Authentic Assessment in Indonesian Language Learning

Atmazaki
Indonesian Language Education
Universitas Negeri Padang
Padang, Indonesia
atmazaki@fbs.unp.ac.id

Abstract--This study aims at explaining the process of embedding authentic, valid, practical and effective assessment models in Indonesian language learning based on local culture contexts.The development process began with preliminary research, prototyping and assessment phases.The third stage involved three schools in the city of Padang, namely class X SMA 3, 7 and 10 Padang.The expected product is a number of assessment instruments as a model or sample that can measure the ability of the students to participate in the school.The results of this study indicate that all the protocols, prototyping, to the assessment phases have been valid, practical and effective.This means that developed instruments can help students improve their language skills.

Keywords: authentic assessment, performance test, language skill

I. INTRODUCTION

Assessment is an important element in learning because throughthe assessment teachers can know the process and learning outcomes achieved by students. Qualified assessment tools (instruments) will be able to detect student learning outcomes so that teachers can also make choices: continue lessons or carry out remedies. As stated in National Curriculum 2013, assessment of language skills is not sufficient only by multiple choice tests (objective test) and esay test or observation, but must be accompanied by performance tests or alternative assessments (Hamayana, 1995; Wiggins, 1998; Muller, 2012). By using alternative assessments, teachers can find out the students' learning progress in language (Brown (2004: 251).

So far, there have been many studies on assessment in language teaching suggesting the use of authentic assessment. To assess the skills of speaking, holistic assessment is more useful because it is easier to use (Ounis, 2017). In general, teachers see the authentic test (using rubric) in a positive light, even though it is said it needs to consider the age of students in the use of rubric (Pineda, 2014). In high school, many teachers use traditional (assessment) tests in their assessment, both in assessing process and learning outcomes, regardless of whether it measures the cognitive, affective or psychomotor aspects. Assessment is still more focused on concept mastery aspects. It requires a valuation technique that can reveal aspects of products and processes, one of them by applying an authentic assessment (Zaim, 2013).

Many Indonesian high school teachers in West Sumatra have not used authentic tests when assessing the performance of their students in using the Indonesian language, such as table 1.

Table 1. The Tests Used by Indonesian High School Teachers in West Sumatra

No.	Assessment tool used by teachers	N	%
1.	Performance tests / authentic	8	12.12
2.	Multiple / objective Choice Test	13	19.69
3.	Essay Test	22	33.33
4.	Guidelines for observation	7	10.60
5.	Etc	16	24.24
	Total	66	100.00

In the lesson plan used by Indonesian high school teachers (documents are made by teacher association) some learning activities and assessments are found as in table 2, while the interview result of 16 Indonesian language teachers of West Sumatra high school is in table 3.

Table 2. Assessment Activity in Teacher Design Isntruction**l**

No.	Speaking Skills	Learning Activity	Assessment
1.	Listening	Read out text / dialog, Listened to a recording of text/dialogue	Test oral /writing
2.	Speaking	Play a role, Interview, read a poem	Observation, often without instruments
3.	Reading	Read the provided text	Answering questions by text, Melengkap i sentences
4.	Writing	Write short text	Writing

Table 3. Vocational Skills Assessment by High School teachers in West Sumatra

Skills	Type of Assessment
Listening	Listen to a text (news, speeches, interviews, poetry) then students are given questions orally. Furthermore, teachers and students discuss the suitability of student answers with the texts that were played earlier.
Speaking	Oral test without any guidance. Students are asked to give a speech to the front of the class and then scored.
Reading	Reading poetry (tinny), reading ceremonial texts (tinny). Almost no judgment in reading because reading is considered a step towards writing.
Writing	Two kinds of tests are used: (1) knowledge tests, in the form of objective tests; and (2) a writing skill test in the form of a description that tests students' ability to produce various types of texts. Portfolio assessment is also done, but it is not explained how the portfolio is used as an instrument of assessment. Some teachers also assess the writing skills in the form of essays and objectives.

