

Item Analysis Of Arabic Language Examination

Ihwan Mahmudi¹, Afni Nurwardah^{*2}, Siti Nikmatul Rochma³, Agung Nurcholis⁴

^{1,2,3}University of Darussalam Gontor, Indonesia

ihwanm@unida.gontor.ac.id¹, afninurwardah@mhs.unida.gontor.ac.id^{*2},
nikmatul.rochma@unida.gontor.ac³, agung.nurcholis@unida.gontor.ac.id⁴

Abstract

Assessment and evaluation are needed to determine how student competence is compatible with predetermined educational goals. This study aims to analyze the Midterm Exam for odd-semester Arabic subjects at Islamic Junior High. The type of research used in this research is descriptive quantitative research. This study uses the documentation method for data collection. In addition, researchers will use the help of a computer program, namely Anathes 4.0.2. The results showed that: (1) The validity test of 15 multiple choice test questions was declared valid (80%). As well as for the validity test of 5 matching test questions they were declared valid (80%). (2) The reliability test of multiple choice test items is 0.87. In contrast, the results of the reliability of matchmaking test items were 0.76. (3) The difficulty level test of the multiple test questions categorized as hard is 0 items (0%), medium is three items (20%), and easy is 12 items (80%). The difficulty level test of the matching test questions categorized as hard is 0 items (0%), medium is four items (80%), and easy is 1 item (20%). (4) The difficulty level test of the Multiple test questions categorized as bad is 1 item (6,66%), enough is two items (13,3%), good is nine items (60%), and Very Good is three items (20%). The difficulty level test of the matching test questions categorized as bad is 0 items (0%), enough is 1 item (20%), good is 0 items (0%), and Very Good is four items (80%). With these results, the test takers or teachers must increase their attention to the analysis of the items, namely in terms of validity, reliability, difficulty level, and discriminatory power.

Keywords: Item analysis; Validity; Reliability; Difficulty level; Differential Power

INTRODUCTION

Language is an intermediary for conveying the will and ideas contained in the mind. Language is also considered one of the most important elements in social life because people use it to express their purposes, and the types of language are many. Differences exist between tribes, countries, or certain circles (Lee et al., 2020). One of the important languages to understand and learn, especially for Muslims, is Arabic because it is useful for understanding Islamic studies from its sources. that is, through the al-Qur'an and Hadits (Madkur, 1991)(Oueslati, Cambria, HajHmida, & Ounelli, 2020).

Teaching a foreign language in Indonesia may be the intermediate language in some educational units to support students' foreign language skills. Arabic is one of the most important foreign languages taught for a long time. It has become an introductory lesson since the establishment of Islamic institutes and Islamic boarding schools till universities. However, there is a different difficulty in learning and teaching it due to their lack of love towards this subject or not maximizing learning, and the components

of implementing education have yet to be fully achieved (Ridho, n.d.). This problem is one of the problems in the process of learning and teaching the Arabic language.

Evaluation is needed to measure success in education. However, evaluation is one of the three components of learning and teaching performance: teaching purpose, teaching implementation, and evaluation of teaching results (Mustopa, Basri, & Barlian, 2021) (Shaker et al., 2020). Lynch (1996) suggests that evaluation collects all information for orderly judgment and decision-making. Evaluation, according to Regulation of The Minister of Education and Culture No. 66 of 2013, is the process of gathering information to measure the achievement of students' learning results, including authentic assessment, self-assessment, portfolio assessment, daily tests, midterm tests, final semester tests, competency level exams, competency quality level exams, national exams and school exam (*Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional*, 2003). And then evaluation is the collection of information to find out the results of the educational process. Evaluation is associated link with the test because it is a tool for measuring learning and teaching outcomes and also for knowing the extent of students' competence according to their prescribed education goal (Rusdiana, 2014). Therefore, this test offered to the students should be of the best quality and reviewed by analyzing all questions or test items.

Based on the previous paragraphs, the researcher specialized this research in Item Analysis of the Arabic language examination At Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta. The researcher saw that the exam questions for the Arabic language subject in this school could not measure the ability of the students to understand this subject; this problem is possible because the teacher, as the question's author, needed to analyze all the items of the exam questions for the Arabic language subject.

