

LAPORAN AKHIR
PENELITIAN SOSIAL DAN HUMANIORA



ANALISIS BUTIR SOAL URAIAN SUMATIF BAHASA INGGRIS
TINGKAT SMP

Oleh;

Neti Hartati, M.Pd. (NIDN: 0322098002)

Muhammad Thoriqun Naja (NIM: 1901055099)

Windy (NIM: 210105501)

Nomor Kontrak Penelitian: 742/F.03.07/2022.

Dana Penelitian: Rp. 7.000.000,-

FAKULTAS,
PROGRAM STUDI
UNIVERSITAS MUHAMMADIYAH PROF DR HAMKA
JAKARTA
2023



UNIVERSITAS MUHAMMADIYAH PROF. DR. HAMKA
LEMBAGA PENELITIAN DAN PENGEMBANGAN

Jln. Tanah Merdeka, Pasar Rebo, Jakarta Timur
Telp. 021-8416624, 87781809; Fax. 87781809

**SURAT PERJANJIAN KONTRAK KERJA PENELITIAN
LEMBAGA PENELITIAN DAN PENGEMBANGAN
UNIVERSITAS MUHAMMADIYAH PROF DR HAMKA**

Nomor : 742 / F.03.07 / 2022
Tanggal : 1 Desember 2022

Bismillahirrahmanirrahim

Pada hari ini, Kamis, tanggal Satu, bulan Desember, Tahun Dua Ribu Dua Puluh Dua, yang bertanda tangan di bawah ini **Dr. apt. Supandi M.Si.**, Ketua Lembaga Penelitian dan Pengembangan Universitas Muhammadiyah Prof. DR. HAMKA, selanjutnya disebut sebagai **PIHAK PERTAMA**; **Neti Hartati, M.Pd. S.Pd., M.Pd.**, selanjutnya disebut sebagai **PIHAK KEDUA**.

PIHAK PERTAMA dan PIHAK KEDUA sepakat untuk mengadakan Perjanjian Kontrak Kerja Penelitian yang didanai oleh RAPB Universitas Muhammadiyah Prof. DR. HAMKA

Pasal 1

PIHAK KEDUA akan melaksanakan kegiatan penelitian dengan judul : **ANALISIS BUTIR SOAL URAIAN SUMATIF BAHASA INGGRIS TINGKAT SMP** dengan luaran wajib dan luaran tambahan sesuai data usulan penelitian Batch 1 Tahun 2022/2023 melalui simakip.uhamka.ac.id.

Pasal 2

Kegiatan tersebut dalam Pasal 1 akan dilaksanakan oleh PIHAK KEDUA mulai tanggal 1 Desember 2022 dan selesai pada tanggal 30 Mei 2023.

Pasal 3

- (1) Bukti progres luaran wajib dan tambahan sebagaimana yang dijanjikan dalam Pasal 1 dilampirkan pada saat Monitoring Evaluasi dan laporan.
- (2) Luaran penelitian, dalam hal luaran publikasi ilmiah wajib mencantumkan ucapan terima kasih kepada pemberi dana penelitian Lemlitbang UHAMKA dengan menyertakan nomor kontrak dan Batch 1 tahun 2022.
- (3) Luaran penelitian yang dimaksud wajib PUBLISH, maksimal 1 tahun sejak tanggal SPK.

Pasal 4

Berdasarkan kemampuan keuangan lembaga, PIHAK PERTAMA menyediakan dana sebesar Rp.7.000.000,- (Terbilang : *Tujuh Juta*) kepada PIHAK KEDUA untuk melaksanakan kegiatan tersebut dalam Pasal 1. Sumber biaya yang dimaksud berasal dari RAB pada Lembaga Penelitian dan Pengembangan Universitas Muhammadiyah Prof. DR. HAMKA Tahun Anggaran 2022/2023.

Pasal 5

Pembayaran dana tersebut dalam Pasal 4 akan dilakukan dalam 2 (dua) termin sebagai berikut;
(1) Termin I 70 % : Sebesar 4.900.000 (Terbilang: *Empat Juta Sembilan Ratus Ribu Rupiah*) setelah PIHAK KEDUA menyerahkan proposal penelitian yang telah direview dan diperbaiki

sesuai saran reviewer pada kegiatan tersebut Pasal 1.

