Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning

Cos Ma'arif H. Laeli, *Universitas Sebelas Maret, Surakarta, Indonesia, cosmaarifhl@student.uns.ac.id*
ORCID: 0000-0002-6379-5838
Gunarhadi, *Universitas Sebelas Maret, Surakarta, Indonesia, gunarhadi@fkip.uns.ac.id*
ORCID: 0000-0002-9044-6836
Muzzazinah, *Universitas Sebelas Maret, Surakarta, Indonesia, yayin_pbio@fkip.uns.ac.id*
ORCID: 0000-0002-6862-2581

Abstract. This research proposes to describe the development of a three-level multiple-choice diagnostic test (three-level multiple-choice diagnostic test) in science learning for elementary school students. The type of this study used was Sukmadinata Research and Development which consisted of a preliminary survey, model development, and model testing. This research subject was grade four with 161 students. Data collection techniques using observations, questionnaires, and tests. The data that has been collected was then analyzed quantitatively and qualitatively. Quantitative data analysis using paired sample t-test and qualitative analysis used the interactive model of Miles and Huberman. The outcomes demonstrated that the development of a three-stage multiple-choice diagnostic test in a diagnostic test instrument was effective in improving students 'critical thinking skills and was able to analyze the level of students' understanding. The main test obtained ρvalue = 0.002> 0.05, so it can be concluded that H_0 is rejected.

Keywords: Elementary school, critical thinking, light, misconception, three-tier multiple-choice diagnostic tests

INTRODUCTION

Critical Thinking Skills, or the ability to think critically are fundamental skills that every individual must have to solve a problem. Zubaidah (2018: 23) explains that Critical Thinking and problem-solving skills include the ability of individuals to reason effectively, ask questions and solve problems, analyze and evaluate, and critically reflect decisions and processes. The indicators of critical thinking skills according to Facione (2015) include interpretation, analysis, inference, evaluation, explanation, and self-regulation. Based on these descriptions, thinking skills become a life skill that needs to be developed through the educational process because it can determine the success of a person's life.

A person's critical thinking ability is closely related to understanding the basic concepts that a person has in analyzing a problem. A person will find it difficult to develop critical thinking skills if he still experiences misconceptions. A wrong understanding of the concept becomes an obstacle for students to develop High Order Thinking Skills (HOTS) which becomes a requirement for someone to think critically. This is evidenced by research conducted by Silva & Almeida (2017) showing that students who have misconceptions tend to have low critical thinking skills.

Many elementary school students experience low critical thinking skills, especially in Natural Science (IPA) learning. Science learning is a branch of science that has a high level of difficulty to understand. This was further strengthened by the data from the Program for International Student Assessment (PISA) in 2016 which stated that Indonesia was ranked 62 out of 72 other countries in the field of science (IPA). Based on the results of this research, Indonesia is still far behind Singapore, which at that time was ranked first and defeated the education systems of Europe, North America, and South America (Berlianto, 2016).

Many of the concepts contained in the IPA are abstract. In this regard, Abbas (2016: 83) emphasizes that mastery of an abstract concept certainly has a higher level of difficulty when compared to the mastery of a concrete concept. Students at the elementary school level are still in the ¬operational concrete stage. Mastery of abstract concepts in elementary science learning must first be based on a correct understanding of concepts. Understanding the correct concept will be the foundation for students to analyze abstract concepts through their critical thinking skills.

Misconception is a condition where a person is unable to construct the experience, he has the new knowledge he has gained. Based on the theory put forward by followers of constructivism, Narjaikaew (2013: 252) states that if students fail to construct their past experiences with new concepts contained in science learning, then these learners can experience misconceptions.

A profile of misconceptions in education, especially at the elementary school level, has been identified in many studies. Pesman and Eryilmaz (2010: 216) which states that as many as 69% of 124 students experience misconceptions about the electrical circuit material. Not only that, but the facts about students' misconceptions were also explained by Maier, Wolf, & Randler (2016: 91) which stated that 19.7% of students experienced misconceptions about the material's adjustment to living things to the environment. This data is further strengthened by research conducted by Laeli, Gunarhadi & Muzzazinah (2019: 660) which states that elementary school students experience misconceptions in various concepts in science learning. These misconceptions include 74.8% of the concept of force, 74.3% for light and its relation to the sense of sight, 66.7% and alternative energy, 56.7% of sound and its relation to the sense of hearing, and 41.3% for form and body functions in the concept of animals and plants.