METHOD

The type of research used in this research is descriptive quantitative research. This study uses the documentation method for data collection. In addition, researchers will use the help of a computer program, namely Anathes 4.0.2.

The researcher divided the research population into two parts: the target population and the expected population. The target population for this research is all Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta students, with 191 students from the seventh through the ninth grades. The expected population for this research is 82 students from the eighth grade, which consists of four classes: the 8A with 26 students, the 8B with 21 students, the 8C with 16 students, and the 8D with 19 students.

From this population, the researcher took all the students from the expected population, 82 students from the eighth grade, as the sample for this research. This sample is called the saturated sample because the researcher took all students from the expected population to be the sample for this research.

RESULT AND DISCUSSION

Item analysis is one of the educational evaluation studies (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). The test is corrected through analysis activities by determining the level of quality of each item of the question. The process of this

analysis is when the exam has been passed, and the teacher gets the students' questions and answers (Makhisoh & Suroyya, 2022). Also, the analysis of question items is the process of determining the quality of the question items in a test regarding their validity, reliability, difficulty level, and differential power (Mahmudi, 2020a). Therefore, the definition of the analysis of exam question items is the series of activities in determining the level of quality of the exam question items so that the information obtained from these activities can be used to improve the quality of the questions that have been prepared.

The purpose of item question analysis is to improve all questions to be effective and suitable by removing or replacing the item questions because good questions help the teacher to know the actual competence of students. This analysis can be done qualitatively or quantitatively. The qualitative valuable analysis for observing tests in terms of their conformity towards subject, language, culture, and how it evaluates. The use of quantitative question items analysis includes validity, reliability, difficulty level, and differential power (Fikriyah, 2021).

The benefits of analyzing exam question items include assisting test users in assessment. Secondly, proportionality in preparing informal and domestic tests, such as tests designed by the teacher for students in the classroom. Third, support for writing practical question items. Fourth, to improve the test in the school. Fifth, to increase the degree of validity and reliability of the questions. (Nasional, 2008). The research will discuss item analysis of the Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta. The study is based on validity, reliability, difficulty level, and differential power.

Validity

Validity refers to the precision and accuracy of the instrument in performing its evaluating function (Mahmudi, 2020b) (Taylor et al., 2020). A good test is highly valid because it can measure the students' competence. So, the teacher must test the questions before presenting them to the students (Leona, 1997). According to the Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School, Bantul Yogyakarta, they have already reached 80%. Therefore, this test is valid. Because the validity result for fifteen questions items of the multiple test questions, in general, is twelve questions are reasonable (80%). As for invalid questions, there are three questions (20%). The validity result for five questions items of the matching test questions, in general, is that four questions are valid (80%). As for invalid questions, there is one question (20%).

So, the teacher must save the valid questions to the Questions Bank and be able to reuse them in future tests. As for invalid questions, they must be corrected before using them in the exam by increasing knowledge about the method of preparing the question items. In this case, the teacher must use a good tool or program. The teacher can also provide the evaluation tool to the professional to correct the validity of the questions so that the test has good validity to use as an evaluation instrument. As for these values based on the results of the analysis below:

1. Validity analysis of multiple test questions

The researcher used the Anathes program to analyze the validity of the Arabic language examination at Luqman Hakim International Integrated Islamic Junior High

School Bantul Yogyakarta. After the previous calculation, the result is matched with the r table at the significance level of 5% and with fifteen questions, so the r Table is 0.482. And if $r_{table} \leq r_{account}$, then the question items are valid. But if the $r_{table} \geq r_{account}$, the question items are invalid.

In general, the validity result for fifteen items of the multiple test questions is that twelve questions are valid (80%). It is item number 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 14. As for invalid questions, there are three questions (20%), which are item number 3, 13, and 15.

