The Feasibility of TOAFL Instrument as a Measuring Tool for Reading Skills of Students Faculty of Education

Marsekal Rahman Hakiem[1]*, Muhammad Uwais Abdurrohman[2], Muhammad Irfan Faturrahman[3]

[1] Universitas Madani, Yogyakarta, Indonesia
[2] Islamic University of Madinah, Saudi Arabia
[3] Universitas Ahmad Dahlan, Yogyakarta, Indonesia
*Corresponding E-mail: marsekalrh@umad.ac.id

Keywords:
Arabic;
Instrument;
Reading ability;
TOAFL

Abstract

Evaluation of Arabic language learning requires a valid and reliable instrument to measure students' abilities accurately. This study aims to analyze the feasibility of the ITMAM (Ikhtibar Tahdid Mustawa al-'Arabiyyah lil Muta'allimin) instrument as a measuring tool for the Arabic reading skills of STITMA Yogyakarta students. The research method uses a descriptive quantitative approach by analyzing 60 questions tested on 311 students using SPSS version 25 software to test validity, reliability, level of difficulty and discriminatory power. Analysis is needed to ensure the quality of the test instrument that can provide accurate measurement results and become the basis for effective learning evaluation. The results showed that all questions were valid and reliable, with 65% having a moderate difficulty level, 37.7% at an easy level and 3.3% at a difficult level. The results of the Discriminatory Power Test showed that 60% of the questions were accepted well and 40% needed revision so the ITMAM instrument is suitable for use as a standard evaluation tool for Arabic reading ability with several improvements to certain questions.

Kata kunci:
Bahasa Arab;
Instrumen;
Kemampuan membaca;
TOAFL

Abstrak

Evaluasi pembelajaran bahasa Arab membutuhkan instrumen yang valid dan reliabel untuk mengukur kemampuan mahasiswa secara akurat. Penelitian ini bertujuan untuk menganalisis kelayakan instrumen ITMAM (*Ikhtibar Tahdid Mustawa al-'Arabiyyah lil Muta'allimin*) sebagai alat ukur keterampilan membaca bahasa Arab mahasiswa STITMA Yogyakarta. Metode penelitian menggunakan pendekatan kuantitatif deskriptif dengan menganalisis 60 butir soal yang diujikan kepada 311 mahasiswa menggunakan software SPSS versi 25 untuk menguji validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Analisis diperlukan untuk memastikan kualitas instrumen tes yang dapat memberikan hasil pengukuran yang akurat dan menjadi dasar evaluasi pembelajaran yang efektif. Hasil penelitian menunjukkan bahwa seluruh butir soal valid dan reliabel, dengan 65% soal memiliki tingkat kesukaran sedang, 37,7% berada di tingkat mudah dan 3,3% di tingkat sukar. Hasil Uji Daya Pembeda menunjukkan 60% soal diterima baik dan 40% perlu adanya revisi,

	sehingga instrumen ITMAM layak digunakan sebagai alat evaluasi standar kemampuan membaca bahasa Arab dengan beberapa penyempurnaan pada butir soal tertentu.
Article Information	**Submitted 2024-11-24. Received 2024-11-24. Revised 2025-01-02. Accepted 2025-01-02. Published 2025-01-13.**

INTRODUCTION

As a system of sound symbols used to communicate, language continues to develop for its users. Language development is influenced by culture, technology and social changes that occur. With these changes, an assessment instrument for a language is needed for its speakers. Various languages in the world have language competency tests such as, Japanese language proficiency test, Hanyu shuiping kaoshi, Test of proficiency in Korean, diplome d'etude en langue francaise, Test of English as a foreign language, etc. So in Arabic a language assessment instrument is also needed. Arabic language proficiency test tools can be one of the evaluation requirement in learning (Wahab, 2015b).

Arabic language competency test is very necessary for language teachers to measure the level of success of their students. The success of Arabic language learning is greatly influenced by the implementation of the appropriate curriculum and evaluation methods (Amrullah, 2021). Arabic, as an international language, has a strategic role in both scientific, cultural, and religious contexts. The effectiveness of Arabic language learning is also supported by the use of competency-based methods and evaluations that are adjusted to the needs of students and global challenges (Noor, 2018).

