

Development of Description Test Instruments Based Scientific Approach to Measure Ability to Create Critical Participants in Class V Basic School

Dwi Widia Sari M.Thoha BS Jaya Rochmiyati Lilik Sabdaningtyas
Fakultas Keguruan dan Ilmu Pendidikan Universitas Lampung

Abstract:

This study aims to develop a description test instrument based on a scientific approach to measure students' critical thinking skills in grade V elementary school and to know the differences in students' critical and reliable critical thinking abilities. The type of research used is research and development that refers to the theory of Borg and Gall. The population in this study were all fifth grade elementary school students at SDN Metro Pusat cluster nusa indah. The sample was taken using simple random sampling technique, taking 2 classes from 14 classes in the population, which consisted of one class as a trial class totaling 22 students from the VB class of SD Negeri 2 Metro Pusat and one class as a sample. totaling 37 students in VA class at SD Negeri 1 Metro Pusat. Data was collected through questionnaires and test questions. The results showed that the test instruments developed were valid and reliable. Based on the assessment of 3 experts, the average is 81.75 including in the very good category. Whereas based on the results of the empirical test, 20 description questions developed were declared to be 0.90 valid and 0.989 reliability was included in the very strong category.

Keywords: Critical Thinking, Description Test Instrument, Scientific Approach.

INTRODUCTION

Today's world education challenges demand high-level thinking skills, namely through the ability to think critically and creatively. Critical thinking allows students to utilize their potential in seeing problems, solving problems, and creating new things. The ability to think critically is the ability to evaluate, compare and contrast. Increased critical thinking skills will certainly be closely related to improving the quality of education can be achieved through improving the quality of learning and the quality of the assessment system (Mardapi, 2012: 12). When the learning process is done well, it will also produce good quality learning. Learning quality is good, it can be seen from the results of the assessment. Teachers as facilitators in a learning, must plan learning, strategies, and construction based questions that can facilitate students to think high-level, creative thinking, and critical thinking. In connection with the construction of the question, the assessment must be carried out well starting from the determination of the instrument, preparation of instruments, review of instruments, implementation of assessment, analysis of assessment results, and follow-up program of assessment results.

In relation to students' critical thinking skills, the scientific learning approach provides space for students to continue to develop critical thinking skills. This is consistent with the scientific approach approach, one of the approaches used in the 2013 curriculum. Competency-based 2013 curriculum focuses on a number of competencies and a set of learning objectives, so that their achievement can be observed in the form of students' behavior or skills as a criterion. The scientific approach (scientific approach), hereinafter referred to as the scientific approach to learning as intended, includes observing, asking, trying, processing, presenting, concluding, and creating for all subjects.

Permendikbud No. 23 of 2016 concerning the Educational Assessment Standard states that: Learning activities in the 2013 curriculum are the process of gathering and processing information to measure the achievement of student learning outcomes including authentic assessment, self-assessment, portfolio-based assessment, repeat, daily repetition, midterm repeat, final test semester, repeat competency level, competency level quality test, national examination, and school or madrasah examination.

Based on the needs analysis questionnaire that had been distributed to fifth grade teachers in nine elementary schools in Nusa Indah cluster in Metro Pusat sub-district, the following results were obtained, the learning activities were carried out following the teacher's book and student books. As much as 55.57% of teachers have used a scientific approach, but it is still not optimal, the teacher does not apply learning or assessment of learning outcomes that train students' thinking skills, they only learn concepts by memorizing. Learning by memorizing causes students not to develop their thinking skills, so that they have not been able to organize, communicate and link information and facts in nature to various learning. Less than 70% of teachers make their own test instruments to help develop students' critical thinking skills. They make problems by imitating the collection of questions in the students' books and other handbooks. Less than 60% of teachers who use description test instruments using a scientific approach indicate that the development of test instruments to measure students' critical thinking skills is still not optimal. Students are only given a problem taken from the subject in the student book. Teachers usually only give questions in textbooks or just follow the collection of

questions in other handbooks, and questions collected in the question bank. The questions given are limited to short entries. Though these questions tend to measure cognitive abilities to remember, understand and apply. These three abilities are low-level thinking skills, not reflecting students' critical thinking skills. Students in answering problems or answering questions given by the teacher are only a short answer and are not accustomed to developing their answers, whereas critical thinking skills are characterized by developing critical thinking to ask questions, solve problems and then evaluate problems in the learning process.

