Journal of English Language Teaching Volume 9 No. 1
Journal of English Language Teaching
ISSN 2302-3198
Published by English Language Teaching Study Program of
FBS Universitas Negeri Padang
available at http://ejournal.unp.ac.id/index.php/jelt

An Evaluation of Assessment Designed by English Language Education Student Taechers During Teaching Practice

Rintan Munirah[1], Refnaldi[2]
English Department
Faculty of Languages and Arts
State University of Padang
email: rintanmunirah@gmail.com

Abstract

Assessment is the process which is done during teaching and learning process by the teachers in order to know sudents' ability and monitor students' development. An assessmnet supposed to have good content validity value. This research is aimed to evaluate the formal formative assessment designed by Englsih Language Education Student Teachers during teaching practice. The evaluation was done by considering the content validity of the assessment. The evaluation will be done on the formal formative assessment which is intended to assess speaking, reading, and writing skill. The instrumentation of this research is Content Validity Evaluation Rubrics. This is a descriptive research. In collecting the data, the researcher has collected the documents in the form of Daily Test. The findings of the research is the speaking, reading, and writing assessment designed by student teachers during teaching practice are categorized as valid seen from its content validity.

Keywords: Student-teachers, formal formative assessment, content validity.

A. INTRODUCTION

Assessment becomes an interesting topic to be discussed in recent past years. It helps examiners and also the teachers to make decision during teaching and learning process. Refnaldi, Zaim, and Moria (2017) believe through the assessment teacher can determine whether they are succeeded or not in teaching their students. Tosuncuoglu (2018) states that assessment is a long-term procedure and it involves information and data regarding the development of the students. For students, assessment is needed to help them monitoring their own development and improvement at school.

In general, language assessment can be put as two. They are summative and formative (Brown, 2004). For further, the assessment can be put as diagnostic, placement, and need analysis assessment based on the goal of conducting the assessment itself . Formative assessment is one of the types of assessment which is done to find out what progress learners are making and to see which areas should be improved. Based on the Centre for Educational Research and

[1] English ELTSP of English Department of FBS Universitas Negeri Padang graduated on March 2020

[2] Lecturer of English Department of FBS Universitas Negeri Padang

Innovation (2008) formative assessment at schools refers to frequent interactive assessments of student advancement and comprehension in order to identify learning requirements and adapt teaching properly. Brown (2004: 19-30)has mentioned that an assessment should be practical, reliable, valid, authentic, and give wash back. Teachers have to know the purpose of the assessment that they are going to use. They also need to focus on the target that they want to achieve.

The language assessment in recent curriculum is influenced by constructivism learning. According to this statement, Saefurrohman (2015) revealed that Indonesian Junior High School English Teachers main purposes was assessment for learning as the first preference. Most of basic competence require students to understand both written and spoken text. Even speaking cannot be separated from listening, teachers never teach listening in contrast during teaching and learning. The listening skill sometimes does not tested in Ujian Semester, Ujian Sekolah, and Ujian Nasional.

Some of researches which have correlation with this topic focus on the implementation and the use of the assessment. Rukmini and Saputri (2017) focus on the implementation of authentic assessment to measure student's productive skills based on 2013 curriculum. Refnaldi et al., (2017) investigate the teachers' needs for authentic assessment, specifically to assess writing skill of grade VII students of Junior High School in Teluk Kuantan. The other researches which found discuss about the evaluation of the assessment and the assessment implementation. Ali Saad Al-Yaari, Saleh Al Hammadi, and Ayied Alyami (2013) evaluate the language testing seen from applied linguistic perspective. Alfallaj and Al-Ahdal (2017)evaluate the EFL tertiary examination system. In that case they focus on the authentic assessment as the tertiary test. Alshakhi (2018) revises the writing assessment process at Saudi Language Institute. Nurdin, Zaim, and Refnaldi (2019) evaluate the implementation of authentic assessment for speaking skill at junior high school level.

