on Newton's Law material.

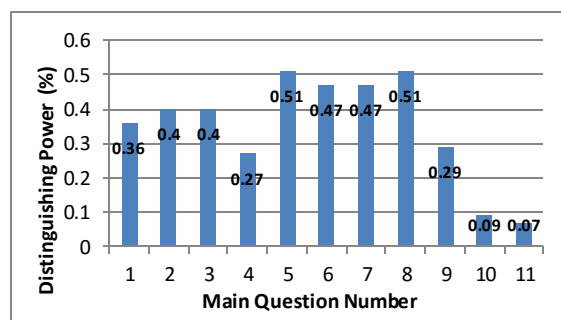
### (5) Practical

The diagnostic test with testlet model that developed is practical to use because of the ease of use of the instrument both in terms of its preparation or use by others. The average achievement in percent of the analysis of the results of the questionnaire responses from teachers and students was 98.96% and 80.25% with very good categories in both. The teacher and students stated that the diagnostic test with testlet model was prepared with clear instructions for the work and was equipped with the presentation of images/graphs/ tables on the questions that needed. In terms of teacher arrangement, it is stated that the diagnostic test with testlet model has an easy level of compiling, examining, and interpreting the result data. In terms of workmanship, students stated that the diagnostic test with testlet model had a good level of difficulty and sufficient time to do on it. Arini, Susilaningsih, & Dewi (2017) who explain that a practical instrument means that it is easy to use

both administratively and technically. Administratively, it means that the use of these instruments is not complicated.

### (6) Distinguishing Power

Distinguishing power test 11 items of the main question of the diagnostic test with testlet model are carried out from the results of the trial use. The results of the analysis of the distinguishing power test resulted in 2 items in the bad category, 6 items in the sufficient category, 3 items in the good category, and there were no questions in the very good category. The results of the distinguishing power test are presented in Figure 1.



**Figure 1.** Distinguishing Power Test Achievements

D'sa & Dionaldo (2017) in their research explained that discrimination ability is an important index that provides information about the ability of items to distinguish between high and low ability. Items that have distinguishing power with the bad category, namely items 10 and 11, mean that the questions are lacking or cannot differentiate the abilities of students. Less discriminating items should be examined to determine possible errors so they can be modified or discarded. This is in line with Rusilowati's (2014) statement, namely, if the item cannot distinguish the two abilities of students, it is necessary to analyze the possibilities, namely: (1) the answer key to the item is incorrect, (2) the item has 2 the correct answer key or more, (3) the measured competence is not clear, (4) the distractor does not function, and (5) most students think that there is misinformation in the item in question.

### (7) Specific

The diagnostic test with testlet model is said to be specific because it is prepared and used specifically for an object that is measured and does not trigger double speculation. The average achievement of the questionnaire responses of teachers and students by 100% and 90% with a very good category. The results of the responses of teachers and students show specific instruments because the diagnostic test model testlet has questions that focus on the temperature and its changes material. In addition, the diagnostic test with testlet model is prepared using communicative language and does not cause double meaning both in the question text and answer choices. Gadbury-Amyot *et al* (2003) explain that the preparation of the instrument or assessment that will be used must be in accordance with a specific concept that has been determined previously.

Specific characteristics also lie in the characteristics and usefulness of the diagnostic test model that distinguishes it from other tests. The diagnostic test with testlet model that was developed has a function to diagnose learners' learning difficulties on temperature and its changes material. The learning difficulty profile of students can be revealed from the achievement of the diagnostic approach. The position of the diagnosis is to find the location of the learning difficulties of students and determine possible ways to overcome them. The diagnostic test with testlet model that was developed specifically localizes the location of learning difficulties in verbal abilities, concept mastery, and ability to count of students on temperature and its changes material.

The diagnostic tests that developed Specifically are arranged in the form of testlets. The form of diagnostic test is generally developed in the form of multiple choice. There are two-tier, three-tier and four-tier multiple choice diagnostic tests.