The above problems are in accordance with the facts found in the preliminary research carried out in the 5th grade of the Central Metro District Elementary School, namely the Metro 1 Public Elementary School and Metro Pusat 2 Public Elementary School as the sample school on March 14, 2017, through the interview activities found the fact that teachers are more many use test instruments in the form of questions in the book to measure students' critical thinking skills. Though the questions in the textbook are not paying attention to the measurement of students' critical thinking skills. The teacher has actually developed test questions on thematic learning which are usually used for daily tests. However, based on the analysis of test instruments conducted, the questions developed did not refer to students' critical thinking skills so that the problem tended to measure low-level thinking skills alone. Existing questions only refer to cognitive knowledge between C1 to C3 only. Teachers pay less attention to demands from C1 - C6. This proves that the development of critical thinking skills of students gets less attention so that in the end the ability of critical thinking of students has not been able to increase.

According to Hosnan (2014: 387) explained that assessment is a teacher's activity which is intended to measure certain competencies or abilities of activities that have been carried out in learning activities. According to Mangiante (2013: 222) assessment is a tool to measure the extent to which students have improved their learning based on standards. Meanwhile, according to Harlen (2013: 6) Assessment is defined as the process of collecting and using facts for a specific purpose about the learning outcomes of students.

According to Subali (2010: 114) procedures for carrying out assessments are (a) preparing grids, (b) arranging instruments, (c) reviewing the quality of instruments qualitatively, (d) testing instruments, to investigate the feasibility and validity empirically, and (e) implementation of measurements.

Assessment not only evaluates results but looks at how students are actively involved in the learning process. The teacher will be more objective in assessing students with performance appraisals. As expressed (Popham, 1995: 139) that in the assessment of student performance, the teacher wants an "authentic" response or the original form of observable activity. Tasks that are given can be in the form of oral or written, whose types of assignments are tailored to the learning objectives.

Rufina (2015) said that the use of assessment provides an opportunity for teachers to learn to identify students' strengths and weaknesses, thereby monitoring their growth and development. Student learning must be monitored appropriately from time to time to ensure the growth process goes well (Zulkifli, 2016: 67).

It is appropriate for teachers to understand and have skills in assessing student learning outcomes, making the teacher able to compile assessment instruments that are in accordance with certain rules. Assessments arranged in accordance with the rules of preparation will produce valid and reliable assessments. So that it will produce data and information about the level of achievement of student competencies in a valid and accurate manner.

In accordance with the above problems, the purpose of research and development is to produce a description test instrument based on scientific approach to measure the valid and reliable ability of critical thinking of grade V students of elementary schools.

METHOD

Types of Research and Procedures

This research is research and development or Research and Development (R & D). Research and development is a research that aims to produce a product for the development of a test instrument in the form of a description. The resulting product is then tested for its validity and reliability. The research and development used was Borg & Gall's design model (2003: 569-575) consisting of 10 steps. The steps that must be followed to produce the product are: initial research and information collection, planning, development of initial product formats, initial trials, product revisions, field trials, product revisions, operational field trials, final product revisions, implementation.

Population and Sample

This research and development population is Elementary School Elementary School V grade students at the Nusa Indah Cluster Metro Center with a total of 402 students.

Samples were taken using the Random sampling technique. This sampling technique is used with the consideration that the characteristics of the schools being sampled are almost the same, namely using the 2013 curriculum with the achievement of different school KKM's, so the sample in this study is 35 students in grade V

SDN 1 Metro Pusat.