The observations show that the learning activities and assessment are very few and varied. There is no clear standard used to determine student achievement because there is no rubric as a reference (Lovron & Razei, 2011; Muller, 2012; Mertler, 2001; De Silv & Radhika. 2013). This indicates that the teaching and assessment process is still monotonous and feedback can not be given properly. Unavailability of assessment tools resulted in the unavailability of accurate data in determining the student's ability so that the process of improvement of learning also can not be done properly.

The 2012 PISA report shows that the Indonesian language of 15/16 years of age (reading and writing) is 396 (world average 496) and ranked 60th out of 61 countries surveyed (http://gpseducation.oecd.org/, accessed, April 12, 2016). The data show that the reading and writing ability of Indonesian children are still very low. Therefore, systematic upgrading is needed both at school and outside the school.

Indonesian students should be accustomed to working on high-ranking issues such as authentic tests or performance appraisals. It will make them more well-trained in using Indonesian language, rather than just knowing aspects of grammar, spelling, punctuation, and diction. It is necessary to develop performance tests that also simultaneously train teachers in making performance tests so that they are able to judge students accurately, accurately, and authentically (Frey, 2012). Authentic assessment results are more likely to be used as a reflection for improved learning. Brown (2004: 251) calls an alternative assessment with the term beyond tests, a test that is capable of detecting not only knowledge, but also the skills and even student attitudes. With an alternative assessment, teachers can ask for new and challenging ideas for critical thinking (Khoshaim & Rashid, 2016: 129).

Based on the above background the formulation of this research problem is "how is the process of developing an authentic assessment instrument that can help improve students' learning ability. This study describes the process of developing authentic assessments associated with local culture so as to produce valid, practical and effective instruments.

II. METHOD

The development research (Gall, Borg, and Gall, 1996) is done following the Plomp (1997) model with three stages: preliminary research, prototyping and assessment phase. In the early stages identified and analyzed the needs, curriculum, students, and concepts to improve learning outcomes of Indonesian language skills. The second phase is designed authentic and validated test prototypes through experts, and students / teachers or third stage, instruments are tested to high school students to determine their effectiveness and efficiency.

Several questionnaires were prepared to analyze field conditions and to determine suitable sizes in authentic prototype tests. The subjects of the trial were the high school students of grade X in Padang, West Sumatera, Indonesia with three schools of SMAN 3 Padang School-1), SMAN 7 Padang 9(School-2) and SMAN 10 Padang 9 (School-3) with 83 students. Data were analyzed descriptively with product feasibility level above 70%.

III. FINDING AND DISCUSSION

1. Findings
a. *Needs Analysis*

The need of assessment tools in high school is known through observation as in table 4. *First*, needs analysis. Data obtained through interviews with high school teachers of Indonesian subjects in Padang as many as three schools. The result of requirement analysis can be seen in table 4.

Table 4. The need for an Indonesian language skills assessment tool

No.	Questions	School-1	Scool-2	School-3
1.	The questions in the student's book / LKS are related to the daily life of the students.	Already	Not yet	Not yet
2.	Level of difficulty problems in the student book / LKS used	Medium	Medium	Medium
3.	Tests validity that are in the text book	Not yet	Not yet	Not yet
4.	Conformity with context	Already	Already	Not yet
5.	Accustomed to making questions for authentic assessment	Not yet	Not yet	Already

Table 4. Cont

6.	An assessment form used to measure students' Indonesian skills	Based on the difficulty level of the problem	Based on the difficulty level of the problem	Based on curriculum demands
7.	The suitability of tests with competencies	Corresponding	Already appropriate	Already appropriate
8.	Constraints using performance test assessments	Time making performance tests	Assessment format	Create a scoring rubric
9.	Performance tests assist students in performing tasks.	Yes	Yes	Yes
10.	Performance tests assist teachers in assessing student performance.	Yes	Yes	Yes

Table 4 shows that the assessment tools contained in textbooks are not authentic yet, and teachers are convinced that authentic judgment helps students do their work. The problem is **the teachers are not accustomed to using authentic assessment.**

Second, analysis of students. By identifying a number of students from 3 high schools (14 students of SMAN 10 Padang, 29 students of SMAN 3 Padang and 29 students of SMA N 7 Padang) randomly selected students' characteristics as in table 5.