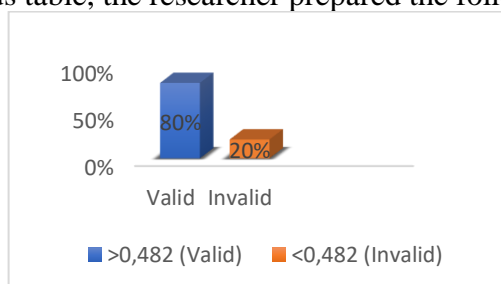
No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0.525	Signifikan
2	2	0.559	Signifikan
3	3	0.452	-
4	4	0.634	Sangat Signifikan
5	5	0.525	Signifikan
6	6	0.664	Sangat Signifikan
7	7	0.598	Signifikan
8	8	0.626	Sangat Signifikan
9	9	0.560	Signifikan
10	10	0.683	Sangat Signifikan
11	11	0.624	Sangat Signifikan
12	12	0.697	Sangat Signifikan
13	13	0.401	-
14	14	0.561	Signifikan
15	15	0.140	-

Picture 1: Validity Analysis Of Multiple Test Question
 Source: Anathes Program

Tabel 1: A Percentage Of The Validity Of The Multiple-Test Questions

No	Validity Criteria	Item Number	Total	Percent
1	>0,482 (Valid)	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14	12	80%
2	<0,482 (Invalid)	3, 13, 15	3	20%

Based on the previous table, the researcher prepared the following chart:



2. Validity analysis of matching test questions

The researcher used the Anathes program to analyze the validity of Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta. After the previous calculation, the result is matched with the r table at the level of significance of 5% and with five questions, so the r Table is 0,576. And if $r_{table} \leq r_{account}$, then the question items are valid. But if the $r_{table} \geq r_{account}$, then the question items are invalid.

The validity result for five questions items of the matching test questions, in general, is four questions are valid (80%). It is item number 1, 3, 4 and 5. As for invalid questions, there are one question (20%), which are item number 2.

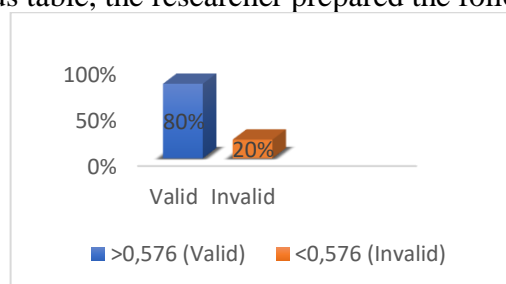
No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0.786	Sangat Signifikan
2	2	0.500	-
3	3	0.799	Sangat Signifikan
4	4	0.771	Sangat Signifikan
5	5	0.786	Sangat Signifikan

Picture 2: Validity Analysis Of Matching Test Question
 Source: Anathes Program

Table 2: A Percentage Of The Validity Of The Matching Test Questions

No	Validity Criteria	Item Number	Total	Percent
1	>0,576 (Valid)	1, 3, 4, 5,	4	80%
2	<0,576 (Invalid)	2	1	20%

Based on the previous table, the researcher prepared the following chart:



Reliability

In addition, a good test must be valid. The test also must have good reliability. Although reliability is when an instrument is used repeatedly to measure the same thing, the results are relatively stable and consistent (Kumar, 2011). And for analyze the reliability of Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta, the researcher used the Anathes program. And the criteria used to interpret reliability are as follows (Firmansyah, 2018):

Table 3: The Description Of Reliability Criteria

The criteria of reliability	The description of reliability
0,801-1,000	very high
0,601-0,800	High
0,401-0,600	Medium
0,201-0,400	low
0,000-0,200	very low

The reliability of the multiple test questions is 0.87, so the test is in the very high-reliability group. As for the reliability of the matching test question is 0.76, this test is considered in the high reliability group. With these two results, the reliability of the exam question items is good. As for these values based on the results of the analysis below:

1. Reliability analysis of multiple test questions

The Anathes program used the Split-Half Method, which is in the Spearman-Brown formula. Based on the previous table, it can be concluded that the reliability of multiple test questions is 0.87, so the test is in the very high reliability group.

Rata2= 11.81
Simpang Baku= 3.36
KorelasiXY= 0.77
Reliabilitas Tes= 0.87

Picture 3: Reliability Analysis Of Multiple Test Question
 Source: Anathes Program

2. Reliability analysis of matching test questions

The Anathes program used the Split-Half Method, which is in the Spearman-Brown formula. Based on the previous table, it can be concluded that the reliability of matching test questions is 0.76, so the test is in the high reliability group.