(2) Termin II 30 % : Sebesar 2.100.000 (Terbilang: *Dua Juta Seratus Ribu Rupiah*) setelah PIHAK KEDUA mengunggah laporan akhir penelitian dengan melampirkan bukti luaran penelitian wajib dan tambahan sesuai Pasal 1 ke simakip.uhamka.ac.id.

Pasal 6

(1) PIHAK KEDUA wajib melaksanakan kegiatan tersebut dalam Pasal 1 dalam waktu yang ditentukan dalam Pasal 3.

(2) PIHAK PERTAMA akan melakukan monitoring dan evaluasi pelaksanaan kegiatan tersebut sebagaimana yang disebutkan dalam Pasal 1. Bila PIHAK KEDUA tidak mengikuti Monitoring dan Evaluasi sesuai dengan jadwal yang ditentukan, tidak bisa melanjutkan penyelesaian penelitian dan harus mengikuti proses Monitoring dan Evaluasi pada periode berikutnya.

(3) PIHAK PERTAMA akan membekukan akun SIMAKIP PIHAK KEDUA jika luaran sesuai pasal 3 ayat (3) belum terpenuhi.

(4) PIHAK PERTAMA akan mendenda PIHAK KEDUA setiap hari keterlambatan penyerahan laporan hasil kegiatan sebesar 0,5 % (setengah persen) maksimal 20% (dua puluh persen) dari jumlah dana tersebut dalam Pasal 4.

(5) Dana Penelitian dikenakan Pajak Penghasilan (PPh) dari keseluruhan dana yang diterima oleh PIHAK PERTAMA sebesar 5 % (lima persen).

(6) PIHAK PERTAMA akan memberikan dana penelitian Termin II dalam pasal 5 ayat (2) maksimal 30 Mei 2023.

Jakarta, 1 Desember 2022

PIHAK PERTAMA
Lembaga Penelitian dan Pengembangan
Universitas Muhammadiyah Prof. DR. HAMKA
Ketua,

PIHAK KEDUA
Peneliti,



Dr. apt. Supandi M.Si.



Neti Hartati, M.Pd. S.Pd., M.Pd.

Mengetahui
Wakil Rektor II UHAMKA

Dr. ZAMAH SARI M.Ag.



LAPORAN PENELITIAN

UNIVERSITAS MUHAMMADIYAH PROF DR. HAMKA Tahun 202X

Judul : Analisis Butir Soal Uraian Sumatif Bahasa Inggris Tingkat SMP

Ketua Peneliti : Neti Hartati, M.Pd.

Skema Hibah : Penelitian Publikasi Nasional Madya (PPNM)

Fakultas : FKIP

Program Studi : Pendidikan Bahasa Inggris

Luaran Wajib

No	Judul	Nama Jurnal/ Penerbit/Prosiding	Level SCIMAGO/ SINTA	Progress Luaran
1	Item Analysis of Short-essay Questions of Pre-service English Teachers' Formative Test	Jurnal Penelitian dan Evaluasi Pendidikan	Sinta 2	Submitted

Luaran Tambahan

No	Judul	Nama Jurnal/ Penerbit/Prosiding	Level SINTA/SCIMAGO	Progress Luaran
1	-	-	-	-

Jakarta, 31 Maret 2023

Mengetahui, Ketua Program Studi	Ketua Peneliti
Silih Warni, Ph.D. NIDN. 0302128002	Neti Hartati, M.Pd. NIDN.0322098002
Menyetujui, Dekan Fakultas Keguruan dan Ilmu Pendidikan	Ketua Lemlitbang UHAMKA
Dr. Desvian Bandarsyah, M.Pd. NIDN.0917126903	Dr. apt. Supandi, M.Si NIDN. 0319067801

LAPORAN AKHIR

Item Analysis of Short-essay Questions of Pre-service English Teachers' Formative Test

Latar Belakang (Background)

Evaluation is an integral part of education. It is a systematic process of collecting and analyzing information or data to make a judgment about a specific program (EDC, 2013). In the classroom setting, teachers usually evaluate the success of their teaching-learning process through assessment, and administering a test is a common instrument that teachers often use in this process. A test is commonly used by teachers to measure students' learning outcomes that will enable them to measure students' knowledge, skills, ability, attitudes, and performance (Brown, 2003, Popham, 2003).

