JURNAL PEMBELAJARAN FISIKA

http://jurnal.fkip.unila.ac.id/index.php/JPF

Vol 10 (1), 2022, 011-022 ISSN: 2302-0105 (p); 2684-9828 (e)

Implementation of Written Assessment Higher Order Thinking Skills in Physical Learning with a Scientific Approach based Blended Learning

Chandra Ertikanto*, I Wayan Distrik, I Putu Dewa Nyeneng

Physics Education Under Graduate Program, University of Lampung, Bandarlampung

** e-mail: chandrafkipunila@gmail.com*

Received: April 30, 2021 *Accepted: June 8, 2022* *Published: June 11, 2022*

Abstract: This study aims to describe: (1) the effectiveness of HOTS in physics learning with a blended learning-based scientific approach, (2) the efficiency of the suitability of learning objectives of physics through a blended learning-based scientific approach with HOTS written assessment, and (3) safety through teacher responses to written assessments. HOTS through a scientific approach based on blended learning in physics learning. The design of this research is the implementation stage of continually development research from 2019 which has gone through the stages of product testing and extensive testing, where focus on Senior High Schools in Bandar Lampung City, Lampung Province during the Covid-19 pandemic. The sample used in this study were 36 students of SMAN 9 Bandar Lampung with a purposive sampling technique. The research instrument used was the HOTS Written Assessment with multiple choices of 20 questions. Data analysis techniques carried out qualitatively. Conclusion: (1) The results of the effectiveness test were 83,33% of students passed the KKM, and the highest score was 97,5. (2) The efficiency of the suitability of the learning objectives of Physics through a blended learning-based scientific approach with HOTS written assessment, is very efficient in use, and (3) The safety of HOTS written assessments through a blended learning-based scientific approach in physics learning is very meaningful.

Keywords: Assessment, Higher order thinking skills, Scientific Approach

DOI: http://dx.doi.org/10.23960/jpf.v10.n1.202202

INTRODUCTION

Learning is an activity which facilitates interaction between teacher and student (Nitko, 1996). The best quality interaction between students and teachers, depends on the creativity and enthusiasm of the teacher in conducting the learning process (Sola & Ojo, 2007). Learning in the revised 2013 curriculum for all levels is expected to use a scientific approach (Anonim-B, 2013). The implementation of learning that uses a scientific approach, encourage students to learn to find themselves and transform information in a complex manner, to check new information with what is already in their mind, and to develop it into abilities that are in accordance with the environment and current developments (Buck & Gayle, 2007) .

Learning with a scientific approach that is based with blended learning through a particular learning model requires students to be active starting from the process of observing the environment, formulating / making questions, formulating / answering problems, designing / trying, reasoning / processing data, and communicating (Anderson et all, 2001), so it is necessary to use a system a session which is authentic based on the learning process (Neumeier, 2005).

Learning with a scientific approach based on blended learning which is oriented towards three domains, namely the realm of attitude (affective), the realm of knowledge (cognitive), and the realm of skills (psychomotor). Students are directed to compare the results of predictions and theories through experiments using a scientific approach (Dwiyogo, 2016: Husamah, 2014). Learning like this is carried out by emphasizing direct experience with the aim of developing competence, students are also expected to be able to understand natural phenomena around them through the process of finding out, this is expected to help students to gain a deeper understanding (Suhandana; 2012).

Authentic assessment has a fairly strong relevance to the scientific approach to learning in accordance with the demands of the 2013 revised curriculum, for example environmental observation, formulation/questioning, formulation/ finding of solutions to problems, design, experimentation, data reasoning/ processing, and communication. Assessment This kind of thing can be realized in the form of a written assessment with higher order thinking skills, it can also describe the ability of student learning outcomes in terms of observing, asking questions, answering problems, designing / trying, and building networks to be communicated (Nitko, 1996).