Another interesting fact in the study of the misconceptions is that misconceptions are not only experienced by students. Teachers who should be a transfer of knowledge and scientific concepts also still have the potential to experience misconceptions. The statement was further strengthened by the research conducted by Mutlu & Sesen (2014: 634) which states that 151 Science Teachers in Turkey experienced a misconception in understanding the concept of chemical elements.

Misconceptions are difficult to identify. Someone who experiences misconceptions, is not the same as someone who does not understand the concept. If students do not understand the concept, then after the teacher explains good learning, students will be able to understand the concept. But if students experience misconceptions, even though a good explanation, students will still find it difficult to accept the correct concept. This was stated also by Kurniawan (2018: 68) who stated that misconceptions would be difficult to eliminate even with quality lectures.

The occurrence of misconceptions in students can cause negative effects in the learning process. Based on the results of research presented by Uzun, Alev, & Karal (2013: 145), it was found that some students could not explain a phenomenon related to their knowledge in language and scientific reasons. Many students can answer questions with the right answers, but not accompanied by scientific reasons. So in that case, students can be said to experience misconceptions.

Many factors cause a person to have a misconception. Mursalin (2013: 2) argues that a person can experience misconceptions if he has the wrong pre-conception, has associative thinking, has humanistic thinking, does wrong or incomplete reasoning, and has wrong intuition. These factors cause one to be unable to explain a phenomenon based on scientific concepts.

Keles, Cepni, Aydin, & Hasiloglu (2011: 1) underline an important point, that the basic concepts in students play an important role in realizing effective concept learning. In this regard, Aydin (2013: 34) asserts that teachers should be able to know and correct misconceptions experienced by students before giving an instruction. Therefore, special instruments are needed to first map and identify misconceptions among students.

Student misconceptions can be identified through tests called diagnostic tests. This test is used to assess students' understanding of the concepts of key concepts (key concepts) on a particular topic. Diagnostic tests can be done utilizing interviews, multiple-choice tests, open-ended questions, two-tier tests and three-tier tests (Djanettea & Fouad, 2014: 583).

A three-tier test is an instrument that is effective enough to be used in identifying students' misconceptions. This is in line with Mubarak (2016: 102) which states that the Three-tier Multiple Choice Diagnostic Test is an instrument used to identify and measure students' misconceptions. Through this instrument, the profile of students' misconceptions can be identified in more detail.

Three-tier Multiple Choice Diagnostic Test is a development test instrument from the two-tier multiple-choice test (Ardiansyah & Rahardjo, 2018: 1). The difference between the two-tier multiple-choice test and the Three-tier Multiple Choice Diagnostic Test is the ability of the instrument in the identification aspect. Three-tier tests use a simple and easy way to identify misconceptions and distinguish them from a lack of knowledge. Judge (2012: 3) explained that the instrument three-tier Multiple Choice Diagnostic Test added the level of confidence in the answers chosen by students, which serves to differentiate the profile of misconceptions and conceptual understanding of students.

The three-tier multiple-choice diagnostic test instrument has three main parts on each item. Cetin (2011: 2) explains that the first level consists of conventional multiple-choice, the second level consists of possible reasons for answers given for the first level, and the third level consists of the level of confidence given in selecting answers in the first and second levels. Through analysis of these three levels, the Three-

278 | COS MA'ARIF HIDAYATUL LAELI

Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning

Tier Multiple Choice Diagnostic test instrument can categorize the level of understanding of students based on points obtained for each item.

The ability of the Three-Tier Multiple Choice Diagnostic test instrument that can categorize the level of understanding of students becomes its advantage for this instrument to identify misconceptions. Adding to this opinion, Mubarak, Susilaningsih & Cahyono (2016: 102) stated that other advantages possessed by the Three-tier Multiple Choice Diagnostic Test instrument include: 1) can diagnose deep misconceptions, 2) can determine the material that must be emphasized more in the process learning, 3) and can help teachers to plan learning better. The advantages possessed by these instruments become an opportunity for teachers to minimize the misconceptions experienced by students in understanding the concept of learning.

Based on some of the results of previous research, the facts about the misconceptions experienced by students in understanding the concept of science is an interesting discussion to study. Through this research, the researcher can analyze the misconception profile of fourth grade students. The difference between this study and previous research is the location of the research, the object of the study, the selection of teaching materials, and the specification of the misconception indicators.