In language classrooms, an achievement test is usually designed to measure students' language and skill progress in relation to the syllabus being used to measure how well students have learned what they have been studying, and to determine what still needs to be done for further learning (Harmer, 2015. P. 411).

A formative test measures learners' abilities as part of a process and is part of the learning process itself. It looks to the future of what needs to be done to help students progress to the next level. Due to this reason, formative assessment is also called an assessment for learning or a progress test (Nicol and Mcfarlane-Dick, 2006). Houston and Thompson (2017, p. 2) explained that "formative assessment was attached to improvement of learning in progress." It serves a feedback purpose to guide subsequent or future learning for students. Based on the pupils' performance in the formative test, teachers can make a decision about what needs to be done in future learning (Harmer, 2015).

Teachers have to ensure that the test they construct has good quality by conducting item analysis. It is a process of analyzing the quality of a test instrument done based on certain steps and procedures to sort out good items from bad ones that need to be eliminated or revised for future use (Brown and Hudson, 2002; Musial, et al. 2009). Item analysis is done through item-response analysis which is an analysis of the quality of each test item based on students' responses to each item. The purpose is to improve the quality of test items by identifying which items are good or bad that need to be revised or rejected for further use. However, studies found that teachers at school rarely conduct item analysis and most teachers and lecturers are less familiar with the procedure of item analysis, especially for essay questions.

Studies on item analysis of English tests are mostly done for multiple-choice items but are rarely found for essay questions. Essay questions, on the other hand, are rarely used in English tests. The time-consuming practice of checking and scoring the students' answers to essay questions may be the main reason for its rare use. This condition may be the main reason for the scarcity of literature on item analysis of essay questions mainly for English language tests. Therefore, a study on item analysis of essay questions is required.

This research used the scheme of Penelitian Publikasi Nasional Madya (PPNM) in Social Humaniora.

Tujuan Riset (Objective)

This study aims to analyze the quality of a formative test constructed by two pre-service EFL teachers conducting their teaching practice program in a junior high school in Jakarta, Indonesia. In this program, the student-teachers are given opportunities to have real teaching practice and assessments just like real teachers. This program is a very good opportunity for them to experience actual teaching and assessing students. To assess their students' learning achievement, they also have to construct a test which is also a very good opportunity for them to apply the knowledge they have previously learned of how to construct a good test. This study aims to analyze the Facility Value (FV), the Discriminating Power (DP), Test Reliability, and item validity.

Metodologi (Method)

The formative test was constructed by two pre-service English teachers designed for Class 8 of a private junior high school in Jakarta, Indonesia. It involved 31 students' answer sheets. An interview was conducted with the pre-service EFL teachers to gain information on the material being tested, the number of students involved, and the scoring system.

The test consisted of 10 short-essay questions and the students were given 15 minutes to do the test. The scoring system was Dichotomy that was 10 scores for the correct answer and 0 for the wrong answer, but no score for half right or wrong. Five (50%) questions aimed to test students' mastery of the use of the degree of comparison, and 5 questions for the Present Continuous tense.

1. Diego made chocolate ... than Adi's made. (Good)
2. Danny is the ... boy in the class. (Clever)
3. Kenny is ... than Kate. (clever)
4. The cost of living in Surabaya is ... than in Jakarta. (Cheap)
5. I feel ... than I did yesterday. (Happy)
6. They ... Lazy today. Do you see it? (Work)
7. Brandon and Rudi ... football in the yard right now. (play)
8. We ... in Tarakan City now. (Live)
9. They ... bread in their kitchen right now. (Make)
10. Her mom ... a vegetable in the market today. (Buy)

Facility Value (FV)

To determine the FV or the level of difficulty of the essay question, Sudijono's (2012: 134) formula was used.