Arabic language learning must focus on applicable communication skills. In this case, the indicators of success include four maharah: listening, speaking, reading, and writing (Roisyah, Nurhidayati Inayati & Saifudin, 2021). On the other hand, research by Qomariyah (2022) emphasizes the importance of the role of evaluation to assess the extent to which learning objectives are achieved, as well as being a diagnostic tool to improve the learning process in the future (Qomariyah, 2022).

As a concrete step in supporting the achievement of learning objectives, the selection of quality evaluation tools is crucial (Setiadi, 2016). Good evaluation tools not only measure students' abilities but also provide diagnostic information to improve the learning process (Satyaningrum dkk., 2024). A study by Harahap shows that analysis of the level of difficulty and discriminating power are two main components in determining the quality of a test instrument (Harahap, 2018). This is reinforced by Arikunto, who

explains that effective test items must have a balance between the level of difficulty and discriminating power to ensure valid and accurate evaluation results (Arikunto, 2018).

Test are the most commonly used evaluation tools to measure students' Arabic language skills. Bloom stated that the main purpose of the test is to identify students' abilities based on certain competencies (Bloom, 1956). In this context, TOAFL (*Test of Arabic as a Foreign Language*) is an important tool to measure students' Arabic language skills. This test is designed not only to measure language proficiency, but also to evaluate specific skills such as listening, reading and understanding language rules (Hikma & Hamim, 2023).

TOAFL is a standard Arabic language graduation test for UIN Jakarta students organized by the UIN Jakarta Language Development Center and was also pioneered in 1999 (Wahab, 2015a). The naming of standard Arabic language tests outside UIN Jakarta is also different; including IKLA (*Ikhtibar Kafa'ah al-Lughah al-'Arabiyyah*) organized by UIN Sunan Kalijaga Yogyakarta (UINSUKA, 2024), ILAA (*Ikhtibar al-Lughah al-'Arabiyah al-Mi'yari*) organized by UIN Maulana Malik Ibrahim Malang (Malang, 2024), ILAIKA (*Ikhtibar Lughah Al'arabiyah Ka Lughatin Ajnabiyyatin*) organized by UIN Salatiga (Salatiga, 2024), IMALAH (*Ikhtibarat Maharah Lughah Arabiyah*) organized by UPI Bandung (Bandung, 2024) and similar TOAFL tests at other universities.

ITMAM is a standard Arabic language test developed by the Language Research and Development of the Majelis At-Turots Al-Islamy Foundation. This test was given to students of STITMA (Sekolah Tinggi Ilmu Tarbiyah Madani) Yogyakarta as a form of Arabic language ability mapping test for students of Arabic language education and Islamic religious education study programs. This test consists of 150 multiple-choice questions with details of 40 questions to measure listening skills, 50 questions to measure Arabic language method skills and 60 questions to measure Arabic text reading skills.

The TOAFL test called ITMAM (*Ikhtibar Tahdid Mustawa al-'Arabiyyah lil Muta'allimin*) is still under development and testing. The results of the student test cannot show that students have not understood the Arabic material that has been studied, but it could be because the questions do not meet the criteria for quality questions (Qomariyah, 2022). Item analysis activities are activities that must be carried out by a teacher in order to improve the quality of the questions that have been

composed. This also aims to identify good, bad, and poorly phrased questions. By analyzing questions, information can be obtained about the shortcomings of a question and instructions for making improvements (Beerepoot, 2023).

A similar opinion was expressed by Fitrianawati, that the activity of analyzing the quality of questions was carried out to obtain information on learning achievement which became a guideline for teachers to determine the next steps, namely maintaining the questions that had been prepared, improving them, fixing them and perfecting them so that the questions could be used as assessment instruments that were suitable for use in subsequent learning activities (Fitrianawati, 2017).

This study was conducted to analyze the quality of the TOAFL Arabic test questions that had been prepared which focused on reading skills in terms of their quantitative aspects, namely the level of difficulty and their differentiating power, totaling 60 questions. This test is needed in the future for academic purposes, namely as a requirement for graduation, further study and scholarship applications.

Similar research by Nawang Wulandari regarding the TOAFL test at IAIN Metro shows the analysis results of 50 questions; (1) 26 questions were suitable for use, (2) 3 questions needed revision, (3) 21 questions were not suitable for use (Wulandari, 2023). In Qomariyah's research related to the TOAFL test conducted at Hasyim Asy'ari Tebuireng University Jombang, it was also found that out of 50 questions, the results were; (1) the level of difficulty is quite good with details, 23 easy questions, 22 moderate questions, and 5 difficult questions, so revision is needed and (2) the discriminatory power of the questions is said to be good at more than 50% with details of 44 questions maintained and 6 questions need revision (Qomariyah, 2022).

