D O I : 10.18860 /ijazarabi. V8i2.29595
I S S N (p r i n t) : 2 6 2 0 - 5 9 1 2 | I S S N (o n l i n e) : 2 6 2 0 - 5 9 4 7
e j o u r n a l . u i n - m a l a n g . a c . i d / i n d e x . p h p / i j a z a r a b i / i n d e x

Analysis Of Online Arabic Language Proficiency Test Questions Based On Bloom's Cognitive Taxonomy

Ihwan Mahmudi*[1], Hanifa Bujanah[2], Moh. Ismail[3], Ahmad Saifulloh[4], Izzul Fatawi[5], Maghfirotul Hasanah[6]
[1,2,3,4,6]Pendidikan Bahasa Arab, Fakultas Tarbiyah, Universitas Darusalam Gontor, Ponorogo, Indonesia, [5]Universitas Terbuka, Indonesia
ihwanm@unida.gontor.ac.id*[1], hanifabujanah48@student.pba.unida.gontor.ac.id[2], ismail@unida.gontor.ac.id[3], saif@unida.gontor.ac.id[4], izzul.official@ecampus.ut.ac.id[5], maghfirotulhasanah45@student.pba.unida.gontor.ac.id[6]

Abstract
One of the visions of the University of Darussalam Gontor is to become a center for the language of the Qur'an. An online Arabic language test is conducted to measure its students' Arabic proficiency. This study aims to analyze the difficulty level of the online Arabic language test at the University of Darussalam Gontor based on Bloom's Cognitive Taxonomy. The research employs a descriptive quantitative method. The data was collected from students' test scores, which cover three language skills: reading (qira'ah), listening (istima'), and writing (kitabah). The sample was randomly selected, comprising 135 students for the writing test, 127 for the listening test, and 129 for the reading test. The results indicate that the difficulty level of the questions based on students' test results for the reading skill shows that 10% of the questions were classified as difficult, 42% as middle, and 48% as easy. For the listening skill, 17% of the questions were classified as difficult, 44% as middle, and 39% as easy. There were no difficult questions in the writing skill, with 83% being middle and 17% easy. Regarding the cognitive domains of Bloom's taxonomy, 21% of the questions in reading test (qiraah) were categorized as HOTS, 52% as MOTS, and 17% as LOTS. For the listening test (istima), 11% of the questions were classified as HOTS, 67% as MOTS, and 11% as LOTS, while for the writing test, 10% were categorized as HOTS, 55% as MOTS, and 25% as LOTS.
Keywords: Evaluation; Difficulty Level; Bloom’s Taxonomy; HOTS; Arabic Proficiency Test.

INTRODUCTION

The evaluation of Arabic language learning in higher education plays an important role in equipping students with optimal language skills. However, the effectiveness of Arabic learning evaluation still faces various significant challenges. The evaluation tends to focus on theoretical aspects without assessing comprehensive language skills (Yakin, 2022), limitations in evaluation methods, lack of technological integration in the assessment process (Al-Batineh & Al-Tenaijy, 2024), as well as the low validity and reliability of tests (Alkhamra & Al-Jazi, 2016), (Fekih-Romdhane et al., 2024), are the main factors hindering evaluation effectiveness. In addition, limited time and frequency of test implementation (Khodeir et al., 2018), also contribute to the suboptimal mapping of students' competencies. This is in line with the opinion of Muhammad Habib Thaha, as cited in the article by Moh. Ismail and Aufa Alfian Musthafa, who stated that the tests

given to students serve as a standard for the success of Arabic language learning and as a means to assess students' abilities (Ismail & Musthafa, 2018).

The success of learning evaluation is achieved through quality standardization in the creation of each test item, because test item is a systematic procedure designed in the form of standardized tasks and given to individuals or groups to be completed, answered, or responded to whether in written, oral, or performance form (Ramadhan et al., 2024), this was also conveyed by Azwar, who stated that a test is a systematic procedure in which the items are arranged according to specific methods and rules (Azwar, n.d.), This view is also supported by Silverius, as cited in an article written by Zulkifli, who defined a test as a systematic procedure used to observe and describe one or more characteristics of an individual using a numerical scale or a categorical system (Matondang, 2009). One widely used concept is Bloom's Cognitive Taxonomy (Mahmudi, n.d.). Cognitive Taxonomy or intellectual abilities of Bloom has six levels, ranging from the lowest to the highest levels (Adams, 2015). These are generally categorized into three thinking classifications: HOTS, MOTS, and LOTS (Ritonga et al., n.d.)