The researches mentioned above, mostly focus on the assessment implementation and the evaluation of the assessment of certain language skills. Unfortunately, the evaluation of assessment still needs to be deeply studied. It is because there are gaps in variety of skills which is evaluated, the subject of the research, and which part of the assessment that needs to be evaluated (practicality, reliability, validity, authenticity).

The problem mentioned above encourages the researcher to conduct an evaluative research entitled "*An Evaluation of Assessment Made by English Language Education Students Teachers During Teaching Practice".*

B. RESEARCH METHOD

This research is evaluation research. "Evaluation research is the systematic process of collecting and analyzing data about the quality, effectiveness, merit, or value of programs, products, or practices" (Gay, 2012, p.17). This research also can be said as summative evaluation because it is done to evaluate the assessment which have been made and used by the student teachers. This research is aimed to analyze and judge the quality of formal formative assessment which has been

designed by the student-teachers who had completed their field teaching practice seen from its content validity. In this research the formal formative assessment that is evaluated is Daily Test. Quantitative data is used in this research.

The source of data in this research is the Daily Test designed by student teachers during their field teaching practice. To collect the data, the researcher use evaluation rubrics. The evaluation rubrics were used to evaluate the formative assessment that they have made seen from its content validity. In the evaluation rubrics, the researcher used scale 1-5 to do the scoring.

Table 1. Grid of Content Validity

No	Indicators	Sub-indicators
1	Learning Objectives	a. Learning objectives in general that is supposed to assessed
		b. Learning objectives in specific that is supposed to assessed
2	Basic Competence	a. The basic competence needed to be mastered
		b. The language functions need to be mastered
		c. The learning topics need to be mastered
		d. Genre-based text need to be mastered

Developed from Kadir et al.(2019)

The same evaluation rubrics used for all three skills. It is because all of the three skills still require the compatibility of learning objectives and basic competence to the test item and designs.

In collecting and analyzing the data, the researcher has dones everal steps. First, the researcher grouped the assessment based on the skill that is intended to be assessed. Next, the researcher did the evaluation by filling the rubrics. At the same time the researcher distributed the rubrics to two experts. Those two experts consist of teacher and lecturer. The expert also did the evaluation by filling the same rubrics. This step is intended to get more relevant data based on the theory inter-rater reliability. Fourth, the researcher calculate the score on the evaluation rubrics by finding the average score of each skill and assessmnet. Lastly, the researcher convert the score gained into several categories mentioned bellow:

Table 2. Conversion of Content Validity Rubrics

Average Score	**Category**
1.00-1.50	Very invalid
1.51-2.50	Invalid
2.51-3.50	Fairly valid
3.51-4.50	Valid
4.51-5.00	Very valid

Source. Sudjana

To check the inter-rater reliability, the researcher used Intraclass Correlation Coefficient formula in SPSS.

C. RESULT AND DISCUSSION

ISSN: 2302-3198

1. Research Findings

There were 5 speaking assessment, 21 reading assessment, and 14 writing assessment. To reveal the detail result of each skill, the validity result will be discussed separately based on the intended skill to be assessed.

a. Content Validity Evaluation Result of Speaking Assessment

Based on the content validity rubrics about speaking skill that have been filled by 3 validator which consist of the researcher and experts, the score gained can be seen as follows.

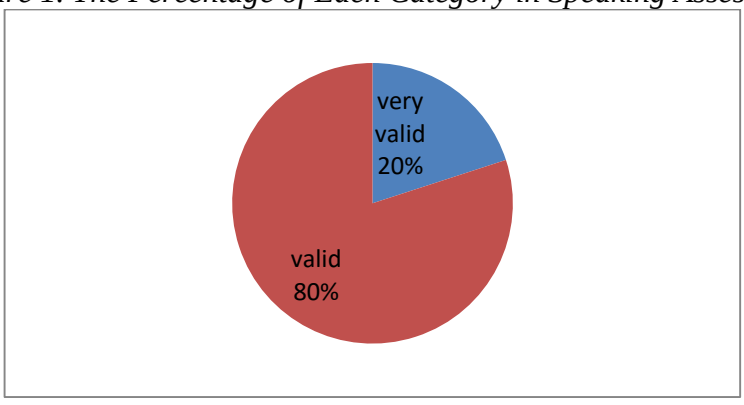
Table 3. The Content Validity Result of Speaking Assessment