This research study is different from previous studies. This research needs to be conducted to obtain results on the quality of ITMAM questions which are expected to help identify the quality of the questions, both per item or as a whole, so that they can be used as a guideline for improving Arabic language questions.

METHOD

This study uses a descriptive quantitative approach (Azizah dkk., 2024). The quantitative approach is used to analyze the level of difficulty, validity, reliability, and discriminatory power of ITMAM test items, while the descriptive approach aims to

illustrate the reading abilities of STITMA Yogyakarta students. The subjects of this study were 311 STITMA Yogyakarta students who had taken the ITMAM reading ability test.

Data collection was carried out through the implementation of the ITMAM test. This test is designed to measure the reading ability of students of the education faculty by presenting various types of questions that include low, medium, and high levels of difficulty. The test results were then analyzed using SPSS version 25 software. The analysis was carried out to test the validity of the test items using the Pearson correlation technique, measure the reliability of the questions using the Alpha Cronbach method, and determine the level of difficulty and discriminatory power based on the calculation of the difficulty and discrimination index (Riduwan, 2015).

The quantitative approach is very appropriate for research that aims to evaluate test instruments because this approach allows researchers to obtain numerical data that can be analyzed statistically (Sugiyono, 2016). Who emphasizes the importance of validity and reliability tests in evaluating the quality of test items in order to produce accurate and reliable measuring instruments (Azizah dkk., 2024).

RESULTS AND DISCUSSION

Here are 10 out of 60 ITMAM questions that were given to students of the Faculty of Education.

Questions
اجتمع أكثر من ١٠٠ رجل من رجال الأعمال، يمثلون أكثر من ٣٠ قطرًا من أقطار العالم؛ وذلك لمناقشة أخطر جريمة في العصر الحديث، أو ما يسمى الجرائم الإلكترونية، التي تُعد سببا من الأسباب الرئيسة التي تواجه نموّ ما يُعرَفُ بالتجارة الإلكترونية. وهذه الجريمة لم تسلم من خطرها أيّ دولة من دول العالم.

1-2

١. ما العنوان المناسب لهذا النص ؟

أ. خطر الجريمة الإلكتروني   د. مناقشة ١٠٠ رجل حول أخطر جريمة

ب. المشكلات في العصر الحديث   هـ. الجرائم في العالم

ج. أسباب الجريمة الإلكتروني

٢. كم رجلا اجتمعوا في هذه المناقشة ؟

أ. ١٠٥   د. ٩٥

هـ. ١٠	ب. ١٠٠
	ج. ٩٠

تسبّبت الحرب العالميّة الأولى، التي استمرت من عام ١٩١٤م حتى عام ١٩١٨م، في خسائر عظيمة في الأرواح والممتلكات أكثر مما سبّبته أي حرب أخرى قبلها، وقد قضت على الأخضر واليابس. وقد كان طرفا هذه الحرب كان كل قوات الحلفاء بقيادة بريطانيا، وقوات الوسط بقيادة رومانيا وصربيا. جانب يتوقع نصرًا سريعًا، لكن الحرب استمرت أربع سنوات، قُتل فيها كثير من الجنود من كلا الجانبين.

٣. ما معنى الجملة " قضت على الأخضر واليابس" ؟

د. قضت سنوات طويلة	أ. لم يبقى على شيء
هـ. مسبّب في هذه الحرب	ب. قتل فيها كثير من الجنود
	ج. لم يستمرّ بعد ذلك أي حرب

٣-٥

٤. من قاد قوات الوسط في الحرب العالمية الأولى ؟

د. بريطانيا	أ. صربيا ورومانيا
هـ. إسبانيا	ب. رومانيا وبريطانيا
	ج. بريطانيا وصربيا

٥. كم مدة الحرب العالمية الأولى ؟

د. ثماني سنوات	أ. أربع سنوات
هـ. ثمانية سنوات	ب. أربعة سنوات
	ج. سنتين

بيروت العاصمة السياسية للبنان، واشتق اسمها من كلمة «بيريث» وهي كلمة فينيقيّة تعني الآبار. وتُعد بيروت المركز التجاري والثقافي الرئيس للبنان. يسكن في بيروت أكثر من ٢,٠٠٠,٠٠٠ شخص وفق إحصائيات عام ٢٠٠٧م. ويعيش فيها حوالي رُبع سكان لبنان، ويزداد معدل السكان بها بنسبة ١٣٪ حاليا.