This research includes quantitative data and qualitative data. Quantitative data are presented through the results of misconceptions diagnostic tests. Meanwhile, qualitative data are presented in the form of descriptions of test results, explanations of misconceptions based on scientific concepts, and factors that influence the occurrence of misconceptions. The formulation of the problem in this study is what is the profile of the misconceptions experienced by fourth grade elementary school students in science learning the concept of the nature of light? This study aims to analyze the profile of misconceptions experienced by fourth-grade elementary school students in science learning the concept of the nature of light.

METHODS

Type of this research is the Model Research and Development (R & D) according to Sukmadinata (2013) that consists of a preliminary study, model development, and model testing. The product of this study was the Three Level Multiple Choice Diagnostic Test instrument in science learning the concept of light for elementary school students. The steps in this research and development follow the model development stages consisting of a preliminary survey, design validation, field testing, and product effectiveness testing. The prototype of the instrument was then validated by the experts to get input and suggestions it was approved by the expert team. There are two experts involved, namely light concept experts and learning instrument evaluation experts. While the field effectiveness test in this study used a pre-experimental research method with a pretest-posttest one group design.

Sample

This research was conducted in the city of Surakarta, Central Java, Indonesia. The samples in this study were SDN 1 Mojo, SDN Wiropaten, SDN Semanggi Lor, SDN Tegalrejo, and SDN Gurawan. The number of students is 161. The determination of the sample is done by purposive sampling, which is to define the sample with specific objectives and considerations.

Data Collection

Research data were collected using test, questionnaires, and observations techniques. The research procedure in this study is as follows:
1. Preliminary Study
   At this stage, Conducting Literacy Studies reviews about misconceptions and scientific concepts in science learning on light material are carried out. Analysis of conceptual understanding of the three-tier multiple-choice diagnostic test instrument used in this study uses a reference from Pacala (2018: 4) which divides into five categories. The five categories are: a) complete understanding, b) partial understanding without misconception, c) partial understanding with misconception, d) misconception, e) not understanding the concept
2. Development Stage
   Before the study was conducted, researchers first developed the Three Tier Multiple Choice Diagnostic Test instrument. The steps taken in developing the instrument are as follows:
   a. Conduct a preliminary study
      The research method carried out at the preliminary study stage is a descriptive method. The purpose of this stage is to collect data are about 1) the existing science learning assessment instrument as a basic comparison material for the product to be developed; 2) responses and information sourced from teachers and students of the learning assessment instrument that has been used so far; 3)

Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning

analyze the need for a three-tier multiple-choice diagnostic test instrument to be developed appropriately; 4) description of findings of needs on a three-tier multiple-choice diagnostic test instrument.

b. Determine the competency achievement indicators (GPA) to be measured

Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 57 of 2014 concerning Curriculum 2013 SD / MI states that grade IV elementary school has nine themes in a year. The theme chosen in this study is the fifth. Theme five for fourth grade is about "My Heroes". This theme was chosen because the material contained in the theme has a broad and complete coverage. Thus, researchers will find it easier to make the test questions to identify misconceptions. The contents of the lesson used in this fifth theme are Science, Nature. The basic competence contained in the content of the Natural Sciences (IPA) lessons on this theme is BC (3.7) Applying the properties of light and its relation to the sense of sight. These basic competencies can be reduced to four competency indicators, namely 1) light sources, 2) nature of light and its use in daily life, 3) the process of the formation of shadows in the sense of sight 4) mirror and optic.

c. Determine clustering indicators for misconceptions

Analysis of conceptual understanding of the three-tier multiple-choice diagnostic test instrument used in this study uses a reference from Pacala (2018: 4) which divides into five categories. The five categories are: 1) complete understanding, 2) partial understanding without misconception, 3) partial understanding with misconception, 4) misconception, 5) not understanding concepts.

