# ANALYSIS OF ARABIC MORPHOLOGY QUESTIONS FOR STUDENTS OF ARABIC LANGUAGE EDUCATION PROGRAM

**Ahmad Mubaligh**

Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia

**Abstract:** The study in this research is aimed at explaining information about the quality of the questions used in mid-semester exams in courses for students of the Arabic Language Education Study Program. This research uses quantitative descriptive research, where data was obtained through documentation techniques. The research results showed that all the question items were declared valid. However, the level of Validity was still in the medium category with nine questions (36%) and the low category with 16 questions (64%). The level of reliability was in the high category (0.786), the level of difficulty of the questions was not balanced, there are four questions (16%) items in the difficult category, 19 questions (76%) in the medium category, and two questions (0.8%) in the easy category, and the level of ability is different, all the items differ in the ability of students in the upper group and lower group

## Introduction | مقدمة

The UU Sisdiknas Number 20 of 2003 stated that learning is a process of interaction between students and educators, besides learning is a source of learning in a learning environment. Educators carry out learning systematically and continuously to help students learn so that they experience changes in behavior towards a positive and better direction by their potential, both in cognitive, affective, and psychomotor aspects.

Changes in behavior indicator of the achievement of the objectives of the learning carried out. According to Robert F. Mager (Hamzah, 2006), changes in behavior can be achieved or can be carried out by students under certain conditions and levels of competence, which are expected to be achieved as a result of learning. What learning outcomes mean here is the amount of knowledge and skills that students have after they receive their learning experiences (Supriyadi, 2013) or after going through learning activities (Agustiana et.al, 2019). According to Rahman and Nasryah (2019), these learning outcomes become a benchmark for an educator's success or a measuring tool to what extent the learning process he implements can develop students' potential. If the learning outcomes are good, then the learning process can be said to be successful and vice versa.

To find out the learning outcomes that each student has achieved in learning, it is necessary to carry out a learning evaluation. This evaluation that will be carried out has the intention to determine the extent of students' achievement levels in relation to the educational goals that have been set (Haryanto 2020). In line with that, Ropii and Fahrurrozi (2017) explained that learning evaluations are carried out by teachers where it can provide various information continuously and thoroughly about the process and results that have been achieved by students. Meanwhile, according to Febriana (2019), learning evaluation is not only carried out to determine student learning outcomes but also to determine how this assessment can improve students.

Evaluation is a very important and main part of every learning process because it can measure the level of progress, development, and achievement of student success (Idrus, 2019), as well as the effectiveness of lecturers in teaching. Evaluation functions as a tool to find out the success of the process and learning outcomes of students (Supriyadi, 2013).

Learning evaluation is carried out using tools or instruments in the form of tests and non-tests. Generally, tests can be interpreted as a tool to measure knowledge, besides that tests are also mastery of measuring objects of a certain material and content (Riinawati, 2021). The test usually consists of several questions that are given to students to answer either orally or in writing (Sudijono, in Syahrundin, 2021). Tested by carrying out tests in the form by answering the questions that have been provided, questions that must be responded to, or tasks that must be carried out by the person being tested. In addition, tests are objective and systematic procedures or tools to make it easier to obtain the desired information or data about a person in a precise and fast way (Syahrundin 2021).

The test as a measuring tool must be contain questions that that are valid and reliable(Qodir 2017), so that all the questions can provide clear information regarding the level of mastery students towards learning material. For this reason, it is necessary to analyze the question items to determine the quality of each question item both logically and empirically, as well as to know the level of difficulty and differentiating power so that the substance of the questions is maintained and the instructional objectives can be achieved (Komarudin and Sarkadi 2017). This analysis is mandatory for the teacher to determine the appropriateness of the test or its suitability to the expected learning outcomes (Kurniawan, 2022).

The purpose of the review is to study and examine each question item in order to obtain quality questions before the questions are used. In addition, question item analysis also has the aim of helping to improve tests that can be done through revision or removing ineffective questions, besides that question item analysis also has the aim of knowing diagnostic information on students. Whether the material that has been taught has been understood or not understood (Sumiati, et. al. 2018).

According to Thorndike and Hagen (in Qodir, 2017), analysis of test questions that have been answered by students has two important objectives, namely: (a). The answer to the question given is diagnostic information which has the purpose of examining the lessons of the class and failures in learning. (b). Answers to separate questions and reviews of questions based on those are the basis for better answer of tests performance for the following year.