6

٦. ماذا تعني كلمة "بيروت" ؟

د. العاصمة السياسية	أ. الآبار
هـ. بيريث	ب. عاصمة لبنان

ج. فينيقية

تقع مدينة جدة، عروس البحر الأحمر، على الساحل الشرقي للبحر الأحمر. تطورت هذه المدينة ونمت بسرعة، وتعتبر الآن ثاني أكبر المدن السعودية بعد الرياض.

أنشأت مجموعة من الصيادين مدينة جدة قبل نحو ٣,٠٠٠ سنة، كان هؤلاء الصيادون يستقرون فيها بعد رحلات الصيد. ثم جاءت بعض القبائل واستقرت فيها. وحدث التحول التاريخي لها عام ٦٤٧ م عندما تحولت جدة إلى ميناء لاستقبال الحجاج وحتى اليوم. وهي الآن المرفأ الرئيس الذي تصدر منه السعودية البضائع غير النفطية، وتستورد من خلاله الاحتياجات المحلية.

٧. ما الفكرة الرئيسية للفقرة الثانية ؟

أ. تاريخ نشأة مدينة جدة	د. موقع مدينة جدة
ب. القبائل مستقرون بجدة	هـ. تطور مدينة جدة
ج. مدينة للصيادين	

٨. ما هي أكبر المدن السعودية ؟

7-10

أ. الرياض	د. المدينة
ب. جدّة	هـ. المملكة
ج. مكة	

٩. ما مفرد الكلمة "بضائع" ؟

أ. بضاعة	د. بضع
ب. بعض	هـ. بضعة
ج. باضعة	

١٠. متى أنشئت مدينة جدّة ؟

أ. قبل ثلاثة آلاف سنة	د. قبل ثلاث سنوات
ب. قبل ثلاثمائة سنة	هـ. قبل ثلاثين ألف سنة
ج. قبل ثلاث عشرة سنة	

Validity Test

Validity testing aims to determine the extent to which a measurement instrument can measure what should be measured. According to Arikunto (2018), validity is the

level of accuracy of a measuring instrument in measuring a particular concept. Validity can be tested using various methods, one of which is the Pearson correlation technique, which connects the score of each question item with the total score (Arikunto, 2018). Validity testing with decision-making criteria if r count ≥ r table then the question is "valid" and if r count ≤ r table then the question is "invalid". The r table value which is the reference with a 99% confidence level N = 42 is 0.393 (Deden, 2023).

Based on the calculation of the validity test of the question instrument to students using SPSS version 25, it was concluded that the ITMAM test trial with 60 questions to 311 students produced r count ≥ r table then the question was said to be **valid**.

Reliability Test

Reliability measures the consistency of the measuring instrument in producing data. According to Sugiyono (2016), an instrument is said to be reliable if it gives the same results in repeated measurements with the same conditions (Sugiyono, 2016). Reliability test with decision-making criteria if the calculated Cronbach's Alpha ≥ Cronbach's Alpha reference then the instrument is "reliable" and if the calculated Cronbach's Alpha ≤ Cronbach's Alpha reference then the instrument is "not reliable". The reference Cronbach's Alpha value used is > 0.80 with a reliability level of "very reliable" (Hair dkk., 1998).

Based on the results of the reliability test of questions using SPSS version 25, the following results were obtained.

Table 1. Recap of Reliability Test Results Data for the First 30 Questions

Calculated Cronbach's Alpha	Cronbach's Alpha Reference	Decision
0,871	0,8	Reliable

Source : Data Processing Results

Table 2. Recap of Reliability Test Results Data for the Last 30 Questions

Calculated Cronbach's Alpha	Cronbach's Alpha Reference	Decision
0,894	0,8	Reliable

Source : Data Processing Results

From the table of data recapitulation of the results of the reliability test of the test instrument above, it is concluded that the ITMAM test trial with 60 questions on 311 students obtained a calculated Cronbach's Alpha ≥ reference Cronbach's Alpha, so the questions are said to be **reliable**.

Difficulty Level Test

The level of difficulty of the questions aims to determine to what extent the questions can be worked on by students. The difficulty index is calculated using the proportion of the number of participants who answered correctly compared to the total number of participants(Arikunto, 2018). Effective questions have a moderate level of difficulty, which means they are not too easy or too difficult.