HOTS (Higher Order Thinking Skills) consist of three levels C4 (analyze), C5 (evaluate), and C6 (create) (Listiani & Rachmawati, 2022). MOTS (Middle Order Thinking Skills) corresponds to C3 (apply), and LOTS (Lower Order Thinking Skills) corresponds to C2 (understand) and C1 (remember) (Rais & Ramadhani, 2023). These three thinking categories, covering Bloom's six cognitive levels, are often used to determine the difficulty level of test items in an exam. In creating tests, the difficulty level of each question should include easy, middle, and difficult items. The difficulty level is based on students' ability to answer, not the perspective of the question creator (Yadnyawati, 2019). The standard commonly used is 25% difficult, 50% middle, and 25% easy questions (Khoirunnisa & Widodo, 2022), difficulty is classified as difficult (less than 0.30), middle (0.31 – 0.70), and easy (greater than 0.70) (Mahmudi, 2020).

The online Arabic test at UNIDA Gontor is one of the tools used to measure the achievement of the university's vision as a center for Qur'anic language development. The test is designed based on the CEFR (Common European Framework of Reference for Languages), a European framework for language proficiency standardization for non-native speakers (Musthofa, 2022). Language standardize test according to CEFR consist of three main criteria: comprehensiveness, transparency, and coherence, aiming to identify various inclusive language knowledge and skills that are explicitly described and easily accessible, ensuring that the different components of a language program complement each other to achieve coherence (Mohamed, 2023).

The language proficiency standards set by the University of Darussalam Gontor, in accordance with CEFR guidelines, consist of six levels (Ismail et al., 2023): Beginner (A1), Elementary (A2), Intermediate (B1), Upper Intermediate (B2), Advanced (C1), and Proficient (C2). The online Arabic test used the platform alikhtibar.com as the official language test platform. This platform offers various features for language learning, including traditional tests and game-based assessments to reduce boredom in learning foreign languages. However, there are still some difficulties in answering the presented questions, which affect the students' success in the exams.

Based on the Arabic test results, when classified according to the standards of the University of Darussalam Gontor's Language Development Directorate, 31% of students were at the Pre-A1 (Pre-Beginner) level, 25% at A1 (Beginner), 27% at A2 (Threshold), 14% at B1 (Intermediate), and 3% at B2 (Advanced).

This issue is likely due to the difficulty level of each test item. In any test, item difficulty analysis significantly influences test outcomes. Previous studies by Rita Wahyu Utami (2018) on the analysis of Arabic Foreign Language Test items from the perspective of Bloom's Taxonomy (Utami, 2018), by Iis Makhisoh and Evi Nurus Suroyya (2022) on the analysis of Arabic exam items in Islamic schools (Makhisoh & Surayya, 2022), and by Ihwan Mahmudi et al. (2023) on Arabic language exam item analysis (Mahmudi & Nurwardah, 2023) show a correlation between the difficulty level of test items and student scores. Additional studies linking test item difficulty with Bloom's taxonomy levels include works by Iqbal Faza Ahmad and Sukiman (Ahmad & Sukiman, 2019) and Ihwan Mahmudi et al. on HOTS-based assessment models for Arabic learning (Mahmudi, Naqiyah, et al., 2023; Mahmudi, Masturoh, et al., 2023). Based on the above background, this research aims to analyze the difficulty level of test items in the online Arabic language test at the University of Darussalam Gontor from the perspective of Bloom's Cognitive Theory.

METHOD

This research employs a descriptive quantitative method. The data were collected from the Arabic language test scores of students, covering three skills: reading (qira'ah), listening (istima'), and writing (kitabah). The sample was randomly selected, with 135 students for the writing test, 127 for the listening test, and 129 for the reading test. The data analysis technique used the formula for calculating item difficulty levels.

The reading test consisted of 29 items, including 6 multiple-choice questions, 18 true/false questions, 2 completion questions, 1 matching question, and 2 short-answer questions. The listening test included 18 items, comprising 6 short-answer questions, 8 items are true/false questions, and 4 multiple-choice questions. The writing test contained 29 items, which included 5 matching questions, 8 true/false questions, 14 multiple-choice questions, and 3 completion questions.

