Evaluating the Validity, Reliability and Authenticity of English Achievement Test for the Twelfth Grade Students of SMAN 4 Tebo, Jambi

Muncar Winarti✉, Abdurrachman Faridi, Fahrur Rozi

Universitas Negeri Semarang, Indonesia

Article Info	Abstract
	This study aims to investigate the implementation of the English achievement test for the twelfth-grade students at SMAN 4 Tebo, Jambi in the academic year 2019/2020. This study was a contenr analyis. The objects of this study were English achievement test items. It consisted of 40 multiple-choice questions and 5 essay questions. The data collection method was a document analysis checklist. In this study, the researcher analyzed the data from English achievement test for validity aspects such as content validity and construct validity, the degree of reliability, and the implementation of language authenticity criteria. The findings of this study revealed that; (1) the content validity shows 76% valid, and the construct validity shows 60% valid; (2) the reliability shows coefficient 0.281 for multiple-choice items and 0,554 for essay items, and it is reliable; (3) the results of analyzing the authenticity shows that the listening items, reading test and essay writing are authentic. However, each part has some weaknesses. Especially in the reading test, most of the passages use in reading tests failed to represent the world context even though the topics of the passages are rational and based on the real context. Nevertheless, the English teacher who constructed the English achievement test did not mention the sources from which the passages were taken. Next, the samples of the format letter, announcement, and pamphlet look unnatural view from the format and design. Moreover, the English Achievement Test for twelfth-grade students of SMAN 4 Tebo, Jambi has fulfilled the characteristics of validity, reliability, and authenticity as a good test or standardized test.

✉Correspondence Address:
KampusPascasarjanaunnes, semarang. Jl. KeludUtara III Semarang
50237, Indonesia
E-mail: muncarwinarti10@gmail.com

INTRODUCTION

English is one of the foreign languages as the main subject at secondary level to university level in Indonesia. Due to the importance of English, the Indonesian government is strongly committed to the success of English language teaching and learning. The success of the English learning process can be seen in its assessment aspect. Cowie and Bell (1999) stated that an assessment has a critical effect on the education process. The functions of assessment are to inform and to improve the learning process.

The process starts with planning, teaching, and learning process in the classroom, evaluation, and the last one is assessment. According to Brown (2003, p. 4), assessment is a going process that covers a much wider area. To assess the learning process, a teacher should consider several aspects in deciding the final student scores. Arikunto (2005) mentioned that the test is a procedure or appliance used to know or measure something with particular steps. In carrying out a test, a teacher should follow a structured process such as planning the test, usually in the form of a specification table or test specification, constructing test items appropriately, trying to ensure the reliability of the test items, administering the test, objectively scoring the test, and assessing the consistency of the test.

Testing is one of the essential aspects of the teaching and learning process. Usually, it should have done to assess students understanding of the materials. By using a test, we can assess students' abilities. In the learning process, a test is a tool of evaluation that has an important role to measure the teaching-learning process at schools. In this case, there are some functions of a test. For example, it is measuring the student's ability and measuring the efficacy of the teaching-learning process. The types of test that usually used at the end of the semester for twelfth-grade High School students is an achievement test.

An achievement test is a test that aims to get data about students' knowledge or capability in one subject, Aisyah, (2015). It can identify the students' strengths and weaknesses in one subject. Brown, (2003) stated that achievement test is the most frequent purpose for which a classroom teacher will use a test is to measure learners' ability within a classroom lesson, unit, or even total of the content curriculum. An achievement test is a summative test because it should have administered at the end of a lesson, unit, or term of the study.

English Achievement tests should fulfill the principles of a good test. A test should be valid, reliable, and authentic. A valid test means an instrument to measure what is aimed to measure, Fuwana (2019). The test should measure what the teacher wants to be measured. For example, if the teacher wants to measure speaking ability, the teacher should give the test in the form of an oral test, not giving the text to read or audio to listen to. Usually, there are two kinds of validity that construct a good test. They are content validity and construct validity. From the two kinds of validity, content validity plays an important role in interpreting the test as a tool of evaluation, so that the teacher can assess the students' ability effectively. Construct validity refers to the degree of the test that should have to be measured. While reliability refers to the consistency of score, Sugianto (2016). It means if the teacher gives the test repeatedly, the result should be approximately the same. According to the regulation of the Minister of Education and Culture No. 81, the year 2013 about the implementation of the 2013 curriculum, an authentic assessment is an assessment that significantly focuses on measuring student's learning process dealing with their behavior, knowledge, and skill.