Table 3. Classification of the level of difficulty of questions

Classification of Difficulty Questions	Interpretation
0,71 - 1,00	Easy
0,31 - 0,70	Medium
0,00 - 0,30	Difficult

Source : *(Zainuri dkk., 2021)*

Based on the results of the test of the level of difficulty of the questions using SPSS version 25, the following results were obtained,

Table 4. Recap Difficulty Level Test Results of Questions

Classification of Difficulty Questions	Interpretation	Number of Questions	Question Number
0,71 - 1,00	Easy	19	5, 8, 10, 11, 15, 18-20, 22, 24, 29, 31, 33, 35, 38, 55-57, 60
0,31 - 0,70	Medium	39	1-4, 6, 7, 9, 12-14, 17, 21, 23, 25-28, 30, 32, 34, 36, 37, 39-49, 51-54, 58, 59
0,00 - 0,30	Difficult	2	16, 50

Source : Data Processing Results

From the summary table of the test instrument difficulty level test results above, it can be concluded that the ITMAM test trial with 60 questions for 311 students, namely

the easy category had 19 questions, the medium category had 39 questions and the difficult category had only 2 questions.

Discriminatory Power

The discriminatory power of questions is used to determine the extent to which questions are able to differentiate between students with high and low abilities. According to Riduwan (2015), discriminatory power is calculated based on the difference in the proportion of correct answers between the upper and lower groups. Effective questions have high discriminatory power (Riduwan, 2015).

Table 5. Interpretation of Discriminatory Power

Classification of Discriminatory Power	Interpretation
0,40 - 1,00	Accepted
0,30 - 0,39	Accepted but needs improvement
0,20 - 0,29	Needs Improvement
0,00 – 0,19	Not used or discarded

Source : *(Zainuri dkk., 2021)*

Based on the results of the test of the discriminating power of questions using SPSS version 25, the following results were obtained.

Table 6. Recap Interpretation of Discriminatory Power Results of Questions

Classification of Discriminatory Power	Interpretation	Number of Questions	Question Number
0,40 - 1,00	Accepted	36	8, 10, 12-14, 17-22, 24-26, 28, 29, 31-36, 38-40, 43, 44, 48, 51, 54-60
0,30 - 0,39	Accepted but needs improvement	17	1-4, 6, 7, 9, 11, 23, 30, 37, 41, 42, 45, 46, 49, 53
0,20 - 0,29	Needs Improvement	7	5, 15, 16, 27, 47, 50, 52
0,00 – 0,19	Not used or discarded	0	-

Source : Data Processing Results

From the summary table of data on the results of the test instrument's differential power test above, it can be concluded that the ITMAM test trial with 60 questions for

311 students, namely 36 questions were accepted, 17 questions were accepted but needed to be improved, 7 questions were improved and no questions were discarded.

DISCUSSION

The results of the analysis show that the ITMAM test instrument has good overall psychometric quality. In terms of validity, all 60 questions were proven valid based on the Pearson correlation test with a calculated r exceeding the table r (0.393). The reliability of the instrument also showed very satisfactory results, with a Cronbach's Alpha value of 0.871 for the first 30 questions and 0.894 for the last 30 questions, both exceeding the reference value of 0.8 which indicates a very high level of reliability. These findings confirm that the ITMAM instrument has strong internal consistency and can be relied on for repeated use.

In terms of difficulty level and discriminatory power, the ITMAM instrument shows a fairly proportional distribution but still requires some adjustments. The majority of questions (39 out of 60) are at the ideal medium level of difficulty, although there are 19 questions in the easy category and only 2 questions in the difficult category. Meanwhile, the discriminatory power of the questions shows promising results with 36 questions (60%) being accepted well and 17 questions (28.3%) being accepted with notes for improvement. However, there are still 7 questions that require significant improvement in terms of discriminatory power. This distribution indicates that although the instrument is generally good, revisions are still needed, especially to increase the level of difficulty of several questions and improve the discriminatory power of certain questions to achieve a more optimal balance.