Table 5. Characteristics of Students

Student Characteristics		School-1	Scool-2	School-3
Gender	Male	10	6	9
	Female	14	23	20
Age	15	9	7	5
	16	15	22	24
	≤ 16	0	2	0
Origin	West Sumatra	20	27	27
	In addition to West Sumatra	4	2	2

Gender is not an important issue in this study, but it merely shows comparison. The average age of students is 16 years and dominant comes from West Sumatra (Minangkabau local tribe).

Third, analysis of the concept and curriculum. The results of the analysis show that the operational keywords or words used to ask students to do something or produce something. Thus, authentic assessment is relevant to Indonesian language skills. For example, the results of the analysis of the Indonesian High School Curriculum for K-X as in table 6.

Table 6. The Concept of Test and Revelancy of Indonesian Curriculum

Text learned	Comprehension	Skills
1. Observation report	Identify, analyze	Interpret, construct
2. Exposition	identify, analyze	Develop, construct
3. Folklore	Evaluate, analyze	Construct, create
4. Anecdotes	identify, compare	Telling, developing,
5. Overview	Identify	Compile an overview
6. Negotiation	Evaluate, analyze	Convey, construct
7. Debate	connect, analyze	Construct, develop
8. Biography	assess, analyze	Reveal, tell
9. Poetry	Identify, analyze elements / content	Demonstrating, writing, presenting

b. Design of Test Prototype

Based on the analysis in the preliminary stage it can be concluded that teachers need an authentic assessment that enables them to assist students in performing tasks and facilitate them in assessing student assignments.

Each authentic assessment tool developed consists of 3 sections, namely (1) context, (2) instructionsand (3) rubric (Atmazaki, 2013). Context is a general explanation of the test that is based on the text to be written. The command contains an explanation about the rules in doing the test, but it can also be added to the aspects required in the writing of the test. Finally, the rubric consists of criteria, weight of value, level of performance, and descriptor.

Based on the template, nine examples of performance appraisal are developed that should be learned by students. The next process is the validation of experts, teachers and students in various formats.

The resulat of self evaluation can be seen in table 7.

Table 7. Results of *Self Evaluation* Performance Tests (FT)

No.	Criteria	FT-1	FT-2	FT-3	FT-4	FT-5	FT-6	FT-7	FT-8	FT-9
1.	Performance tests are designed according to the scoring structure.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.	The context of performance tests related to local culture.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.	Complete and detailed performance test instructions.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4.	Full performance test rubric: there are criteria, weights, performance levels, descriptors, and scores.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5.	The language in the context, instructions, and assessment rubrics is easy to understand and there is no typing error.	Yes	No	Yes	Yes	No	No	Yes	No	Yes
6.	Images, tables, or graphs (if any) are designed with the right placement.	No	No	No	No	No	No	No	No	No

Note: FT (Performance Test). FT-1: Observation Report; FT-2: Exposition; FT-3: Folklore; FT-4: Anecdotal; FT-5: Overview; FT-6: Negotiation; FT-7: Debate; FT-8: Biography; FT-9: Poetry

There are certain parts that need to be revised, that is, first, the placement of images, graphics, and tables. Because it is not used then it is ignored. Secondly, the test structure, some of which do not match the template so it is still difficult to understand.

After all the shortcomings were corrected the expert validation was done as shown in table 8. This validation is focused on content Tabel 8), constructs (Table 9), and language (Table 10).