Rata2= 3.70
Simpang Baku= 1.59
KorelasiXY= 0.61
Reliabilitas Tes= 0.76

Picture 4: Reliability Analysis Of Matching Test Question
 Source: Anathes Program

Difficulty level

A good test is certainly neither the easiest nor the most difficult. Because if the test is too easy or too difficult, the ability of the respondents cannot be known. The reason is the exams that are too easy do not interest students, and exams that are too difficult make students inactive in judging. The list of the good degree of difficulty level for the questions is on the ratio of 2.5:5:2.5 or it means the degree of good difficulty is if it reaches 25% of the easy question, 50% of the medium question, and 25% of the hard question (Firmansyah, 2018). The difficulty level test of the multiple test questions categorized as hard is 0 items (0%), the medium is 3 items (20%), and the easy is 12 items (80%). And the difficulty level test of the matching test questions categorized as hard is 0 items (0%), the medium is 4 items (80%), the easy is 1 item (20%). And the criteria used to interpret reliability are as follows (Mahmudi & Ahmad, 2020):

Table 4: The Description Of Difficulty Level Criteria

The criteria of difficulty level	The description of difficulty level
0,00-0,30	Hard
0,31-0,70	Medium
0,71-1,00	Easy

So the degree of difficulty level for the items of the exam questions for the Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta was not good because it did not fit into the list of good difficulty level for the questions that were on the ratio of 2.5:5:2.5. As for these values based on the results of the analysis below:

1. Difficulty level analysis of multiple test questions

The results of the difficulty level test used the Anathes for Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta, are as follows, the difficulty level test of the multiple test questions categorized as hard is 0 items (0%), the medium is 3 items (20%) that are number 2, 6 and 9, the easy is 12 items (80%) that are number 1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14 and 15.

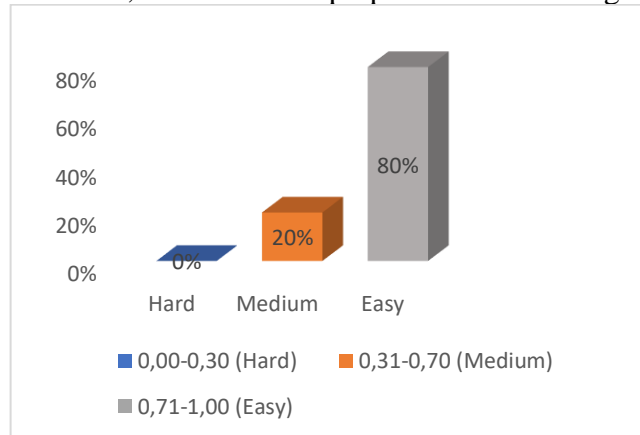
Table 5: A Percentage Of The Difficulty Level Of The Multiple Test Questions

No	Difficulty level Criteria	Item Number	Total	Percent
1	0,00-0,30 (Hard)	-	-	0%
2	0,31-0,70 (Medium)	2, 6, 9	3	20%
3	0,71-1,00 (Easy)	1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15	12	80%

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)
1	1	66	82.50
2	2	50	62.50
3	3	74	92.50
4	4	65	81.25
5	5	66	82.50
6	6	52	65.00
7	7	64	80.00
8	8	68	85.00
9	9	54	67.50
10	10	66	82.50
11	11	60	75.00
12	12	59	73.75
13	13	67	83.75
14	14	59	73.75
15	15	75	93.75

Picture 5: Difficulty Level Analysis Of Multiple Test Questions
 Source: Anathes Program

Based on the previous table, the researcher prepared the following chart:



2. Difficulty level analysis of matching test questions

The results of the difficulty level test used the Anathes for Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta, are as follows, the difficulty level test of the matching test questions categorized as hard is 0 items (0%), the medium is 4 items (80%) that are number 1, 3, 4 and 5, the easy is 1 item (20%) that is number 2.

