



Available online:

<http://journal.imla.or.id/index.php/arabi>

Arabi : Journal of Arabic Studies, 7 (2), 2022, 198-207

DOI: <http://dx.doi.org/10.24865/ajas.v7i2.523>

TEST THE VALIDITY AND RELIABILITY OF ARABIC LEARNING QUESTIONS

Muhammad Faris Imaduddin¹, Hikmah Maulani², Ijlal Haziel Taufik³

^{1,2} Indonesia University of Education, Indonesia

³ Islamic University of Madinah, Saudi Arabia

Corresponding E-mail: farismuhammad@upi.edu

Abstract

This study aims to determine the validity and reliability of Arabic language test items for grade II students of an elementary school without a pesantren (Islamic Boarding School) background in Bekasi City. The research method used is descriptive with a qualitative approach. Data collection techniques used in this research were non-test instruments in the form of observation and test instruments in the form of 10 multiple-choice Arabic language questions, adapted to teaching materials at the existing learning level. Data analysis techniques include reducing, presenting, and drawing conclusions from existing data. Based on the data found, the level of validity of the Arabic language test questions from question numbers 1 to 10 provided 60% is in the valid category and 40% is not. While the level of reliability based on Cronbach's Alpha testing technique is at 0.82 which indicates very reliable. It is hoped that the findings of this research have positive implications for teacher knowledge.

Keywords: *evaluation, validity, reliability, Arabic language education*

Abstrak

Penelitian ini bertujuan untuk menentukan validitas dan reliabilitas butir-butir soal tes kebahasaan bagi siswa kelas II salah satu sekolah dasar tanpa latar belakang pesantren yang ada di Kota Bekasi. Metode penelitian yang digunakan adalah deskriptif dengan pendekatan kualitatif. Partisipan yang diambil pada penelitian ini melibatkan lima belas orang siswa yang terdiri dari 7 orang laki-laki dan 8 orang perempuan. Teknik pengumpulan data melalui instrumen non-tes berupa observasi dan instrumen tes berupa 10 butir soal kebahasaan pilihan ganda, yang disesuaikan dengan bahan ajar pada tingkatan pembelajaran yang ada. Teknik analisis data dengan mereduksi, menyajikan data serta menarik kesimpulan dari data yang ada. Berdasarkan data yang ditemukan, tingkat validitas tes soal kebahasaan dari nomor soal 1 sampai 10 yang disediakan 60% masuk kategori valid dan 40% tidak. Sedangkan tingkat reliabelnya berdasarkan Teknik pengujian Cronbach's Alpha berada pada angka 0,82 yang menunjukkan sangat reliabel. Temuan riset ini diharapkan berimplikasi positif pada pengetahuan pengajar.

Kata Kunci: evaluasi; validitas; reliabilitas, pendidikan bahasa Arab

Introduction

Learning is a series of activities that have a purpose and direction. In addition, learning is also a process with goals that must be achieved by teachers and students (Fahmi et al., 2022). Goals make it easier for teachers to direct the learning process and support the achievement of learning objectives. Various teacher efforts are expected to enable students to achieve their learning goals by using different strategies and methods, different learning media, different materials, and various other methods. In the process, assessment is an inseparable part. Assessment is a systematic and continuous activity that collects information related to learning processes and outcomes so that teachers receive information about student development. The success or failure of the learning program that has been implemented is to provide direction for teachers to improve the quality of learning activities. A scoring system must be developed and implemented properly. The success of this assessment depends on the accuracy of the tools used. Whether the instrument can provide the right information based on the learning objectives that have been set or not.

Some of the instruments commonly used by teachers in learning are validity and reliability testing. Testing is a tool or procedure that functions as a measuring tool. This will encourage educators to form better learning strategies, and encourage students to study harder. The efforts and efforts built and carried out by the teacher need to be analyzed further through an assessment, whether they are appropriate and have achieved the objectives or not. Therefore, evaluation is necessary to see the student's ability in learning and see the achievement of the learning objectives. An analysis of the items is needed to obtain good quality items. Item analysis is an activity that examines the quality of each item. This aims to identify good, good enough, and not good questions.

Knowing the quality of each question gives a signal to a teacher to improve the quality of questions that do not meet the standards. Item analysis shows which questions the students answered correctly and incorrectly. This analysis allows a teacher to identify and classify students who are successful and not successful. According to Bagiyono in Fahmi (2022) if, a question can, not distinguish between the abilities of two students, the question may not meet the criteria for a good question. Validity comes from the word Validity which means the level of accuracy and accuracy of a test tool or metric in carrying out its measurement function (Rizal et al., 2019). The test can be said to be very valid if the measuring instrument used performs the measurement function and provides accurate results. While reliability according to Sugiono in M. Rizal (2019) is a series of measurements or a series of measuring instruments that has consistency if the measurements made with the measuring instrument are repeated or the extent to which the measurement results can be trusted or not.