Table 1. *Interpretation of Three-tier Multiple Choice Diagnostic Test results*

Answer Combination			Classification of students' answers	Score
1st level	2nd level	3rd level		
Right	Right	Sure	Complete understanding	5
Right	Right	Not sure	Partial understanding without misconception	4
Right	False	Sure	Partial understanding with misconception	3
False	Right	Sure		
False	Right	Not sure		
Right	False	Not sure		
False	False	Sure	Misconception	2
False	False	Not sure		
Right	Right	To guess	Don't understand the concept	1
Right	False	To guess		
False	Right	To guess		
False	False	To guess		

Source: Pacala (2018: 4)

d. Draft a Three-tier Multiple Choice Diagnostic Test Instrument

The product produced in this study is a Three Tier Multiple Choice Diagnostic Test instrument. The use of this instrument is integrated with the TPACK approach during the learning process to follow up the results of the analysis of the instrument.

e. Generate answer keys and assessment guidelines

f. Validate the draft Three-tier Multiple Choice Diagnostic Test instrument by relevant experts

The data analysis technique in this study was carried out using an interactive data analysis model (Cohen, Manion, & Morrison, 2007: 108) with the following steps: (a) analyzing data, data analysis in this process involves data about learning instruments that usually used in thematic learning, (b) data revision (revised data) which includes linkages or linkages between linguistic aspects (network), summaries or conclusions (structured summaries), and learning model scripts (material scripts), (c) data verification (verification data) was conducted using peer-debriefing techniques to obtain valid research findings.

The data validity technique used in this study was the content and construct validity. The first validity is done by testing the instrument by an expert judgment. This validity test aims to assess the dimensional accuracy of the construct description. Validation through expert judgment was carried out by 8 experts. There are several experts in this research, experts who are experienced in science education materials, experts in teacher training, and instrument experts who assess the feasibility of developing instruments. The content validity index can be known using the Aiken's V. Index Formula to calculate the coefficient of the validity of Aiken's V is as follows:

$V = S/[n(c-1)]$ where is: $S = \Sigma ni (r-\ell o)$

280 | COS MA'ARIF HIDAYATUL LAELI

Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning

The calculated V value is then compared with the Vtable value at the 0.05 significance level. If Vcount> Vtable, then the test item is declared valid.

Table 2. *Results of Aiken's Validity calculation*

Butır Problem	Tier	V	Information
1	I	0.92	Valıd
	II	0.92	Valıd
2	I	0.96	Valıd
	II	0.96	Valıd
3	I	0.92	Valıd
	II	0.83	Valıd
4	I	0.92	Valıd
	II	0.96	Valıd
5	I	0.96	Valıd
	II	0.83	Valıd
6	I	1.00	Valıd
	II	1.00	Valıd
7	I	0.92	Valıd
	II	0.83	Valıd
8	I	1.00	Valıd
	II	0.92	Valıd
9	I	0.96	Valıd
	II	1.00	Valıd
10	I	0.96	Valıd
	II	1.00	Valıd
11	I	0.96	Valıd
	II	0.83	Valıd
12	I	0.83	Valıd
	II	1.00	Valıd
13	I	1.00	Valıd
	II	1.00	Valıd
14	I	1.00	Valıd
	II	0.92	Valıd
15	I	1.00	Valıd
	II	1.00	Valıd

g. Trial draft Three-tier Multiple Choice Diagnostic Test instrument
Trying out the draft is the second validation stage. The second validity is to test all the items that have been validated by the expert empirically. The second validity was taken from a sample with a wider coverage. Instrument testing was carried out at İslam Diponegoro Elementary School.

h. Conduct evaluation and analysis of the results of trials
To determine the validity of the question items used the biserial point correlation formula. With a significant level of 5%, the problem is said to be valid if r arithmetic> r table. The validity test results show that all test items are valid. The validity of items is determined using the formula:

$$r_{pbis} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Reliability test is used to measure the reliability level of critical thinking skills. The formula used is Alpha Cronbach. (Cohen, Manion, & Morrison, 2018). Value is obtained at a price with a significance level of 5%. Problem is said to be reliable if r count> r table.

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

The questionnaire instrument is said to be reliable if the value of r arithmetic> 0.7; whereas if the r count <0.7 then the item is said to be reliable (Arikunto, 2013: 231). The result of testing the reliability of the

test instruments obtained the value of r at the 95% confidence level or the significance of 5% is 0.784, thus indicating that the test instrument was declared reliable.

Conduct a final revision of the Three-tier Multiple Choice Diagnostic Test Instrument

3. Evaluation Stage

At the evaluation stage, an independent sample t-test can be tested. An Independent sample t-test was conducted to determine the significance of differences in the average critical thinking skills of the experimental group and the control group. The scores compared are the gain scores of the experimental group and the control group. The gain score is a measure of the difference between the post-test and pretest data in each group.

