

The Development of Assessment Instruments Based on Multiple Intelligence in Elementary School

Dodo Septiawan* Lilik Sabdaningtyas Darsono Alben Ambarita
Faculty of Teacher Training and Education, University of Lampung
St. Soemantri Brojonegoro No.1 Gedung meneng Bandar Lampung 35145

Abstract

This study aims to develop theoretically and empirically appropriate multiple intelligences based test instruments to measure students' knowledge in elementary schools. The type of research used is research and development that refers to the theory of Borg & Gall. The population in this study included fourth grade students in elementary schools in central Lampung. The sampling technique uses a purposive technique. The sample of this study was fourth grade students, amounting to 44 elementary schools in Central Lampung. Data was collected through questionnaires and tests. The feasibility of the instrument is theoretically based on the assessment of 3 experts with an average of 84.32 very good categories. While the feasibility of the instrument empirically based on the analysis there are 30 decent and quality MCQs from the results of validity, reliability, distinguishing power, level of difficulty, and distractor test. The results of the study showed that the test instruments developed had met the criteria for writing proper and quality test instruments to measure students' knowledge.

Keywords: Instrument, Multiple Intelligences, Elementary School

DOI: 10.7176/JEP/10-12-18

Publication date: April 30th 2019

1. Introduction

Entering the 21st century competition and challenges in all aspects of life are getting bigger. Life in the 21st century requires a variety of knowledge and skills that must be mastered by someone, so that education can be expected to prepare individuals to master various skills in order to become successful individuals in life. Efforts to realize the goals of national education require learning that emphasizes the processes through which it is implemented, because it will indirectly have an impact on student learning outcomes. The relationship between educational goals, learning activities / processes, and evaluation is so close that it cannot be separated. Evaluation is needed to monitor the extent to which learning activities succeed in their efforts to achieve educational goals.

Assessment in the 2013 curriculum (K13) or better known as authentic assessment has strong relevance to the scientific learning approach in accordance with the demands of K13. The assessment is able to describe an increase in student learning outcomes, in order to observe, reason, try, build networks, etc. Therefore, assessment of learning outcomes must be carried out by starting from the determination of instruments, preparation of instruments, review of instruments, implementation of assessments, analysis of the results of the assessment, and follow-up program on the results of the assessment. Assessment of student learning outcomes in basic education and secondary education includes aspects of attitude (affective), knowledge (cognitive), and skills (psychomotor).

Assessment carried out by the teacher must be able to make each student achieve, and discover the unique potential he has. When educators have understood the plan and the purpose of making question instruments that are in accordance with the indicators in Basic Competence that must be mastered by students, educators will easily make test instruments in measuring the achievement of the learning process.

Based on the results of the PISA analysis reported by the Organization for Economic Cooperation and Development (OECD), Indonesia is only ranked 64th out of 65 countries (OECD, 2012). The results of this study indicate that the average Indonesian student is only able to recognize a number of basic facts but cannot communicate, relate, or even apply some abstract and complex concepts. This is because Indonesian students are less stimulated to improve High Level Thinking Skills (HOTS).

Assessment is the activity of the teacher intended to measure certain competencies or abilities for activities that have been carried out in learning activities (Hosnan, 2014, p. 387). Assessment information is needed to make decisions about students' abilities, based on their abilities and achievements (Kankam Boadu, et al. 2015, p. 62). This study refers to the results of Klenowski's research (2009, p. 51) which shows that the existence of assessment in learning is proven to help develop students' thinking skills. According to Bentri, et. Al (2016, p.1010) assessment generally includes cognitive, affective, and psychomotor aspects. This is so that the purpose of the teacher's assessment is authentic (comprehensive).

Regulation of the Minister of Education and Culture No. 23 of 2016 concerning Educational Assessment Standards states that: Learning activities in the 2013 curriculum are the process of collecting and processing information to measure student achievement including authentic assessment, self-assessment, assessment, repetition, daily examinations, mid-term examinations, final semester exams, examinations competency level, competency level quality test, national exam, and school or madrasah exam.

Based on the results of a needs analysis conducted on 8 grade IV elementary school teachers on September 27, 2018, 75% of the results showed that the assessment instruments used by educators only came from students' books regardless of students' suitability contextual material. The assessment instrument used as an evaluation tool must be prepared by the educator himself and actually make guidelines in the evaluation. The results of other questionnaires show that 75% of educators have difficulty in making assessment instruments. The results of the next questionnaire showed that 87.5% of the instruments did not contain linkages based on competencies and indicators. Grid making is one of the most important things that can be used as a guide before making an assessment instrument.