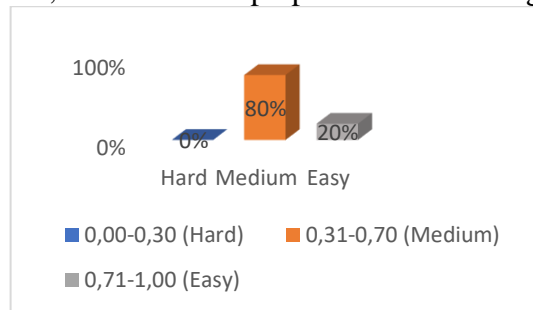
No Butir Baru	No Butir Asli	Tkt. Kesukaran(%)
1	1	63.64
2	2	88.64
3	3	52.27
4	4	59.09
5	5	59.09

Picture 6: Difficulty Level Analysis Of Matching Test Questions
 Source: Anathes Program

Table 6: A Percentage Of The Difficulty Level Of The Matching Test Questions

No	Difficulty level Criteria	Item Number	Total	Percent
1	0,00-0,30 (Hard)	-	-	0%
2	0,31-0,70 (Medium)	1, 3, 4, 5	4	80%
3	0,71-1,00 (Easy)	2	1	20%

Based on the previous table, the researcher prepared the following chart:



Differential Power

Differential power analysis Questions benefit to know which students have understood the subject and the students who did not understand the subject and also to know the smart student and the weak student (Rahman & Nasryah, 2019). The difficulty level test of the Multiple test questions categorized as bad is 1 item (6,66%), enough is 2 items (13,3%), good is 9 items (60%), Very Good is 3 items (20%). The difficulty level test of the matching test questions categorized as bad is 0 items (0%), enough is 1 item (20%), good is 0 items (0%), Very Good is 4 items (80%). And the criteria used to interpret differential power are as follows (Mahmudi, 2020a):

Table 7: The Description Of Difficulty Level Criteria

The criteria of differential Power	The description of differential Power
0,00-0,20	Bad
0,21-0,40	Enough
0,41-0,70	Good
0,71-1,00	Very Good

In view of the previous result, the questions have reached 80% so the differential power of this test is good and be able to differentiate the students' competence in terms of their comprehension of the subject also to differentiate between the smart students and the weak students. As for these values based on the results of the analysis below:

1. Differential Power analysis of multiple test questions

The results of the differential Power test used the Anathes for Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta, are as follows, the difficulty level test of the Multiple test questions categorized as bad is 1 item (6,66%) that is number 15, enough is 2 items (13,3%) that are number 3 and 13, good is 9 items (60%) that are number 1, 2, 4, 5, 7, 8, 10, 11 and 14, Very Good is 3 items (20%) that are number 6, 9 dan 12.

No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	21	11	10	45.45
2	2	21	6	15	68.18
3	3	22	17	5	22.73
4	4	22	11	11	50.00
5	5	22	12	10	45.45
6	6	22	5	17	77.27
7	7	21	10	11	50.00
8	8	22	10	12	54.55
9	9	22	6	16	72.73
10	10	22	9	13	59.09
11	11	22	8	14	63.64
12	12	22	6	16	72.73
13	13	22	14	8	36.36
14	14	22	8	14	63.64
15	15	22	20	2	9.09

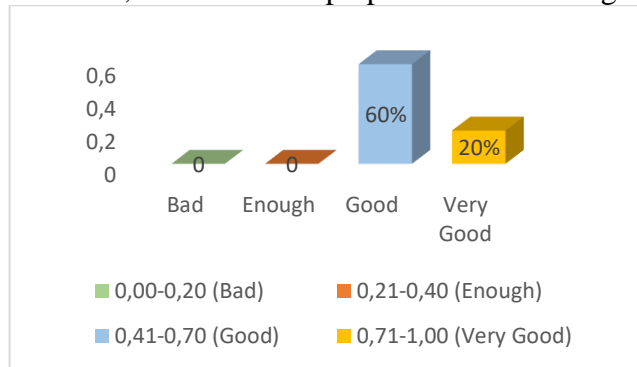
Picture 7: Differential Power Analysis Of Multiple Test Questions