The instrument itself does not determine the validity and reliability of the instrument. According to Sugiyono in Febriyanto Yusup (2018), the user of the measuring instrument who takes the measurement and the object being measured are the factors that affect the validity and reliability of a measuring instrument other than the device itself. However, these issues can be resolved by conducting adequate validity and reliability assessments on the instrument. This test is conducted to maintain validity and reliability. In addition, to overcome the influence of the user of the measuring instrument, the user must improve his ability to use the measuring instrument. Another factor that is no less important that affects the validity and reliability of the instrument is the factor of the subject being measured. To overcome this, the researcher must be able to control the subject.

According to Surucu (2020), the concepts of validity and reliability are closely related but describe different characteristics of measuring instruments. In general, a measuring instrument can be trusted or reliable even though the measuring instrument is not valid, but if the measuring instrument is valid, the instrument can also be trusted or reliable. However, reliability alone is not sufficient to guarantee efficacy. Even if the test is reliable, it may not accurately reflect the desired behavior or characteristics.

In this section, the researcher includes various previous research findings related to the research to be carried out, then provides a summary, of both published and unpublished research.

Previous research is one of the references for researchers to enrich the theory used in reviewing the research that has been done. Previous research related to the validity and reliability test has been done a lot. So that it became a reference and a spur for researchers to participate in researching similar things, with different results. The following are some of the previous studies that the researcher used as reference material; First, the research conducted by Deni Mulana and Anwar Sanusi in 2018 with the title "Analysis of Arabic Language Questions for the Regional Joint *Madrasah* Final Examination (UAMBD) of *Madrasah Ibtidāiyah* Years 2017-2018". The study stated that the questions used for the Regional Joint *Madrasah* Examination showed high validity and reliability. Second, the research conducted by Bagusradityo Aryobimo in 2021 with the title "Analysis of Practice Questions in the *an-Nahw al-Wāḍiḥ* Book for *Madrasah* Aliyah Students". The results show that the testing items are relatively changing based on student understanding, the running time of a test or exam, and other factors (Aryobimo, 2021). Third, the research conducted by Ardiansyah in 2022 with the title "Analysis of Explanation Test Items for SKI Class IX Subjects at MTsS TI Candung for the 2020/2021 Academic Year". Showing similar results, that the level of validity and reliability of the items tested is very high. This can be seen from the suitability of the questions with the specified grid (Ardiansyah, 2022). Fourth, the research conducted by Bahrudin Fahmi, Syahrul Rizqi, and Nurul Elmira H., in 2022 with the title "Analysis of Arabic Language Questions at MAS Pondok Pesantren Assalam Kampar Riau" (Fahmi et al., 2022). The four studies concluded that the validity and reliability test was very helpful for teachers in analyzing the level of difficulty of the items. So that it becomes a reference in development or improvement when making or compiling questions.

Based on the literature above, the researcher found the urgency of the importance of analyzing the items to see the value of each question. In order to answer the things that have not been answered. The focus of this research is the quality of the questions in Arabic subjects which were made specifically for test questions in one of the elementary schools in the city of Bekasi.

In terms of validity and reliability, the research focus is the core that is obtained from the researcher's experience or through the knowledge gained from studying the scientific literature (Moleong, 2014). This study examines a new view of the validity and reliability test. So that it becomes a reference for teachers in determining the level of quality of the questions that will be used as material for assessing the learning process that has been created and implemented. It should also be remembered that an instrument needs to be retested periodically every time it will be used (Yusup, 2018). Even though an instrument is valid and reliable, it does not immediately make the instrument usable anywhere, anytime, to any subject (Yusup, 2018). The findings in this research have implications for teachers' knowledge in understanding what matters relating to the development, preparation, and use of tests that have been good and need to be maintained.

Method

This study uses a descriptive qualitative method. The descriptive qualitative method is to express an empirical fact objectively and scientifically based on scientific logic, and procedures and supported by strong methodologies and theories according to the scientific discipline being occupied (Musthafa & Hermawan, 2018). The qualitative method analytical callical study approach used in this study, as expressed by Sugiyono in Agustinus (2012) is a qualitative way to gather comprehensive, meaningful data. Qualitative approaches might considerably impact the focus of the investigation. This means that the nature of the relationship between researchers and informants, research objects, and research subjects is directly shown through qualitative approaches. Qualitative research has characteristics or characteristics that should be a guideline by researchers, as confirmed by Bogdan and Biklen in Agustinus (2012) that the characteristics of qualitative research include: First, the researcher himself as the main instrument to go directly to data sources. Second, implementing the data collected in this study are more likely to be words than numbers. Third, explained that the results of the study emphasized more on the process not solely on the results.

Fourth, through inductive analysis, researchers reveal the meaning of the circumstances that occur. Fifth, expressing meaning is an essential part of a qualitative approach.