Based on the thoughts above, in this article, the researcher wants to explain the results of the analysis of Arabic morphology questions used in the mid-semester exam for students majoring in Arabic language education

## Method | منهج

This research was conducted using descriptive-quantitative research, because this research aims to describe quantitative data, in the form of numbers, in detail and clearly. The data for this research is in the form of answers from 134 students to 25 questions and the answer key for each question on the Mid-Semester Examination (UTS) for the Arabic morphology course at the Department of Arabic Language Education. This data was obtained by researcher using documentation techniques.

The data that has been collected is then analyzed using descriptive-quantitative data analysis techniques to calculate the level of validity, reliability, level of difficulty, and distinguishing power using SPSS 16.0 software and the following formulas:

1. Analysis of the Validity and reliability of the question items is calculated using SPSS 16.0 software.

2. The difficulty level of the questions is calculated using a formula (Sridadi, 2002):

$$TK = \frac{U + L}{T}$$

Information:
TK: Index of the difficulty level of the questions sought
U:  Number of students in the upper group who answered correctly for each question
L: Number of students in the lower group who answered correctly for each question
Q: Number of students in the smart group and the poor group

3. The level of differentiating power of the questions is calculated using a formula (Sridadi, 2002):

The calculation of differentiating power is carried out using the following formula:

$$DP = \frac{U - L}{½\,T}$$

Information:
DP: Differential power index of the questions being searched for
U: Number of students in the upper group who answered correctly for each question
L: Number of students l students in the lower group who answered correctly for each question
Q: Number of students in the smart group and the poor group

## Result | نتائج

**Validation of Question Items**

Validity is one of the aspects of analysis that must first be carried out to measure the level of accuracy of the items to be used so that the items are truly able to assess what is supposed to be assessed (Qodir, 2017). Question items are said to be valid if they are able to measure what they want to measure accurately (Rahman & Nasryah, 2019) or if they are able to judge what should be judged (Febriana, 2019). Conversely, a question item is said to be invalid if it is unable to measure the concept to be measured.

According to Thomas L. Good & Jere E. Brophy (Komarudin & Sarkadi, 2017), validity is the most basic test and includes several considerations as a reference for reliability. This means that if a test does not have high Validity, then the Validity of the test is still in doubt.

The Validity of the 25 questions in this research was calculated and analyzed using SPSS 16.0 software. The results of this calculation obtained the following results:

**Table 1 Validity level of Question Items**

| Item | Validity Coefficient (r) | Sig (2 tailed) | Validity Level |
|---|---|---|---|
| 1 | 0.471 | 0,000 | Currently |
| 2 | 0.302 | 0,000 | Low |
| 3 | 0.278 | 0.001 | Low |
| 4 | 0.364 | 0,000 | Low |
| 5 | 0.319 | 0,000 | Low |
| 6 | 0.346 | 0,000 | Low |
| 7 | 0.334 | 0,000 | Low |
| 8 | 0.387 | 0,000 | Low |
| 9 | 0.567 | 0,000 | Currently |
| 10 | 0.272 | 0.001 | Low |
| 11 | 0.295 | 0.001 | Low |
| 12 | 0.397 | 0,000 | Low |
| 13 | 0.486 | 0,000 | Currently |
| 14 | 0.357 | 0,000 | Low |
| 15 | 0.354 | 0,000 | Low |
| 16 | 0.515 | 0,000 | Currently |
| 17 | 0.265 | 0,000 | Low |
| 18 | 0.353 | 0,000 | Low |
| 19 | 0.467 | 0,000 | Currently |
| 20 | 0.486 | 0,000 | Currently |
| 21 | 0.527 | 0,000 | Currently |
| 22 | 0.387 | 0,000 | Low |
| 23 | 0.583 | 0,000 | Currently |
| 24 | 0.439 | 0,000 | Currently |
| 25 | 0.550 | 0,000 | Currently |

The data in the table above shows that all question items are declared valid. This means that all question items are able to measure the concepts measured in the UTS for the Shorof course. All question items that are declared valid are then classified into several categories so that the level of Validity of each question item is known, which question items are very strong, strong, medium, low, and very low. The determination of the level of Validity of the question items is carried out using the following provisions:0.00 - 0.199 (Very Low), 0.20 - 0.39 (Low), 0.40 - 0.59 (Currently), 0.60 - 0.79 (Strong), 0.80 - 1.00 (Very strong).