Data collection instruments are tools used to collect all data, which are needed in a study. Data collection instruments in this study were used to collect data from the preliminary study, product development and product testing stages. In the preliminary study phase, the instruments used were documentation, test instruments and questionnaire sheets for expert validation and teacher validation.

Data analysis used in this research activity is the analysis of qualitative and quantitative descriptive data. Analysis of qualitative descriptive data in this study was used to process data sourced from comments and suggestions obtained from material experts, linguists and evaluation experts contained in the validation questionnaire. The results of the analysis of qualitative descriptive data will later be used as a condition to meet valid theoretical criteria.

Quantitative descriptive data analysis was used to analyze the data obtained in the form of validation assessment scores of material experts, linguists and evaluation experts to assess the feasibility of the developed instrument content. As well as student test results to measure the level of instrument validity and reliability.

A data or information can be said to be valid if in accordance with the actual situation (Arikunto, 2013: 72). Then the instrument is said to be valid if the instrument used can measure what should be measured. Data from the validation sheet results, giving an overview or quality exposure of the assessment instruments developed. The reliability of the description form questions by using the Product Moment Correlation formula as follows:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Information:

- rx_y = correlation coefficients X and Y
 - N = number of respondents
 - $\sum XY$ = the total multiplication of X and Y scores
 - $\sum Y$ = number of variable scores Y
 - $\sum X$ = number of variable scores X
 - $\sum X^2$ = total squared variable score X
 - $\sum Y^2$ = total squared variable score Y
- (Arikunto, 2010: 213)

The question is said to be valid if r count is greater than r table

Reliability test of the description test instrument was done by Cronbach Alpha method using the help of a Microsoft Office Excel computer program. The reliability value of the written test was in the form of a description according to Kusaeri (2014: 65) for tests designed to measure heterogeneous learning outcomes using the alpha coefficient formula. Therefore, to calculate the reliability of the description test in this study, researchers used the alpha coefficient formula

$$r_{11} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right)$$

Information:

- r₁₁ = reliability sought
 - $\sum \sigma_b^2$ = number of grain variants
 - σ_t^2 = total variant
 - k = many questions
- (Arikunto, 2006: 196)

Next interpret the magnitude of the reliability value with the correlation index as follows:

Table 1. List of coefficient interpretations“r”

Coefficient r	Reliability
0,80 – 1,00	Very strong
0,60 – 0,79	Strong
0,40 – 0,59	Medium
0,20 – 0,39	Low
0,00 – 0,19	Very low

Source : Sugiyono (2010: 257).

The instrument is said to be reliable if it provides a consistent or consistent result if it is tested repeatedly.

The level of difficulty of the problem is the opportunity to answer correctly a problem at a certain level of ability which is usually expressed in the form of an index. The difficulty level index of the problem is expressed in the form of proportions ranging from 0 to 1. The greater the difficulty level index obtained from the calculation results, the easier the problem is. Calculation of the difficulty level index is done for each question.

Calculating the level of difficulty of the description problem is used in the following formula:

$$\text{Difficulty} = \frac{\text{mean}}{\text{skor maksimum yang ditetapkan}}$$

Sumber: Kusaeri (2014: 106)

Table. 2. Difficulty Criteria Level Item Problem

No.	Difficulty Index	Difficulty
1	0,71 – 1,0	Easy
2	0,31 – 0,70	Medium
3	0,00 – 0,30	Difficult

Source: Kusaeri, (2010: 107).

If a problem has a difficulty level of 0, it means that no student is able to answer correctly, so it can be said that the problem is difficult. But on the contrary if a problem has a difficulty level 1, then the question is said to be easy and all students are able to answer it correctly

RESULTS AND DISCUSSION

Research result

The results of this research and development are in the form of a test instrument in the form of a description to measure critical thinking ability through a valid and reliable scientific approach, for fifth grade students of elementary school in odd semester. Theme 1 Organ Motion Animal and Human Sub-theme 1 Organ Animal Motion.

The procedure used in this study refers to the Borg & Gall development model which contains the main steps of the study as follows:

Research and Information Collection Phase.