Previous studies that focused on validity and reliability have been conducted by Mistar, J. (2011), Umam (2011), Sugianto (2011), Akib and Ghafar (2015), Ali and Sultana, (2016), Putri (2017), Jayanti, Husna, and Hidayat (2019), and Furwana, (2019). Moreover, Bentri, Hidayati, and Rahmi (2016) conducted a study that only focused on applying the authenticity assessment to English class. Meanwhile, some studies focused on the realization of authenticity have been analyzed by Fitriani (2017), Hidayati (2016), Moria, Refnaldi, and Zaim (2017), Rizavega (2018), Rukmini and Saputri (2017), Muthohharo, Bharati, and Rozi (2020). Lastly, Widyaningrum (2016) focused her study on the content validity and authenticity of the 2012 English test in the Senior High School National Examination.

From the previous studies above, the researcher intends to analyze the realization of validity, reliability, and authenticity of English achievement test for twelfth-grade students in senior high school level in the academic year 2019/2020. The main purposes of this study are, firstly, to investigate the implementation of validity tests in terms of content validity and construct validity in English achievement test Or *Ujian Satuan Pendidikan (USP)* 2020. Secondly, this study investigates the reliability test of the English achievement test *(USP)* 2020. Thirdly, this study attempts to describe the authenticity of the English achievement test *(USP)* 2020. The writer focuses only on the achievement test for twelfth grade. The research focuses on the validity, reliability, and authenticity, practicality, and the washback of the 40 multiple-choice questions and 5 essay questions of the English achievement test in SMAN 4 Tebo, Jambi.

This research is hopefully supporting the concept of English achievement test. It is also can be beneficial for English teachers as additional knowledge in developing their technique of making a good English achievement test. It also may improve their ability for assessing students' in the English achievement test realization, especially for Senior high school English teachers.

METHODS

This study used descriptive methods in a qualitative approach. According to Fraenkel and Wallen (2012), a descriptive method is a method used to explain, analyze, and classify something through. It is descriptive because the objectives of this study are to find information as many as possible. The researcher had to survey, collect, and explore data from a different source, a book, and other types of documents. This research aimed to explain the implementation and realization of validity, reliability, and authenticity in the English achievement test. To achieve these research objectives, the content analysis method used to find out the realization of content validity, construct validity, reliability, and authenticity in English achievement test items.

The objects of this research were the validity, reliability, and authenticity of paper booklet test items of English achievement test at SMAN 4 Tebo in the academic year 2019/2020. The English achievement test items proposed by the school. There are 40 multiple-choice questions and 5 essay questions. The standard of the tests was from teacher's association or *Musyawarah Guru Mata Pelajaran (MGMP)* in Tebo Regency and Minister Education policy.

After the researcher gathered all the data, the researcher analyzed the data that was taken from the documentation analysis to find out the realization of validity such as content validity was analyzed by comparing the materials in the syllabus to the items of the test, and construct validity was analyzed by comparing the indicators in the syllabus to the items of the test. While the reliability was analyzed by using the Kuder-Richardson Formula (KR20). Then, authenticity was analyzed by comparing the syllabus, English achievement test items, and authenticity criteria.

RESULTS AND DISCUSSIONS

The first research question elaborated the validity result of English achievement test for the twelfth-grade students so SMAN 4 Tebo, Jambi in academic year 2019/2020. The results of data analysis can be seen on the following table 1.:

Table 1. Distribution of item validity of English Achievement Test 2020

No	Criteria	Number	Percentage
1.	Valid items	34	76
2.	Invalid items	11	24

Table. 1 it can be seen that there are 34 items or 76 % that fulfil the requirements of validity. Those are 29 multiple-choice items number 1, 2, 6, 7, 9, 11, 12, 14, 15, 17, 19, 20, 21, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, and 40. For the essay items number 41, 42, 43, 44, and 45 fulfil requirements of the validity. As a result, the whole test is categorized as a valid test if the value of the test percentage is started from 60% and more than it. So, since the percentage is 76%, it could be categorizing that the English achievement test is valid in the level of high validity. Furthermore, there are 11 invalid items. Those are items number 3, 4, 5, 8, 10, 13, 16, 18,

22, 23, and 25. Those items do not fulfil the requirements of validity.