Most teachers do not know the analysis of items to see validity, reliability, level of difficulty, difference in strength, and troublemakers. Ideally the instruments used can encourage students to think about developing their intelligence, but 87.5% indicate that the instruments used by educators have not referred to this. Smith's research (2014) proves the validity and reliability needed for valid accreditation by higher education than is needed in any field. Clarity about the learning objectives needed requires the development of a valid and reliable protocol. So the first step in a valid and reliable comparison of learning outcomes that is integrated with work is clarity in articulating the objectives of the WIL curriculum learning. Although these components can stimulate students' thinking skills to highlight the plurality of intelligence. Based on the "needs analysis" questionnaire, 100% of teachers agreed to develop instruments that refer to the concept of multiple intelligences to evaluate student learning outcomes. The ability to think is closely related to intelligence. According to Gardner (2011, p. 64), a person's intelligence is not measured by standard psychological tests, but can be seen from their habits two things. First, a person's habits to solve problems. Second, a person's habits to create new products that have cultural values (creativity).

Based on the description above, a teacher must be able to plan learning concepts, strategies, and problem-based construction as instruments that facilitate students to develop their knowledge. According to Mardapi (2012: p.12), improving the quality of education can be achieved through improving the quality of learning and the quality of the assessment system. Both are related to each other. A good learning system will produce good quality learning. The quality of good learning can be seen from the results of the assessment. The solution offered in overcoming the problem above is to make the application of the concept of assessment that can encourage students to give up their ability to answer. According to Ibrahim in Istianah (2013: p.45) to realize learning that can develop thinking skills must depart from learning and assessment that makes students active. Multiple Intelligence is one of the appropriate learning strategies to be developed in terms of assessment, along with the demands of learning in the application of integrated thematic needs in the 21st century. The characteristics of multiple intelligences are also able to encourage and develop students' knowledge and intelligence.

Hosnan (2014: p.387), explains that assessment is the activity of teachers intended to measure certain competencies or abilities of an activity that has been carried out in learning activities. According to Mangiante (2013: p.222), assessment is a tool to measure where students have improved their learning based on standards. Popham (1995: p.3) explains that educational assessment is a formal effort to determine the status of students in relation to interesting educational variables ".

This study is in accordance with the results of a study conducted by Yalmanci, et al. (2013: p.27) concerning the influence of multiple intelligence-based learning theory on student achievement and retention in science. The results showed that multiple intelligence-based learning had a positive and significant effect on students' scientific achievement and retention of enzyme materials compared to using conventional methods. Lunenburg (2014: p.14) about the application of Multiple Intelligences in class is an innovation in teaching planning.

This study discusses how activities can be carried out in implementing Multiple Intelligences in classroom learning. Sugiarti, I Kaniawati and L Aviyanti (2017) Development of Instruments for Critical Thinking Assessment in Middle School Physics. This instrument supports several indicators consisting of argument analysis, deduction thinking, induction thinking, and viewing information in the form of scenarios, texts, graphs and tables. Furthermore, in a study from Kusuma et al. (2017, p.145), Their ability to think, develop or not, depends on how the teacher familiarizes them in the learning process. It is stated that the assessment of HOTS Instruments is an assessment for effective learning to train students HOTS and measure its effectiveness. students' thinking ability according to the level of thinking of each student. The results of the assessment carried out by the teacher, can be trusted if the teacher evaluates the same subject several times, and still obtains relatively the same or reliable results (Arifin, 2014, p. 78). It is appropriate for teachers to solve and have skills in assessing their students' learning outcomes, allowing teachers to make instruments to assess according to certain rules. Assessments arranged according to the rules prepared will produce valid and reliable ones. Thus it will produce data and information about student achievement levels that are valid and accurate.

Therefore, this study aims to study the development of valid and valid multiple intelligences-based instruments in measuring the knowledge of students in grade IV of elementary school. In accordance with the above problem, the purpose of this research and development was to develop an instrument on the theme of thematic learning 8 "the area where I live" sub-theme 1 "the environment in which I live" in grade IV of elementary

school.

2. Methods

This type of research is Research and Development (R & D). The research and development used is the design model Borg & Gall (2003: 569-575) which consists of 10 steps, namely: initial research and information gathering, planning, development of initial product formats, initial trials, product revisions, field trials, revisions products, operational field trials, final product revisions, implementations. This study only made the first step up to the seventh step, namely the preliminary study phase until the field trial stage.