Table 3. *Results of Independent Sample t-Test of Critical Thinking Skills*

Levene's Test for Equality of Variances	F	Sig.	T	Df	Sig.(2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	1.635	.204	5.525	100	.000	12.294	2.225
Equal variances not assumed			5.525	96.278	.000	12.294	2.225

The test results obtained a probability value of 0.000, which is the 95% level of confidence (α = 5% and sig > 0,05) It can be concluded that H$_0$ is rejected. Therefore, it is proven that the 95 percent confidence level the ordinary post-test score of students' critical thinking skills in the two population is not the same. The test results establish that the Critical Thinking in the experimental group is better than the control group

RESULTS

The readiness of the prototype of the Three-Tier Multiple Choice Diagnostic Test instrument begins with examining the course of study which aims to determine the material to be included in the Three Tier Multiple Choice Diagnostic Test instrument. This is so that the instruments developed can help students think critically and master the material being studied properly. The scope of light concept in elementary school students includes light sources, the nature of light from the seeing process, and optical devices.

The resulting product is a Three-Tier Multiple Choice Diagnostic Test instrument. The instrument developed has 7 main components consisting of question grids, question work instructions, three-tier multiple-choice diagnostic test questions, answer keys, answer sheets, scoring guidelines, and results interpretation guidelines. Of the seven components, four are held by the teacher, and the other three are given by students. The components held by the teacher include question grids, answer keys, scoring guidelines, and results interpretation guidelines. Meanwhile, the components given to students include manual work, a three-tier multiple-choice diagnostic test, and an answer sheet.

The three-tier multiple-choice diagnostic test itself consists of 3 parts. The three-tier multiple-choice diagnostic test instrument has three main parts on each item. Tier I is the first level consisting of a question statement and conventional multiple-choice answer choices. Tier II is the second level, which consists of the possible answer choices for the answer choices given at the first level. Meanwhile, Tier III is the third level, which consists of the level of confidence in choosing answers and reasons in the first and second levels. The following is an example of a three-tier multiple-choice diagnostic test item

Tier I

When the lights go out, we will light a candle to light the room. This can be illustrated through the following picture.
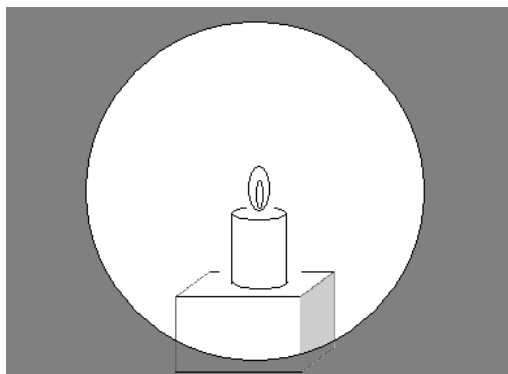

Fıgure 1. *Candle in dark area*

In the picture above, there is a room that looks bright, but there are some others still dark. Based on these images, then we can draw the conclusion that ...
a. *Light only travels to a part of the room*
b. *Light spread throughout the room*
c. *Candlelight does not propagate*
d. *Radiant light from a candle that has been short only radiates around the candle*

Tier II
What is your reason for choosing that answer?
a. *Light has limited mileage to illuminate a room*
b. *The light emitted by the candle will travel straight across the room, except that light rays close to the light source have a higher line density compared to those far away.*
c. *Light cannot propagate. Light is a static energy (silent)*
d. *Light can only be seen in bright areas and cannot be seen in dark areas*

Tier III
Are you sure about your answer?
a. *Yes, I am sure*
b. *No, I'm not sure*
c. *I don't know the answer and just guessing*

The results of the assessment of the instrument prototype by experts gained an average of 3.41. The results of the expert's assessment indicate that the Three Level Multiple Choice Diagnostic Test Instrument can help critical thinking process. This is because students are actively taken in learning such as finding data, finding, analysing, and working out problems. Illustrations related to the characteristics of light need to be improved because statements that are only theoretical have not been able to improve students' critical thinking skills.

Meanwhile, from assessment result of the instrument prototype by the learning instrument evaluation expert, it is known that the presentation of evaluation questions through the three-tier multiple choice diagnostic test instrument is considered good. The instrument has an average of 3.23. However, the model was revised in various ways, namely the level of difficulty, the divergence in the interrogative sentences and the disturbing function. Evaluation questions need to present questions that are contextual and applications. The illustration of the light phenomenon presented in; the question is a phenomenon of the properties of light that occurs in everyday life. In addition, in choosing an answer option, the function of the cheater must be owned by every wrong answer. Through these disturbing answers, students are required to think critically in finding the correct answers and reasons according to existing scientific theories.