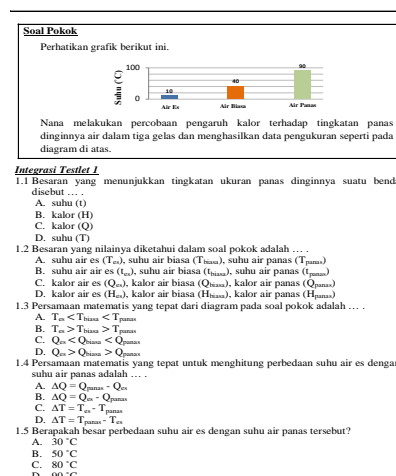


Figure 2. Form of Diagnostic Test Model Testlet

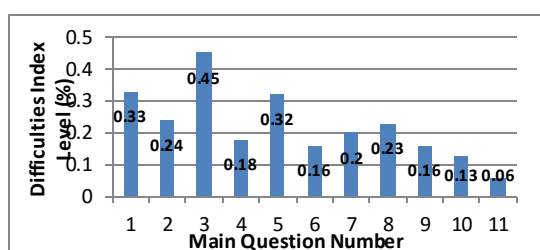
Testlet is a newly developed question form that combining the pedagogical advantages of the Constructed-Response and procedural forms of the Multiple Choice test (Slepokov & Shiell, 2015). Wainer, Sireci, & Thissen (1991) stated that testlets are included in the super test type that produces many hierarchical responses. The developed testlet diagnostic test model is used to detect the learning difficulty profile by scoring analysis using the Graded Response Model (GRM) method because of the graded level of problem solving. For example, the use of the GRM method for scoring, if students who answer correctly on the testlet integration items number 1, 2, 3, 4, and 5 will get the maximum score. Students who answer all testlet integration items and answer incorrectly in points 3 and 5 will get a score of 2. However, if students answer incorrectly on the testlet integration item number 1 then automatically the score is 0 even though they can answer all or some of the questions right afterwards.

### (8) Proportional

A good test instrument has proportional characteristics, meaning that the instrument has a level of difficulty or difficulty that is balanced between difficult, medium, and easy category. Arifin (2012) states that in order to obtain good learning outcomes data, it is better that the proportion of the difficulty level of questions is



normally distributed with a proportion of difficult: medium: easy questions of 25%: 50%: 25% or 1: 2: 1. According to Arikunto (2013) the test questions are too easy or too difficult does not mean they should not be used. The item difficulty level function is related to the test objective. Wahidmurni, Mustikawan, & Ridho (2010) explained that for the purposes of the diagnostic test, items with medium difficulty were used. Based on the theoretical review shows that the proportional comparison of difficult: moderate: easy questions of 2.7: 1: 0 on the diagnostic test with testlet model is not as it should be. The difficulty test results are presented in Figure 3.



**Figure 3.** Achievement of Difficulties

The diagnostic test with testlet model is expected to have difficulty in the medium category but the results of the analysis show the average achievement of the difficulty level in the difficult category. The number of problems with the difficulty level of the diagnostic test with testlet model because the temperature and its changes material that asked are not yet complete learning. Muna, Noer, & Linda (2017) in their research explained that if the items were categorized as difficult, the prediction of the information was that the material in question had not been taught or had not yet been thoroughly studied.

Clarification was carried out through the interview method to randomly selected students and science teachers. Information obtained from the clarification process is that the teacher is carried out briefly and does not make special assessments for temperature and its changes material due to time constraints. The absence of a special assessment indicates the ability of students

to temperature and its changes material have never been measured. Other information obtained was that students felt less accepting of the material explanation at school and did not try to find additional teaching materials independently at home. Students have an initial view of the temperature and its changes material are quite difficult because there is a calculation. Students find it difficult to do the test questions for the given testlet model so that it affects the difficulty level test results.