RESULTS AND DISCUSSION

Analysis of Item Difficulty Levels in the Online-Based Arabic Proficiency Test

Analysis of item difficulty level of each test item in the online Arabic language test for the *Maharah Qira'ah* section, there are 29 items, which can be classified according to their difficulty levels as shown in the following table:

Table 1. Analysis Of The Difficulty Level For The *Maharah Qira'ah* Section

No.	Difficulty Level Indicator	Test Items	Quantity	Percentage
1.	0,00 – 0,30 (Difficult)	11, 12, 28	3	10 %
2.	0,31 – 0,70 (Middle)	1, 3, 5, 10, 16, 21, 22, 23, 25, 26, 27, 29	12	42 %
3.	0,71 – 1,00 (Easy)	2, 4, 6, 8, 9, 13, 14, 15, 16, 17, 18, 19, 20, 24	14	48 %

Analysis of the difficulty level for the *Maharah Istima'* section, there are 18 items, which can be classified according to their difficulty levels as shown in the following table:

Table 2. Analysis Of The Difficulty Level For The *Maharah Istima'* Section

No.	Difficulty Level Indicator	Test Items	Quantity	Percentage
1	0.00 – 0.30 (Difficult)	10, 11, 12	3	17%
2	0.31 – 0.70 (Middle)	5, 6, 8, 9, 13, 16, 17, 18	8	44%
3	0.71 – 1.00 (Easy)	1, 2, 3, 4, 7, 14, 15	7	39%

Analysis of the difficulty level for the *Maharah Kitabah* section, there are 29 items, which can be classified according to their difficulty levels as shown in the following table:

Table 3. Analysis Of The Difficulty Level For The *Maharah Kitabah* Section

No.	Difficulty Level Indicator	Test Items	Quantity	Percentage
1	0.00 – 0.30 (Difficult)	-	0	0 %
2	0.31 – 0.70 (Middle)	1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28	24	83 %
3	0.71 – 1.00 (Easy)	4, 6, 17, 26, 29	5	17 %

Analysis of Item Difficulty Levels in the Online-Based Bloom's Cognitive Taxonomy

When analyzing each test item of reading skill (*Maharatul Qira'ah)* by classifying them according to Bloom's cognitive taxonomy, the items are classified based on the levels of thinking skills: HOTS (Higher Order Thinking Skills), MOTS (Middle Order Thinking Skills), and LOTS (Lower Order Thinking Skills), as follows:

Table 4. The Difficulty *Maharah Qira'ah Based On* Bloom's Cognitive Taxonomy

No.	Level of Thinking Skills	Test Items	Quantity	Percentage
1	HOTS (Higher Order Thinking)	21, 22, 26, 27, 28, 29	6	27 %
2	MOTS (Middle Order Thinking)	4, 5, 6, 7, 8, 9, 10, 16, 17, 18, 19, 20, 23, 24, 25,	15	52 %
3	LOTS (Lower Order Thinking)	1, 2, 3, 11, 12, 13, 14, 15	8	21 %

Difficulty level of listening skill *(maharatul Istima)* according to Bloom's cognitive taxonomy:

Table 5. The Difficulty *Maharah Istima' Based On* Bloom's Cognitive Taxonomy

No.	Level of Thinking Skills	Test Items	Quantity	Percentage
1	HOTS (Higher Order Thinking)	10, 11, 12, 13	4	22 %
2	MOTS (Middle Order Thinking)	5, 6, 7, 8, 9, 14, 15, 16, 17, 18	10	56 %
3	LOTS (Lower Order Thinking)	1, 2, 3, 4	4	22 %

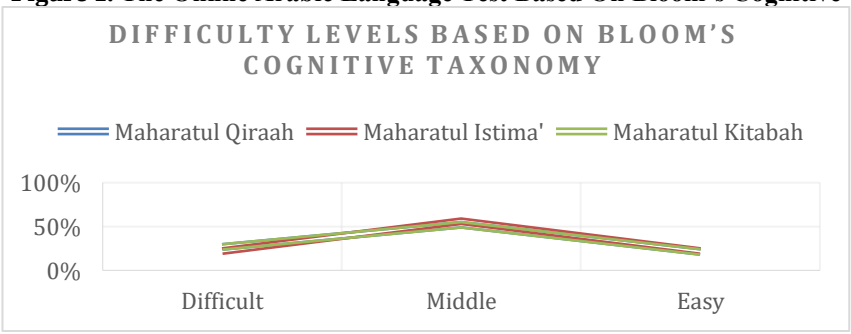
The difficulty level of writing skill *(maharatul kitabah)* test according to Bloom's cognitive taxonomy:

Table 6. The Difficulty *Maharah Kitabah Based On* Bloom's Cognitive Taxonomy

No.	Level of Thinking Skills	Test Items	Quantity	Percentage
1	HOTS *(Higher Order Thinking)*	1, 3, 8, 9, 11, 26	6	27 %
2	MOTS *(Middle Order Thinking)*	2, 4, 5, 6, 7, 10, 12, 13, 14, 15, 16, 22, 23, 24, 25	15	52 %
3	LOTS *(Lower Order Thinking)*	17, 18, 19, 20, 21, 27, 28, 29	8	21 %

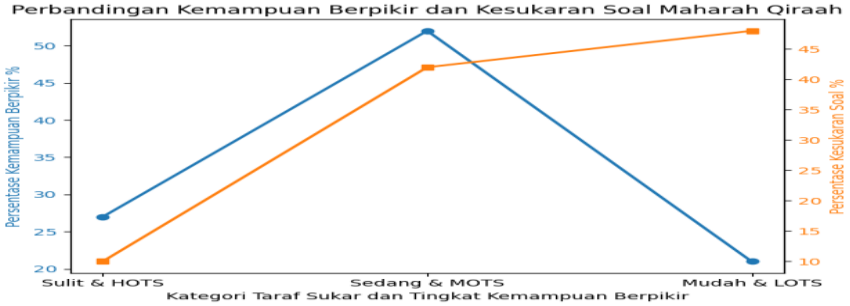
Generally, the level of difficulty of the items in the online Arabic Language test based on Bloom's Cognitive shown in the following picture:

Figure 1. The Online Arabic Language Test Based On Bloom's Cognitive



From the results of the item difficulty analysis based on students' scores and Bloom's cognitive taxonomy, a slight difference is observed, as shown in the graph below:

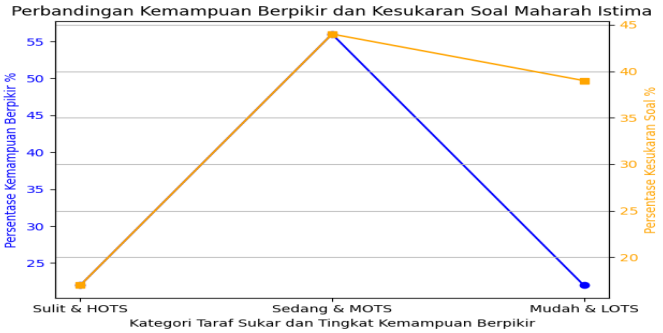
Figure 2. Comparison Of The Two Analyses In The *Maharah Qira'ah* Section



Based on the graph, it can be observed that the test items in the *Maharah Qira'ah* section, when viewed in terms of difficulty levels (indicated by the yellow line), show an imbalance across the three categories (Easy, Middle, and Difficult). The easy-level questions have significantly higher values, far exceeding those of the difficult-level questions.

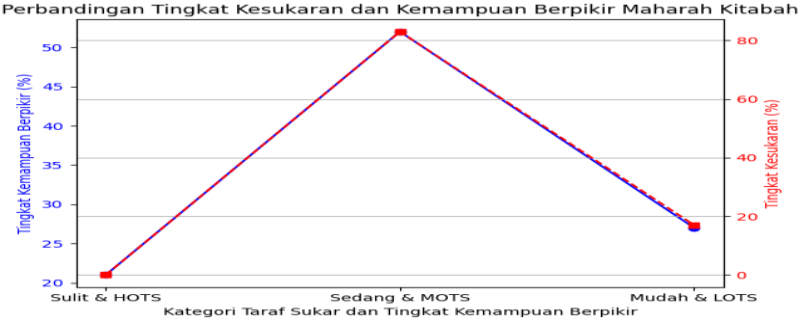
When viewed from the perspective of Bloom's Cognitive Taxonomy (indicated by the blue line), the composition of the *Maharah Qira'ah* questions is more balanced among the three difficulty levels (Easy, Middle, and Difficult) based on the standardized composition of 25% Easy, 50% Middle, and 25% Difficult.

Figure 3. Comparison Of The Two Analyses In The *Maharah Istima'* Section



Based on the graph, it can be observed that the test items in the *Maharah Istima'* section, when viewed in terms of difficulty levels (indicated by the yellow line), show an imbalance across the three categories (Easy, Middle, and Difficult). The middle-level questions have significantly higher values compared to the difficult-level questions, indicating a substantial gap. On the other hand, when viewed from the perspective of Bloom's Cognitive Taxonomy (indicated by the blue line), the *Maharah Istima'* questions show a more balanced composition among the three difficulty categories (Easy, Middle, and Difficult), aligning more closely with the standardized distribution of 25% Easy, 50% Middle, and 25% Difficult.