In learning, it generally ends with an evaluation process, with the aim of knowing the level of attainment of students' knowledge and abilities. Process evaluation is very important to note, because evaluation is a tool to assess and measure the level of student ability (Arikunto & Jabar. 2012). The evaluation process which aims to assess students' abilities. designed and implemented by the teacher in accordance with the lesson plan. The assessment process is expected to be in line with the development of models and learning approaches being developed (Akinoglu: 2008). This assessment process is used by the teacher with the aim of seeing student learning outcomes. Learning with the revised 2013 curriculum, pays great attention to a session on all aspects in a comprehensive, meaningful way assessment conducted in the realm of attitudes (affective), the realm of knowledge (cognitive), and the realm of skills (psychomotor), which also means that from input, process, to output in learning, otherwise known as a session authentic (Anonim-C, 2014).

Written authentic assessment with high-order thinking skills is also commonly implemented in the 2013 revised curriculum, especially in the multiple choice written test on the Final Semester Examination (UAS), School Examination (US) and on the National Examination (Kemendikbud; 2014), this is none other than because This form of plural choice written tests can reveal students' abilities in remembering, understanding, organizing, implementing, analyzing, synthesizing, evaluating, and communicating (Kunandar; 2013: Ruiz-Primo & Furtak; 2007). The plural choice written assessment is also able to reveal a lot of concepts with higher-order thinking skills, if designed correctly.

In the field, it turns out that most of the teachers have not made the test as mentioned above. Based on filling out the questionnaire, the physics subject teacher at one of the public high schools in Bandar Lampung still uses written tests to assess student learning outcomes. Most of the written tests made by the teacher are only used to measure the level of knowledge (cognitive), and are not yet suitable as the scientific approach demands a session in the revised 2013 curriculum. Implementation a session at the end of the semester at SMAN, most teachers used multiple choice assessment tests, and most of the teachers argued that the multiple choices assessment only measured cognitive abilities. Meanwhile, a small proportion of teachers thought that a session the multiple choices test used does not cover all aspects. The reason most teachers think that the assessment is written multiple choices designed to cover only the C1 to C3 domains (Sanjaya; 2010). This happens because of the limitations of the teacher in making the test, it also shows that a session written plural choices implemented in schools have not been based on the scientific approach referred to in the 2013 revised curriculum.

Based on the preliminary description above, a study was conducted with the title "Implementation of Higher Order Thinking Skills (HOTS) Written Assessments in Physics Learning with a Scientific B Approach based on Blended Learning". The main problem in this research is: "How is the HOTS Written Assessment Product for Learning Physics with a Scientific Approach tested?" In order to focus more on solving these problems, several research questions were raised as follows: (1) How is the effectiveness HOTS Physics learning with a scientific approach based on blended learning for high school? (2) How is the efficiency of conformity learning objectives of Physics through a scientific approach with a HOTS written session for High school? and (3) How meaningful is the teacher's response to a written HOTS session on Physics learning.

The main objective of this research is to test the product a written session of HOTS for learning Physics with a scientific approach based on blended learning. In detail, this study aims to describe: (1) the effectiveness of HOTS in high school physics learning with a scientific approach based on blended learning, (2) suitability efficiency Physics learning objectives through a scientific approach based on blended learning with a HOTS written session for SMA, and (3) existence through teachers' responses to a HOTS written session through a scientific approach based on blended learning on learning Physics.

METHOD

1. Research Design & Procedures

In 2020 an Experimental Research was conducted as a continuation of the R & D research step, and aims to test the p instrument the set of written assessment on higher order thinking skills (HOTS) in physics learning with a scientific approach based on blended learning in Senior High Schools in Bandar Lampung City, Lampung Province.

2. Population and Sample

This research was conducted in one of the high schools in the city of Bandar Lampung, Lampung Province. The selection of SMA was carried out purposively, namely SMA that had implemented the 2013 curriculumrevision, and the high school certainly applies a scientific approach, during the Covid-19 pandemic, learning was carried outbased on blended learning.

3. Data Collection and Instrument

The research is declared successful if the assessment instrument product from the result of research which are developed has the following values: (1) Effective; (2) Efficient, and (3) Beneficial.