Table 8. Content Validity

No.	Aspects of Content	Validation score								
		FT-1	FT-2	FT-3	FT-4	FT-5	FT-6	FT-7	FT-8	FT-9
1.	To compliance with the learning objectives.	93.75	87.50	75.50	75.50	93.75	93.75	93.75	87.50	81.25
2.	Suitability i ndikator (criterion on the rubric) with text and competence.	87.50	87.50	75.50	75.50	93.75	87.5	93.75	93.75	87.50
3.	Relevance k Context matter	81.25	87.50	68.75	68.75	75	68.75	81.25	81.25	81.25
4.	Compliance with the high school students' cognitive level.	87.50	81.25	75.50	75.50	87.5	81.25	81.25	87.50	87.50
	Average	**87.50**	**85.94**	**73,44**	**73,44**	**87.5**	**82.82**	**87.50**	**87.50**	**84.38**
	Interpretation	**MV**	**MV**	**V**	**V**	**MV**	**MV**	**MV**	**MV**	**MV**

Note: V = valid; M = most

Table 9. Construct Validity

No.	Aspects of Construct	Validation score								
		FT-1	FT-2	FT-3	FT-4	FT-5	FT-6	FT-7	FT-8	FT-9
1.	Represented one of the high school materials.	97.75	75.00	75.00	75.00	100	100	87.25	81.25	81.25
2.	The test consists of three elements: context, instruction, and rubric.	93.75	75.00	75.00	75.00	100	87.5	93.25	93.75	87.50
3.	There is a scoring rubric.	93.75	81.25	75.00	75.00	100	81.25	93.75	87.50	87.50
4.	When the problem of using images, graphics, or tables can be read clearly.	87.50	81.25	75.00	75.00	81.25	75.00	75.00	87.50	93.75
5.	Overall this test is easy to understand.	81.25	75.00	75.00	75.00	100	87.5	81.25	81.25	87.50
	Average	**90.00**	**77.5**	**75.00**	**75.00**	**96.25**	**86.25**	**86.25**	**86.25**	**87.50**
	Interpretation	**MV**	**V**	**V**	**V**	**MV**	**MV**	**MV**	**MV**	**MV**

Table 10. Language Validity

No.	Aspects of Language	Average of Assessment of Tests								
		FT-1	FT-2	FT-3	FT-4	FT-5	FT-6	FT-7	FT-8	FT-9
1.	Use of spelling.	75.00	81.25	68.75	68.75	75.00	81.25	75.00	81.25	87.50
2.	Accuracy of diction	87.50	87.5	75.00	75.00	93.75	87.5	81.25	87.50	87.50
3.	Truth sentence structure.	87.50	87.5	75.00	75.00	81.25	87.5	75.00	81.25	81.25
4.	Communicative.	87.50	81.25	81.25	81.25	93.75	87.5	81.25	87.50	81.25
5.	There are no words that give rise to a double interpretation.	87.50	81.25	81.25	81.25	93.75	81.25	75.00	81.25	81.25
6.	The context of the performance test is dipah ami	87.50	87.5	81.25	81.25	100	81.25	81.25	93.75	93.75
7.	The instructions are easy to understand if	87.50	81.25	81.25	81.25	100	81.25	68.75	87.50	87.50
8.	Descriptor rubrics are easy to understand	81.25	81.25	81.25	81.25	100	81.25	81.25	93.75	93.75
	Average	**85.15**	**83.59**	**78.13**	**78.13**	**92.19**	**83.59**	**78.13**	**87.72**	**87.75**
	Interpretation	**MV**	**MV**	**V**	**V**	**MV**	**MV**	**V**	**MV**	**MV**

Based on expert validation, almost no tests require revision, an average assessment of 5 experts above 80 with good / valid category (V = valid; SV = Most Valid).

Then one to one evaluation is done, asking students to answer some questions based on the test script (27 people from 3 high schools), then asked some questions. The results are listed in Table 11.

Table 11. Validation of Individual *(One Two One)*

No.	Questions	Students answer	
		Yes	**No**
1.	Have you ever performed a performance test like this?	22	5
2.	Do you understand these performance test instructions?	27	0
3.	Do you understand the terms on this performance test?	27	0
4.	Can you understand the statement on this performance test?	26	1
5.	Does the context help and facilitate you in doing this performance test?	27	0
6.	Can you analyze the problem in the given context?	23	4
7.	Do you know the rubric before?	8	19
8.	Does your teacher ever provide an assessment rubric on a performance test?	6	21
9.	Have you seen the rubric attached to this performance test?	5	22
10.	Can you understand the rubric attached to this performance test?	27	0
	Amount	**198**	**72**
	%	**73.33**	**26.66**

Although many students are not familiar with performance tests, they are not difficult to understand what they mean. Thus, this test can be tested in Indonesian language learning activities in school to determine the practicality and effectiveness.

c. Evaluation Process

Trials were conducted at School-1, School-2 and School-3. After test administration then the students are asked to fill out the practicality questionnaire of this test, then conducted interviews with teachers and students to ask their opinion about this performance test. Practicality tests according to the students are listed in table 12 and 13 (Each questionnaire for one test is assessed by 10 students).