The content validity analysis of English achievement test was done by comparing the material in the standard examination or lattice with the English test item. From the analysis, it found, there are 58% of materials are representing the English standard examination. Based on the result, it means the English achievement test *(USP)* in the academic year 2019/2020 is valid in terms of content validity. According to Brown (2003, p. 22) a test actually samples of the subject matter about which conclusions are to be drawn, and if it requires the test-taker to perform the behavior that is being measured, it can claim content-related evidence of validity, often popularly referred to as content validity.

This research is supported to research that was conducted by Putri (2017), she had already proved that from 50 questions of the test only 10 items are invalid. Then, supported by Jayanti, Husna, and Hidayat (2019), they found that most of the test items of English National Final Examination for Junior High School 2017/2018 matched with competence standard and the English Syllabus and graduation standard. Additionally, Sugianto (2017), he analyzed the validity based on test items, and the validity of the whole test was determined based on the percentage of all valid items.

Construct Validity Items of English Achievement Test 2020

According to Brown (2003, p. 25) any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions is called construct validity. From the result analysis of construct validity of the English achievement test for SMAN 4 Tebo, Jambi achievement test in 2020 was done by comparing the indicator in the lattice to the content of each item of the test that made by English teacher and then calculating the percentage of the learning indicator in the content of each item included in the test. The result shows the English achievement test *(USP)* 2020 contains 60% valid items and 40% invalid items. There is little significant difference. It found that not all items represent the indicators in the lattice. Furthermore, the English

achievement test *(USP)* 2020 is valid in terms of construct validity.

Additionally, from the analysis, it found that materials included in the test are grade X is 47%, grade XI is 40%, and grade XII is 13%. As a result, the materials were mostly taken from grade X is 21 items, grade XI is 18 items, and grade XII are 6 items. There are functional texts and transactional texts related to the 2013 curriculum. For listening skill consist of transactional text such as; asking and giving information, complementing and thanking, expressing regret, expressing congratulate, expressing an opinion, expressing inviting a person, expressing a feeling, expressing encouragement, the dialogue of expressing gratitude, the monologue of procedure text, the monologue of narrative text, the monologue of report text, and monologue of descriptive text. Then for reading skill contains functional text, such as; recount text, announcement, procedure text, a letter, explanation text, analytical text, and descriptive text. The last, for writing skill contains transactional dialogue of giving an opinion, functional text of procedure text, recount text, descriptive text, and song.

Reliability Analysis of English Achievement Test

Reliability is one of the important parts of the assessment. Reliability refers to the consistency of scores if the test is given to the students on two occasions or more. The reliability level is important to be analyzed. The reliability of the test was measured by using Kidder Richardson's formula (KR 21) for multiple-choice items and Alpha Cronbach for essay items. The reason chose different formula is because both of the two English school examinations are consisting of multiple-choice and essay. KR 21 used to analyze Multiple Choice Questions (MCQs). MCQs is a kind of dichotomy scoring. However, Alpha Cronbach used to analyze the essay items.

In coincidence with the finding, the result of reliability analysis is in a reliable test as an English achievement test. From the results of Microsoft excel computation, it found that the coefficient of reliability test items of the English achievement test *(USP)* is 0,281 for multiple choice. Then, for essay items is 0,554, it interpreted that the English achievement test *(USP)* 2020 is

reliable in the level of low reliability for multiple-choice items and fair reliability for essay items.

The next researcher studied on the field of reliability of the test reported by Muslaini (2016), she revealed that the reliability aspect of the whole test was highly reliable according to the K-20 formula. These findings showed that the quality was not good enough as a prediction test and needed to be reformulated to gain a qualified English prediction test. The other findings of the current research complied with previous findings reported by Sugianto (2017). This study aims to analyze statistically the validity and reliability of the English summative test for the second semester of the tenth graders. The result of the summative test was also reliable. The coefficient of reliability was 0,89. Therefore, reliability was at the level of excellent reliability.

Furthermore, it could be inferred that the findings of the previous studies may differ from this research finding. Validity and reliability are the two kinds of characteristics of a good test that are concerned in the English achievement test as a standardized test at the secondary level, especially for twelfth-grade students.

