

# DEVELOPING INSTRUMENTS FOR EVALUATING VALIDITY, PRACTICALITY, AND EFFECTIVENESS OF THE AUTHENTIC ASSESSMENT FOR SPEAKING SKILL AT JUNIOR HIGH SCHOOL

Jean Susan Kadir<sup>1</sup>, M. Zaim<sup>2</sup>, and Refnaldi<sup>3</sup>

<sup>1</sup> FBS Universitas Negeri Padang, Padang, Indonesia, ✉ [jeanenglish37@gmail.com](mailto:jeanenglish37@gmail.com)

<sup>2</sup> FBS Universitas Negeri Padang, Padang, Indonesia, ✉ [mzaim\\_unp@yahoo.com](mailto:mzaim_unp@yahoo.com)

<sup>3</sup> FBS Universitas Negeri Padang, Padang, Indonesia, ✉ [refnaldi@yahoo.com](mailto:refnaldi@yahoo.com)

## Abstract

An authentic assessment can provide a real-life experience to the student and accommodate the need of assessing productive skills, particularly for speaking skill. The authentic assessment instrument has to fulfill the requirements of being valid, practical and effective to make it a good device. This research focuses on developing tools for evaluating validity, practicality, and effectiveness of authentic assessment for speaking skill at junior high school. The instruments were produced by using theoretical based approach and were validated by experts in language assessment and English teachers of the junior high school. The developing tools were in the form of the questionnaire. There were five indicators for validity consisting of 2 to 7 sub-indicators, 4 indicators for practicality comprised 3 to 4 sub-indicators, and 6 indicators for effectiveness composed of 2 to 8 sub-indicators. The instruments used a 5–point evaluation scale.

**Key words:** *developing instruments, validity, practicality, effectiveness, authentic assessments,*

---

## Introduction

The teaching and learning process based on the curriculum 13 in Indonesia expects students to be productive, creative, innovative and affective. The curriculum is designed to focus on developing attitude (spiritual and social), knowledge, and skills of the students. In other words, it reinforces authenticity in learning because its priority to develop local wisdom, encourage self-development activities and mastery of learning and life skills (Retnawati, Hadi & Nugraha, 2016:34). Nevertheless, a traditional assessment which mostly is in form of multiple choices is not able to assess the development of student's attitude, knowledge and life skill. Therefore, the need of using authentic assessment rises up as it can come in the multiple forms of assessment that are able to reflect students learning, achievement, motivation, attitudes on instructionally-relevance classroom activities( O'Malley & Valdes-Peirce , 1996: 4). In other words, it assesses what students have learned and practice during the teaching and learning process throughout the course.

Authentic assessment also offers a meaningful and productive way of assessing students (Gronlund and Cameron, 2004:10 in Case, 2013: 319). In addition, it offers students to be assessed of being in a real-life circumstance that can be related to either their background or their future career as it can comes in many forms. For instance, to assess speaking skill can be assessed in a form of *Oral Interview, Story of Text Retelling, Picture cued description or stories, Radio Broadcast* and so on O'Malley and Pierce (1996:77).

However, it may have a problem with the cost, time, and limitation of skill and knowledge that need to be assessed while a good assessment must be valid, practical and effective. Evaluating the validity, practicality and effectiveness of the authentic assessment is crucial concerning its drawbacks which might be related to those before it is actually used to measure how far the attitude, knowledge and skill of students have changed along or after the course.

An instrument must be used to check its validity, practicality and effectiveness. Developing the instrument before actually evaluating the authentic assessment is essential to make sure that the evaluation can outcome a fine authentic assessment which is actually measure what need to be measured without any time constraint, inaccuracy or fund issue. Therefore, this research is aimed to develop an instrument to evaluate the validity, practicality and effectiveness of the authentic speaking assessment for junior high school. The research question in this research is formulated as follows:

How to develop an instrument for evaluating the validity, practicality and effectiveness of the authentic assessment for speaking skill at junior high school?

## **Theoretical background**

### **Validity**

If a test is measured what needs to be measured from the students, then it is a valid test. For instance, we want to measure the writing ability of students, we might ask students to write as many words as they can in 15 minutes then we simply count the words for the final score and also consider the comprehensibility, rhetorical discourse elements and the organizations of ideas (Brown & Abeywickrama, 2010:30). Validity can be categorized based on the strategies which have been used to investigate validity itself. There are content validity, construct validity, criterion validity, consequential validity and face validity.