Table 12. Assessmen practicality according to the students

No.	Indicators	Practicality									Amount	Average
		TK1	TK2	TK3	TK4	TK5	TK6	TK7	TK8	TK9		
1.	The instructions are easy to understand.	87.5	75	80	70	82.5	77.5	82.5	82.5	85	722.5	80.27
2.	Language is easy to understand.	85	77.5	82.5	75	82.5	75	85	85	75	722.5	80.27
3.	Understandably difficult to do.	92.5	80	90	70	82.5	72.5	80	80	77.5	725	80.56
4.	Presenting problems that can be found in everyday life.	82.5	75	87.5	77.5	82.5	70	75	75	85	710	78.89
5.	Provide benefits for students if able to do it.	80	80	80	82.5	85	75	85	85	85	737.5	81.95
6.	Increase student motivation in learning.	82.5	90	92.5	72.5	75	80	85	85	75	737.5	81.95
7.	Relates to previously learned material.	80	85	77.5	70	75	82.5	77.5	77.5	80	705	78.33
8.	Make it easy for students to learn even if there are no teachers.	82.5	82.5	82.5	72.5	70	77.5	80	80	85	712.5	79.17
9.	Use easy-to-understand terms.	90	80	87.5	77.5	72.5	67.5	90	90	72.5	727.5	80.83
10.	Pictures, tables, and graphics can be read clearly.	85	80	87.5	80	80	80	87.5	87.5	82.5	750	83.33
11.	Provide enough time to complete.	80	77.5	87.5	77.5	75	70	80	80	77.5	705	78.33

Based on the results of the analysis of the aspects of practice, the performance test charged local culture of West Sumatera for K-X can be used to assess student performance.

Based on interviews conducted with one of the Indonesian language teachers in School-1, School-2 and School-3 regarding the development of performance tests using local culture concluded as follows.

Table 13. Assessment practicality according to the teachers(Interview)

No.	Questions about	School-1	Scool-2	School-3
1.	Indonesian language skills test based on local culture is good to use.	Nice to use	Nice to use and very necessary	Nice to use
2.	This test is useful in everyday life.	Can provide an overview of the benefits	Can provide an overview of the benefits	Can provide an overview of the benefits
3.	Interested to develop Indonesian test using complete and detailed instructions.	Interested	Interested	Very interested

4.	Convenience for students in the test	Can provide convenience	Can provide convenience	Can provide convenience
5.	Does the assessment rubric need to be attached to the test?	No need	Very necessary	Very necessary
6.	Rubrics can maximize student performance.	Can maximize student performance	Can maximize student performance	Can maximize student performance
7.	This rubric can assess the overall aspects of the assessment being tested on the students	Can assess overall	Can assess overall	Can assess overall
8.	Rubrics can provide the same rating if used by different assessors	Can provide the same assessment	Can provide the same assessment	Not too same but leads

Based on interviews to 39 students regarding the development of performance tests using local culture concluded at tabel 14.

The effectiveness of performance tests can be known from student learning outcomes. Based on the test, the overall effectiveness value is 85 with the predicate B. Based on the minimum exhaustiveness criteria established by the school, the average value has been met. Student learning outcomes are in table 15.