Authenticity Analysis of English Achievement Test

Authenticity is an essential part of assessing students' abilities, knowledge, or skill. Brown (2004) states that the authenticity of the test is presenting in the following ways; the language in the test is as natural as possible, items are contextualized rather than isolated, topics are meaningful (relevant, interesting for the students, the sonic thematic organization to items is provided, such as through a storyline or episode and tasks represent, or closely approximate, real-world tasks. To analyze authenticity, it will be used as the five criteria of authenticity from Brown, (2004). Then, the object which was the focus of this research was divided into authenticity listening test items, the authenticity of the test tasks, and the authenticity of the test texts. The English achievement test was composed of three main sections namely, listening, reading, and writing. In addition, the researcher attempted to analyze the criteria of authenticity in order to obtain the findings of this study, including relevant topics, thematic organization, natural language, contextualized items, and real-world representativeness.

First, authenticity for listening test items has not a significant problem related to the naturalness of language used in the listening, contextualization of the test items, thematic organization, relevance of the test topics to the learner, and real-world representativeness. The language used in the listening test is similar to the real-world conversations and there is also some word reduction that makes the conversation natural. For the listening test question number 4, the man reduces the word is and *not into isn't* and the word *do not into don't*. Next, the listening test question English achievement test *(USP)* number 2, for instance, the man and woman reduce the word *we will,* and *we would* into *we'll* and *we'd*. Although, there is not hesitations and white noise found in the listening conversations. There are two of three features that can be used to express the natural language use in the listening section; they are hesitations and white noise, Brown (2004).

Furthermore, all listening test items of the English achievement test *(USP)* are considered as contextualized items because the test developed from two learning topics integrated, namely transactional and interpersonal expressions and monologue texts. Moreover, there are fifteen listening questions in the listening test item that are relevant for the students' senior high school. The learning topics used in the conversation are about asking and giving information, complementing and thanking, expressing regret, expressing congratulate, and expressing an opinion. There are 30 questions of the listening test item. 15 questions in the English achievement test *(USP)*. However, there are some questions that are missing in the listening items. And the topics material of questions and standard competence are not matched, the topics material such as expressing a feeling, expressing gratitude, and expressing encouragement.

As a result, all topics dialogue in the listening test is applicable in daily-life situations. On another side, the example of description about Angkor Wat, and King Tutankhamun from the spoken monologue text is the kind of functional text. The texts are not natural because they were

not relevant to daily life. Angkor Wat is not from Indonesia's historical building. Angkor wat is a historical building from Cambodia. Then, King Tutankhamun is a king from the Egyptian dynasty. These texts are not familiar to the students. Therefore, it will make students not easy to understand the question. Because the texts are not appropriate with students' daily life knowledge.

Second, Authenticity analysis on instruction/ task of English achievement test *(USP)*. The data shows that there are 45 test items and 10 passages employed in it. Most of the test task problems are to fulfill the naturalness of language used in the test instructions. Moreover, the language test was not intended to test some grammatical or lexical items, the test-designers should avoid a linguistic mistake in order to make the test a highly authentic reading test.

The visible characteristic of an authentic test was the true language, Richard (2001). It means that the test should not contain linguistic mistakes, lexical morpheme, word orders and grammar (syntactic matter), diction, and meaning (semantic matters) in the test task, in order to avoid misunderstanding among students in understanding the instructions. Authenticity is not only about the quality of the text at all, but authenticity is achieved when students understand the purposes of the teacher. In addition, it will be convenient for students to answer the questions.

Moreover, it can be concluded that the English achievement test from the academic year 2019/2021 is contextualized. The test task was designed from certain learning topics namely transactional text, functional text, and essays. For the thematic item organization, it finds that there 36 test tasks in the English achievement test *(USP)* constructed thematically while there are 9 items in the English achievement test *(USP)* constructed independently.

Third, authenticity analysis of the test texts of the English achievement test (*USP)*. The authenticity of the test texts means the naturalness of language used in the test passages and the real-world representativeness as well as the relevance of the test topics to the students. The result of the analysis shows that there is 55% English achievement test *(USP)* of the text which met the indicator of naturalness of the language used in the test text. Then, the indicator of test passages is not applicable as the authentic test texts. It is caused by the existence of linguistic facets like a typographical mistake, (inconsistent use of the bold word, capital letter, font style, and missed spacing). An example of a typographical mistake such as; The teacher tried to rewrite the number using a pen. This mistake will disturb concentration when students read the instruction. Next, other problems are inconsistent to use the bold word, capital letter, font style, and missed spacing. Most of the English achievement test items in *(USP)* are very often of missed spacing, inconsistency to use font style, and found error format of a good paragraph.