Content validity is a conclusion and measured performance based on the subject matter that is being tested (Bachman, 1990: 22, Brown, 2010:30). In other words, investigating content validity of a test is examined whether the test is a representative sample of the content which the test was designed to measure.

Reaching the criteria over the test means that you have a valid test in a form of criterion-related evidence. Correlation of the scores being validated with some other, well-respected measure of the same construct is also regarded as the concept of criterion validity (Brown, 2009:233). In other words, criterion validity is about how performance on the assessment accurately predicts future performance or estimate present performance on some other valued measure.

Construct validity as Bachman (1990:26) mentioned can be concluded exists in a test if it is demonstrated that it measures just the ability which it is supposed to measure. Accuracy in measuring the intended criteria impact on the test takers preparation, effect on the learner, social consequences of the test interpretation and use is being considered as the consequential validity (Brown & Abeywickrama (2010: 34) .

Face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnels who decide on its use, and other psychometrically unsophisticated observers (Mousavi, 2009:247 as cited in Brown & Abeywickrama, 2010:35).

All types of validity actually can be explained that a valid test is a test that its content include all the things that students have learned, practiced and trained in the class.

### **Practicality**

An expensive, time –consuming test, vague procedure and administration of test is impractical. In other words, a practical test is a test that considers time, budget, resources, and administration issues such as how to evaluate the students' work. It is in line with what Brown & Abeywickrama (2010: 26) states that a test can be practical when it stays within budgetary limits, it can be completed by the test taker within appropriate time constraints, the test also should has clear distinction for administration, the test should appropriately utilizes available human resources, the test does not exceed available material resources and also consider the time and effort involved for both design and scoring.

Being practical in assessment also means that if the extent to which the demands of the particular test specification can be met within the limits of time and existing human and material resources, it can be concluded as practical assessment (Bachman & Palmer, 2009). Similarly, Harris (1969 ) also points out the economy and the ease of administration and scoring and it means that a consideration of cost per copy, how many administrators or scorers need the budget along the process( p.21-22).

All these theories can be illustrated in several examples when it becomes impractical. A test of language proficiency that takes a student five hours to complete is impractical-it consumes more time (and money) than necessary to accomplish its objective (Brown & Abeywickrama, 2010: 26). A test that requires individual one-on-one proctoring is impractical for a group of several hundred test-takers and only a handful of examiners (ibid). Another example in defining a practicality in assessment is while longer tests can increase validity because they capture more measurement data they may be impractical to administer. In addition, an overly long exam could induce fatigue in candidates, which in turn could introduce errors into the measurements (Cambridge, 2010:30). In addition, in making sure that a test is practical, administrative details such as printed material or information about time duration and how to do it, reasonable time in doing it, technical issues ( equipment), scoring system and how to report it should be overlooked before doing the test (Brown & Abeywickrana, 2010: 40) .

Practicality is an integral part of the concept of test usefulness and affects many different aspects of an examination. It can be defined as the extent to which an examination is practicable in terms of the resources necessary to produce and administer it in its intended context and use. A practical examination is one that does not place an unreasonable demand on available resources. Therefore, analyzing the practicality of an

authentic assessment will overlook its budget, time of designing, implementing and scoring the assessment itself, administration issues, and material resources.

### **Effectiveness**

Effectiveness is another important consideration in analyzing the authentic assessment. Braskamp & Engberg (2014) states that an effective assessment should have these criteria; having a clear purpose and readiness for assessment, involving stakeholders throughout the assessment process, teacher have to know about what they are going to assess and the way they will assess.

Furthermore, an effective assessment is an assessment which considers the effect of the assessment on students' learning behaviour and outcome, aligns to the learning outcomes and curriculum, provides and explains the structure of the assessment for their course, and provides feedback afterward (Swanson, 2002: 3). The authentic assessment is effective when it has a clear conception of all intended learning outcome, a variety of assessment procedures be used, the instructional relevance of the procedures be considered, an adequate sample of student performance, the procedures be fair to everyone, the specification of criteria for judging, feedback to students that emphasizes strengths of performance and weakness to be corrected a comprehensive grading and reporting system (Gronlund & Waugh, 2009:19-23):

What Gronlund & Waugh (2009) means by having a clear conception of all intended learning outcomes in an effective assessment is that teachers should know what types of knowledge, understanding, application and performance skills to determined whether the learning and teaching process has been conducted successfully. Teachers need to categorize all intended learning outcomes in terms of student performance and plan of how to assess them all.