**Table 2 Classification of Question Items based on Validity Level**

| No | Different Power | Amount | Percentage | No. Question Items |
|----|-----------------|--------|------------|--------------------|
| 1 | Very low | - | - | - |
| 2 | Low | 16 | | 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 15, 17, 18, 21, |
| 3 | Currently | 9 | | 1, 9, 13, 16, 19, 20, 23, 24, 25 |
| 4 | Strong | - | - | - |
| 5 | Very strong | - | - | - |

The data in the table above shows that there are 9 questions with medium validity (36%) and 16 questions with low validity (64%). To guarantee the Validity of a test, the test maker needs to create a grid as a guide for preparing the test so that the questions created do not deviate from the measurement objectives and are representative of the entire teaching material that will be revealed (Inanna et.al, 2021).

## Reliability

Reliability is a coefficient used to indicate the degree to which an instrument or measuring tool can be trusted and obtain relatively stable or consistent results (Farida & Musyarofah 2021) on repeated trials over time or across raters (Warju et.al, 2020). For this reason, a test is said to be reliable if it always gives the same results when used for the same group at different times (Azwar, in Fitrianawati, 2017). In short, reliability relates to the stability or consistency of scores over time or across raters (Warju et.al, 2020).

If the results of measurements are carried out using the test repeatedly on the same subject. Thus, an exam is said to have reliability (constant measuring power) if the scores or grades obtained by examinees for their exam work are stable anytime, anywhere, and by whomever the exam is carried out, checked, and assessed (Qodir, 2017).

**Table 3 Reliability Statistics**

| Cronbach's Alpha | N of Items |
|------------------|------------|
| ,786 | 25 |

Seen from the table above, it is known that the reliability coefficient value of the questions used for the UTS for the short-term course is 0.786. This value shows that the test items are reliable questions because they have a reliability, and based on the interpretation guidelines put forward by Basuki and Hariyanto (2014: 119), the test reliability value is included in the high category, because it is in coefficient 0.70 -0.89. In

this way, the questions have measurement results that are reliable and can be used. This is in accordance with the reliability calculation proposed by Basuki and Hariyanto (Fauziyyah 2019) as follows that 0.00 – 0.19 (Very low), 0.20 – 0.39 (Low), 0.40 – 0.69 (Enough), 0.70- 0.89 (High), 0.90 – 1.00 (Very High).

For the assessment to obtain a high level of reliability, according to Komarudin and Sarkadi (2017), There are two things that need to be considered, namely: (a) preparing assessment tools and formulating questions or tasks with simple and clear sentences. Avoid ambiguous questions or assignments (questions or assignments that allow for multiple interpretations). The number of questions is sufficient because a sufficient number of questions or assignments will increase the reliability of the assessment, (b) The implementation of the assessment must be conditioned as well as possible, among other things, the time provided must be sufficient and strictly scheduled, the situation in the place is conducive to administering the exam, such as calmness, Arrange the seating with sufficient distance so that copying and cheating can be avoided. The implementation of the assessment requires good supervision.

## Discussion | مناقشة

### Difficulty Level

The process of reviewing test questions in terms of difficulty is one of the processes in analyzing the level of difficulty of the question items so that from this can be obtained which questions are easy, medium and difficult (Qodir, 2017). Analysis of the level of difficulty of each question item was carried out by counting the number of students who answered the question item correctly. If more students who answer the question items correctly, then the questions are included in the easy category. Conversely, if there are fewer students who answer the questions correctly, then the questions are included in the difficult category.

Calculating the level of difficulty of a question is a measurement of how big the level of difficulty is. If a question has a proportional level of difficulty, then the question is said to be good. A test question should be neither too difficult nor too easy(Arifin 2012). A test is said to be good if the question items on the test are neither too difficult nor too easy. Questions that are too easy do not stimulate students to increase their efforts to solve them. On the other hand, questions that are too difficult will cause students to become discouraged and not have the enthusiasm to try again because it is beyond their reach(Arikunto, in Elviana, 2020).