Table 14. Assessment practicality according to the students(Interview)

No.	Question	Students answer	
		Yes	No
1.	Do you like Indonesian language tests based on local culture?	22	0
2.	Are local culture-based tests useful in everyday life?	19	3
3.	Do you like authentic tests with different contexts?	19	3
4.	Does the context contained in this test help you with the test?	21	1
5.	Are the commands in this test easy to understand?	19	3
6.	Do you need specific instructions?	20	2
7.	Does the illustration need to be added as a limitation for doing the test?	17	5
8.	Do the rubrics need to be attached to the test?	15	7
9.	Does the rubric help you to maximize the tasks you create?	15	7
10.	Can you rate your own assignments using rubrics?	20	2
	Amount	**187**	**33**

Table 15. Student Score Using Performance Tests

No.	Performance Test	Value of Effectiveness
1.	Exposition	86.1
2.	Negotiation	83.0
3.	Debate	86.4
4.	Anecdotal	85.8
5.	Biography	91.1
6.	Short Story	82.2
7.	Observation Result Report	84.1
8.	Overview	84.4
9.	Poetry	81.2
	Amount	**764.3**
	Average	**85.0**
	Predicate	**B**

2. Discussion

Preliminary analysis of school needs and condition (students, curriculum, and concept of Indonesian language skills) shows that authentic assessment is a necessity. Indonesian subjects, as seen in the Curriculum 2013 (core competencies and basic competencies) focus on language skills and text as the basis (Regulation of Minister of Education and Culture No. 24/2016).

In this preliminary analysis phase it is found that teachers and students are generally not familiar with authentic assessment because they have not been able to make it (Gao & Grisham-Brown, 2011). When K-13 demands an authentic assessment, the teacher should try to make it, but it is not in accordance with the theory and needs of an authentic assessment to assess students' language skills.

This requires that teachers need to be re-trained through upgrading or workshops so they can develop assessments that are relevant to the needs of teaching language skills. Teachers need to be re-trained through upgrading or workshops so they can develop assessments relevant to the needs of language skills teaching. In this study, teachers were not involved in the design of authentic assessment prototypes, but were made by researchers in accordance with the theory and learning needs (curriculum demands).

In accordance with theory, authentic judgments are organized into three parts: context, instruction, and rubric (Muller, 2012). In accordance with the kind of text that 10th grade high school student must study, nine authentic scoring tools are designed-for example-to assess students' ability to produce text, ie writing.

After self assessment, the instruments are validated by experts and the results are valid in terms of constructs, content, and language. To ensure that the instrument can be understood by students then performed one to one evaluation. As seen in the results, 73.33% of students stated that the instrument was understandable. This means that the instrument can already be tested in the school (class) to determine the level of practicality and effectiveness.

The results showed that 79.45% of students stated that the instrument is practical. It means that the instrument is easy to understand, context relevant, useful, helpful in tasks, motivation, and time provided enough (Hulela, 2017). According to the teacher, this instrument is also practical to use because it is easy to understand both the instructions and the rubric so as to help students do the task.

The last thing to do is to ask the students to do the tasks according to the instructions in the performance test instrument. The average student is able to do the job well with an average score of 85. From the high learning outcomes, it can be interpreted that the authentic assessment developed is effective.

IV. CONCLUSION

Based on the discussion it can be concluded that in general Indonesian language teachers in West Sumatra have not been familar with authentic assessment. They use traditional judgments so that skills aspects cannot be accurately assessed. Generally, they have not been able to make it even though they believe that authentic assessment is more suitable for assessing students' speaking abilities than traditional (multiple-choice) assessments.

Through a series of development activities, this study produced nine authentic local culture-based assessment tools (Minangkabau / West Sumatra) to assess writing skills, according to the type of text they had to learn. Development activities start from making prototype design tests, validating and testing in the classroom. The test context in this Research yields nine valid assessment tools, practical and effective for assessing Indonesian language skills of high school students.

The validity of this authentic assessment is enhanced by analyzing and adjusting the curriculum content with assessment materials, namely basic competencies and essential indicators that students must master. The construct validity is enhanced by constructing a scoring tool structure consisting of three parts, namely context, instruction and rubric. Context is important in communicating because there is no communication without context, whereas instruction is the stages and limits of activities that must be done so that students know the limits of the tasks they must do. Then, the rubric that serves as a quality guideline for both students and for teachers.