Furthermore, the topics material of the English achievement test found some topics that do not match the authenticity and English syllabus. These are 2 test text topics in English achievement test 2020 are not relevant to senior high school students. The topic is not relevant to senior high school students because the passage used specific terms related to pizza pamphlets, a photocopy machine, and song lyrics.

The result shows that three of the passages used in the reading test failed to represent the world context even though the topics of the passages are rational and based on the real context. On another side, the English teacher constructs the English achievement test did not mention the sources where the passages are taken from. Then, the samples of the format letter, announcement, and pamphlet look un-natural viewed from the format and design.

Based on the Indonesian Ministry of Education and Culture published in the 2013 Curriculum, teachers should implement authentic assessment as the method of assessing students' competencies. The results of the study revealed that the English teachers of the school should have implemented an authentic assessment to measure students' English productive skills. In doing so, the teachers asked the students to described picture cues and retell the study as the performance assessment, to write a text for the portfolio assessment, and to produce a comic for the project assessment. However, the implementation has not been conducted properly yet. Fitriani (2017) stated that teacher's difficulties included excessive marking loads, managing valid assessments,

monitoring academic dishonesty, and maintaining quality and consistency of marking. In other words, the English teacher still experienced some constraints during instructional activities so that the assessment process did not run effectively.

Findings of the present research are in line with Rukmini and Saputri's study (2017) arguing that the English teachers of the school should have implemented the authentic assessment to measure students' English productive skills. In doing so, the teachers asked the students to describe picture cues and retell the story as the performance assessments, to write a text for the portfolio assessment, and to produce a comic for the project assessment. Whereas, the implementation has not been conducted properly yet.

Other research reported by Zaim and Moria (2017), they concluded that Authentic assessment is the process of gathering information by teachers about students' progress and achievement. It is done by using several activities that are relevant and closely related to daily life. The use of authentic assessment cannot be separated from teachers' needs on it. They found that (1) there are several types of authentic assessment needed by the teachers; writing sample, process writing, portfolio, performance assessment, journal, and project/exhibition; (2) the topics needed were factual and familiar topics for students such as family, famous people, things around them; and (3) teachers need simple analytical scoring rubrics. Clearly, the teachers need several types of authentic assessments that are appropriate to assess students' writing skill.

CONCLUSION

This research aimed to investigate the realization of English Achievement Test *(USP)* in academic year 2010/2020 in form of validity, reliability and authenticity. There are two important part of the validity to build a standardized test; the content validity and construct validity. The result from the analysis shown that the content validity and construct validity of the English achievement test is valid in the level of high validity for the content validity, and the English achievement test is valid in

construct validity. With the percentage are the content validity is 76%, the construct validity is 60%. Furthermore, for reliability analysis, it is also being in a reliable test as the English achievement test. From the Microsoft excel computation, it was found that the coefficient of reliability of the test items English achievement test *(USP)* 2020 is 0,281. Then, for the multiple-choice and essay is 0,554, it is interpreted that the English achievement test *(USP)* 2020 is reliable in the level of low reliability for multiple-choice items and fair reliability for the essay items. Lastly, the result of authenticity has been analyzed from the relevant topics, thematic organization, natural language, contextualized items, and real-world representativeness. The results of analyzing the authenticity show that the listening items, reading tests, and essay writing are authentic. However, each part has some weaknesses.

Teaching and learning process could not measure without a test. A good test or standardized test can measure the level of students' achievement from the teaching and learning process. Furthermore, it is needed for the next researchers to conduct this research related to the principles of assessment in the form of validity, reliability, authenticity, practicality, and washback to build a good test that standard with language competence for students.

REFERENCES

Aisyah, A. (2018). Evaluating students' achievement test in reading for interpretation. Academic Journal Perspective: Education, Language, and Literature, 2(2),269-274.

Akib, E., & Ghafar, M. N. A. (2015). The validity and reliability of assessment for learning (AfL). Education Journal, 4(2), 64-68.

Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). Language teaching, 35(2), 79.

Ali, C. M., & Sultana, R. (2016). A Study of the validity of English language testing at the higher secondary level in Bangladesh. International Journal of

Applied Linguistics and English Literature, 5(6),64-75.

Arikunto, S. (2005). Dasar-dasar evaluasi pendidikan (Edisi revisi, cetakan ke-5). Jakarta: Bumi Aksara.

Brown, D. H (2003). Teaching by principles: An interactive approach language pedagogy. USA: Longman.