The prototype of the Three-Tier Multiple Choice Diagnostic Test instrument that has been revised based on expert input and suggestion is then tested on preliminary testing or a small scale. The purpose of small-scale trials in this study was to determine the estimated time required to carry out the Three Tier Multiple Choice Diagnostic Test. The limited trial was carried out at SDN Tegalrejo, Surakarta. The limited trial involved 32 fourth grade students of SDN Tegalrejo. The execution of the limited trial was carried out cooperatively with grade IV teachers at SDN Tegalrejo Surakarta. The number of items used in this small-scale trial was 20 items with three tiers for each item. The time it takes students to work on the questions is about 70 minutes, so each student's item can work for approximately 3.5 minutes.

After passing a limited trial, this instrument was continued to the broad stage of testing. The student assessment questionnaire for the Three Tier Multiple Choice Diagnostic Test questions on a small-scale test showed that the Three Tier Multiple Choice Diagnostic Test questions were categorized as good enough so that the 20 revised questions could be directly tested on a wide-scale test. The wide-scale trial in this study was used to determine the reliability, level of difficulty, differentiation, the functioning of the Three-Tier Multiple Choice Diagnostic Test destructor, and the analysis of student misconceptions. This wide-scale trial was carried out at SDN Semanggi involving 27 students and SDN Mojo 1 involving 23 students. Also, interviews with the teacher and student responses are also conducted at this stage. Interviews at this stage were conducted to determine the responses and opinions of teachers to the Three Tier Multiple Choice Diagnostic Test instrument.

The product testing phase aims to determine the effectiveness of the Three-Tier Multiple Choice Diagnostic Test instrument development to increase the critical thinking of elementary students in science with the light concept. The experimental group consisted of 23 grade IV students of SDN Mojo I and 28

283 | COS MA'ARIF HIDAYATUL LAELI

Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning

grade IV students of SDN Wiropaten. While the control group was 24 grade IV SDN Gurawan students and 27 grade IV SDN Semanggi Lor students. So that the sample involved was 102 students, each group was 51 students. The design used during the testing phase was a one-group pretest-posttest design.

Before testing using the independent sample t-test, a prerequisite test is needed, which consists of a normality test, a homogeneity test, and a balance test. The normality test uses the Kolmogorov-Smirnov analysis technique, the homogeneity test uses Levene's test, and the balance test by comparing the pretest averages between the groups and the experimental group.

Table 4 *Results of Normality Test Analysis of the Independent Sample t-test*

Tests of Normality							
Score	Kelas	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Critical Thinking Skill	Pre-Test Eksperimen	.093	51	.200*	.975	1	.344
	Post-Test Eksperimen	.086	51	.200*	.977	1	.427
	Pre-Test Kontrol	.081	51	.200*	.973	1	.305
	Post-Test Kontrol	.100	51	.200*	.986	1	.819
*. This is a lower bound of the true significance.							
a. Lilliefors Significance Correction							

Based on Table 4, it can be seen that the data normality tests the 5% significance level, obtained a probability of 0.200 in the pretest data for the experimental class, and 0.200 in the posttest data for the experimental class. As for the controller class, a probability of 0.200 is obtained in the pretest, and 0.200 in the posttest. The probability that there is a whole normality test shows that it is greater than $\alpha = 0.05$, so it can be concluded that the test data in the experimental class and in the control class are normally distributed.

Table 5. *Homogeneity Test Results Independent Sample t-test Analysis*

Test of Homogeneity of Variances			
Critical Thinking Skill			
Levene Statistic	df1	df2	Sig.
7.318	1	100	.052

The homogeneity test of the experimental class and control class using Levene's test resulted in a probability value of 0.052. This acquisition is greater than $\alpha = 0.05$ (sig.> 0.05). It can be concluded that the test data in the experimental class and control class have homogeneous variants, or the data comes from populations with the same variants.

Table 6. *Results of the Analysis of the Balance of the Independent Sample t-test*

Levene's Test for Equality of Variances	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	7.318	.052	1.544	100	.178	3.490	2.261
Equal variances not assumed			1.544	89.440	.178	3.490	2.261

The balance test was carried out by comparing the pretest average of critical thinking skills between the experimental group and the control group. The test results obtained a probability value of 0.178, which is the 95% confidence level ($\alpha = 5\%$ and sig.> 0.05), it can be concluded that H_0 is accepted. Thus, it is evident that the 95 percent confidence level, there is no significant difference between the average Critical Thinking Skill in the experimental group and the control group. These results indicate that the two groups have equal abilities.