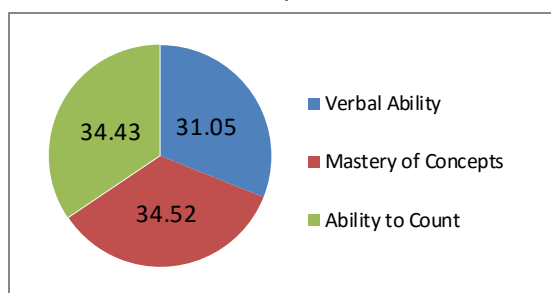
The difficulty level in this study was analyzed using classical theory. Hanbleton, Swaminathan, & Rogers (1991: 2-5) explain that in classical theory, the difficulty level category depends on the ability of students who are given the test. If students have high abilities, the difficulty of the questions will be very easy. If students have low abilities, the difficulty of the questions will be very difficult. The level of difficulty becomes very difficult to estimate precisely because the estimation of the level of difficulty is biased by the sample. The level of difficulty items depends on the ability of the research subject being tested and the general ability of all research subjects (not only individuals). The characteristics of the different research samples made it difficult to estimate the difficulty level of the questions accurately.

### Student Learning Difficulty Profile

Profile of learning difficulties obtained from the scoring process from a combination of answers from students. The diagnostic test with testlet model on temperature and its changes material has eight competency achievement indicators, namely: (1) analyzing the concept of temperature, (2) determining the characteristics of various types of thermometer, (3) determining the temperature scale on the thermometer, (4) converting the thermometer scale, (5) determining the scale of the non-scaled thermometer by comparing the scaled thermometer, (6) analyzing the length expansion, (7) analyzing the area expansion, and (8) analyzing the volume expansion. Each competency achievement indicators becomes a reference for the preparation of indicators of items that each question has 3

indicators of learning difficulties, namely, verbal ability, mastery of concepts, and ability to count. A recapitulation of the level of learning difficulties for each indicator is presented in Figure 4.

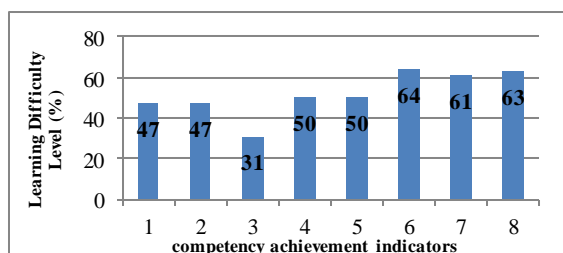
The level of learning difficulties in each indicator of learning difficulties has quite high criteria. Achievement of the level of learning difficulties from highest to lowest is 34.52% in indicators of mastery of concepts, 34.43% in indicators of ability to count, and 31.05% in indicators of verbal ability.



**Figure 4.** Achievement of Learning Difficulty Levels for Each Indicator (%)

### (1) Verbal Ability Indicator

Purwanto *et al* (2015) states the verbal ability in Physics includes the ability to understand and remember the meaning of words, symbols, physics terms contained in concepts and problems. The highest level of learning difficulty indicators of verbal ability on competency achievement indicators number 6 is analyzing the length expansion of 64%. The lowest level of learning difficulty indicators of verbal ability is found in competency achievement indicators number 3, which is determining the temperature scale on the thermometer by 31%. A profile of learning difficulties indicators of verbal ability at each competency achievement indicators is presented in Figure 5.



**Figure 5.** Learning Difficulties Profile

Verbal ability has the lowest level of learning difficulty of the three indicators of learning difficulty in this study. Anugraheni & Handhika (2018) in their research showed that students' verbal representation ability was the highest compared to other representational abilities in understanding material. Apart from the learning process, the verbal ability of students is known to come from daily habits that are often done so that students are easier in terms of verbal (written/oral). Students rarely realize long expansion in their daily habits so that competency achievement indicators number 6 is the highest verbal ability. Determining the temperature scale on the thermometer becomes the lowest verbal ability difficulty because students often hear about the temperature scale and the use of a thermometer.