No	ST	Researcher's result							Expert 1's result							Expert 2's result						
		1		2				av	1		2				av	1		2				av
		a	b	a	b	c	d		a	b	a	b	c	d		a	b	a	b	c	d	
1	ST 2	4	4	4	4	4	4	4	4	4	4	3	4	4	3,83	4	4	4	4	4	4	4
2	ST 4	4	4	4	4	4	4	4	4	5	4	4	4	4	4,17	5	5	4	4	4	4	4,33
3	ST 6 A	4	4	4	4	4	4	4	5	5	5	4	5	5	4,83	5	4	4	4	4	4	4,17
4	ST 10	5	5	5	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	ST 14	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
average		4,4	4,4	4,4	4,4	4,4	4,4	4,4	4,4	4,6	4,4	4	4,4	4,4	4,37	4,6	4,4	4,2	4,2	4,2	4,2	4,3

Seen from the score gained in each sub-indicator from each expert. The highest score is gained by student teacher 14 in assessment scored by all of the three raters. The score is 5 which categorized as very valid. The topic of that assessment is giving prohibition. For further, seen from the score given per sub indicator, the highest score is 4.6 which is categorized as very valid. This score is taken from Expert 1's result to sub indicator b in indicator 1 and Expert's 2 result to sub-indicator a in indicator 1. The lowest score is 4 for sub-indicator b in indicator 2. This sub-indicator is about language function needed to be mastered.

The percentage of each category in speaking assessment can be drawn as follow:

Figure 1. The Percentage of Each Category in Speaking Assessment



The data is gained through the statistical analysis by finding the average score from the raters. Most of the assessment is categorized as valid. The percentage is 80% from the total of assessment. It means that 4 of 5 assessment gained the score below 4.51. The very valid assessment is only 20%. It means that the very valid assessment is only 1 of 5 assessment gained the score above 4.51. That assessment is the assessment designed by student teacher 14.

b. Content Validity Evaluation Result of Reading Assessment

Based on the content validity rubrics about speaking skill that have been filled by 3 validator which consist of the researcher and experts, the score gained can be seen below.

Table 4. The Content Validity Result of Reading Assessment

No	ST	Researcher's result							Expert 1's result							Expert 2's result						
		1		2				av	1		2				av	1		2				av
		a	b	a	b	c	d		a	b	a	b	c	d		a	b	a	b	c	d	
1	ST 1	4	4	4	4	3	4	3, 83	5	5	5	5	5	5	5	4	4	4	4	4	4	4
2	ST 2	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	ST 3	3	3	4	4	4	4	3, 67	4	3	3	3	4	4	3, 5	3	4	4	4	4	3	3, 67
4	ST 4	5	4	5	4	4	4	4, 33	5	5	5	4	4	4	4, 5	5	5	4	4	4	4	4, 33
5	ST 5	4	4	4	4	4	4	4	4	4	5	4	5	5	4, 5	4	4	5	4	4	4	4, 17
6	ST 6 A	5	4	4	4	4	4	4, 17	5	5	5	4	5	4	4, 67	5	5	5	4	4	4	4, 5
7	ST 6 B	4	4	4	4	4	4	4	5	5	5	5	5	5	5	4	4	4	4	4	4	4
8	ST 7	3	3	2	3	3	3	2, 83	3	3	3	3	2	3	2, 83	2	2	3	2	3	3	2, 5
9	ST 8	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
10	ST 9	4	4	4	3	4	4	3, 83	4	4	4	4	4	4	4	4	4	3	3	3	4	3, 5
11	ST 10	3	3	3	3	3	3	3	4	4	4	4	4	4	4	3	3	3	3	3	3	3
12	ST 11	4	4	4	4	4	4	4	5	5	5	4	5	4	4, 67	4	4	4	4	4	4	4
13	ST 12	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4
14	ST 13	4	4	4	4	4	4	4	5	5	5	5	5	5	5	4	4	4	4	4	4	4
15	ST 14	4	4	5	5	5	5	4, 67	5	5	5	5	5	5	5	4	4	4	4	4	4	4
16	ST 15	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
17	ST 16	3	3	4	4	4	4	3, 67	4	4	4	4	4	4	4	3	3	3	3	3	3	3
18	ST 17	4	4	4	4	4	4	4	5	4	5	5	5	4	4, 67	4	4	4	4	4	4	4
19	ST 18	4	4	4	4	4	4	4	5	5	5	5	5	5	5	4	4	4	4	4	4	4
20	ST	4	4	4	4	4	4	4	4	5	5	5	5	5	4,	4	4	4	4	4	4	4