4. Data Analysis

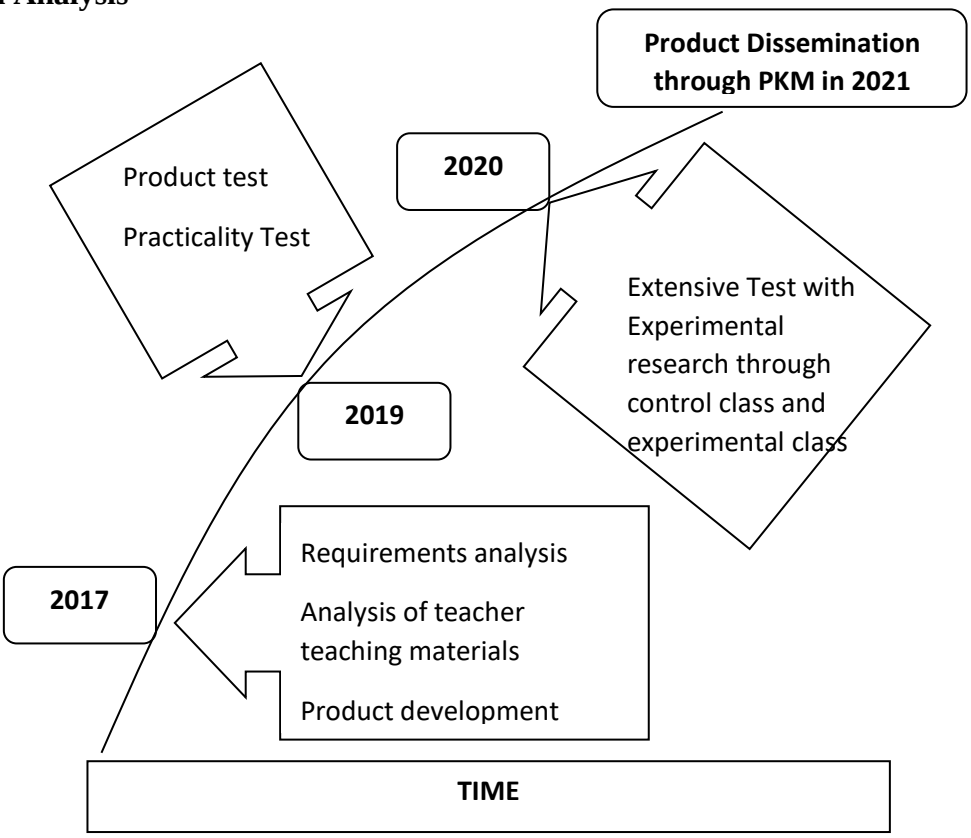

Figure 1. Flowchart of Research Stages

The HOTS written test assessment tool for Physics learning with a scientific approach before being used, a series of tests were carried out in 2020, including: (1) Validity: the test results were analyzed with Anates V4, obtained all items had a discrepancy index of more than 0.80 into good category; (2) Reliability: the reliability of the multiple-choice test was 0.96. The reliability value of this test is in the good category; (3) Difficulty Level; Based on the results of the trial, it was obtained that the difficulty level was between 0.25 - 0.75, with the level of difficulty being included in the good category. According to Fernandes in Koyan (2012); (4) Distinguishing Power: all items have a difference index of more than 0.5. Meanwhile, Kartowagiran (2012), explains that an item is said to be of high quality if its distinguishing power is at least 0.41; and (5) Quality of Detractors (Distractors): The results of the multiple-choice test test obtained distractors' quality in the range of 0.04 – 0.53. It means that the distractor quality in this test is categorized as good or functioning. According to Kartowagiran (2012), a distractor is said to be good if it is chosen by at least 2% of all test takers.

RESULT AND DISCUSSION

1. Research Result

The results of research conducted at one of the SMAN in Bandar Lampung are in accordance with the objectives of the study, to achieve the research objectives obtained data according to the research steps. The quantitative data of the pretest and posttest results obtained at the beginning and end of the lesson are shown in Table 1.

Based on the data in Table 1, the average pretest value of the experimental class is greater than the average value of the control class. The average posttest score of the experimental class is greater than the average posttest score of the control class.