The average score for each question = $\frac{\text{the total score of all students for each question}}{\text{The total number of students}}$

$FV = \frac{\text{The average score of each item}}{\text{The maximum score of each item}}$

To determine the qualification of the Difficulty Level, Arikunto's (2018: 225) criterion was employed.

0.71 - 1.00 = Easy

0.31 - 0.70 = Medium

0.00 < 0.30 = Difficult

Discriminating Power (DP)

The formula used to analyze the DP was adapted from Arikunto (2003, p. 238).

$$DP = \frac{G_u - G_l}{U - L}$$

G_u = The number of upper-level students (UG) who answered the item correctly

U = The total number of students in the upper-level group (UG)

G_l = The number of the lower-level (LG) students who answered the item correctly

L = The total number of students in the lower-level group (LG)

Arikunto's (2003, p. 232) classification of the score of the DP was used.

0.70 – 1.00 = excellent

0.40 - 0.69 = good

0.20 – 0.39 = satisfactory

0.00 – 0.19 = poor

Reliability

In this study, to test the reliability of the test, the split-half method of reliability test was employed. The items were divided into two halves which consisted of the odd number items and the even number items. The scores of the first half are considered as variable x and the scores of the other half are considered as variable y . The formula of Pearson Product Moment Correlation was employed to get the Coefficient Correlation of the half items (r_{gg}).

$$r_{gg} \text{ or the } r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{\{N \sum x^2 - (\sum x)^2\} \{N \sum y^2 - (\sum y)^2\}}}$$

The score of the r_{gg} was 0.819216. To calculate the reliability of the whole test, Spearman Brown formula was then used:

$$r_{tt} = \frac{2 \times r_{gg}}{1 + r_{gg}}$$

Notes:

r_{tt} = Coefficient reliability of a test

r_{gg} = even and odd Correlation Coefficient (half the test with the other half)

Item validity

Because this formative test used a dichotomy scoring, which was 0 for the wrong answer

and 10 for the correct answer, to analyze the validity of each item, the Point Biserial Correlation formula was used (Sudijono, 2012, p. 185),