ISSN: 2302-3198

	19														83							
21	ST 20	4	4	4	4	4	4	4	5	5	4	5	5	5	4, 83	4	4	4	5	4	4	4, 17
Average		3, 86	3, 76	3, 9	3, 86	3, 86	3, 9	3, 86	4, 29	4, 24	4, 29	4, 14	4, 29	4, 19	4, 24	3, 67	3, 71	3, 71	3, 62	3, 62	3, 62	3, 66

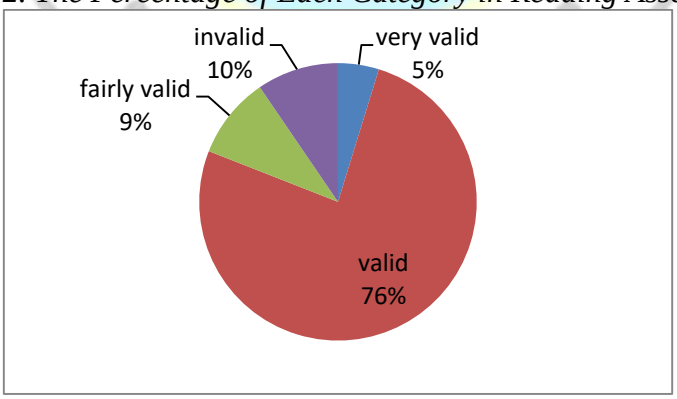
From 21 assessment collected, all of those assessment contain reading assessment. Seen from the average score gained in each sub-indicator from each expert, the lowest is gained by student teacher 2 and 8 scored by expert 1 and 2. The score is 2 it can be categorized as invalid. The assessment designed by student teacher 2 is intended to assess the topic about giving and asking the information related to date, day, and month. The assessment designed by student teacher 8 is intended to assess eighth grade students. The topic is about present continuous tense.

Even though the lowest score is given towards student teacher 2 and 8 is categorized as invalid, the highest score is gained is categorized as very valid. The score is given by researcher and expert one to student teacher 12. The score is 5. The topic which is intended to be assessed is asking attention and checking understanding. Through this assessment the student teacher 12 provides 40 objective questions which the answer lays on the text given.

More detail, seen from the score given per sub indicator, the lowest score is 3. 62 which is categorized as valid. This score is taken from Expert 2's result to sub-indicator b,c, and d in indicator 2. The researcher and expert 1 also categorize this sub indicator as valid. The highest score is gained from expert 1's result. The score is 4. 29 for sub-indicator a in indicator 1, sub indicator a and d in sub-indicator 2.