Source: Anathes program

Table 8: A Percentage Of The Differential Power Of The Multiple Test Questions

No	Differential Power Criteria	Item Number	Total	Percent
1	0,00-0,20 (Bad)	15	1	6,66%
2	0,21-0,40 (Enough)	3, 13	2	13,3%
3	0,41-0,70 (Good)	1, 2, 4, 5, 7, 8, 10, 11, 14	9	60%
4	0,71-1,00 (Very Good)	6, 9,12	3	20%

Based on the previous table, the researcher prepared the following chart:



2. Differential Power analysis of matching test questions

The results of the differential Power test used the Anathes for Arabic language examination at Luqman Hakim International Integrated Islamic Junior High School Bantul Yogyakarta, are as follows, the difficulty level test of the matching test questions categorized as bad is 0 items (0%), enough is 1 item (20%) that is number 2, good is 0 items (0%), Very Good is 4 items (20%) that are number 1, 3, 4 and 5.

No	No Btr Asli	Rata2Un	Rata2As	Beda	SB Un	SB As	SB Gab	t	DP(%)
1	1	1.00	0.27	0.73	0.00	0.46	0.10	7.48	72.73
2	2	1.00	0.77	0.23	0.00	0.43	0.09	2.49	22.73
3	3	1.00	0.05	0.95	0.00	0.21	0.05	2.11	95.45
4	4	1.00	0.18	0.82	0.00	0.39	0.08	9.72	81.82
5	5	1.00	0.18	0.82	0.00	0.39	0.08	9.72	81.82

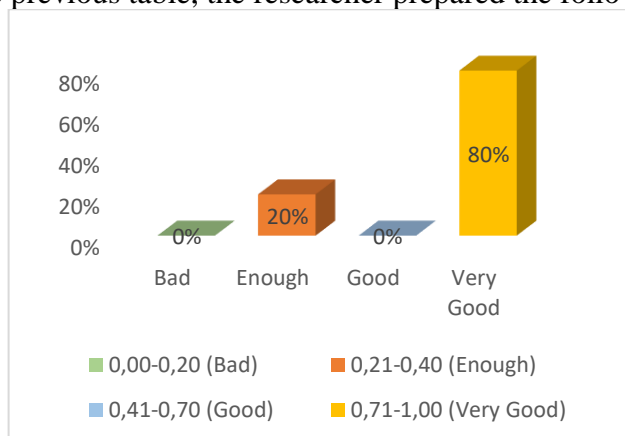
Picture 8: Differential Power Analysis Of Matching Test Question

Source: Anathes program

Table 9: A Percentage Of The Differential Power Of The Matching Test Questions

No	Differential Power Criteria	Item Number	Total	Percent
1	0,00-0,20 (Bad)	-	-	0%
2	0,21-0,40 (Enough)	2	1	20%
3	0,41-0,70 (Good)	-	-	0%
4	0,71-1,00 (Very Good)	1, 3, 4, 5	4	80%

Based on the previous table, the researcher prepared the following chart:



CONCLUSION

The results showed that: (1) The validity test of 15 multiple choice test questions was declared valid (80%). As well as for the validity test of 5 matching test questions they were declared valid (80%). (2) The reliability test of multiple choice test items is 0.87. At the same time, the results of the reliability of matchmaking test items were 0.76. (3) The difficulty level test of the multiple test questions categorized as hard is 0 items (0%), medium is three items (20%), and easy is 12 items (80%). The difficulty level test of the matching test questions categorized as hard is 0 items (0%), medium is four items (80%), and easy is 1 item (20%). (4) The difficulty level test of the Multiple test questions categorized as bad is 1 item (6,66%), enough is two items (13,3%), good is nine items (60%), and Very Good is three items (20%). The difficulty level test of the matching test questions categorized as bad is 0 items (0%), enough is 1 item (20%), good is 0 items (0%), and Very Good is four items (80%).

The school should hold a program or workshop to reveal the teacher's information about evaluation, especially in analyzing the question items, so that the teacher can make good questions in terms of their validity, reliability, difficulty level, Differential Power, and also increase the teacher's observation of the necessity of analyzing the question items raises.