Since the intended learning outcomes might be various, the assessment procedure also shares the same pattern. Therefore, an effective assessment also provides a various assessment procedure. For example, having an assessment of performance skill will require teachers to depend mainly in judgment, rating scales or checklists.

On the other hand, the requirement of the instructional relevant assessment means that the intended outcomes of instruction, the domain of leaning task, and the assessment procedures are in closely related.

An adequate sample is also needed to fulfill the criteria of effective assessment. It means that since the learning and teaching may include knowledge and skills too many to be presented in a single test due to the time limitation and other constrains, an assessment only can measure some knowledge and skill as Gronlund and Waugh (2009:22) illustrates that in preparing a classroom test, for example, there may be 100 terms that the students should know but we only have room for 20 terms in our test, thus, we must select a representative sample from the 100 words because we want to be able to generalize from performance on the 20 terms as to hi well students know the 100 terms of the sample is adequate can be estimated that 18 correct answers on the 20 terms indicates that a student knows about 90 percent of the 100 terms( 19-21).

Another consideration of an effective assessment is the procedures be fair to everyone. It can be said that fairness means an overlook when preparing and using assessment procedures, clearly communicating our intentions to students and using the results to improve learning. In other words, including racial, gender stereotypes, having ambiguous direction, or inappropriate level of material might in the assessment material may affect the result of the test from certain test takers.

In addition, criteria for judging the successful performance are also essential in a way that students need to be informed what is expected of them and clearly specified criteria of successful performance can be used to clarify the learning tasks.

Finally, feedback and a comprehensive grading and reporting system should be provided in order to reflect what students have done along the course and it is hope can improve their performance, skill and knowledge. For grading, it is recommended to use both letter grades and more elaborate report which reflect and support the assessment procedures. Students also must be informed before the course and they are provided with periodic feedback relate to their learning progress.

### **Methodology**

The development of the instrument was based on theoretical based approach of validity, practicality and effectiveness of an assessment. After conducting the literature review, it was decided that the instrument is in form of questionnaire due to its capacity in gathering data about abstract ideas or concepts that are otherwise difficult to quantify such as opinion, attitudes and belief (Artino et al, 2014:463). The indicators were developed from the theory of validity, practicality and effectiveness. The indicators were developed to sub-indicator to be more specific and precise to evaluate the authentic speaking assesment. Afterward, they were validated by experts in language assessment and English teachers of the junior high school. In validating the questionnaire, a Likert scale (table 1) was used (Riduwan, 2004:87).

**Table 1. The Likert Scale Options**

Criteria	Score
Very bad	1
Bad	2
Fair	3
Good	4
Very good	5

## Findings and Discussion

After doing a library research, it was concluded that for the validity, only content, construct and face validity theory were developed in the questionnaire. Criterion validity is not suitable for evaluating authentic assessment since it focuses to evaluate the standardized test such as a placement test (Brown, 2009:235) . For the consequential validity, it was excluded since the research did not focus on evaluating the effect of the authentic assessment as shown in table 2.

**Table 2. Questionnaire of Validity**

No.	Indicators	Sub-indicators
<b>Content Validity</b>		
1.	Learning objectives	a. The authentic speaking assessment is represented the learning objectives in general. b. The authentic speaking assessment is represented the learning objectives in specific.
2.	Basic competence	a. The authentic speaking assessment is assessed the basic competences needed to mastered. b. The authentic speaking assessment is assessed the language functions needed to mastered. c. The authentic speaking assessment is assessed the learning topics needed to mastered. d. The authentic speaking assessment is assessed genre-based texts needed to be mastered.
<b>Construct Validity</b>		
3.	Speaking ability	a. The authentic speaking assessment is assessed students' speaking ability in general. b. The authentic speaking assessment is assessed students' fluency to express ideas in oral. c. The authentic speaking assessment is assessed students' accuracy in the use of appropriate words in expressing ideas in oral. d. The authentic speaking assessment is assessed students' accuracy in the use of appropriate grammar in expressing ideas in oral. e. The authentic speaking assessment is assessed students' accuracy in tone and pronunciation while expressing ideas in oral. f. The authentic speaking assessment is assessed students' speaking ability in having monologue. g. The authentic speaking assessment is assessed students' speaking ability in having dialogue.
<b>Face Validity</b>		
4.	Types of speaking assessment	a. The authentic speaking assessment is required students to express ideas in monolog. b. The authentic speaking assessment is required students to interact with others. c. Activities or tasks given in the authentic speaking assessment can be generally understood by students. d. Activities or tasks given in the authentic speaking assessment have already done by students in the teaching and learning process. e. Activities and tasks provided are varied in assessing speaking ability.
5.	Test procedure	a. Instruction in the authentic speaking assessment is given comprehensibly of how to do the test. b. Information of test duration is provided in the authentic speaking assessment.