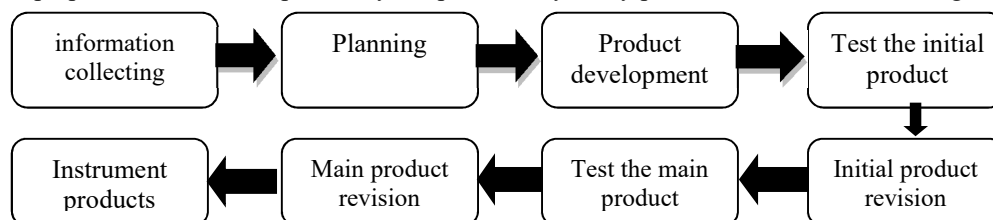


Figure 1. Design Model of Borg and Gall

2.1 Information collecting

The first stage of research is a preliminary study through observation. Observations were made to obtain data and information about the problems that exist in the school population.

2.2 Planning

Conduct planning that includes identification of initial conditions, setting goals, reviewing theory, setting goals, and designing research instruments. At this stage, curriculum analysis is carried out to find out basic competencies in grade 4 in Indonesia according to the research schedule

2.3 Product development

Developing products from research in the form of assessment instruments. The result of the design of this product is a prototype assessment instrument. Basic development of initial products, namely researchers designed multiple intelligence based assessment instruments by referring to the grid that had been compiled. The products produced in this study are multiple intelligences based assessment instruments. Furthermore, the researcher conducted product validation with experts and practitioners to see the feasibility of the instrument theoretically. Validated initial products cover 3 aspects, namely material aspects, evaluation, and language. Validation is carried out by experts or experts. The three experts validated product design, especially in the preparation of evaluations, material, and language in MCQs. The assessment of experts is intended to get an assessment and input in the form of suggestions and criticisms of the assessment instruments made by the researcher. Product validation is calculated using the formula:

$$V_{ah} = \frac{T_{sc}}{T_{sh}} \times 100\%$$

Information:

V_{ah} = Expert Validation

T_{sc} = Total of Empiric score

T_{sh} = Total of maximum score

2.4 Test the initial product

This trial was conducted to determine the level of instrument feasibility empirically. The empirical feasibility analysis of instruments is obtained from the calculated results with the product moment. Reliability testing is done by calculating the reliability coefficient. Selannutnya test the level of difficulty, power difference, and proportion of distractors.

2.5 Initial Product Revision

After conducting an initial trial by revising the main product, based on the suggestion of analyzing the items found some errors or shortcomings, then the next step is to revise the product. These results are used to improve the items to be developed. Items that have been corrected are then made into test instruments in the form of multiple choice questions that will be tested in the main class

2.6 Test the main product

This final prototype was tested on the subject of the study. The instrument trial was carried out at SDN 1 Binakarya

Buana and SDN 1 Binakarya Putra with its address at Desa Binakarya Buana, Rumbia District, Central Lampung Regency. Large group trials were held on March 4-9 in class IV A SDN 1 Binakarya Buana and class IV B SDN 1 Binakarya Putra. Rumbia District, Central Lampung Regency. This stage is carried out to test multiple intelligence based assessment instrument products on aspects of empirical feasibility.

2.7 Revised the main product

Empirical criteria are used to analyze question items. Analysis of question items was used to test the quality of questions tested in large group trials. Quantitative data checks empirical criteria using tests of validity, reliability, level of difficulty, power difference, and distractors. The instrument is said to be valid if it gives results that are static or consistent after being tested

3. Results and Discussion

This research is a development research that begins with conducting preliminary research on test instruments used by students in the elementary school in Rumbia Subdistrict, Central Lampung Regency, Lampung. Based on the results of the study, it was found that the teacher did not understand the components of the concept of intelligence possessed by students, so that the assessment of cognitive aspects was less attention. Assessment is the activity of the teacher intended to measure certain competencies or abilities for activities that have been carried out in learning activities (Hosnan, 2014, p.387). Assessment information is needed to make decisions about students' learning abilities, based on their abilities and achievements (Kankam Boadu, et al., 2015, p.62). However, as long as this test instrument is made by the teacher it has not referred to the level of multiple intelligences that students have by paying attention to the quality of the material.