Table 1. Data on Average Student Pretest and Posttest Results

No	Parameter	Experiment Class		Control Class	
		Pre	Post	Pre	Post
1	Total students	36	36	36	36
2	Lowest score	2.5	35	5	40
3	The highest score	32.5	100	25	100
4	Maximum Value	100	100	100	100
5	Average value	13.6	80.94	12.5	70

Furthermore, the test results were analyzed using the Anates V4 program, with 36 subjects and 20 multiple choice items and 5 description questions. The results of the analysis of this item are to find out about the sign correlation, level of difficulty, distinguishing power and quality of distraction.

2. Research Discussion

(2.1) The effectiveness of the HOTS instrument for learning Physics with a scientific approach based on blended learning for

SMA. The results of the ability assessment are taken through tests both at the beginning and after the students have finished participating in learning, then students are given evaluation questions to determine the level of student understanding of the subject of Elasticity and Hooke's Law. A total of 36 students of class XI IPA4 (experimental class), 31 students scored above the KKM and five students scored below the KKM. In class XI IPA4 (experimental), the average ability before applying learning with a blended learning-based scientific approach was only 13.60, after being given treatment the student's ability increased to 80.94. There was an increase in the average student's ability of 67.34 after the application of the blended learning-based scientific approach.

Based on the percentage of data 83.33% of class XI IPA4 (experimental) has been completed. Arikunto (2012) states that if 75% of students who learn are then carried out a thorough test above the KKM score, then the test instrument product is said to be effective and suitable for use as a test instrument (Haryati: 2013), thus, the HOTS Instrument for learning Physics with a scientific-based approach Blended development is said to be feasible and effective as a test instrument. The results of the HOTS effectiveness test in physics learning with a blended-based scientific approach.

Implementation of a written assessment of higher order thinking skills in physics learning, with a blended learning-based scientific approach oriented to a scientific model. in which there is an experimental method at each stage of its activity. In the activity, analysis questions are presented, experimental procedures, and illustration pictures related to Elasticity and Hooke's Law, in each activity students are helped in finding the desired concept (Windschitl: 2004), referring to the achievement of learning objectives, so that the learning developed is effective for finding the desired concept (Windschitl: 2004) used.

The results of the effectiveness test on class XI IPA4 students have been carried out, and the results obtained indicate that the implementation of a written assessment of higher order thinking skills in physics learning with a blended learning-based scientific approach, which was developed and tested at a high school in Bandar Lampung, is effective, as an instrument. written assessment of higher order thinking skills with the acquisition of learning outcomes of 83.33% of the total number of students as many as 31 students have passed the KKM, out of 36 students with the highest score of 97.5.

(2.2) Efficiency of suitability of learning objectives of Physics through a scientific approach based on blended learning with HOTS written assessment.

First, the main purpose of research to produce written assessment of higher order thinking skills in physics learning with a scientific approach based on blended learning feasible and effective. The material described relates to the situation or context of the student's real world

Second, The written assessment of higher order thinking skills in physics learning with a scientific approach based on blended learning material Elasticity and

Hooke's Law has gone through several stages of the process, one of which is the formative evaluation process. Formative evaluation includes: material expert test, design expert test, and one on one test in the previous year.

All three have been passed and there are suggestions for improvements for a written assessment of higher order thinking skills in physics learning with a blended learning-based scientific approach on Elasticity and Hooke's Law, as explained in the previously described development research stage. The written assessment of higher order thinking skills in physics learning with a scientific approach based on blended learning was revised in accordance with the recommendations, so that a written assessment was obtained that was ready to be tested, in accordance with the 2013 Revised Curriculum standards.

The concepts of Elasticity and Hooke's Law in everyday life are visualized attractively through pictures and a series of experimental questions contained in a written assessment of higher order thinking skills in physics learning with a blended learning-based scientific approach to Elasticity and Hooke's Law; The written assessment of higher order thinking skills in physics learning with a blended learning-based scientific approach, the material of Elasticity and Hooke's Law is arranged systematically, making it easier for students to do;

(2.3) The effectiveness of HOTS's written assessment through a scientific approach based on blended learning in physics learning.

At the end of the implementation of the written assessment, higher order thinking skills in physics learning with a scientific approach based on blended learning. In terms of the characteristics of the test instrument, it is determined by taking into account the characteristic indicators of the questions used in the test, namely validity, reliability, distinguishing power, difficulty level, and distracting quality.

a. Test Validity

The results of the validity test that have been carried out from the aspects of material, construction and language are illustrated in a graph as in Figure 2.