Information gathering and needs analysis is done by observation, interviews and distributing questionnaires in the form of questionnaires to teachers in the Nusa Indah Cluster, Metro City Center, who have implemented the 2013 curriculum. After that, researchers conducted literature studies to study the theories and results of previous research related with the product to be developed. The product developed will have a theoretical basis and be supported by strong empirical facts. Field surveys were conducted to collect data related to needs.

Based on a series of research and development literature studies and needs analysis obtained information that, the real conditions that exist in the Nusa Indah Cluster Metro City Center District are: 1) The learning activities carried out are monotonous and only in accordance with the books of teachers and students; 2) The implementation of learning is not maximized using scientific approaches conducted in the classroom; 3) Teachers have not made their own test instruments that can help students' critical thinking abilities; 4) The low development of description test instruments through the scientific approach carried out by teachers; 5) Teachers only provide questions in textbooks that tend to measure low thinking skills; 6) Teachers only provide multiple choice questions and short entries as test instruments; 7) Development of test instruments conducted by the teacher in the form of questions that are used for daily tests, midterm exams do not pay attention to the measurement of critical thinking abilities.

The next step is to evaluate the psychomotor domain assessment instrument in the 2013 curriculum teacher's book, the rubric in the teacher's book is still general and has not been developed into an easy assessment instrument with indicators that are in accordance with the performance assessment and there are no usage instructions. In the teacher's book and the fourth grade students' 2013 curriculum semester book.

Based on the needs analysis, it is necessary to develop an assessment instrument that involves the ability of students to think critically in solving problems. The assessment instrument developed is a description test

instrument to measure the ability to think critically in learning using a contextual approach in elementary school.

Planning

This stage is carried out by designing a prototype instrument for student assessment based on the suitability of KI, KD and indicators in the teacher's book used in the lesson.

Drafting the Initial Product

The researcher developed an assessment instrument adapted to the steps to make a performance appraisal according to Harsiyati (2011: 200) with the following descriptions. 1) Analysis of KI / KD to determine the characteristics of KD; 2) Determine the description of skills or behaviors that should be done in order to be able to make the product required in KD; 3) Determine the product indicator which is determining the indicator; 4) Determine assessment techniques or tools; 5) Determine the indicator of the problem, in the realm of question indicator skills is a statement that describes in detail the student responses demanded, types of stimuli and criteria; 6) Develop contextual tasks that show mastery of students' skills; 7) Prepare assessment rubrics / guidelines.

Initial Trial

This initial trial is to find out the validity of the instrument theoretically. The theoretical validation includes expert judgment. The assessment of experts was used as the basis for revising and refining the prototype. Expert assessments are carried out by submitting an assessment instrument, student performance assignments, assessment rubrics, assessment instruments, and scoring guidelines. Then the expert is asked to assess the suitability of the instrument with the indicators on the validation sheet. Aspects assessed include 4 things, namely: (1) material; (2) evaluation; (3) language.

Table 3. Validation Results by Experts

Validator	Value
Material expert	78,50
Linguist	91,75
Evaluation Expert	75,00
Average	81,75

The results of the validation by the validator that has been carried out are used to revise the product of the performance appraisal instrument developed which will then be used for the main field test

Revision

Instrument revisions were carried out based on input from experts / validators. In general, suggestions for input obtained during the validator process and trial readability of the questions made to students. In general, suggestions or input obtained during the validation process from experts. According to material experts, indicators need to be developed so that instruments can measure achievement of Basic Competencies. Whereas according to the construction expert, the question instructions need to be corrected, the problem editor, the pictures to be adjusted to the measurement of critical thinking skills. Furthermore, according to linguists there are still errors in the use of punctuation marks and inappropriate word selection. So, the test instrument was improved based on the advice given by the expert.