The results of the analysis show that the three competency achievement indicators with the highest level of verbal ability difficulty are length expansion, volume expansion and area expansion. The difficulty level of learning at the three competency achievement indicators for expansion has a slight difference. In sequence, the achievement of the difficulty level of learning at the three competency achievement indicators of expansion from the highest was 64%, 63%, and 61%. Expansion material is one of the materials considered difficult by students because students often mistakenly compare each expansion. Students say that it is difficult and often wrong to compare the symbols used in the expansion of length, area, and volume, especially the expansion coefficient symbols. Competency achievement indicators number 3, which is determining the temperature scale on the thermometer, is a learning difficulty at the lowest verbal ability. The main problem displayed for competency achievement indicators number 3 is a picture of the various thermometers and their scales. Students can know the components and symbols that are known from the scale shown on each thermometer because they have understood the concept of the upper fixed point and the lower fixed point on the thermometer in question.

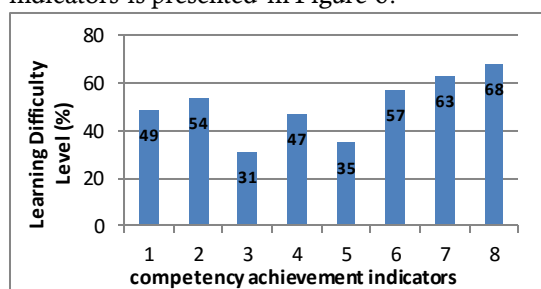
Identifying the level of difficulty verbal ability will lead teachers to develop appropriate



academic programs as a solution. Remedial is a special form of teaching that is often carried out by teachers with the aim of correcting some or all of the learning difficulties faced by students. Dewi *et al* (2019) Combination of various media and storytelling also has a positive influence on learning activities. This learning activity is able to increase the learning motivation and become a guide for students in conducting learning activities. At the same abilities, opportunities and conditions to achieve learning goals, the performance and results achieved by motivated students will be better than students who are not motivated.

## (2) Concept Mastery Indicators

The ability to master concepts is the ability of students to grasp understandings and be able to express a material to be understood. Competency achievement indicators number 8 is analyzing volume expansion to be an indicator with the highest level of difficulty learning mastery of concepts at 68%. The lowest level of learning difficulty in the mastery concept indicator is found in competency achievement indicators number 3, which is determining the temperature scale on the thermometer by 31%. A profile of the learning difficulties of the concept mastery indicators at each competency achievement indicators is presented in Figure 6.



**Figure 6.** Profile of Learning Difficulties with Concept Mastery Indicators

The results of using the diagnostic test with testlet model showed that the learning difficulty indicator of concept mastery was the highest of the other indicators of learning difficulties. Students are known to tend to memorize formulas and how to solve problems rather than trying to understand the concept correctly.

The learning process of science must develop concept changes so that learning science requires mastery of the correct concepts. Mastery of concepts in volume expansion is the achievement of the highest learning difficulty level. The three competency achievement indicators with the highest difficulty level of mastery of concepts are volume, area, and length expansion. Expansion material of substances is one of the materials considered difficult by students because students often experience errors in mastering the concept of expansion. Mistakes in mastery of the concept of expansion lead to errors or errors in solving the problem of expansion. Mistakes in mastering a concept can lead to other concepts that are not in accordance with the concepts used by scientists so as to create a misconception. Mastery of concepts by students is not only about recognizing a concept but can connect between concepts with other concepts in various circumstances.

The level of mastery of the concept of a material is needed as a prerequisite for the continuation of the next level of learning. Students who have mastery of concepts will easily solve problems in the learning process and vice versa. Budiarti *et al* (2017) in their research explained that the mastery of the concept of low temperature and heat was caused by students who tended to memorize formulas more than to understand the concepts correctly. The factors that can cause low concept mastery in students due to misconceptions.