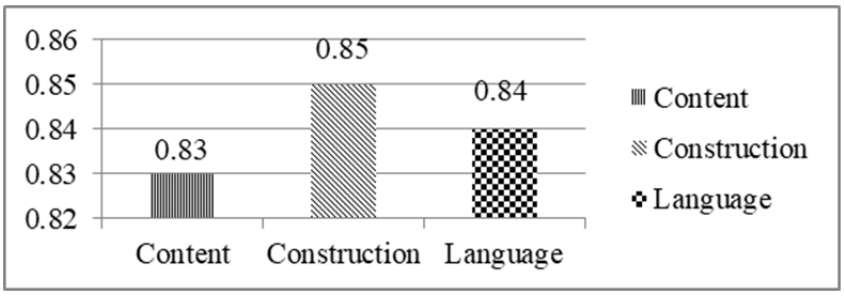

Figure 2. Graph of Validation Results

HOTS assessment validation scores from each aspect: material aspects with a score of 4.16 or 83%, construction aspects with a score of 4.23 or 85% and language aspects with a score of 4.22 or 84%. This score shows qualitatively the instrument is categorized as very valid. In line with Khabibah (2006), that the scores in the $4 \leq VR \leq 5$ interval are very valid. A valid instrument means that the measuring instrument used to obtain the data is also valid. Valid means that the instrument can be used to measure what should be measured.

b. Reliability

The measurement results obtained Test Reliability = 0.96. Determination of the degree of reliability of multiple choice items, using the criteria proposed by Sugiono (2013). The reliability value in the range $0.81 < r \leq 1.00$ is in the very high category. The reliability value of this test is included in the very significant question category, this is in line with Guilford's opinion in Kartowagiran (2012), which states that a measuring instrument that has a reliability coefficient of 0.8 is good.

Furthermore, the reliability of the description questions used the AnatesV4 program. The number of subjects = 36 and 5 items with the results of the reliability test = 0.81. The reliability value of this essay test is in good category. This value is in line with the opinion of Nunnally in Kartowagiran (2012), that the description questions which have a reliability coefficient of 0.6 - 0.7 and for multiple choice questions that have a reliability coefficient of 0.75 - 0.90 can be said to be good. The same thing is explained by Feldt and Brehmman in Kartowagiran (2012), that an instrument that has a reliability coefficient of ≤ 0.7 is said to be reliable.

A reliable instrument is an instrument that, when used several times to measure the same object, will produce the same data. A test is said to have high reliability, if the test provides consistent result data even though it is given at different times to the same respondent. If the test results are constant or if the changes are not significant, then the test is said to be reliable. Therefore, reliability is often referred to as trustworthiness, reliability, consistency, consistency, and stability of the instrument. Reliability concerns the problem of measuring instrument accuracy. An instrument is considered reliable if the instrument can be trusted as a measuring tool for research data.

c. Level of Difficulty

The level of difficulty is needed to find out how difficult the instrument being tested is based on the test results carried out by students. The calculation of the level of difficulty in this study uses the Anates V program. The results of the difficulty level test for the assessment instrument are presented in the table below.

Table 2. The sample of table format

No	Form of Problem	Level of Difficulty	Percentage
1	Multiple Choice	Easy	10%
		Moderate	80%
		Difficult	10%
2	Description	Easy	20.00%
		Moderate	60.00%
		Difficult	20.00%

Judging from the difficulty level of the questions, the number of questions is proportional, with details of most of the questions in the medium category as many as 75%, 15% difficult and as many as 10% easy. The assumption used to obtain good quality, in addition to meeting validity and reliability, is the balance of the difficulty level of the question. The balance that is meant is the existence of questions that include easy, medium and difficult proportionally. The level of difficulty of the questions in terms of the ability of students to answer questions. There are several basic considerations in determining the proportion of the number of questions in the easy, medium and difficult categories. The first consideration is the existence of balance, that is, the number of questions is the same for the three categories.