**Table 2. Cont**

	c. The test duration given in doing the activity or task is reasonable.
	d. The level of difficulty in doing the test is reasonable.

There were two indicators for the content validity. The two content validity indicators focuses on evaluating learning objectives and basic competences since content validity is aimed on measuring whether the test actually represents all the items of the subject matter (Kimberlien & Winstertein,2008:2276). The learning objectives consists of two sub-indicators of general and specific of the learning objective while the basic competences were divided into four of basic competence, language function, learning topics and genre-based texts needed to be mastered.

On the other hand, measuring the ability that should be acquired by students along the course is related to construct validity. Therefore, the indicators of construct validity is only for speaking ability which was composed of seven sub-indicators which were developed based on theory of speaking ability of being fluency and accuracy. The last part of the validity is face validity. Face validity measures the test look right in the eyes of examinees, administrative personnel and observers. In this research, the development of the face validity indicators was developed from point of view of administrative personnel who were teachers. There were two indicators which were types of speaking assessment and test procedure. Types of speaking assessment covers five sub-indicators which focuses whether the task and activities given in the class is included in the authentic assessment. The test procedure copes four sub-indicators of instruction, information of test duration, the duration of the test, the level of difficulty of the test. In other words, the validity consists of five indicators of two to seven sub-indicators as it can be seen in table one.

Time, media, test procedure and scoring were the four indicators of practicality (Abeywickrama ,2010: 26). Therefore, the questionnaire consists of four indicators for practicality comprised of three to four sub-indicators (Table 3).

**Table 3. Questionnaire of practicality**

No	Indicators	Sub-indicators
1.	Time	a. The practical level of the authentic speaking assessment related to time duration in conducting the test.
		b. The practical level of the authentic speaking assessment related to time given to students in doing the test.
2.	Media	a. The practical level of the authentic speaking assessment related to the media needed in conducting it.
		b. The practical level of understanding media used in the authentic speaking assessment.
3.	Test procedure	a. The practical level of the authentic speaking assessment related to understand the procedure of conducting it in the class.
		b. The practical level of the authentic speaking assessment related to the procedure of conducting it in the class.
		c. The practical level of understanding instruction given in doing the authentic speaking assessment by students.
		d. The practical level of the authentic speaking assessment related to the procedure of having students doing the test.
4.	Scoring	a. The practical level of the authentic speaking assessment related to scoring instrument.
		b. The practical level of the authentic speaking assessment related to scoring.
		c. The practical level of the authentic speaking assessment related to giving feedback.

Both time and media have two sub-indicators which focuses on how practical the authentic assessment to be conducted, done and understood either by teachers or students. The indicators of test procedure in the practicality theory are quite different with the validity which focused on the availability of information of conducting the test. Test procedure in the practicality is covered of the actual execution of the test. There were four sub-indicator were expanded from the indicators. The last one is scoring. Since the authentic speaking assessment is in multiple forms, the scoring system is also different for each. That is to say, the scoring systems, scoring and feedback were evaluating specifically in the scoring indicator.

There are six indicators for effectiveness composed of 2 to 8 sub-indicators as shown in table 4. Indicators of effectiveness based on the theories have similarities with the indicators of validity and practicality.