$$\gamma_{pbi} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

γ_{pbi} = Biserial Correlation Coefficient

M_p = Mean score of the students who responded correctly to the analyzed item

M_t = Mean of the total score of all students

S_t = Standard Deviation of the total score

P = The proportion of students who answered the item correctly

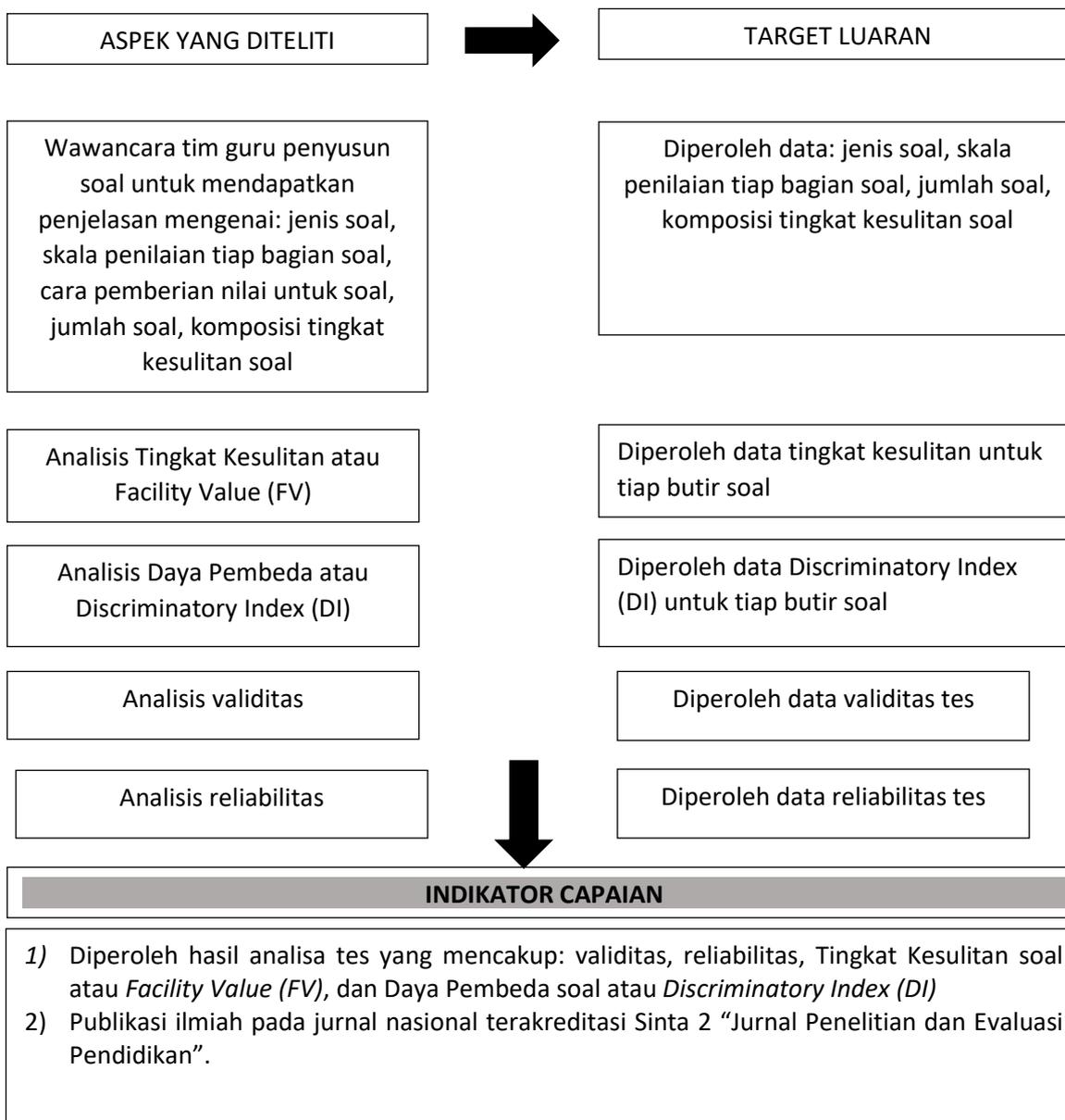
q = The proportion of students who answered the item incorrectly ($q = 1-p$)

The gained Point Biserial Correlation index (γ_{pbi}) was then consulted with the r table at the level of significance of 5% based on the number of students involved in this study.

(Sudijono, 2012, p. 185)

Diagram Alir Penelitian

Pada tahap awal, peneliti berusaha mendapatkan kelengkapan data untuk proses Analisa yang berupa: soal, kunci jawaban, hasil jawaban siswa, hasil skor jawaban siswa.



Jadwal Kegiatan

No	Kegiatan	Bulan Ke-					
		1	2	3	4	5	6
1	Pemerolehan kelengkapan data untuk proses Analisa yang berupa: soal, kunci jawaban, hasil jawaban siswa, hasil skor jawaban siswa.						
2	Wawancara tim guru penyusun soal untuk mendapatkan penjelasan mengenai: jenis soal, skala penilaian tiap bagian soal, cara pemberian nilai untuk soal, jumlah soal, komposisi tingkat kesulitan soal						
3	Analisis Tingkat Kesulitan atau Facility Value (FV)						
4	Analisis Daya Pembeda atau Discriminatory Index (DI)						
5	Analisis validitas						
6	Analisis reliabilitas						
7	Penulisan laporan						
8	Penulisan artikel						
9	Proses submit artikel pada jurnal nasional terakreditasi						