The assessment instrument developed has gone through several stages, starting from fulfilling the rules of writing questions, theoretical validation and empirical validation. Based on the results of the field test (field test), the quality of the test instruments developed that met the standards of feasibility was obtained. The product of the test instrument developed was used to measure students' knowledge. Based on the preparation of the test instruments carried out in accordance with the steps of the preparation of the test instruments proposed by Hendriana and Soemarno (2014: 84-85) which states that the preparation of a test should start from the preparation of the grid to the analysis of items to see the validity of the questions to MCQs. Then carry out follow-up in the form of revision of the test items that are still possible, or replace the test items that are not good / invalid with the new test items in order to get a decent and quality test instrument. The following is a discussion of the feasibility and quality of the test instruments that have been developed.

3.1 Validation of Assessment Instruments

Validation is done to find out the content validity of the instrument using expert judgment. Aspects assessed include 3 things, namely: (1) material; (2) evaluation; and (3) language. The test is carried out by submitting a set of assessment instruments consisting of grids, question cards, instrument forms, answer keys and assessment guidelines. Then experts are asked to assess the suitability of the instrument with the indicators on the questionnaire sheet. Based on expert reviews, the instruments developed were declared eligible for the trial after revision. The results of the construct validity test obtained from expert judgment can be seen in Table 1. Based on the results shown in Table 1 it can be concluded that the assessment instruments developed include categorical validations.

Tabel 1. Expert Test Results

No	Aspect	Validator 1	Validator 2	Average	Description
1	Evaluation	81,25	90,62	85,94	Very good
2	Language	80,35	83,93	82,14	Very good
3	Material	73,68	96,05	84,90	Very good
Validation of Construct				84,32	Very good

Based on the results of expert judgment and practitioners, the average expert validation score obtained was 84.90% (very good), the evaluation aspect of the evaluation score was 85.94% (very good), and the score aspect validation score was 82.14% (very good). These results indicate that the instrument has very good validity in measuring student knowledge. However, there are several questions that need to be corrected according to the suggestions and suggestions of 5 validators, such as; indicators that are not in accordance with the item questions, writing sentence formulas, selecting images, completing the introductory information (stimulus), and answer choices. Thus, the instrument is declared feasible to be tested after the revision.

3.2. Quality of Instrument Assessment In the initial product test

After doing the validation of the experts, then proceed with a trial to see the quality of items developed questions including reliability, distinguishing power, level of difficulty, and interference. Limited testing was conducted at Restu Buana 3 Elementary School with a sample of 18 people. The number of initial assessment instruments developed was 40 questions. However, after the validity test, only 30 questions are valid and according to the

criteria. The results of these tests include.

3.2.1 Instrument Validity

Table 2. Instrumen Validity

Item's Number	Total	criteria	Percentage (%)
1,2,3,5,6,7,8,10,11,12,13,14,16,18,19,20,21,22,24,25,27,28,29,31,32,33,34,36,37,38,40	31	Valid (r Calculate > r table)	77,5 %
4,9,15,17,23,26,30,35,39	9	Invalid (r calculate < r table)	22,5 %

In testing the validity of this instrument using the Product Moment Correlation formula. The question is said to be valid if r count is greater than r table. The value of the R table in this group of questions is 0.468. Questions that are valid are 31 questions (77.5%) while the questions that are invalid are 9 questions (22.5%).

3.2.2. Instrument Reliability

The purpose of developing this research is not only to produce a valid instrument, but also to produce a reliable instrument. Reliability test results are presented in table 3.

Table 3. Instrument Reliability

Item's Number	Value r Calculate	Criteria
1-40	0,895	Very strong

Based on the table it can be seen that the reliability level of the question is in the very strong category with the number of r calculated obtained is 0.895.

3.2.3 Difficulty level

The next step of analysis is to determine the level of difficulty for each question item. To find out the difficulty level of the instruments developed, can be seen in table 4.

Table 4 Difficulty Level

Category	Item's Number	Total	Percentage (%)
0,00-0,30 (Hard)	1,7,9,11,13,17,22,35,37	9	22,5 %
0,31-0,70 (medium)	2,3,4,5,6,8,10,14,15,16,18,20,21,23,24,26,29,30,34,36,38, 39, 40	23	57,5 %
>0,71 (Easy)	12,19 ,25,27,28,31,32,33	8	20 %

Based on the table it can be concluded that 20% of the test questions were developed in the easy category, 57.5% in the moderate category, and 22.5% in the difficult category.

3.2.4 Distinguishing Power

The next step of the analysis is to determine the distinguishing strength for each question item. The following reliability test results are presented in table 5.