In order to determine the level of difficulty of the question items, all students are grouped based on the overall ranking of the test results or scores they obtained into three groups, namely: the clever group (Upper Group), the middle group (Middle Group), the poor group (Lower Group). Of the three groups, the clever group (Upper Group) and the poor group (Lower Group), each consisting of 34 (25%) of all 134 test takers, were used to analyze the difficulty level of the questions.

The results of calculating the level of difficulty of the questions are shown in the following table:

**Table 4 Difficulty Level for Each the Question Item**

| No | Upper Group | Lower Group | Difficulty Level | Category |
|----|-------------|-------------|------------------|----------|
| 1  | 28 | 7  | 0.51 | Medium |
| 2  | 23 | 8  | 0.46 | Medium |
| 3  | 28 | 14 | 0.62 | Easy |
| 4  | 24 | 9  | 0.49 | Medium |
| 5  | 19 | 7  | 0.38 | Difficult |
| 6  | 29 | 10 | 0.57 | Medium |
| 7  | 26 | 14 | 0.59 | Medium |
| 8  | 25 | 10 | 0.51 | Medium |
| 9  | 28 | 3  | 0.46 | Medium |
| 10 | 20 | 10 | 0.44 | Medium |
| 11 | 22 | 11 | 0.49 | Medium |
| 12 | 25 | 9  | 0.5  | Medium |
| 13 | 27 | 7  | 0.5  | Medium |
| 14 | 15 | 3  | 0.26 | Difficult |
| 15 | 25 | 7  | 0.47 | Medium |
| 16 | 30 | 6  | 0.53 | Medium |
| 17 | 19 | 5  | 0.35 | Difficult |
| 18 | 20 | 5  | 0.37 | Difficult |
| 19 | 29 | 11 | 0.59 | Medium |
| 20 | 28 | 9  | 0.54 | Medium |
| 21 | 33 | 12 | 0.66 | Medium |
| 22 | 30 | 18 | 0.71 | Easy |
| 23 | 31 | 7  | 0.56 | Medium |
| 24 | 27 | 11 | 0.56 | Medium |
| 25 | 27 | 4  | 0.46 | Medium |

The data in the table is then classified into several categories so that the quality of each question item is known, including which questions are difficult, medium, and easy. The determination of the level of difficulty of the questions is carried out using the following conditions: $0.00 \leq IK < 0.20$ (Expert), $0.20 \leq IK < 0.40$ (Difficult), $0.40 \leq IK < 0.60$ (Medium ), $0.60 \leq IK < 0.90$ (Easy), $0.90 – 1.00$ (Very Easy).

**Table 5 Classification of Question Items based on Difficulty Level**

| No | Different Power | Amount | Percentage | No. Question Items |
|----|-----------------|--------|------------|--------------------|
| 1 | Hard | 4 | 0.16 | 5, 14, 17, 18 |
| 2 | Currently | 19 | 0.76 | 1, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 19, 20, 21, 23, 24, 25 |
| 3 | Easy | 2 | 0.08 | 2, 22 |

The data in the table shows that the level of difficulty of the questions out of the total number of questions tested in the Sharaf course can be classified into three categories, namely 4 (16%) items in the difficult category, 19 (76%) in the medium

category, and the easy category was 2 (0.8%). This data shows that the question items are not good because seen from the distribution of the composition of the level of difficulty, the question items are not meaningful or disproportionate, the question items in the medium category are more than 50%, the question items in the easy category are less than 30%, and the difficult category is less than 20%. %. According to Kunandar (Sudi Prayitno, 2019), a set of tests is said to be good if it has questioned whose level of difficulty is balancedor proportional, namely easy (30%), medium (50%), and difficult (20%). Thus, the assumption used to obtain effective question quality for measuring good learning outcomes is the proportionality of the difficulty level of the questions.

**Different Power**

The differentiating power of a question is the ability of a question to differentiate between students who are clever and students who are less clever(Rahman and Nasryah 2019). Counting differential power will be known as the ability of each question item to differentiate students who are in the smart group (Upper Group) from students who are in the lower group (Lower Group). For this reason, the calculation is carried out by sorting all student test results from the highest score to the lowest score, then grouping them into the upper group and the lower group. Each group consisted of 34 (25%) of all 134 test participants.