Brown, D. H. (2004). Language assessment principles and classroom practices. New York: Pearson Education, Inc.

Bentri, A., Hidayati, A., & Rahmi, U. (2016). The problem analysis in applying instrument of authentic assessment in 2013 curriculum. International Journal of Science and Research (IJSR), 1008-1012.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. Assessment in Education: Principles, Policy & Practice, 6(1), 101-116.

Djiwandono, S. (2011). Tes bahasa pegangan bagi pengajar bahasa. Jakarta: PT indeks.

Fitriani, F. (2017). Implementing authentic assessment of curriculum 2013: teacher's problems and solutions. Getsempena English Education Journal, 4(2).

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). How to design and evaluate research in education. New York: McGraw-Hill Humanities/SocialSciences/Languages.

Furwana, D. (2019). Validity and reliability of teacher-made English summative test at second grade of Vocational High School 2 Palopo. Language Circle: Journal of Language and Literature, 13(2).

Ghosh, S., Bowles, M., Ranmuthugala, D., & Brooks, B. (2017). Improving the validity and reliability of authentic assessment in seafarer education and training: a conceptual and practical framework to enhance resulting assessment outcomes. WMU Journal of Maritime Affairs, 16(3), 455 472.

Hidayati, N. (2016). The Authenticity of English language assessment for the twelfth graders of SMK (Vocational High School)

Negeri 4 Surakarta. Premise: Journal of English Education, 5(1), 140-159.

Jayanti, D., Husna, N., & Hidayat, D. N. (2019). The validity and reliability analysis of english national final examination for junior high school. VELES Voices of English Language Education Society, 3(2), 127-135.

Mistar, J. (2011). A study of the validity and reliability of self-assessment. TEFLIN Journal, 22 (1), 45-58.

Moria, E., Refnaldi, R., & Zaim, M. (2017). Using authentic assessment to better facilitate teaching and learning: the case for students' writing assessment. In Sixth International Conference on Languages and Arts (ICLA 2017). Atlantis Press. http://doi.org/10.2991/https://icla-17.2018.54

Muthohharoh. S.R., Bharati. D.A.L and Rozi. F (2020). The implementation of authentic assessment to assess students' higher order thinking skills in writing at MAN 2 Tulungagung. English Education Journal.,10 (3), 374-386.

Ramadani, M., Supahar, S., & Rosana, D. (2017). Validity of evaluation instrument on the implementation of performance assessment to measure science process skills. Jurnal Inovasi Pendidikan IPA, 3(2), 180-188.

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. In Fifth International Seminar on English Language and Teaching (ISELT 2017). Atlantis Press.

Richards, J. C. (2001). Curriculum development in language teaching. Ernst Klett Sprachen.

Rizavega, I. H. (2018). Authentic assessment based on curriculum 2013 carried by efl teacher. Jurnal Profesi Keguruan, 4(2), 142-149.

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), 263-273.

Sugianto, A. (2011). Analysis of validity and reliability of English formative tests. Journal on English as a Foreign Language (JEFL), 1(2), 87-94.

Sugianto, A. (2016). An analysis of English national final examination for junior high school in terms of validity and reliability. Journal on English as a Foreign Language, 6(1), 31-42.

Susandari, Warson, and Faridi, A. (2018). Evaluation of exercises compatibility between revised Bloom's taxonomy and 2013 curriculum reflected in English textbook. English Education Journal, 10 (2), 252-265.

Umam, C. (2011). National examination of english in indonesia: a validity and reliability-based elucidation. Universum:

Jurnal KeIslaman dan Kebudayaan, 5(1), 1-14.

Uswatunnisa, U. (2020). An analysis of english national exam: test of english proficiency for student. ELT WORLDWIDE, 7(1), 63-69.

Tosuncuoglu, I. (2018). Importance of assessment in ELT. Journal of Education and Training Studies, 6(9), 163-167.

Wangid, M. N., Mustadi, A., Senen, A., & Herianingtyas, N. L. R. (2017). The evaluation of authentic assessment implementation of Curriculum 2013 in Elementary School. Jurnal Penelitian Dan Evaluasi Pendidikan, 21(1), 104-115.

Widyaningrum, F. A. D., & Prabandari, C. S. (2016). Content validity and authenticity of the 2012 English test in the Senior High School National Examination. LLT Journal: A Journal on Language and Language Teaching, 16(1), 23-29.