Lampiran Format Susunan Organisasi Tim Pengusul dan Pembagian Tugas

No.	Nama/NIDN	Instansi Asal Bidang Ilmu	Alokasi Jam/minggu	Uraian Tugas
1.	Neti Hartati, M.Pd.	FKIP, Program Studi Pendidikan Bahasa Inggris	24 jam/minggu	Menyusun proposal penelitian, mengunjungi sekolah tempat penelitian, mewawancarai mahasiswa pada program magang mengenai soal formatif yang dipakai, menganalisa jawaban siswa, melakukan analisis butir soal; uji FV, uji DP, uji Reliabilitas, Uji validitas. Menulis artikel, mengirimkan artikel pada jurnal, membuat laporan penelitian.
2.	Muhammad Thoriqun Naja (NIM: 1901055099)	FKIP, Program Studi Pendidikan Bahasa Inggris	4 jam	Mengumpulkan jawaban siswa, memeriksa ulang jawaban siswa.
3.	Windy (NIM: 210105501)	FKIP, Program Studi Pendidikan Bahasa Inggris	4 jam	Memfotokopi lembar jawaban siswa, memeriksa ulang jawaban siswa.

Hasil dan pembahasan

Facility Value (FV)

The analysis of the FV of each question is shown in the following table:

Table 2. The Result of Facility Value Analysis

Difficulty Category	Number of Items	Number of Questions	Percentage
Easy	1, 4, 5, 9	4	40%
Medium	2, 3, 6, 7, 8, 10	6	60%

As shown in the table, the difficulty level of the short-essay questions only consisted of easy and medium levels with the ratio of 2:3, but there is no difficult question.

Discriminating Power (DP)

Discriminating Power (DP) is the ability of each item to distinguish the students who master the tested material from the students who do not. The index of the DP ranges from 0.00 to 1.00. All of the 10 items have excellent, good, and satisfactory levels of DP which means that all items can differentiate students' mastery of the learning materials.

Table 3. The Result of Discriminating Power Analysis

DP Category	Number of Items	Number of Questions	Percentage
Excellent	7, 8, 9	3	30%
Good	1, 2, 5, 6, 10	5	50%
Satisfactory	3, 4	2	20%

Test Reliability

Reliability of the test refers to the level of consistency of an instrument that is whether it measures consistently when being tested and re-tested to the same subjects or testees. Using the formula of Spearman Brown, the r_{tt} was then gained with the result of 0.900625. According to Sudijono, (2011, p. 209), if the $r_{tt} > 0.7$, it means the test has high reliability but if $r_{tt} < 0.7$, the test is unreliable. This finding indicates that the test has high reliability.

Item Validity

Table 2. The Result of Each Item Validity Analysis

Difficulty Category	Number of Items	Number of Questions	Percentage
Valid	1, 2, 4, 6	4	40%
Invalid	3, 5, 7, 8, 9, 10	6	60%

The table indicates that there were only 4 (40%) valid items, while the rest 6 (60%) are invalid. According to Sudijono (2012, p. 183), the invalidity or the low score of validity of test items is an indicator that there is something wrong with the test and that the test makers should be cautious. It is an indicator that the test items failed to measure what are supposed to measure. Sudijono (2012, p. 182) further explained that the low validity or the invalidity of each test item is determined by the number of testees who could(not) answer the item correctly. Sudijono's (2012)

formula to test item validity above also indicates that the higher the number of students who could answer each item correctly, the higher the score of the validity score. In other words, the more students who master the learning material being tested in the item, the higher the score of the validity of the test item. This discussion suggests that the invalidity of the six items indicates that the students have not mastered the materials being tested in those six items.

A deeper analysis of the students' responses to question number 3, 5, 7, 9, 10 revealed that students could not answer the questions because they have not mastered the formula of Present Continuous Tense, and word derivatives.

However, a grammatical analysis of item 8 and 10 revealed that these two items have grammatical errors that influenced the quality of those items which, to some extent, influenced the score validity of those items, the DP, and the FV which indicates that the test maker's English grammar mastery may influence the quality of a test item and a test as a whole.

Item 8 "We ... in Tarakan City now. (Live)" revealed its grammatical error. This item was intended to test the students' mastery of the Present Continuous Tense. However, the use of the verb "Live" for temporal action is inappropriate because it is usually used to refer to general truth but not for temporal action. The word "stay" is more appropriate in this context. Two students used Simple Present Tense to answer this item although they used the incorrect verb "lives" for the subject "We". Thus, this item failed to measure, at least, the two students' mastery of the use of the Present Continuous Tense.