**Table 4. Questionnaire of Effectiveness**

No	Indicators	Sub-indicators
1.	Learning objectives	a. The effective level of the authentic speaking assessment with the learning objectives in general.
		b. The effective level of the authentic speaking assessment with the learning objectives in specific.
2.	Basic competence	a. The effective level of the authentic speaking assessment in assessing the basic competences needed to be mastered.
		b. The effective level of the authentic speaking assessment in assessing the language functions needed to mastered.
		c. The effective level of the authentic speaking assessment in assessing the learning topics needed to mastered.
		d. The effective level of the authentic speaking assessment in assessing genre-based texts needed to be mastered.
3.	Speaking ability	a. The effective level of the authentic speaking assessment in assessing speaking ability.
		b. The effective level of the authentic speaking assessment in assessing students to express their ideas in monolog.
		c. The effective level of the authentic speaking assessment in assessing students to interact with others.
		d. The effective level of the authentic speaking assessment in assessing students' fluency in expressing ideas in oral.
		e. The effective level of the authentic speaking assessment in assessing accuracy in choice of words to express ideas in oral.
		f. The effective level of the authentic speaking assessment in assessing accuracy in grammar to express ideas in oral.
		g. The effective level of the authentic speaking assessment in assessing accuracy in pronouncing words to express ideas in oral.
		h. The effective level of the authentic speaking assessment in assessing students' speaking ability in monologue.
		i. The effective level of the authentic speaking assessment in assessing students' speaking ability in dialogue.
4.	Types of speaking assessment	a. The effective level of the authentic assessment in including speaking activities or tasks to express ideas in monolog.
		b. The effective level of the authentic assessment in including speaking activities or tasks to express ideas in dialogue.
		c. The effective level of the authentic assessment in including speaking activities or tasks which are practiced in the class.
		d. The effective level of the authentic assessment in providing variety of speaking assessment.
5.	Test Procedure	a. The effective level of the authentic speaking assessment in providing a comprehensible instruction to do the test.
		b. The effective level of the authentic speaking assessment in time duration for conducting it.
		c. The effective level of the procedure in conducting the authentic speaking assessment of role play.
		d. The effective level of the procedure in conducting the authentic speaking assessment of pair dialogue.
		e. The effective level of the procedure in conducting the authentic speaking assessment of brief question and answer.
		f. The effective level of the procedure in conducting the authentic speaking assessment of picture cues description.

**Table 4. Cont**

		g. The effective level of the procedure in conducting the authentic speaking assessment of story of text retelling.
		h. The effective level of the procedure in conducting the authentic speaking assessment of oral report.
6.	Scoring	a. The effective level of scoring instrument for role play in evaluating students' speaking ability.
		b. The effective level of scoring instrument for pair dialogue in evaluating students' speaking ability.
		c. The effective level of scoring instrument for brief question and answer in evaluating students' speaking ability.
		d. The effective level of scoring instrument for picture cues description in evaluating students' speaking ability.
		e. The effective level of scoring instrument for story of text retelling in evaluating students' speaking ability.
		f. The effective level of scoring instrument for story of oral report in evaluating students' speaking ability.
		g. The effective level of scoring the test result.
		h. The effective level of giving feedback to students

The indicator number 1 to number 4 shares the same idea as the indicator of validity. The difference is that in the effectiveness it more focuses on how well the assessment can assess what should be assessed not to how well the assessment covers the material, activity or practice in the class as Garson (2013:8) states that a valid test measures what actually to be measured. For indicator number five, test procedure, unlike the sub-indicators in validity focusing on the availability of information of conducting the test and in practicality focusing on how smooth the execution of test, it merely focuses to see how well can the assessment assess students with the time and instruction given. It also specifically evaluates each form of the assessment that includes in the authentic assessment. For the scoring, it also shares differences with one of the indicators in the practicality. The scoring in the effectiveness indicator is more specific to see how each form of the scoring from different types of assessment work effectively to measure students' ability in oral communication since an effective assessment should elaborate report which reflect and support the assessment procedure (Gronlund and Waugh,2009:22) .

## Conclusion

All of the questionnaires have been carefully designed based on the related theories in order to guarantee that the authentic speaking assessments used in the school will be closely evaluated before it is actually used.

The examination also has been done by the experts of language assessment. Their validation has concluded that these questionnaires are valid to be used and the result can be reliable to judge whether the validity, practicality and effectiveness of the authentic speaking assessment have been fulfilled. The English teachers of Junior High School were also involved in judging the validity of the questionnaire. They are as the practitioners who will use the authentic assessment need to be involved in order to make sure the theories and practices are included in the questionnaire to represent the valid data of the validity, practicality and effectiveness of the authentic assessment for speaking skill at junior high school.