With involvement in the development of authentic assessment it is expected that teachers are able to develop their own authentic assessment in accordance with the competence of language they should train to students.

References

Brown, H. D. (2012). *Language Assessment: Principles and Classroom Practices*. New York: Longman.

De Silva, R.. (2013) **"**Rubrics for Assessment: Their Effects on ESL Students' Authentic Task Performance" in CLC Symposium taken from: https://www.researchgate.net/profile/Radhika_De_Silva3/publications

Earl, L. & Katz, S. 2006. *Rethinking Classroom Assessment with Purpose in Mind. Manitoba Education, Citizenship and Youth.* Accessed on 13 June 2013 from http://www.edu.gov.mb.ca/k12/assess/wncp/.

Frey, Bruce B., Schmitt, V. L., & Justin P. A. (2012). "Defining Authentic Classroom Assessment". *Practical Assessment, Research & Evaluation*, 17(2). Available online: http://pareonline.net/getvn.asp?v=17&n=2

Gall, Borg, and Gall. (1996). *Educational Research An Introduction*. Sixth Edition. New York : Longman Publishers.

Gao, X., & Grisham-Brown, J. (2011). The use of authentic assessment to report accountability data on young children's language, literacy and pre-math competency. *International Education Studies*, *4*(2), 41–53. https://doi.org/10.5539/ies.v4n2p41

Hamayan,ElseV. (1995)."ApproachtoAlternative Assessment".*AnnualReview of Applied Linguistic*; 15-212-226. Cambridge: Cambridge University Press.

Hulela, K. (2017). The Practice of Scaling Down Practical Assessment Components of Agriculture in Junior Secondary Schools Curriculum: A Synthesis of Teachers Perceptions. *Journal of Education and Training Studies*, *5*(3), 43. https://doi.org/10.11114/jets.v5i3.2197

Khoshaim, H. B. & Rashid, Saima (2016). "Assessment of the Assessment Tool: Analysis of Items in a Non-MCQ Mathematics Exam", *International Journal of Instruction Vol.9, No.1, January 2016, e-ISSN: 1308-1470 www.e-iji.net p-ISSN: 1694-609X* .

Kizlik, B. (2009). *Measurement, Assessment, and Evaluation in Education*. Available online http://www.adprima.com/measurement.htm diakses tanggal 20-01-2013.

Mertler, C. A. 2001. "Designing Scoring Rubrics For Your Classromm." Dalam *PracticalAssessment:ResearchandEvaluation,* No7(25)diambil tanggal 4 Mei 2009 dari: *http//:PAREonline.net/getvn.asp?v=7@n=25*

Muller, J. (2012)."WhatisAuthenticAssessment?."Accessed on 10 March 2012 from http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm#names.

OECD (2016), PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial *Literacy,* PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264255425-en

Plomp, Tj. (1997). Educational Design: Introduction. From Tjeerd Plomp (eds). Educational & Training System Design: Introduction. Design of Education and Training (in Dutch).Utrecht (the Netherlands): Lemma. Netherland. Faculty of Educational Science and Technology, University of Twente.

Saefurrohman & , Elvira S. B. (2016) .English Teachers Classroom Assessment Practices in *International Journal of Evaluation and Research* in Education (IJERE) Vol.5, No.1, March 2016, pp. 82-92; ISSN: 2252-8822; *Journal homepage: http://iaesjournal.com/online/index.php/IJERE*

Wiggins, G. (1994). "Toward more authentic assessment of language performances."DalamHancock,C.R.(Ed.), *Teaching,Testing,and Assessment: Making the connection.* Northeast conference reports. Lincolnwood, IL: National Textbook Co.

Zaim, M. (2013). Asesmen otentik: implementasi dan permasalahannya dalam pembelajaran bahasa Inggris di sekolah menengah. *Proceeding ISLA-2*, Padang:

FBS UNP

Ounis, M. (2017). A omparison between holistic and anakytic assessment of soeaking, 8(4), 67—690.

Pineda, D. (2014). The feasibility of assessing teengers'orang English language performace with a rubric. Profile, 16(1), 1657—790