After the analysis, prerequisite test is fulfilled, an independent sample t test can be tested. Independent sample t - test was taken to find out the significance of differences in the average critical thinking skills of the experimental group and the control group.

Table 7. *Independent Sample t-test results*

Levene's Test for Equality of Variances	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	1.635	.204	5.525	100	.000	12.294	2.225
Equal variances not assumed			5.525	96.278	.000	12.294	2.225

The test results obtained a probability value of 0.000, which is the 95% level of confidence (α = 5% and sig.> 0.05), so it can be concluded that H_0 is rejected. Thus, it is proven that the 95 percent confidence level the average post-test score of students' critical thinking skills in the two populations is not the same. The test results show that the Critical Thinking Skill in the experimental group is better than the control group.

DISCUSSION AND CONCLUSIONS

The stages of developing the Three Tier Multiple Choice Diagnostic Test instrument in learning the properties of light begin with the preparation of the instrument product. The analysis of the scope of the learning material was carried out to organise the particulars to be exhibited on the instrument so that they could be used to pick out misconceptions and strengthen students' critical thinking skills (Mubarak, 2016). In general, the preparation of the Three Tier Multiple Choice Diagnostic Test instrument in this study refers to the aspects of a good instrument. A good instrument is consisting of the aspects of having high validity and reliability, having sufficient distinguishing power, the level of difficulty of the tests for all the questions, consisting of 30% easy questions, 50% moderate questions, and 20% of difficult questions (Kunandar, 2013)

The prototype of this instrument was evaluated by experts and at the same time received input as a prototype fix substance. The overhauled prototype was then applied at the limited test phase. The outcomes of the realization of the limited test show that the prototype is the good category, but it needs to be improved according to user input in the limited test. The results of the repair of the instrument prototype were then tested on a wider scale.

Through testing the Three Tier Multiple Choice Diagnostic Test prototype instrument on a broader scale, there are many advantages when the teacher uses the Three Tier Multiple Choice Diagnostic Test instrument in learning. Besides being able to improve critical thinking skills, the Three Tier Multiple Choice Diagnostic Test instrument is also able to identify the level of misconceptions in students. The level of understanding of the concept of grade IV students in the five elementary schools is grouped into 5 categories. The category consists of complete understanding, partial understanding without misconception, partial understanding with misconception, misconception, not understanding concepts. Based on data processing, it can be observed that the level of understanding of students in learning science in light material has differences. The majority of students belong to the category of partial understanding with misconception, which is 38%. Meanwhile, port students with the complete understanding only reached 11%. Out of the two categories, students with partial understanding without misconception reached 31%, misconception category reached 16% and did not understand the concept of 4%.

Besides, based on the results of research data on the broader trial stage, it was proven that students who were given the Three Tier Multiple Choice Diagnostic Test instrument had a higher student's critical thinking ability. Strengthening audiences' statement (2018), this study proves that the use of the Three Tier Multiple Choice Diagnostic Test instrument can help students collect, categorize, analyze, and evaluate information or evidence so that students can conclude to solve problems following existing scientific theories. The Three Tier Multiple Choice Diagnostic Test instrument can encourage students to have a deep understanding, evaluate inappropriate thinking, express logical assumptions, and have strong confidence in making decisions (Jones, 2015).

The development of the Three Tier Multiple Choice Diagnostic Test instrument in learning the nature of light in science can identify misconceptions and improve students' critical thinking skills. Product improvement during the development stage resulted in a Three Tier Multiple Choice Diagnostic Test instrument suitable for use by grade 4 elementary students to improve critical thinking in science. This instrument conforms to the views of experts and users

285 | COS MA'ARIF HIDAYATUL LAELI

Development of Three Tier Multiple Choice Diagnostic Test to Identify
Misconception and Improve Critical Thinking Skill in Science Learning

REFERENCES

Abbas, M. (2016). Pengembangan Instrumen Three Tier Diagnostic Test Miskonsepsi Suhu dan Kalor. Ed-Humanistics, 01, 83–92.