Table 5. Distinguishing Power

Distinguishing Power Index	Item's Number	Annotation
0,71 – 1,00		Very good
0,41 – 0,70	2,5,6,14,15,18,20,21,29,34,39,40	Good
0,21 – 0,40	1,3,8,9,10,11,12,13,16,19,22,23,24,25,26,27,28,31,32,33,36,37,38	Pretty good
0,00 – 0,20	7,17,30	Bad
Negative	4,35	Not good

Based on the table above, it can be concluded that the test questions developed have different power with good categories 12 items, quite good 23 items, not good 3 items, and 2 items in the very bad category (discarded / replaced).

3.2.5 Distracting Power (Distractor)

The distribution of choice / answer choices (distractors) from multiple choice questions items can be seen in table 6 below.

Table 6. Distractor

Category	Item's Number	Total	Percentage (%)
rpbis positive answer key, Response >5% , and rpbis negative distractor	1,2,3,4,5,7,8,9,10,11,12,13,14,15,17,20,22,23,25,27,29,31,32,35,37,40	26	65 %
rpbis negative answer key, Response < 5%, and rpbis positive distractor	6,16,18,19,21,24,26,28,30,33,34,36,38,39	14	35 %

Based on the table, it is known that the problem with the spread of answer choices / options generally works well even though there are some deception that must be replaced and revised.

Based on the analysis of multiple choice test instruments and the description above, a decision was made that only 30 questions were tested at the field test stage, namely questions number 1,2,3,5,6, 8, 10, 11, 12, 13, 14, 16, 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, 31, 32, 33, 34, 36, 37, 38, and 40. The prototype produced in the small group stage is then called Prototype III.

3.3. Quality of Assessment Instruments in Field Test

The final prototype was tested on research subjects in 4th grade students of Binakarya Buana 1 Elementary School and 1 Binakarya Putra Public Elementary School, as many as 44 students. At this stage, the analysis of questions included validity, reliability, difficulties, different powers, and distractors. The item analysis for the multiple choice form also looks as follows.

3.3.1 Validity

Table 7. Instrument Validity

Item's Number	Total	Criteria	Percentage %
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28, 29,30	30	Valid (r Calculate > r table)	100 %
-		Invalid (r Calculate < r table)	-

In testing the validity of this instrument using the Product Moment Correlation formula. Item questions are said to be valid if r count is greater / equal to r table. The value of the R table in this group of questions is 0.297. Questions that are valid are 30 questions (100%).

3.3.2 Reliability

Table 8. Instrument Reliability Test Results

Item's Number	Value r Calculate	Criteria
1-30	0,861	Very strong

Based on the results of the reliability test items in the instrument as many as 30 questions that were tested obtained a value of r count 0.861 with very strong criteria.

3.3.3 Difficulty level

Table 9. Test the level of difficulty of the question

Category	Item's Number	Total	Percentage (%)
0,00-0,30 (Hard)	1,3,6,9,11,17,22,24,28,	9	30 %
0,31-0,70 (medium)	2,5,7,10,12,13,15,16,18,20,21,23,26,27,30	15	50 %
>0,71 (Easy)	4,8,14,19,25,29	6	20 %

3.3.4 Distinguishing Power

Table 10 Distinguishing Power of Multiple Choices Questions Product

Distinguishing Power Index	Item's Number	Annotation
0,71 – 1,00		Very good
0,41 – 0,70	1,3,15,20,23,28	Good
0,21 – 0,40	2,4,5,6,7,8,9,10,11,12,13,14, 16,17,18,19,21,22,24,25,26, 27,29,30	Pretty good
0,00 – 0,20		Bad
Negative		Not good

3.3.5 Distracting Power (Distractor)

Table 11 Distractor

Category	Item's Number	Total	Percentage (%)
rpbis positive answer key, Response >5%, and rpbis negative distractor	1-30	30	100
rpbis negative answer key, Response < 5%, and rpbis positive distractor	-	-	-

Based on the table above, it can be seen that the items with the spread of the answer choices / options that work well are 30 items (100%).

Based on the analysis of multiple choice test instruments and descriptions above, then the multiple choice quality questions are included in the question bank as many as 30 questions, namely item number 1-30 because all questions meet the criteria of validity, reliability, level of difficulty, Distinguishing Power and distractor after through field trials.

3.4 Revision of Main Products

The final product revision was conducted based on the results of field trial and findings when the product was tested. Based on the results of this theoretical and empirical validation, the developed instrument product was not revised and feasible to be implemented.