Figure 4. Comparison Of The Two Analyses In The *Maharah Kitabah* Section



Based on the graph, it can be seen that the test items in the *Maharah Kitabah* section, when analyzed in terms of difficulty level (indicated by the red line), show an imbalance among the three categories (Easy, Middle, and Difficult). The middle-level questions have significantly higher values, and no difficult-level questions were found in this analysis. When analyzed based on Bloom's Cognitive Taxonomy (indicated by the blue line), the *Maharah Kitabah* questions demonstrate a more balanced composition among the three difficulty categories (Easy, Middle, and Difficult), adhering more closely to the standard composition of 25% Easy, 50% Middle, and 25% Difficult (Khoirunnisa & Widodo, 2022). The analysis if item difficulty is an integral part of assessing test quality, (Bichi & Talib, 2018), making the evaluation and assessment process more precise and accurate (H. Akhtar & K. Kovacs, 2023) in measuring the test-takers' abilities.

CONCLUSION

Based on the data analysis, this research indicate that the difficulty level of the test items for students' tests in the *Maharah Qira'ah* section shows that 10% of the questions fall into the difficult category, 42% are in the middle category, and 48% are in the easy category. For *Maharah Istima'*, 17% of the questions are categorized as difficult, 44% as middle, and 39% as easy. In the *Maharah Kitabah* section, there are no difficult questions, with 83% classified as middle and 17% as easy. In terms of Bloom's cognitive taxonomy, the difficulty level analysis reveals that, for the *Maharah Qira'ah* test, 21% of the questions are categorized as HOTS, 52% as MOTS, and 17% as LOTS. For the *Maharah Istima'* test, 11% of the questions are categorized as HOTS, 67% as MOTS, and 11% as LOTS. And or the *Maharah Kitabah* test, 10% as HOTS, 55% as MOTS, and 25% as LOTS.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Directorate of Research, Technology, and Community Service, Ministry of Higher Education, Science, and Technology of The Republic Indonesia, for funding this research through the 2024 fiscal Year National Competitive Applied Research Grant Scheme under the Contract number 059/SP2H/PT/LL7/2024.

REFERENCES

Adams, N. E. (2015). *'Bloom's Taxonomy of Cognitive Learning Objectives.' 103, No, 3*, 152–153. https://doi.org/10.3163/1536-5050.103.3.010.

Ahmad, I. F., & Sukiman. (2019). *Analisa Higher Order Thinking Skills (HOTS) Pada Soal Ujian Akhir Siswa Kelas 6 Kmi Dalam Kelompok Mata Pelajaran Dirasah Islamiyah Di Pondok Modern Tazakka Batang*. *2, No, XVI*. https://doi.org/<https://doi.org/10.14421/jpai.2019.162-02>

Al-Batineh, M., & Al-Tenaijy, M. (2024). *Adapting to technological change: An investigation of translator training and the translation market in the Arab world. 10, No, 7*, 28535.

Alkhamra, RA., & Al-Jazi, AB. (2016). *Validity and reliability of the Arabic Token Test for children*. *51, No, 2*, 183–191.

Azwar. (n.d.). *Tes Prestasi, Fungsi, dan Pengembangan Pengukuran Prestasi Belajar* (II Cetakan VI). Pustaka Pelajar 2007.

Bichi, AA., & Talib, R. (2018). *Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development*. *7, No, 2*, 142.

Fekih-Romdhane, F., Mouzakzak, F. A., Abilmona, G., Moustafa, A., Dahdouh, O., & Hallit, S. (2024). *Arabic translation and validation of the clinician administered Staden schizophrenia anxiety rating scale (S-SARS)*. *10, No, 21*, 39792.

H. Akhtar & K. Kovacs. (2023). *Measurement precision and user experience with adaptive versus non-adaptive psychometric tests*. 225. https://doi.org/10.1016/j.paid.2024.112675

Ismail, Moh., & Musthafa, A. A. (2018). *Pengembangan Materi Tes Keterampilan Menyimak Bahasa Arab Menggunakan Moodle Untuk Meningkatkan Kemampuan Menyimak Mahasiswa Universitas Darussalam Gontor*. *13, No, 2*, 28–49. http://dx.doi.org/10.21111/at-tadib.v13i2.2467

Ismail, Moh., Zaid, A. H. bin, & Rochma, S. N. (2023). *Tathwir Ma'ayir Ikhtibarat al-Lughah al-'Arabiyah al-Muqananah fi Barnamaj alikhtibar.com wa Atsrauha fi Qiyas Kafaah al-Thulab al-Nathiqin bi Ghairiha. 14, No, 1*, 65–81.