Initial Field Test

Initial field testing was carried out at VB SD Negeri 2 Metro Pusat with a sample of 22 people. The number of description test instruments that were developed early were 20 questions. Students are asked to work on the questions that have been provided. The following is the result of the analysis of the questions tested at the initial trial stage. In testing the validity of this instrument using the Product Moment Correlation formula at Microsoft Excel. The question is said to be valid if r count is greater than r table. The R table value in this problem group is 0.423. All questions or 100.00% are declared valid. The level of reliability of the question is in a very strong category with the amount of r count 0.99. The level of difficulty of the instrument developed is 100.00% of the test questions developed in the medium category.

Product Revision

Product revisions are carried out based on the results of field trials and findings when the product is tested. Based on the results of this theoretical and empirical validation, the instructional products that were developed were not revised and feasible to proceed to the main field test.

Main Field Test

The Main Field Test was conducted on the VA Class of SD Negeri 1 Metro Pusat as many as 35 students. The

number of description test instruments developed early as many as 20 questions overall have good quality of validity, reliability, level of difficulty and different power. Then all the questions enter the product implementation stage. At this stage students are asked to work on the questions that have been provided.

Ahir revision

The final product revision is based on the results of hypothesis testing and findings in the field when the product is tested. Based on the results of hypothesis testing that has been carried out, students' learning outcomes are increased. Furthermore, based on the results of consultations with material experts, linguists and evaluation experts, it was concluded that this description test instrument was not revised and was feasible to be implemented.

DISCUSSION

Based on a series of research and development literature studies and needs analysis obtained information that, the real conditions that exist in the Nusa Indah Group Metro City Center City are: 1) The learning activities carried out are monotonous and only in accordance with the teacher's and student's books; 2) The implementation of learning is not maximized using scientific approaches conducted in the classroom; 3) Teachers have not made their own test instruments that can help students' critical thinking abilities; 4) The low development of description test instruments through the scientific approach carried out by teachers; 5) Teachers only provide questions in textbooks that tend to measure low thinking skills; 6) Teachers only provide multiple choice questions and short entries as test instruments; 7) Development of test instruments carried out by the teacher in the form of questions used for daily tests, midterm examinations do not pay attention to critical thinking ability measurement. Development of student performance assessment is intended to improve students' performance skills in conducting experiments or activities trying and communicating in scientific learning by giving a number of material assistance and instructions to students. So that the teacher can develop an authentic assessment that is feasible and meet certain criteria in accordance with the response assessment of students, teachers and two expert validators.

This is in accordance with the learning theory of Vygotsky, Scaffolding. Scaffolding means giving a large amount of assistance to a student during the initial stages of learning and then the student takes on increasingly large responsibilities as soon as he can. The assistance can be in the form of instructions, warnings, encouragement, problem solving into steps of solving, giving examples, or others so as to enable students to grow independently. (in Trianto. 2013: 76-77).

Performance appraisal of grade V students is applied to learning in schools that have implemented the 2013 Curriculum. The curriculum uses integrated thematic learning. The point is integrated teaching is intended as a teaching activity by combining several subject matter in the same theme and time. Learning in the 2013 curriculum uses a scientific approach consisting of activities to see, ask, try, reason, and communicate, this is in accordance with Dyers' theory in Maryanti (2015: 1) that creativity can be obtained through observing, asking, trying, reasoning, and network.

The level of development of elementary students always starts with the real thinking stage. In everyday life, they do not see subjects on their own. They see objects or events which contain a number of concepts / material in several lessons.

The development of student performance assessment instruments is also based on the advantages possessed by performance appraisal. Performance assessment can be used as an alternative to the tests that have been widely used to measure the success of student learning in school. Therefore the use of performance appraisal is important in the learning process because it can provide more information about students' abilities. in the process or product, not just getting information about the right or wrong answers. Students are more capable of theorizing, but are less skilled at the theory.