The results of calculating the level of difficulty of the questions are shown in the following table:

**Table 6 Calculation Results of Differentiating Power for Each Question Item**

| No | Upper Group | Lower Group | Differentiating Power (PA-PB) | Category |
|----|-------------|-------------|-------------------------------|----------|
| 1  | 28 | 7  | 0.62 | Good |
| 2  | 23 | 8  | 0.44 | Good |
| 3  | 28 | 14 | 0.41 | Good |
| 4  | 24 | 9  | 0.44 | Good |
| 5  | 19 | 7  | 0.35 | Enough |
| 6  | 29 | 10 | 0.56 | Good |
| 7  | 26 | 14 | 0.35 | Enough |
| 8  | 25 | 10 | 0.44 | Good |
| 9  | 28 | 3  | 0.74 | Very Good |
| 10 | 20 | 10 | 0.29 | Enough |
| 11 | 22 | 11 | 0.32 | Enough |
| 12 | 25 | 9  | 0.47 | Good |
| 13 | 27 | 7  | 0.59 | Good |
| 14 | 15 | 3  | 0.35 | Enough |
| 15 | 25 | 7  | 0.53 | Good |
| 16 | 30 | 6  | 0.71 | Very Good |
| 17 | 19 | 5  | 0.41 | Good |
| 18 | 20 | 5  | 0.44 | Good |
| 19 | 29 | 11 | 0.53 | Good |
| 20 | 28 | 9  | 0.56 | Good |

| 21 | 33 | 12 | 0.62 | Good |
| 22 | 30 | 18 | 0.35 | Very Good |
| 23 | 31 | 7 | 0.71 | Good |
| 24 | 27 | 11 | 0.47 | Good |
| 25 | 27 | 4 | 0.68 | Good |

The data in the table is then classified into several categories so that the quality of each question item is known, which questions are good, which questions are bad and need to be revised, and so on. The determination of whether a question item is good or bad is done using the following conditions: 0.00 – 0.20 (Bad), 0.21 – 0.40 (Fair), 0.41 – 0.70 (Good), 0.71 – 1 .00 (Very good).

**Table 7 Classification of Question Items based on the Differential Power Index**

| No | Different Power | Amount | Percentage | No. Question Items |
|---|---|---|---|---|
| 1 | Not good | - | - | - |
| 2 | Enough | 5 | 0.20 | 5, 8, 10, 11, 14 |
| 3 | Good | 17 | 0.68 | 1, 2, 3, 4, 6, 7, 12, 13, 15, 17, 18, 19, 20, 21, 23, 24, 25 |
| 4 | Very well | 3 | 0.12 | 9, 16, 22 |

In the table above, it can be concluded that the number of questions with sufficient differentiating power is 5 (20%), 17 (68%), good and 3 (12%) very good. This means that all the questions used in the Mid-Semester Examination for students majoring in Arabic Language Education in the Sharaf course are accepted, and there is no need for revision. Apart from that, all of these question items are also able to differentiate the abilities of students in the intelligent group (Upper Group) from students in the lower group (Lower Group).

Knowledge of powered differentiation is very important because one of the basic guidelines for compiling learning outcomes test items is the assumption that the abilities of students are different, so one of the analyses that must be carried out to find out whether the items can be said to be good as an evaluation tool is an analysis of the differentiating power (Sudjiono, in Supandi & Farikhah 2016).

# Conclusion

Based on the results of the analysis of the Mid-Semester Examination questions in the Arabic morphology course, the following conclusions can be obtained: 1). Judging from their Validity, all the questions are included in the valid category; only there are no ones in the high category, nine questions (36%) are in the medium category and 16 questions (64%) are in the low category, (2) judging from their reliability, all the questions are reliable questions because they have a reliability coefficient of 0.786 or greater than 0.60, (3) seen from the level of difficulty, the distribution of the difficulty composition of the questions is not proportional, the questions in the medium category are more than

50%, namely 19 questions (76%), (4) seen from the differentiating power, all the questions differ.

The ability of students in the intelligent group and the poor group. Apart from that, all questions can be accepted without revision. Furthermore, the researcher hopes that this research will continue to be carried out by future researchers so that it can add information and references related to the quality of the questions that will be adequately tested, especially questions in the field of Arabic morphology.

## Acknowledgment

## Bibliography

Agustiana, Mega, Hastari Mayrita, and Andina Muchti. 2019. "Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Bahasa Indonesia Kelas Xi." *Jurnal Ilmiah Bina Edukasi* 11 (01): 26–35. https://doi.org/10.33557/jedukasi.v11i01.203.