The level of difficulty of a test is intended to separate students who really learn a lesson from students who have not studied the lesson, so a good test or item is a test or item that can really separate the two groups of students earlier. So, besides having to have a certain degree of difficulty, each item must also be able to distinguish between students who are smart and students who are less intelligent.

According to Fernandes in Kartowagiran (2012), the items that produce an average score of about 50% of the maximum score can be said that these items have the right level of difficulty. Meanwhile, Thomas and Dawson in Kartowagiran (2012) explained that items that had a difficulty level of 0.25 - 0.75 were said to be good.

d. Discernment

Distinguishing Power Analysis of a test question is how the ability of the question is to distinguish students who belong to the smart group from students who are less intelligent groups. By knowing the distinguishing power we can identify the extent to which a student can receive learning and as an evaluation of the learning model applied by an educator. The distinguishing power in this study ranged in the range (0.44 - 1.00), this value indicates that the distinguishing power of the questions is very good.

The index used in differentiating high-ability test takers from low-ability test takers is the discriminating power index. This index shows the suitability of the question function with the overall test function. Thus this question is the same as the distinguishing power of the questions, namely having the power of difference between high-skilled test takers and low-ability test takers.

e. Tricking Quality

The form of multiple choice questions is a question where the answer must be selected from several possible answers that have been provided. Each multiple choice question consists of a question subject and an answer choice. The answer choices consist of an answer key and a distractor.

Distractors are answers that are not correct, but allow students to be fooled into choosing it if students do not master the subject matter well. The results of the instrument trial were selected by at least 6% of the test participants, this means that the distractor on the questions was functioning properly. It is in line with Kartowagiran (2012) that enforcement is said to be good if it is chosen by at least 2% of all test takers. Meanwhile, Nitko (2012) states that a cheat is said to function when it is at least chosen by a test taker from the low group. There should be more voters from the low group than the top group. Distractors can also be said to function when test takers from the upper group can distinguish between the counter and the answer key so that there are more people who choose the answer key than those who choose the distractor.

Based on the description above, the instrument parameters in the testing activity are in the form of a multiple choice test with five answer choices and the essay questions have a high level of stability. This is also supported by a high reliability estimate of 0.96 at the trial stage, this shows that the measurement results with this instrument are reliable. This is in line with Suryabrata (2000), a test that has a reliability coefficient of at least 0.80. The results of the measurement by means of a test with this reliability coefficient can be used to make decisions about individuals.

CONCLUSION

The conclusions from the results of this study are; (1) The results of the effectiveness test in class XI IPA4 students at a high school in Bandar Lampung showed that the implementation of a written assessment of higher order thinking skills in physics learning with a scientific approach based on blended learning was developed effectively, with the acquisition of learning outcomes 83,33 % of students passed the KKM, and the highest score was 97.5; (2) Efficiency of suitability of learning objectives of Physics through a scientific approach based on blended learning with HOTS written assessments, very efficient when viewed from: evaluation, which makes it easier for students to do it; visualization that is displayed in everyday life can be interesting through pictures and a series of experimental questions; systematic, makes it easier for students to do; The product of the implementation of the written assessment of higher order thinking skills in physics learning is very efficient to use; and (3) The reliability of HOTS written assessment through a scientific approach based on blended learning in physics learning in terms of: (a) the validity of the content of the material is 83%, construction is 85% and language is 84%; (b) Reliability has met the requirements, even including high with a reliability coefficient of more than 0.80; (c) The level of difficulty, including both, is in the range between 0.28 and 0.78; and (d) Very good difference, in the range 0.44 to 1.00. construction 85% and language 84%; (b) Reliability has met the requirements, even including high with a reliability coefficient of more than 0.80; (c) The level of difficulty, including both, is in the range between 0.28 and 0.78; and (d) Very good difference, in the range 0.44 to 1.00. construction 85% and language 84%; (b) Reliability has met the requirements, even including high with a

reliability coefficient of more than 0.80; (c) The level of difficulty, including both, is in the range between 0.28 and 0.78; and (d) Very good difference, in the range 0.44 to 1.00.