The government to carry out learning using a scientific approach has a good purpose, one of which is to improve intellectual ability, especially the high-level thinking skills of students (Daryanto, 2013: 54). The ability to think critically of students into the category of high-level thinking skills. This is in accordance with the opinion of Conklin (2012: 14) who said that "critical thinking is a term that is generally associated with high-level thinking skills characterized by careful analysis and consideration". Therefore, the description test instrument was chosen to be developed in this study. According to Sunarti & Selly (2014: 30) the question of the form of description requires the ability of students to organize and formulate answers using their own words. Problem form descriptions should measure various abilities (memory, understanding, application, analysis and other higher abilities).

The description test instrument based on the scientific approach developed in this study follows the steps in making the test instrument proposed by Purwanti, et.al (2008: 4.15-4.19), namely 1) test planning; determine the scope of the material to be measured, determine the form of the test, and determine the length of the test, 2) write the question item; establish content validity; conduct trials, and revise questions, and 3) conduct measurements

with tests; maintain the objectivity of the test, give a score on the test results, and analyze the test results.

In the second stage, writing the question. Writing draft questions is adjusted to the lattice of the questions created. The ability to think critically is one part of a higher order thinking ability (HOTS). So in terms of writing indicators referring to Bloom's Taxonomy cognitive aspects. Based on the revised Bloom Taxonomy, it is divided into six levels, three levels of low level capability namely aspects of remembering (C1), understanding (C2), applying (C3), and three levels of high-level abilities namely analyzing (C4), evaluating (C5), and creating / creation (C6) (Mohammed, et.al., 2015: 14).

In line with Cabrera's (2001) study, this study shows that the instruments developed meet several requirements recommended by the assessment literature, namely (1) meaningful to users, (2) reliable and valid, and (3) observable index behavior to avoid subjective impression.

The instruments developed have gone through several stages, ranging from fulfilling the rules of writing instruments, theoretical validation and empirical validation. The results of the validity are in line with the research conducted by Budhiwaluyo (2016) in his research the results of field trials show that the product has a high value of validity and reliability in measuring performance in student practicum.

The results of the content validity are in line with the research conducted by Putri (2017), the results of her research indicate that the description test instrument developed has fulfilled the content validity based on evaluation by 3 experts and 3 practitioners. The reliability of all rubrics in the assessment is categorized as very high.

The feasibility of product development is in line with the results of Ratnami's (2016) study of the feasibility of description test instruments, namely the quality of the results of the development of performance appraisal according to expert review, namely content expert test is in very good qualification of 90.00%; the learning design expert test is at very good qualification of 92.00%; assessment expert test is in very good qualification learning that is 90.00% and field trial is in very good qualification 90.76%). Usman (2014) in his research produced teacher response data in very good categories.

Mardhapi (2004) based on the compiled article can be concluded that the assessment instrument must have evidence of validity and reliability.

The results of this study in accordance with research conducted by Abosalem (2016: 1-11) states that using high-level assessment will help students to improve and evaluate their thinking abilities such as using multiple choice tests or descriptions.

The development of the description test instrument has passed the validity test, namely theoretical validation and empirical validation. Theoretical validation includes the assessment of experts, namely material experts, evaluation experts and linguists, the instrument is declared feasible to be tested after the revision.

Based on the validity test using the Product Moment formula on Micosoft Excel 20 the question is declared valid. The number of testees as many as 22 students, the r table value is 0.423. The r value of this table is compared to r count. If the calculated r value is greater than r table then the problem is declared valid. All valid questions are then used at the product implementation stage with a total sample of 35 so that the r table is 0.334, stated that all questions are valid.

After validity testing, the reliability of the description test instrument was tested. In the reliability test the field trial phase reliability of the questions in the category is very strong with r count 0.99 and the product implementation stage is still in a very strong category with r count 0.989. This is contrary to the opinion of Widoyoko (2014: 120-121) which states that one of the shortcomings or weaknesses of the description test instrument is low test reliability. This means that the score achieved by the test participants is not consistent if the same test is tested several times. In the field trial, the reliability of the test instrument was also included in the very strong category.