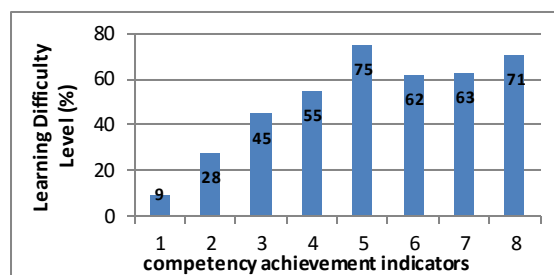
Mastery of concepts needs to be improved so that learning outcomes can be maximized and achieve predetermined learning objectives. Remedial activities can be carried out to fix some or all of the learning difficulties faced by students. Efforts to improve concept mastery can be done through the use of instructional media by teachers in the learning process. Dewi *et al* (2018) in their research stated that the use of Science Digital Storytelling can improve students concept abilities. Research by Kaniawati *et al* (2016) states that the use of audiovisual media has an influence on students' conceptual mastery. Efforts to improve concept mastery do not only come from teachers but students can also contribute. Positive

self-concept and high interest in learning can improve students' mastery of science concepts. Students conceptualize a positive self-condition starting from small things both at home and at school and are supported by teachers, parents, and friends.

### (3) Ability to Count Indicator

The ability to count measured in this study is the mathematical ability and arithmetic of a concept that has been learned. The lowest level of learning difficulty indicators of arithmetic skills of 9% is found in competency achievement indicators number 1, namely analyzing the concept of temperature. Determine the scale of a non-scaled thermometer by comparing it to a scale with a competency achievement indicators of number 5 with the highest difficulty level by 75%. A profile of learning difficulties indicators of ability to count at each competency achievement indicators is presented in Figure 7.

The learning difficulty for the highest numeracy ability indicator is found in competency achievement indicators number 5, namely, determining the scale of an unscaled thermometer by comparing it with a scaled thermometer. The ability to count in determining the scale of an unscaled thermometer by comparing it with a scaled thermometer is the highest indicator because students rarely use a thermometer in the learning process of temperature and its changes material because the number is limited. Physics is classified as physical knowledge so that in learning it requires direct contact (experiment) with what you want to know. The results of research by Subekti & Ariswan (2016) show that there is a significant increase in concept understanding in terms of cognitive physics learning outcomes from the use of guided inquiry learning models through the experimental method. Students in solving calculation problems are based on memorizing mathematical formulas that have been accepted, not because they understand the concept of temperature and its changes.



**Figure 7.** Learning Difficulty Profile Indicators Ability to Count

The achievement of the difficulty level of the ability to count indicator is not much different from the mastery concept indicator. The ability of students to understand the concept influences the ability to solve the count problem. The results of research by Charli, Amin, & Agustina (2018) explain that the factors that cause students to have difficulty solving problems related to formulas are students not mastering the concept of material well. According to the results of research from Kulsum & Nugroho (2014), it is explained that students are only skilled in the material and examples of calculation questions that the teacher has given in class for memorizing. Memorization done by students causes them to not understand the physical meaning of a concept.

The numeracy ability of students in science (physics) learning must be improved so that the learning objectives can be achieved and obtain maximum results. Basic math skills are needed in learning physics because physics material really requires high numeracy and logic skills. Research conducted by Jatiuturo *et al* (2018) provides information that learning physics with the inquiry method can develop students mathematical logic skills. Another study by Dewi *et al* (2020) states that the inquiry method can involve maximally entire ability of learners to search and investigate in a systematic, critical, logical, analytical. Moreover the inquiry learning model is one model of learning that can promote students active learning and improve conceptual understanding. Chasanah & Dewi (2015) has developed the use of contextual-based sciencepolygame media as a science-edutainment media to help students be more motivated to learn and diligently work on questions. In addition,

students need to foster a sense of interest in mathematical problems by being diligent in learning so that they can remember formulas by understanding concepts not by memorizing.