Arifin, Zainal. 2012. *Menganalisis Kualitas Tes*. *Evaluasi Pembelajaran*.

Elviana. 2020. "Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates." *Jurnal Mudarrisuna* 10 (2): 58–74.

Farida, and Anna Musyarofah. 2021. "Validitas Dan Reliabilitas Dalam Analisis Butir Soal." *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab* I (1): 34–44. https://jurnal.lp2msasbabel.ac.id/index.php/AL-MUARRIB.

Fauziyyah, Nisa. 2019. "Kualitas Soal Pilihan Ganda Penilaian Akhir Mata Pelajaran Matematika" 8: 814–24

Febriana, Rina. 2019. *Evaluasi Pembelajaran*. Edited by Bunga Sari Fatmawati. Pertama. Jakarta: Bumi Aksara. https://www.ptonline.com/articles/how-to-get-better-mfi-results.

Fitrianawati, Meita. 2017. "Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik." *JPT : Jurnal Pendidikan Tematik* 2 (3): 316–22. http://publikasiilmiah.ums.ac.id/handle/11617/9117.

Hamzah, B. Uno. 2006. *Perencanaan Pembelajaran*. PT Bumi Aksara.

Haryanto. 2020. *Evaluasi Pembelajaran; Konsep Dan Manajemen*. UNY Press.

Idrus. 2019. "Evaluasi Dalam Proses Pembelajaran Idrus L 1." *Evaluasi Dalam Proses Pembelajaran*, no. 2: 920–35.

Inanna, Rahmatullah, and Muhammad Hasan. 2021. *Evaluasi Pembelajaran: Teori Dan Praktek*.

Komarudin, and sarkadi. 2017. "11_Buku_Evaluasi_Pembelajaran," 294.

Kurniawan, Andri dkk. 2022. *Evaluasi Pembelajaran*. Edited by M.Pd Ari Yanto and S.Pd Tri Putri Wahyuni. *Remaja Rosdakarya*. Pertama. Padang: PT. Global Eksekutif Teknologi.

Qodir, Abdul. 2017. *Evaluasi Dan Penilaian Pembelaiaran*. Yogyakata: Penerbit K-Media.

Rahman, Arief Aulia, and Cut Eva Nasryah. 2019. *Evaluasi Pembalajaran*. *Uwais Inspirasi Indonesia*.

Riinawati. 2021. *Pengatar Evaluasi Pendidikan*. Thema Publishing.

Ropii, Muhammad, and Muhammad Fahrurrozi. 2017. *Evaluasi Hasil Belajar*. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.

Sudi Prayitno, H. 2019. *Evaluasi Pembelajaran Matematika Penerbit Duta Pustaka Ilmu Bersama Menyebar Ilmu*.

Sumiati, Ati, Umi Widiastuti, and Usep Suhud. 2018. "Workshop Teknik Menganalisis Butir Soal Dalam Meningkatkan Kompetensi Guru Di SMK Cileungsi Bogor." *Jurnal Pemberdayaan Masyarakat Madani (JPMM)* 2 (1): 136–53. https://doi.org/10.21009/jpmm.002.1.10.

Supandi, Supandi, and Lailatul Farikhah. 2016. "Analisis Butir Soal Matematika Pada Instrumen Uji Coba Materi Segitiga." *JIPMat* 1 (1): 71–78. https://doi.org/10.26877/jipmat.v1i1.1085.

Supriyadi, 2013. *Evaluasi Pembelajaran Bahasa Indonesia*. *UNG Press Gorontalo*. https://scholar.google.co.id/scholar?cluster=5059232428618740961&hl=id&as_sdt=2005&sciodt=0,5

Syahrundin, Husni. 2021. *Buku Ajar Evaluasi Pembelajaran*. Cetakan Pe. Purbalingga: Eureka Media Aksara. https://www.ptonline.com/articles/how-to-get-better-mfi-results.

Warju, Warju, Sudirman Rizki Ariyanto, Soeryanto Soeryanto, and Rio Adi Trisna. 2020. "Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan." *Jurnal Pendidikan Teknologi Dan Kejuruan* 17 (1): 95. https://doi.org/10.23887/jptk-undiksha.v17i1.22914..