4. Conclusion

Based on the results and discussion, it was concluded that the final product in this study produced a theoretically and empirically feasible assessment instrument to measure the knowledge of 4th grade students in Elementary School especially on theme 8 sub-themes 1. Empirical feasibility was evidenced from the assessment of 3 experts namely evaluation experts, experts language, and material experts who stated that the test instruments developed in the category were very good. In addition, this assessment instrument fulfills empirically feasible criteria as evidenced by the results of the analysis of test instruments in small and large group trials. Of the 40 multiple choice questions tested, 30 quality questions were obtained to measure students' knowledge. Question items meet the level of validity and reliability because when tested on several samples the same results are obtained in the strong category. Then at the level of difficulty in the medium category, the power of the power is different in the good category, and all of the answers to the answers function properly. Thus, the use of multiple intelligences based assessment instruments is an alternative for teachers to measure and train the level of intelligence of students based on the ability to answer the questions they are working on.

References

- Arifin, Zaenal. (2014). Pengembangan Instrumen Pengukur Berpikir Kritis Matematika Peserta didik SMA Kelas X. *jurnal THEOREMS (The Origibal Research Mathematic)*.
- Bentri, Alwen., Abna Hidayati, Ulfia Rahmi (2016). The Problem Analysis in Applying Instrument of Authentic Assessment in 2013 Curriculum dalam *International Journal of Science and Research (IJSR) 10 Oktober 2016 (Volume 5, P.1008-1012)*.
- Brog, Walter R., and Gall, Meredith Damien. (1989). *Education Research*. New York: Longman
- Gardner, Howard. (2011). *Frames of Mind; The Theory of Multiple Intelligence* (e-book). Perseus Book Group. New York.
- Hendriana, H dan Soemarmo, U. 2014. *Penilaian Pembelajaran Matematika*. Refika Aditama. Bandung.
- Hosnan. 2014. *Pendekatan Sainifik dan Kontekstual dalam Pembelajaran Abad 21: Kunci Sukses Kurikulum 2013*. Jakarta: Ghalia Indonesia.
- Istianah, Euis. 2013. Meningkatkan Kemampuan Berpikir Kritis dan Kreatif Matematik dengan pendekatan Mode Eleciting Activities (MEAs) Pada Siswa SMA. *Jurnal Ilmiah Infinity Program Studi Matematika STKIP Siliwangi Bandung*. Vol: 2, No: 1

- Kankam Boadu, Anthony Bordoh, Isaac Eshun, Thoephilus Kweku Bassaw, Fredrick Yaw Korang. (2015). Teachers' perception of authentic assessment techniques practice in social studies lessons in senior high schools in Ghana. *International Journal of Educational Research and Information Science*. 10 Januari 2015. (Volume 1 No 4 pp 62-68).
- Klenowski, V. (2009). Assesment for Learning revisited: an Asia-Fasific perspective. *Assesment in Education: Principles, Policy, Practice*. Vol. 16, No 3, November 2009, pp. 263-268.
- Kusuma, Merta Dhewa., Rosidin, Undang., Abdurrahman, and Agus Suyatna. 2017. about The Development of Higher Order Thinking Skill (Hots) Instrument Assessment In Physics Study inside *IOSR Journal of Research & Method in Education (IOSR-JRME) Volume 7, Issue 1 Ver. V (Jan. - Feb. 2017)*, pp. 26-32.
- Lunenburg, Fred C. 2014. *Applying Multiple Intelligences in the Classroom: A Fresh Look at Teaching Writing*. International Journal of Scholarly Academic Intellectual Diversity.
- 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: P. 222-258.
- Mardapi, D. (2012). *Pengukuran Penilaian Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- OECD. (2012). PISA: Science competencies for tomorrow world volume 1: Analysis. Rosewood. Drive: OECD
- 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.
- Smith, Calvin. 2014. Assessment of Student Outcomes from Work-Integrated Learning: Validity and Reliability. *Asia-Pasific Journal of Cooperative Education*, (volume 6 P. 210-223).
- Sugiarti, T., I Kaniawati and L Aviyanti. 2017. Development of Assessment Instrument of Critical Thinking in Physics at Senior High School dalam *Journal of Physics: Conference Series*. (Volume 9, No. 1, P.20-33).
- Yalmanci, Gurbuzoğlu, et al. 2013. The Effects of Multiple Intelligence Theory Based Teaching on Students' Achievement and Retention of Knowledge. *International Journal on New Trends in Education and Their Implications*, 4(3), P. 27-36.