## CONCLUSION

Based on the explanation above, it can be concluded that the results of the study are in the form of the diagnostic test with testlet model which has the following characteristics: (1) valid, analyzed using the Aiken V formula and each question has a  $V_{\text{count}}$  greater than  $V_{\text{table}} = 0.78$  in the assessment and material expert, (2) reliable, has a value of  $r_{11} = 0.70$  which is greater than  $r_{\text{table}} = 0.36$ , (3) relevant to the 93% achievement in the excellent category, (4) representative with 83% achievement in the excellent category, (5) practical with 88.75% achievement in the very good category, (6) discriminatory, 2 items with bad category, 6 items in enough category, 3 items in good category, and there are no questions with very good category, (7) specific with 95.50% achievement in very good category; has a function to detect learning difficulties with indicators of verbal ability, mastery of concepts, and numeracy skills; specific to the material temperature and its changes; diagnostic tests using a testlet model consisting of 11 main questions which are translated into 55 integrated testlets; and can be accessed online at the address <http://bit.ly/tesdiagnostik-testlet> and (8) proportional with an average score of 0.22 in the difficult category. The level of difficulty is not proportional so it needs to be followed up or can be used by students who have high ability in understanding the material.

Profile of learning difficulties obtained from the use of diagnostic test with testlet model on the temperature and its changes material, namely: (1) verbal ability, has the highest level of learning difficulties of 64% in competency achievement indicators 6 and the lowest 31% in competency achievement indicators 3, (2) mastery of concepts, has the highest level of learning difficulties is 68% at competency achievement indicators 8 and the lowest is 31% at competency achievement

indicators 5, and (3) ability to count have the highest learning difficulty level of 75% at competency achievement indicators 5 and the lowest is 9% at competency achievement indicators 1.

## SUGGESTION

To perfect future research it is necessary to pay attention to data collection techniques and form the right instruments for online research and pay attention to the right data analysis techniques to produce more accurate research results.

## REFERENCES

- Aiken. (1985). *Psychological Testing and Assesment*. Boston: Allyn Bacon.
- Anugraheni, N. S., & Handhika, J. (2018). Profil Kemampuan Multirepresentasi Siswa dalam Materi Fluida. *Quantum: Seminar Nasional Fisika dan Pendidikan Fisika*, pp. 533-537.
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arifin, Z. (2012). *Evaluasi Pembelajaran (Edisi Revisi)*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama RI.
- Arini, F. Z. R., Susilaningsih, E., & Dewi, N. K. (2017). Pengembangan Instrumen Penilaian Proses untuk Mengukur Keterampilan Sains dan Aktivitas Siswa SMP. *Journal of Innovative Science Education*, 6(2), 170-178.
- Budiarti, I. S. (2017). Students' Conceptual Understanding Consistency Of Heat And Temperature. *Journal of Physics: Conference Series*, 795(1), 012051.
- Charli, L., Amin, A., & Agustina, D. (2018). Kesulitan Siswa dalam Menyelesaikan Soal Fisika pada Materi Suhu dan Kalor di Kelas X SMA Ar-Risalah Lubuklinggau Tahun Pelajaran 2016/2017. *Journal of Education and Instruction*. 1(1), 42-50.