The percentage of each category in reading assessment can be seen below

Figure 2. The Percentage of Each Category in Reading Assessment



The data is gained through the statistical analysis by using percentage formula towards the average score from all three raters. Most of the assessment is categorized as valid. The percentage is 76% from the total of assessment. It means that 16 of 21 assessments gained the score below the range 3.51-4.50. The very valid assessment is only 5%. It means that the very valid assessment is only 1 of 21 assessments gained the score with range 4.51-5.00. That very valid assessment is designed by student teacher 12. The

rest of percentage of the assessment is categorized as fairly valid and in valid. Each of category gained 9.5%. It means there are two assessments categorized as fairy valid and two assessment categorized as invalid. The fairly valid assessment is designed by student teacher 7 and 10. The invalid assessment is designed by student teacher 2 and 8.

c. Content Validity Evaluation Result of Writing Assessment

Based on the content validity rubrics about speaking skill that have been filled by 3 validator which consist of the researcher and experts, the score gained can be seen as follows.

Table 5. The Content Validity Result of Writing Assessment

No	ST	Researcher's result							Expert 1's result							Expert 2's result						
		1		2				av	1		2				av	1		2				av
		a	b	a	b	c	d		a	b	a	b	c	d		a	b	a	b	c	d	
1	ST 1	3	3	3	3	3	3	3	3	3	3	3	4	4	3,33	3	3	3	3	3	3	3
2	ST 2	4	4	4	4	4	4	4	5	5	5	5	5	5	5	4	4	4	5	4	4	4,17
3	ST 4	4	4	4	4	4	3	3,83	5	4	4	4	4	4	4,17	5	5	4	4	4	4	4,33
4	ST 6 A	3	3	3	3	3	3	3	2	3	3	3	3	3	2,83	2	2	2	2	2	2	2
5	ST 7	3	3	4	4	4	4	3,67	4	4	3	4	4	3	3,67	4	4	3	4	3	3	3,5
6	ST 8	3	3	3	3	3	3	3	2	2	2	2	2	2	2	3	3	3	3	3	3	3
7	ST 9	4	4	4	5	4	4	4,17	4	4	4	5	4	4	4,17	4	4	4	4	4	4	4
8	ST 10	5	5	5	4	5	5	4,83	5	4	4	3	4	4	4	4	4	4	4	4	3	3,83
9	ST 14	4	4	5	5	5	5	4,67	5	5	5	5	5	5	5	4	4	5	5	5	5	4,67
10	ST 15	4	4	4	4	4	3	3,83	4	4	4	3	4	4	3,83	4	4	4	3	3	3	3,5
11	ST 16	4	4	4	4	4	4	4	4	4	4	5	4	4	4,17	3	3	3	4	3	3	3,17
12	ST 17	4	4	5	4	4	5	4,33	5	5	5	5	5	5	5	4	4	4	4	4	4	4
13	ST 19	4	4	4	3	4	4	3,83	5	5	5	5	5	5	5	4	4	4	4	4	4	4
14	ST 20	3	3	3	4	3	3	3,17	4	4	4	5	4	4	4,17	4	4	4	4	4	4	4
Average		3,71	3,71	3,93	3,86	3,86	3,79	3,81	4,07	4	3,93	4,07	4,07	4	4,02	3,71	3,71	3,64	3,79	3,57	3,35	3,65

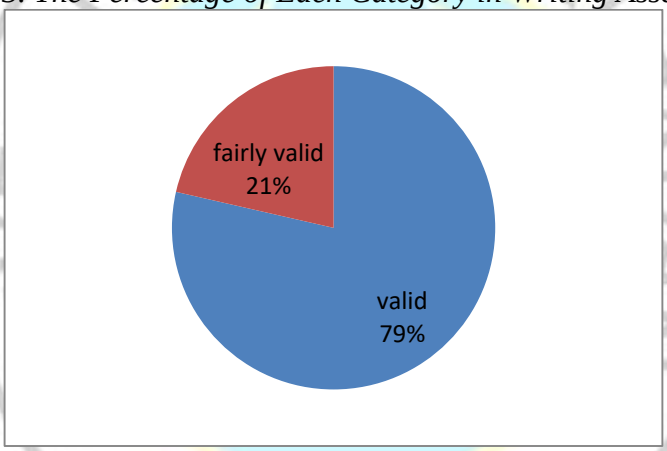
From 21 assessment collected, there are only 14 assessment that contain writing assessment. Seen from the average score gained in each sub-indicator from each expert, the lowest score is gained by student teacher 8 scored by expert 1. The score is 2 it can be categorized as invalid. The assessment

ISSN: 2302-3198

designed by student teacher 8 is done to assess writing skill related using present continuous. The form of the assessment given is multiple choice reading-writing spelling task. The highest score is gained by student teacher 14, 17, and 19 in assessment scored by expert 1. The score is 5 which categorized as very valid. The topic discussed in the assessment designed by student teacher 14 and 19 is the same. It is about giving prohibition. The assessment designed by student teacher 17 asking for ability.