## References

- Artino et all. 2014. *Developing Questionnaire for Educational Research: AMEE Guide No. 87*. Medical Teacher.
- Bachman, Lyle F. 1990. *Fundamental Consideration in Language Testing*. Oxford: Oxford University press.
- Bachman, Lyly F & Palmer, A.S. 2009. *Language Testing in Practice: Designing and Devloping useful Language tests*. New York. Oxford University Press.
- Bernhardt, E.B et all. 2004. *The Practicality and Efficiency of Web-Based Placement Testing for College-Level Language Program*. Foreign language Annal.vol.37.No.3
- Braskamp, L.A & Engberg M.E. 2014. *Guidelines for Judging the Effectiveness of Assessing Student Learning*. Chicago. Loyola University.
- Brown, H.Douglas and Priyanda Abeywickrama.2010. *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education.
- Buzzeto-More, Nicole. 2010. *Assessing the Efficacy and Effectiveness of an E-Portfolio Used for Summative Assessment*. Interdisciplinary Journal of E-Learning and Learning Objects. Volume 6.

- Brown, H.Douglas, 1994. *Teaching Action by Principles. An Integrated Approach to Language Pedagogy (3<sup>rd</sup> Ed)*. New Jersey: Prentice Hall Regents.
- Copeland, H.L & Hawson, M.G. 2000. *Developing and Testing an Instrument to Measure the Effectiveness of Clinical Teaching in an Academic Medical Center*. Academic Medicine- Vol 75. No.2/February 2000.
- Case. R. 2013. *Four Principles of Authentic Assessment*. Blogs.ubc.ca/file. Retrieved February 13<sup>th</sup> 2018.
- Cambridge English. 2016. *Principles of Good Practice*. Cambridge University.
- Garson, David. 2013. *Validity and Reliability*. Asheboro, NC 27205 USAG. David Garson and Statistical Associates Publishing.
- Gay, I. R. and P. Airasian. 2009. *Educational Research: Competencies for Analysis and Applications*. New Jersey: Pearson Education.Inc.
- Genesee, Fred and Upsur, John A.2002. *Classroom-based Evaluation in Second Language Education*. Cambridge: Cambridge University Press.
- Gronlund, E.N & Waugh. K.C (2009). *Assessment of Student Achievement* Upper Saddle River, New Jersey. Pearson Education, Inc.
- Harris, David, p. *Testing English as a Second Language*. New York: mc.Graw Hill Book Company.
- Hughes, Arthur.2003. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kimberlin, C.I & Winterstein, A.G. 2008. *Validity and Reliability of Measurement Instruments Used in Research*. Research Fundamental -Vol 65 Dec 1. 2008.
- Luoma. S. 2004. *Assessing Speaking*. Cambridge: Cambridge University Press.
- Olfos. R & Zulantary. H. 2007. *Reliability and Validity of Authentic Assessment in a Web Based Course*. Educational Technology & Society, 10 (4).156-173.
- Ojung. J & Allida . D. 2017. *A Survey of Authentic Assessment Used to Evaluate English Language Learning in Nandi Central Sub-County Secondary Schools , Kenya*. Baraton Interdiscipline Research Journal, 7 (special issue), pp 1-11.
- O' Malley, J.Michael & Pierce, Lorraine Valdez.1996. *Authentic Assessment For English Language Learners: Practical Approaches For Teachers*. Addison-Wesley Publishing Company.
- Marhaeni. A.A.I.N *et all*. 2014. *Toward Authentic Language Assessment: A Case in Indonesia EFL Classroom*. The European Conference on Language Learning.
- Reynisdottir. B.B. 2016. *The Efficacy of Authentic Assessment*. University of Iceland.
- Riduwan. 2004. *Metode dan Teknik Menyusun Tesis*. Bandung. Alfabeta.
- Swanson. 2002. *Constructing Written Test Questions for the Basic and Clinical Sciences*. Third Edition (Revised) 2002. National Board of Medical Examiners. Available at [http://www.abme.org/PDF/ItemWriting\\_2003/2003IWGwhole.pdf](http://www.abme.org/PDF/ItemWriting_2003/2003IWGwhole.pdf).
- Zaim. M .2013. *Asesmen Otentik: Implementasi dan Permasalahannya dalam Pembelajaran Bahasa Inggris di Sekolah Menengah*. Proceeding of The International Seminar on Language and Arts. FBS Universitas Negeri Padang.