- Chasanah, L., & Dewi, N. R. (2015). Pengembangan Sciencepoly Game Berbasis Kontekstual sebagai Media Science-Edutainment pada Materi Kalor dan Perpindahannya untuk Siswa Kelas VII SMP. *Unnes Science Education Journal*. 4(2), 873-880.
- Dewi, N. R., Savitri, E. N., Taufiq, M., & Khusniati, M. (2018). Using Science Digital Storytelling to Increase Students' Cognitive Ability. *Journal of Physics: Conference Series*, 1006 012020.
- Dewi, N. R., Nurkhalisa, S., Savitri, E. N., Wibowo, S. W. A., & Dwijayanti, I. (2019). Digital Storytelling Based on Natural Environment Exploration to Improve the Mastery of Classification Concepts. *Journal of Physics: Conference Series*, 1321(3), 032096.
- Dewi, N. R., Saputri, E., Nurkhalisa, S., & Akhlis, I. (2020). The Effectiveness of Multicultural Education Through Traditional Games-Based Inquiry Toward mproving Student Scientific Attitude. *Journal of Physics: Conference Series*, 1567(4), 042051.
- D'Sa, J. L & Dionaldo, M. L. V. (2017) Analysis of Multiple Choice Questions: Item Difficulty, Descrimination Index and Distractor Efficiency. *Journal of Nursing Education*, 9(3), 109-114.
- Gadbury- Amyot, C. C., Kim, J., Palm, R. L., Mills, G. E., Noble, E., & Overman, P. R. (2003). Validity And Reliability Of Portfolio Assessment Of Competency In A Baccalaureate Dental Hygiene Program. *Journal of Dental Education*, 67(9), 991-1002.
- Hanlbleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals Of Item Response Theory*. California: Sage Publications.
- Jatiutoro, P., Sunarno, W., & Soeparmi. (2018). Pembelajaran IPA Fisika dengan Model Problem Solving Menggunakan Metode Inkuiri Terbimbing dan Metode Gasing Ditinjau Dari Kemampuan Verbal dan Kemampuan Logika Matematika Siswa. *Jurnal Inkuiri*, 7(1), 133-142.
- Kaniawati, I., Samsudin, A., Hasopa, Y., Sutrisno, A. D., & Suhendi, E. (2016). The Influence of Using Momentum and Impulse Computer Simulation to Senior High School Students' Concept Mastery. *Journal of Physics: Conference Series*, 739(1), 12060.
- Kulsum, U., & Nugroho, S. E. (2014). Penerapan Model Pembelajaran Cooperative Problem Solving untuk Meningkatkan Kemampuan Pemahaman Konsep dan Komunikasi Ilmiah Siswa pada Mata Pelajaran Fisika. *UPEJ Unnes Physics Education Journal*, 3(2), 73-78.
- Muna, A., Noer, A. M., & Linda, R. (2017). Pengembangan Instrumen Tes Diagnostik Model Testlet Sebagai Pendeteksi Kesulitan Belajar Peserta Didik Pada Pokok Bahasan Larutan Penyangga. *Jurnal Online Mahasiswa Bidang Keguruan dan Ilmu Pendidikan*, 4(2), 1-15.
- Purwanto, E., Sunarno, W., & Aminah, N. S. (2015). Pembelajaran Fisika Dengan Contextual Teaching and Learning Menggunakan Media Animasi Flash Dan Video Ditinjau Dari Kemampuan Berpikir Abstrak Dan Kemampuan Verbal Siswa. *Jurnal Inkuiri*, 4(4), 77-86.
- Rachmawati, R. (2018). Analisis Keterkaitan Standar Kompetensi Kelulusan (SKL), Kompetensi Inti (KI), dan Kompetensi Dasar (KD) dalam Implementasi Kurikulum 2013. *Jurnal Diklat Keagamaan*, 12(34), 231-239.
- Rifa'i, A & Anni, C.T. (2016). *Psikologi Pendidikan*. Semarang: UPT UNNES Press.
- Rusilowati, A. (2014). *Pengembangan Instrumen Penilaian*. Semarang: UNNES Press
- Simkin, M. F. & Kuechler, W. L., 2005. Multiple-Choice Test and Student Understanding: What Is The Connection? *Decision Sciences Journal of Innovation Education*, 3(1), 73-96.
- Shiell, R.C. & Slepko, A. (2015). Integrated Testlets: A New Form of Expert-Student Collaborative Testing. *Collected Essays on Learning and Teaching*, 8(2), 201-210.

- Slepko, A. D & Shiell, R. C. (2014). Comparison of Integrated Testlet and Constructed-Response Question Formats. *Physical Review Special Topics-Physics Education Research*, 10(2), 020120-1 – 020120-15.
- Subekti, Y., & Ariswan, A. (2016). Pembelajaran Fisika dengan Metode Eksperimen untuk Meningkatkan Hasil Belajar Kognitif dan Keterampilan Proses Sains. *Jurnal Inovasi Pendidikan IPA*, 2(2), 252-261.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: CV Alfabeta.
- Wahidmurni., Mustikawan, A., & Ridho, A. (2010). *Evaluasi Pembelajaran (Kompetensi dan Praktik)*. Yogyakarta: Nuha Litera.
- Wainer, H., Sireci, S. G., & Thissen, D. (1991). Differential Testlet Functioning Definitions And Detection. *ETS Research Report Series*, 1991(1), 1-42.