CONCLUSION

Based on the analysis of ITMAM (TOAFL) questions on reading skills material that has been conducted by researchers, the following conclusions can be drawn, (1) The results of the Validity Test show that 60 questions worked on by 311 students are declared valid. (2) The results of the Reliability Test show that 60 questions worked on by 311 students are declared reliable. (3) The level of difficulty of ITMAM questions has a percentage of 65% or 39 questions at the moderate level, 31.7% or 19 questions are at the easy level and 3.3% or 2 questions are at the difficult level. (4) The discriminatory power of ITMAM questions is said to be effective because 60% or 36 questions are

accepted as good, 28.3% or 17 questions are accepted but need revision, 11.7% or 7 questions need revision and no questions are discarded. The researcher realizes that this research is still far from perfect. Therefore, the researcher hopes that in the future there will be research developments in terms of language test questions or question quality based on student scores on a global scale.

REFERENCES

Amrullah, A. F. (2021). *Manajemen Kurikulum Pembelajaran Bahasa Arab* (1 ed.). Penerbit Kencana.

Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara.

Azizah, M., Udin, A. F., Munir, D. R., & Faziyah, M. (2024). *Buku Referensi Perencanaan Penelitian Kuantitatif: Panduan Komprehensif Metode Penelitian* (1 ed.). Media Penerbit Indonesia.

Bandung, D. P. B. A. F. U. (2024). *IMALAH iBT*. https://imalahupi.com/

Beerepoot, M. T. P. (2023). Formative and Summative Automated Assessment with Multiple-Choice Question Banks. *Journal of Chemical Education*, 2947–2955.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. David McKay Company.

Deden, Mutakin. (2023). Pemanfaatan Software SPSS untukAnalisis Instrument Penelitian Dalam Penelitian Tindakan Kelas. *Jurnal Inovasi dan Manajemen Pendidikan*, *3*.

Fitrianawati, M. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *Prosiding Seminar Nasional dan Call for Papers Pendidikan*.

Hair, J. F., Andersin, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis Fifth Edition*. Prentice Hall International, Inc.

Harahap, P. (2018). ANALISIS SOAL TOAFL IAIN CURUP. *Ihya Al-Arabiyah*, *4*.

Hikma, N., & Hamim, H. (2023). *Lulus TOAFL: Test of Arabic as Foreign Language*. UIN Press Malang.

Malang, P. P. B. U. (2024). *ILAA*. https://ppb.uin-malang.ac.id/toafl/

Noor, F. (2018). Kurikulum Pembelajaran Bahasa Arab di Perguruan Tinggi. *Arabiyatuna : Jurnal Bahasa Arab*, *2*.

Qomariyah, L. (2022). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal TOAFL Universitas Hasyim Asy'ari Tebuireng Jombang. *Lisanan Arabiya: Jurnal Pendidikan Bahasa Arab*, *6*.

Riduwan. (2015). *Dasar-Dasar Statistika*. Alfabeta.

Roisyah, Nurhidayati Inayati, N. L., & Saifudin, S. (2021). EFEKTIVITAS PEMBELAJARAN BAHASA ARAB MASA PANDEMI PADA MAHASISWA PROGRAM STUDI PENDIDIKAN AGAMA ISLAM (PAI) UNIVERSITAS MUHAMMADIYAH SURAKARTA. *Tahdida: Jurnal Pemikiran dan Gerakan Muhammadiyah*, *19*.

Salatiga, U. P. B. U. (2024). *ILAIKA*. https://uptpb.uinsalatiga.ac.id/ilaika/

Satyaningrum, R., Ihsan, M., & Daroini, S. (2024). Efektivitas Penggunaan Media Jeopardy Gamedalam Evaluasi Maharah Qira'ah. *Kalamuna*, *5*, 157–166.

Setiadi, H. (2016). Pelaksanaan penilaian pada Kurikulum 2013. *Jurnal Penelitian dan Evaluasi Pendidikan*, *20*.

Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

UINSUKA, P. B. (2024). *IKLA*. https://pusatbahasa.uin-suka.ac.id/id/

Wahab, M. A. (2015a). Perkembangan Tes Bahasa Arab Standar di Indonesia. *International Conference of Arabic Language*.

Wahab, M. A. (2015b). Tingkat validitas dan realibilitas TOAFL. *Fakultas Ilmu Tarbiyah dan Keguruan UIN Syarif Hidayatullah*.

Wulandari, N. (2023). TAHLIL IKHTIBARAT AL-ARABIYAH KA LUGHAH AJNABIYAH. *An Nabighoh Jurnal Pendidikan dan Pembelajaran Bahasa Arab*, *25*.

Zainuri, A., Aquami, & AnNur, S. (2021). *Evaluasi Pendidikan: Kajian Teoritik*. CV. Penerbit Qiara Media.