According to Kusaeri (2014: 106) the level of difficulty of the problem is the opportunity to answer correctly a problem at a certain level of ability which is usually expressed in the form of an index. The difficulty level index of the problem is expressed in the form of proportions ranging from 0 to 1. The greater the difficulty level index obtained from the calculation results, the easier the problem is. Of the 20 questions that were produced and tested, both in the preliminary trial and field trials, it was found that all the questions in the medium category.

This is in accordance with the principles that must be considered by the teacher in conducting an assessment, according to Sudaryono (2012: 54-55), namely: (1) Continuous principles, (2) comprehensive principles, (3) Principles of objectivity (objectivity) , (4) Principles of validity (validity) and reliability (reliability), (5) Principle of measuring criteria, (6) Principle of usability.

Products resulting from the development of critical thinking assessment instruments have the following advantages: 1) the instrument of Critical Thinking Assessment is an assessment instrument in accordance with the revised 2013 curriculum; 2) the instrument of critical thinking assessment contains 4 indicators of critical thinking ability, namely giving simple explanations, building basic skills, concluding, and providing further explanations. Each item item contains 10 indicators of critical thinking ability; 3) Critical Thinking Assessment

instruments are presented in a language that is simple, clear, and easy to understand. Besides that, it is equipped with pictures / tables as a stimulus; 4) this critical thinking assessment instrument has fulfilled the rules of question writing and also validity theoretically and empirically; 5) assessment instruments developed effectively in measuring students' critical thinking skills.

CONCLUSION

Based on the results of the research and discussion, the conclusion of a description test instrument based on a scientific approach can be realized because it has been tested realistically and is arranged based on the steps of making the test instrument, namely 1) test planning; determine the scope of the material to be measured, determine the form of the test, and determine the length of the test, 2) write the question item; establish content validity; conduct trials, and revise questions, and 3) conduct measurements with tests; maintain the objectivity of the test, provide a score of test results, and analyze test results. Whereas the substance of the test instrument produced is based on 4 indicators of critical thinking ability which are defined as 1) providing simple explanations, 2) building basic skills, 3) concluding, and 4) providing further explanations.

The description test instrument produced has fulfilled theoretical and empirical validation. Theoretical validation is seen through evaluating expert judgment of 75.00 (good category), material expert at 78.50 (very good category), and linguist at 91.75 (very good category). While empirical validation through the results of trials and analysis as many as 20 questions are declared valid. The reliability of the questions is 0.99 (the category is very strong), the level of difficulty in the category is moderate, and the distinguishing power of 2 questions in the category is very good and 18 questions in the good category. Based on the results of the product implementation it was stated that 20 questions were valid. The reliability of the questions is 0.989 (very strong category), the level of difficulty in the medium category, and the discriminating power of 5 questions in the good category and 15 questions in the good category.