Khodeir, MS., Hegazi, MA., & Saleh, MM. (2018). *Development and standardization of a test for pragmatic language skills in Egyptian Arabic: The Egyptian Arabic pragmatic language test (EAPLT)*. *69, No, (5–6)*, 209–218.

Khoirunnisa, S., & Widodo, E. (2022). *Difficulty Level, Distinguishing Power, Distractor Effectiveness, And Anova (Case Study: Paper Based Test (PBT) Period 2 Questions*. *16, No, 2*, 542.

Listiani, W., & Rachmawati. (2022). *'Transformasi Taksonomi Bloom Dalam Evaluasi Pembelajaran Berbasis2022 HOTS.' 2*, 340.

Mahmudi, I. (n.d.). *Taksonomi Hasil Belajar Menurut Benyamin S. Bloom*. *2, No, 9*, 3507. https://doi.org/10.55927/mudima.v2i9.1132.

Mahmudi, I. (2020). *Evaluasi Pendidikan* (1st ed.). Lintang Books.

Mahmudi, I., Masturoh, F., & Febrianti, W. D. (2023). *Higher Order Thinking Skills (HOTS) – Based Assessment: A Proposed Model for Arabic Learning*. *1*.

Mahmudi, I., Naqiyah, N., & Setiyadi, A. C. (2023). *The Influence Of High Order Thinking Skill (Hots) Based Questions On Arabic Language Learning Outcomes Of Madrasah Tsanawiyah Student*. *18, No, 2*, 49–69. http://dx.doi.org/10.21111/attadib.v18i2.10547

Mahmudi, I., & Nurwardah, A. (2023). *Item Analysis of Arabic Language Examination*. <https://ejournal.uin-malang.ac.id/index.php/ijazarabi/article/view/19821>

Makhisoh, I., & Surayya, E. N. (2022). *Analisa Butir Soal Ujian Madrasah Bahasa Arab Menurut Taksonomi Bloom*. *1*, *No, 1*, 40–52.

Matondang, Z. (2009). *Validitas dan Reliabilitas Suatu Instrumen Penelitian*. *6, No, 1*.

Mohamed, S. (2023). *The Development of an Arabic Curriculum Framework Based on a Compilation of Salient Features from CEFR Level Descriptors*. *51, No, 1*, 33–47. https://doi.org/10.1080/09571736.2021.1923781

Musthofa, T. (2022). *CEFR-Based Policy in Arabic Language Teaching and Cultural Dimension in Indonesian Islamic Higher Education*. *8, No, 2*, 97–107.

Rais, M. R., & Ramadhani, S. (2023). *Analisa Butir Soal Hots (Higher Order Thinking Skills) Mata Pelajaran Pendidikan Agama Islam Di Sma Negeri 1 Peukan Bada Aceh Besar*. *13, No, 1*, 10.

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). *Validitas and Reliabilitas*. *06, No, 02*. https://doi.org/10.31004/joe.v6i2.4885

Ritonga, A. W., Ritonga, M., Septiana, V. W., & Mahmud. (n.d.). *Crossword Puzzle as a Learning Media during the Covid-19 Pandemic: HOTS, MOTS or LOTS? '*. *1933, No, 1*, 012126. https://doi.org/10.1088/1742-6596/1933/1/012126

Utami, R. W. (2018). Tahlil Bunud Asilah Ikhtibarat Al Lughah Al Arabiyah Lighairin Nathiqina Biha Ala Dhoui Nadhariyah Tasnif Bloom: Dirasah Wasfiyah Tahliliyah Fil Jamiah Al Islamiyah Al Hukumiyah Ponorogo. *(unpublished masters, Universitas Islam Negeri Maulana Malik Ibrahim)*. <http://etheses.uin-malang.ac.id/12295/> [accessed 25 February 2024]

Yadnyawati, I. A. G. (2019). *Evaluasi Pembelajaran*. UNHI Press.

Yakin, A. (2022). *Problematika Pembelajaran Bahasa Arab Di Indonesia Dan Implikasinya Terhadap Pengembangan Kurikulum Bahasa Arab*. *7, No, 1*, 57–67.