Seen from the score given per sub indicator, the lowest score is 3.5 which is categorized as fairly valid. This score is taken from Expert 2's result to sub-indicator d in indicator 2. The sub-indicator is about genre based text. However, The researcher and expert 1 categorize this sub indicator as valid. The highest score is gained from expert 1's result. The score is 4. 07 for sub-indicator a in indicator 1 and sub indicator b and c in indicator 2. The percentage of each category in writing assessment can be drawn as follow:

Figure 3. The Percentage of Each Category in Writing Assessment



The data is gained through the statistical analysis by using percentage formula towards the average score from all three raters. Most of the assessment is categorized as valid. The percentage is 79% from the total of assessment. It means that 11 of 14 assessment gained the score below the range 3.51. There is no very valid assessment. Recorded 21% of assessment is categorized as fairly valid. It means 3 of 14 assessment gained score in range 2.51-3.50.

2. Discussion

The Degree of Agrement among the three raters are very good. From the finding, we know that the Daily Test which test speaking skill are valid seen from its content validity. Those speaking assessment designed are able to assess both learning objective and basic competence. Through the assessment designed, all of instruction require students to produce the utterances—dialogue or sentences. According to Malec et al., (2017) an oral test is said to have validity of content only if it contains an adequate sample of the appropriate structures, as the examples are dialog, debate, role play or pair work. Four of five assessments found are assessment in the form of role-play.

Same with Daily Test that assess Speaking skill, Daily Test that assess reading is just categorized as valid. Most of the score are 4 for each sub-indicator. Maisarah (2016) said in assessing Junior High School students in reading the students are expected to understand the meaning in written discourse whether interpersonal and transactional or formal and informal. More than half of reading assessment is intended to assess students understanding the dialogue and monologue text.

As writing skill get the lowest average score, which is 3.76, there are some questions which do not assess the material discussed properly, but still categorized as valid. Maisarah, (2016) stated that for writing, the students are expected to be able to utter the meaning in written form both in simple interpersonal and transactional discourse or formal and informal. Some of the writing assessment required students to re-arrange the sentences and translate a sentence. Those samples of questions are compatible with Maisarah argument.

The result of speaking assessment content validity is supported by the previous study conducted by Nurdin et al., (2019). In that research, the speaking assessments designed by teachers are also categorized as valid seen from its content. The findings in writing result has same result which is found by Trianita, (2019). She find that writing assessment created by teachers are also valid seen from the content.

D. CONCLUSIONS AND SUGGESTIONS

By considering the findings and discussion, there are several conclusions. First, the ability of student teachers in creating an assessment is different from one to another. The speaking, reading, and writing assessment that have been analyzed are categorized as valid. All of those three assessments gained the average score in range 3.51-4.50 in general. Seen from the category, all of those three assessments are applicable seen from their content validity. However, if it seen from individual score, some assessment need improvement and revision. In conclusion, the assessment designed by student teachers already assessed what it supposed to measure.

Seen from the individual score per sub-indicator, most of the low score found in the indicator 2 in sub-indicator b and c. It means, some of the assessment does not challenge students in using appropriate language skill. In addition, some of the assessment also does not discuss appropriate topic.

Referring to the conclusions and implications of the research, there are some suggestions proposed by researcher. There are several assessment questions in assessing Reading and Writing skill that should be revised. It is suggested to do the revision by considering the criteria of content validity evaluation. Based on the evaluation of speaking assessment content validity, the speaking activities about giving information about things position need to be re-evaluated.