ACKNOWLEDGEMENTS

The author would like to thank especially the leadership and staff of FKIP and department staff, who helped facilitate the research so that this article was realized. In this case, the author's research was financed by a grant.

REFERENCES

Akinoglu, Orhan. 2008. "Assessment of the Inquiry-Based Project Implementation Process in Science Education Upon Students' Points of Views". International Journal of Instruction, 1 (1).

Anderson, Lorin W, David R. Krathwohl, Peter W. Airasian, Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, dan Merlin C. Wittrock. 2001. A Taxonomy for Learning, Teaching, and Assessning. New York: Pre-Press Company, Inc

Anonim-A. 2013. Permendikbud Nomor 66 Tahun 2013 tentang Standar Penilaian. Jakarta: Departemen Pendidikan dan Kebudayaan.

Anonim-B. 2013. Permendikbud Nomor 81A Tahun 2013 tentang Implementasi Kurikulum. Jakarta: Departemen Pendidikan dan Kebudayaan.

Anonim-C. 2014. Permendikbud Nomor 104 Tahun 2014 tentang Penilaian Hasil Belajar oleh Pendidik pada Pendidikan Dasar dan Pendidikan Menengah. Jakarta: Departemen Pendidikan dan Kebudayaan.

Arikunto, S. 2012. Prosedur Penelitian Suatu Pendekatan Praktik. Jakarta: PT Rineka Cipta

Arikunto, S. & Jabar, A. 2012. Evaluasi Program Pendidikan. Jakarta: Bumi Aksara.

Buck, and Gayle A. 2007. "Learning How Make inquiry into Science Discernible to Middle Level Teachers". Journal for Science Teacher Education.

Dwiyogo, Wasis D. 2016. Pembelajaran berbasis blended learning. Malang: Wineka Media.

Haryati, Mimin. 2013. Model dan Teknik Penilaian pada Tingkat Satuan Pendidikan. Jakarta: Referensi.

Husamah. 2014. Pembelajaran Bauran (Blended Learning). Jakarta: Prestasi Pustaka.

Kartowagiran, B. 2012. Penulisan butir soal. Makalah disampaikan pada Pelatihan penulisan dan analisis butir soal bagi Sumber daya PNS Dik-Rekinpeg, di Hotel Kawanua Aerotel, Jakarta.

Khabibah, S. 2006. Pengembangan Perangkat Pembelajaran Matematika dengan Soal Terbuka untuk Meningkatkan Kreatifitas Siswa Sekolah Dasar, Disertasi, Program Pasca Sarjana Universitas Negeri Surabaya.

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

Koyan, W. 2012. Konstruksi Tes: Singaraja. Universitas Pendidikan Ganesha Press.

Neumeier, P. 2005. A Closer Look at Blended Learning. Parameter of Designing a Blended Learning Environment for Language Teaching and Laerning ReCALL. 17 (2) 163-178.

Nitko, A. J. 1996. Educational assessment of students. Prentice-Hall Order Processing Center, PO Box 11071, Des Moines, IA 50336-1071.

Ruiz-Primo, Maria Araceli & Furtak, Erin Marie. 2007. "Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry". Journal of Research in Science Teaching, 44 (1)

Sanjaya, Wina. 2010. Perencanaan dan Desain Sistem Pembelajaran. Jakarta: Prenada Media Group

Sola, Agboola Omowunmi & Ojo, Oloyede Ezekiel. 2007. "Effects of Project, Inquiry and Lecture-Demonstration Teaching Methods on Senior Secondary Students' Achievement in Separation of Mixtures Practical Test" Academic Journals. 2 (6).

Sugiyono. 2013. Metode Penelitian Kuantitatif, Kualitatif dan R & D. Bandung: Alfabeta

Suhandana, Anggun. 2012. Belajar dan Pembelajaran Berbasis Komputer. Bandung: CV Alfabeta.

Suryabrata, S. 2000. Pengujian Signifikansi Hipotesis Nol dalam penelitian Psikologis. Buletin Psikologi, 8(2). 23-32.

Windschitl, Mark Ronal. 2004. "Inquiry-Based Science in the Middle Grades: Assessment of Learning in Urban Systemic Reform". Journal of Research in Science Teaching, 41 (10).