Ardiansah, A., Masykuri, M., & Rahardjo, S. B. (2018). Student certainty answering misconception question: Study of Three-Tier Multiple-Choice Diagnostic Test in Acid-Base and Solubility Equilibrium. Journal of Physics: Conference Series, 1006(1). https://doi.org/10.1088/1742-6596/1006/1/012018

Arikunto, S. (2013). Dasar-Dasar Evaluasi Pendidikan (Edisi 2). Jakarta: Bumi Aksara

Aydin, A. (2013). International Journal of Education and Practice Representation of Science Process Skills in The Chemistry. International Journal of Education and Practice, 1(5), 51–63.

Berlianto. (2016, December 6). Singapore Ranked Number One in International Education. Retrieved March 31, 2019, from SINDOnews.com: www.international.sindonews.com

Cetin-Dindar, A., & Geban, O. (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. Procedia - Social and Behavioral Sciences, 15, 600–604. https://doi.org/10.1016/j.sbspro.2011.03.147

Cohen, L & Manion, L,. (2007). Research Methods in Education, Fourth Edition. London and New York

Djanette, B., & Fouad, C. (2014). Determination of University Students' Misconceptions about Light Using Concept Maps. Procedia - Social and Behavioral Sciences, 152, 582–589. https://doi.org/10.1016/j.sbspro.2014.09.247

Facione, P. a. (2015). Critical Thinking : What It Is and Why It Counts. Insight assessment. Retrieved from https://www.insightassessment.com/CT-Resources/Teaching-For-and-About-Critical-Thinking/Critical-Thinking-What-It-Is-and-Why-It-Counts/Critical-Thinking-What-It-Is-and-Why-It-Counts-PDF

H Laeli, C. M. (2020). Misconception of Science Learning in Primary School Students, 397(Icliqe 2019), 657–671.

Jones, L. L., & Estes, Z. (2015). Convergent and divergent thinking in verbal analogy convergent and divergent thinking in verbal analogy. Thinking & Reasoning, 21(4), 473–500. https://doi.org/10.1080/13546783.2015.1036120

Keleş, P. U., Çepni, S., Aydin, S., & Haşiloğlu, M. A. (2011). The effect of conceptual change texts on eliminating the misconceptions of K5 students' alternative views about the birds. Procedia - Social and Behavioral Sciences, 28, 1061–1071. https://doi.org/10.1016/j.sbspro.2011.11.192

Kunandar. 2013. Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013). Jakarta: PT. Raja Grafindo Persada.

Lusiana,Kurniawan, D. T., & Maryanti, S. (2018). Analysis of Math Teacher Candidates' Misconception on the Dynamic Electricity Concept. Scientiae Educatia, 7(1), 67. https://doi.org/10.24235/sc.educatia.v7i1.2481

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. Computers and Education, 95, 85–98. https://doi.org/10.1016/j.compedu.2015.12.002

Miles, MB, & Huberman, AM (1994). Analisis Data Kualitatif. London: Sage.

Mubarak, S. (2016). Journal of Innovative Science Education Pengembangan Tes Diagnostik Three Tier Multiple Choice. Journal of Innovative Science Education, 5(2), 101–110.

Mursalin. (2013). Model Remediasi Miskonsepsi Materi Rangkaian Listrik Dengan Pendekatan Simulasi PhET. Jurnal Pendidikan Fisika Indonesia, 9, 1–7.

Mutlu, A., & Sesen, B. A. (2015). Development of a Two-tier Diagnostic Test to Assess Undergraduates' Understanding of Some Chemistry Concepts. Procedia - Social and Behavioral Sciences, 174, 629–635. https://doi.org/10.1016/j.sbspro.2015.01.593

Narjaikaew, P. (2013). Alternative Conceptions of Primary School Teachers of Science about Force and Motion. Procedia - Social and Behavioral Sciences, 88, 250–257. https://doi.org/10.1016/j.sbspro.2013.08.503

Pacala, F. A. A. (2018). for Conceptual Understanding in Momentum and Collision, (March).

Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. Journal of Educational Research, 103(3), 208–222. https://doi.org/10.1080/00220670903383002

Sukmadinata, N. (2011). Metode Penelitian Pendidikan. Bandung:PT Remaja Rosdakary

Uzun, S., Alev, N., & Karal, I. S. (2013). A cross-age study of an understanding of light and sight concepts in physics. Science Education International, 24(2), 129–149.

Zubaidah, S., Fuad, N. M., Mahanal, S., & Suarsini, E (2017). Improving creative thinking skills of students through differentiated science inquiry integrated with mind map. Journal of Turkish Science Education, 14(4),77-91. http://doi.org/10.12973/tused.10214a