REFERENCES

- Abosalem, Yousef. 2016. Assessment Techniques and Students Higher-Order Thinking Skills. *International Journal of Secondary Education*, 4 (1): 1- 11.
- Arikunto, Suharsimi. 2013. *Dasar-Dasar Evluasi Pendidikan*. Edisi kedua. Jakarta. Bumi Aksara.
- Borg, D. Walter, Joyce P. Gall and Meredith D. Gall. 2003. *Educational Research An Introduction*. Boston: Perason Education, Inc (569-595)
- Budhiwaluyo, Nugroho. 2016. Pengembangan Instumen Penilaian Kinerja pada Praktikum Struktur dan fungsi Sel di SMA Negeri 1 Kota Jambi. *Edu-Sains Jurnal*. Vol 5. No.2.
- Cabrera, F. Aberto. 2001. Developing Performance Indicators for assessing Classroom Teaching Practices and Student Learning. *Research in Higher Education*. 42.(3). Pp 327-352.
- Conklin, W. 2012. Higr Order Thinking Skill to Develop 21st Century learners. Huntington Beach, California: Shell Education.
- Daryanto. 2013. *Pembelajaran Tematik Terpadu Terintegrasi Kurikulum 2013*. Gaya Media. Yogyakarta.
- Harlen, W. 2013. *Assesment & Inquiry based science education*. Triestly Italy: Global Network of Science Academies (IAP) Science Education Programme (SEP).
- Hosnan. 2014. *Pendekatan Sainifik dan Konstektual dalam Pembelajaran Abad 21: Kunci Sukses Kurikulum 2013*. Jakarta: Ghalia Indonesia.
- Mangiante, Elaine Silva. 2013. Planing Science Intruction for Critical Thinking: Two Urban Elementary Teacher' Responses to a State Science Assesment. *Journal Education Science*, Vol 3: 222-258.
- Mardapi, Djemari. 2004. *Teknik Penyusunan Instrumen tes non tes*. Yogyakarta: Mitra Cendikia.
- Maryanti, Endah Febriana. 2015. Instrumen Penilaian Otentik PETASAN GALAU pada Mata Pelajaran Kewira Usahaan. *Jurnal Study Sosial UNILA*. Vol.2 No. 4 Hal 24.
- Mohammed, Gulistan Saido., et.al. 2015. Higer Order Thinking Skill Among Secondary Schooll Students in Science Learning. *The Malaysian Online Journal of Educational Science*. Vol 3 (3): 13 – 20.
- Permendikbud No.23 Tahun 2016: *Standar Penilaian Pendidikan*.
- Popham, W.J. 1995. *Education Evaluation*. Engglewood Cliffs N.J. : Prentice-hall _____. 1978. Criteria Referenced Measurement, Englewood, Cliffs, N.J: Prentice-p://web.ku.edu.
- Purwanti, Endang. Et., al. 2008. *Bahan Ajar Cetak Asesmen Pembelajaran SD*. Direktorat Jendral Pendidikan Tinggi Departemen Pendidikan Nasional. Bandung.
- Putri, Suwandi F & Istiyono, Edi. 2017. The Development of Performance Assessment of Stem-Based Critical Thinking Skill in the High School Physics Lessons. *International Journal Of Environmental And Science Education*. 12.(5) pp. 1269-1281.
- Ratnami, Made V. 2016. *Assesmen kinerja dalam Pembelajaran IPA Pada Siswa Kelas V SD Negeri 1 Banyuning Kecamatan Buleleng Kabupaten Buleleng Tahun Pelajaran 2015/2016*. Universitas Pendidikan Ganesha.

- Rufina C. Rosaroso and Nelson A. Rosaroso. 2015. Performance-based Assessment in Selected Higher Education Institutions in Cebu City, Philippines. *Asia Pacific Journal of Multidisciplinary Research* 3(4): 72-77.
- Subali, Bambang. 2010. *Prinsip Asesmen dan Evaluasi Pembelajaran*. Pustaka. Yogyakarta.
- Sudaryono. 2012. *Dasar-Dasar Evaluasi Pembelajaran*, Yogyakarta. Graha Ilmu.
- Sugiyono. 2010. *Metode Penelitian Tindakan (Pendekatan Kuantitatif, kualitatif, dan R&D)*. Alfabeta. Bandung.
- Sunarti & Selly Rahmawati. 2014. *Penilaian dalam Kurikulum 2013*. Ar-ruzz Media, Yogyakarta.
- Trianto. 2013. *Model Pembelajaran Terpadu, Konsep, strategi, dan Implimentasinya dalam Kurikulum Tingkat Satuan Pendidikan (KTSP)*, Jakarta. Bumi Aksara.
- Usman, Herman. 2014. Pengembangan Perangkat Penilaian Kinerja Praktikum Fisika Pada Peserta Didik SMP UNISMUH Makassar. *Jurnal Sains dan Pendidikan Fisika*. Jilid 10. Nomor 3. Hal 274-284.
- Widoyoko, Eko Putro. 2014. *Penilaian Hasil Pembelajaran di Sekolah*. Pustaka Pelajar, Yogyakarta.
- Zulkifli, Nur Idayu. 2016. The Assessment of Children's Performance at Chile Care Centre. *Procedia – Sosial and Behavioral Science*. 234: 64-67.