ISSN: 2302-3198

For the other researcher, conducting interview have to be considered. The interview can be a good way to help researcher to clarify the data collected and deeper understanding. For student teachers, knowing the purpose of conducting the assessment, the learning objectives, and the final result that is required are very important. This can help student teachers in designing better assessment with high content validity value.

BIBLIOGRAPHY

Alfallaj, F. S. S., & Al-Ahdal, A. A. M. H. (2017). Authentic Assessment: Evaluating the Saudi EFL Tertiary Examination System. *Theory and Practice in Language Studies*, *7*(8), 597. https://doi.org/10.17507/tpls.0708.01

Ali Saad Al-Yaari, S., Saleh Al Hammadi, F., & Ayied Alyami, S. (2013). Evaluation of Language Testing: An Applied Linguistic Perspective. *International Journal of English Language Education*, *1*(2). https://doi.org/10.5296/ijele.v1i2.3059

Alshakhi, A. (2018). Revisiting the Writing Assessment Process at a Saudi English Language Institute: Problems and Solutions. *English Language Teaching*, *12*(1), 176. https://doi.org/10.5539/elt.v12n1p176

Brown, H. D. (2004). Language assessment: Principles and Classroom Practice. In *Longman*.

Centre for Educational Research and Innovation. (2008). Assessment For Learning - Formative Assessment. *OECD/CERI International Conference - Learning in the 21st Century: Research, Innovation and Policy*, 1–24. https://doi.org/10.5959/eimj.3.2.2011.e1

Gay, L. R. (2012). *Educational Research* (Tenth Edit). Pearson. https://doi.org/10.1192/bjp.112.483.211-a

Kadir, J. S., Zaim, M., & Refnaldi, R. (2019). Developing Instruments for Evaluating Validity, Practicality, and Effectiveness of The Authentic Assessment for Speaking Skill at Junior High School. *Advances in Social Science, Education and Humanities Research*, *276*, 98–105. https://doi.org/10.2991/icoelt-18.2019.14

Maisarah, I. (2016). *Developing Reading and Writing Assessment for Seven Grade Sstudents of SMP in Merangin District Based on School-Based Curriculum*. 421–430.

Malec, A., Peterson, S. S., & Elshereif, H. (2017). Assessing young children's oral language: Recommendations for classroom practice and policy. *Canadian Journal of Education*, *40*(3), 362–392.

Nurdin, Hi. R., Zaim, M., & Refnaldi, R. (2019). Developing Instruments for Evaluating the Implementation of Authentic Assessment for Speaking Skill at Junior High School. *Advances in Social Science, Education and Humanities Research*, *276*(Icoelt 2018), 106–111. https://doi.org/10.2991/icoelt-18.2019.17

Refnaldi, R., Zaim, M., & Moria, E. (2017). Teachers' Need for Authentic

Assessment to Assess Writing Skill at Grade VII of Junior High Schools in Teluk Kuantan. *Advances in Social Science, Education and Humanities Research, 110*. https://doi.org/10.2991/iselt-17.2017.32

Rukmini, D., & Saputri, L. A. D. E. (2017). The Authentic Assessment to Measure Students' English Productive Skills Based on 2013 Curriculum. *Indonesian Journal of Applied Linguistics*, *7*(2), 25. https://doi.org/10.17509/ijal.v7i2.8128

Saefurrohman. (2015). Classroom assessment preference of Indonesian junior high school teachers in English as foreign language classes. *Journal of Education and Practice*, *6*(36), 104–110.

Tosuncuoglu, I. (2018). Importance of Assessment in ELT. *Journal of Education and Training Studies*, *6*(9), 163. https://doi.org/10.11114/jets.v6i9.3443

Trianita, M. (2019). *An Anlayisis of Validity, Reliability, Practicality, and Effectiveness of Authentic Assessment for Writing Skill at Junior High School*. Universitas Negeri Padang.

ISSN: 2302-3198