As the author of the question, the teacher should increase his observation towards analyzing the items of the questions in terms of their validity, reliability, difficulty level, and Differential Power. To analyze the question items, the teacher must note the appropriateness and excitement of the test questions because they help attract students' attention when answering the question. Also, it is important note for the teacher to know the types and question forms before compiling the assessment tool. Because it can make it easier for teachers to determine the appropriate result for all items of questions so that the teacher can facilitate the analysis of questions, this research concludes that all teachers must take care of their Arabic writing and speaking so as not to cause corruption of Arabic terminology in the eyes of the students, in their hearing, or their pronunciation by using the Arabic language.

REFERANCES

- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development*, 82(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Fikriyah, N. (2021). Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas VII Semester Ganjil SMP Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019 / 2020. *Maharaat: Jurnal Pendidikan Bahasa Arab*, 3(2), 130. <https://doi.org/10.18196/mht.v3i2.10501>
- Firmansyah, H. (2018). Analisis Butir Soal Tes Pilihan Ganda Mata Pelajaran Jasmani Olahraga dan Kesehatan Kelas VIII SMP N 1 Wonosari Tahun Ajaran 2017-2018. Universitas Negeri Yogyakarta.
- Kumar, R. (2011). *Research Methodology a step-by-step guide for beginners* (3rd editio). New Delhi: Sage Publication India Pvt Ltd.

- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682>
- Leona. (1997). *Al-Ikhtibaaraat wa al-Maqayis*. Kairo: Dar al-Syuruuq.
- Madkur, A. A. (1991). *Tadriisu Funuuni al-Lughoh al-'Arobiyah*. Kairo: Dar al-Syawaf.
- Mahmudi, I. (2020a). *Evaluasi Pendidikan*. Sleman: Lintang Books.
- Mahmudi, I. (2020b). *Pengembangan Instrumen Penelitian Sosial*. Sleman: Lintang Books.
- Mahmudi, I., & Ahmad, R. R. F. (2020). Item Analysis of Islamic Education For Class VIII at Junior High School 1 Jetis Ponorogo. *Educan : Jurnal Pendidikan Islam*, 4(1), 118. <https://doi.org/10.21111/educan.v4i1.4565>
- Makhisoh, I., & Suroyya, E. N. (2022). Analisa Butir Soal Ujian Madrasah Bahasa Arab Menurut Taksonomi Bloom. *Muhadasah: Jurnal Pendidikan Bahasa Arab*, 4(1), 40–52.
- Mustopa, A., Basri, H., & Barlian, U. C. B. (2021). Analisis standar penilaian pendidikan. *Jurnal Manajemen Pendidikan* Vol. 09 No. 01, Januari 2021, P. 24.
- Nasional, D. P. (2008). *Panduan Penulisan Butir Soal*.
- Oueslati, O., Cambria, E., HajHmida, M. Ben, & Ounelli, H. (2020). A review of sentiment analysis research in Arabic language. *Future Generation Computer Systems*, 112, 408–430. <https://doi.org/10.1016/j.future.2020.05.034>
- Rahman, A. A., & Nasryah, C. E. (2019). *Evaluasi Pembelajaran*. In *Uwais Inspirasi Indonesia*. Ponorogo: Uwais Inspirasi Indonesia.
- Ridho, U. (n.d.). *Evaluasi Dalam Pembelajaran Bahasa Arab*. *Jurnal An-Nabighoh* Vol. 20 No. 1, 2018, P. 20.
- Rusdiana, E. ratna W. (2014). *Evaluasi Pembelajaran*. Bandung: Pustaka Setia.
- Shaker, M. S., Wallace, D. V., Golden, D. B. K., Oppenheimer, J., Bernstein, J. A., Campbell, R. L., ... Wang, J. (2020). Anaphylaxis—a 2020 practice parameter update, systematic review, and Grading of Recommendations, Assessment, Development and Evaluation (GRADE) analysis. *Journal of Allergy and Clinical Immunology*, 145(4), 1082–1123. <https://doi.org/10.1016/j.jaci.2020.01.017>
- Taylor, S., Landry, C. A., Paluszek, M. M., Fergus, T. A., McKay, D., & Asmundson, G. J. G. (2020). Development and initial validation of the COVID Stress Scales. *Journal of Anxiety Disorders*, 72, 102232. <https://doi.org/10.1016/j.janxdis.2020.102232>
- Undang-undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. (2003).