The same finding was also found for question number 10. Eight (26%) students did not get the score. The reason for this was because 4 (13%) students did not use 'to be' and used the incorrect -ing form of the word 'buy' of which they used 'buying' and 'buyer'. One of them also used to be 'are' for the subject 'Her mom'. However, the rest 4 (13%) of the students did not get the score caused by the grammatical error of this item. Question number 10. "Her mom ... a vegetable in the market today. (Buy)" is ambiguous in the use of Tenses caused by the ambiguous time signal 'today'. Although the student-teachers intended to test students' knowledge of the use of Present Continuous Tense, this sentence may mean that the event of "her mom" bought a vegetable happened in the earlier time of 'today' (Past Tense). Therefore, this sentence may be answered in Simple Past Tense. Thus, the time signal 'today' should be replaced by the word 'now' to avoid confusion. Because of this confusion, four students answered it in Simple Past tense. Thus, this item failed to measure what is intended to measure, at least for those 4 students, whether they have mastered the use of Present Continuous Tense or not.

Further, there is another question with a grammatical error that is Question number 1 "Diego made chocolate ... than Adi's made. (Good)". The question is confusing in terms of meaning. It was ambiguous what was being compared whether the quality of chocolate or the way Diego and Adi made chocolate. Because of this, four students used 'to be + the comparative degree of good' ("is better"). Meanwhile, the test makers intended to compare the verb or the way Diego and Adi made chocolate. Therefore, the question should be revised into "Diego made chocolate ... than Adi did. (Good)". However, the student-teachers seemed to realize the grammatical error so they considered those four students' answers correct. Because of this decision, it did not affect the score validity of this item.

The grammatical errors of those three items, to some extent, influenced the score validity of those items, the DP, and the FV which indicates that the test maker's English grammar mastery may influence the quality of a test item and a test as a whole.

Daftar Pustaka (Voncoover)

- Arifin, Z. (2012). *Evaluasi Pembelajaran*. PT. Remaja Rosdakarya.
- Arikunto, S. (2018). *Dasar-dasar Evaluasi Pendidikan*, edisi 3. Bumi Aksara.
- Brown, H. D. (2003). *Language Assessment: Principles and Classroom Practices*. Pearson Longman.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced Language Testing*. Cambridge University Press.
- Cijek, G. J. (2010). An Introduction to Formative Assessment. In H. L. (Eds), *Handbook of Formative Assessment* (pp. 3-17). Routledge.
- EDC. (2013, September 22). *Understanding Evaluation: promote Prevent-3 Bold Steps*. Retrieved from Education Development Centre, Inc: <http://positiveschooldiscipline.promoteprevent.org/>
- Harmer, J. (2007). *How to Teach English*, new edition. Pearson Education Limited.
- Harmer, J. (2015). *The Practice of English Language Teaching* 5th edition. Pearson.
- Houston, D., & Thompson, J. N. (2017). Blending Formative and Summative Assessment in a Capstone: 'It's not your tools, it's how you use them'. *Journal of University Teaching & Learning Practice*, Vol. 14, Issue 3, 1-13, <http://ro.uow.edu.au/jutlp/vol14/iss3/2>.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press.
- Musial, D., Nieminen, G., Thomas, J., & Burke, K. (2009). *Foundations of Meaningful Educational Assessment*. McGraw-Hill Higher Education.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher education*, Vo. 31. No. 2, April, 199-218.
- Popham, W. J. (2003). *Test better Teach Better: the Instructional Role of assessment*. Virginia: Association for Supervision and Curriculum Development.
- Sudijono, A. (2012). *Pengantar Evaluasi Pendidikan*. Rajawali Press.
- William, D. (2011). *Embedded Formative Assessment*. Solution Tree Press.

Target Jurnal Internasional (Output)

Jurnal Penelitian dan Evaluasi Pendidikan (Sinta 2)