Participants

The participants in this study were class II students in private schools in Bekasi City who took Arabic lessons totaling 15 (fifteen) students. The students consist of 8 (eight) girls and 7 (seven) boys. The background is a public school and not a boarding school. Those who have not studied Arabic in depth. Below is the participant data which is an important object in this research. They were divided by gender and ability to read the Qur'an, because it relates to language skills.

Table 1. Ability to read the Qur'an

Characteristics	N
Man	7
Woman	8
Ability to read the Qur'an	
Well	8
Enough	5
Not enough	2

Data collection technique

Data collection was carried out by researchers in this study, namely using test and non-test instruments. The test instrument is in the form of giving questions to review student's ability in understanding. While the non-test instrument is an observation made by researchers to see and observe student performance when working on the tested Arabic language test questions and to obtain data that will later be analyzed to produce useful information to answer research questions or improve understanding of a topic. The two instruments above will produce a data or analysis.

Data analysis technique

The data analysis techniques used were (1) analyzing the data obtained from the results of the non-test instrument in the form of observation and the test instrument in the form of an Arabic language test, and (2) describing the data from the analysis. The first stage, the data analysis carried out by the researcher is First, data reduction. At this stage the researcher will sort, select, concentrate and simplify the rough data obtained from field notes, both from the results of non-test instruments in the form of observations and test instruments in the form of Arabic language test questions. Second, data presentation. At this stage, the researcher will present the data in a narrative description adapted to the data obtained in the field. Third, verification and data conclusion. At this stage, the data is re-examined to draw conclusions from the findings in the field that are adjusted to the problem formulation.

Result and Discussion

After analyzing Arabic language material for grade students of one elementary school (SD) elementary school in Bekasi City. Item analysis was carried out to identify which items were good or bad. The items taken by the researchers were Arabic language test questions based on a grid of competency standards and basic competencies. The question is in the form of multiple-choice with 10 (ten) items and 4 (four) answer options. From these items, information will be obtained that can be used to produce more quality questions, so that good information can be obtained for test takers. In terms of validity and reliability, there are the following findings.

According to Wahab in Maulana (2020), validity can be classified into three, namely: content validity, criterion validity, and construction validity. The validity that is examined regarding this Arabic language question is content validity. The term "content validity" refers to the tool's

reliability and the relationship between the scale of the item and the characteristics it is meant to measure. The level of content validity is determined by asking an expert's opinion about one of the pieces (Söylemez, 2018). By thoroughly comparing the domain of the content of the question with the domain of the content of the test grid that serves as the foundation for the preparation, it is possible to assess the authenticity of the content of the Arabic language lesson.

According to Brains and Manhein in Louangrath (2018), the validity test measures how accurately what is observed matches expectations. According to Cronbach and Meehl in Louangrath (2018), validity tests have historically been used in the literature to assess experimental rather than instrumental design.

Based on the discussion above, Arikunto in Febrianawati Yusup (2018) explains further that validity is a measure that shows an instrument's levels of validity or validity. A valid or valid instrument has high validity. On the other hand, an instrument that is not valid means it has low validity.

As for the content validity test results related to the Arabic language question, it can be concluded that the content validity of the Arabic language test item shows a very high level. This is indicated by the 100% conformity of all the items with the grid that has been set.

A grid is a format that contains information about the scope and content or competencies to be assessed or tested. This grid is organized according to the purpose of the rating and serves as a guide for structuring questions. The grid must refer to competency standards and basic competencies and consist of components that must be detailed, clear, and meaningful. The grid is a matrix table that contains the specifications of the questions to be asked. This grid becomes a reference for question makers, so that everyone who makes questions creates questions with relatively the same content and difficulty. The question grid matrix consists of two rows: a column and a row. Columns show competency criteria, core competencies, materials, material descriptions, and metrics. While the line, shows the target to be measured or tested.

This suitability becomes the initial parameter for measuring an item's validity and reliability level. Because students' understanding will be by existing learning. And it is appropriate for a teacher to conduct a test, based on the achievements established since the beginning of the learning period.

The analysis used is based on the differentiating power of the test takers' answers to the Arabic language test for grade II general elementary school (SD) students who do not have a boarding school background or the world of education oriented to Arabic language learning. The students who became the object of this research were 15 (fifteen) students with 10 (ten) multiple-choice questions.

Table 2. Answers for each Arabic language test question

Question										Total
1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	1	1	1	10
1	1	1	1	1	1	1	1	1	1	10
1	1	1	1	1	1	1	1	1	1	10
1	1	1	1	1	1	1	1	0	1	9
1	1	1	1	1	1	1	1	1	1	10
1	1	1	1	1	1	1	1	0	1	9
1	1	1	1	1	1	0	1	1	1	9
0	1	1	1	1	1	0	1	1	0	7
1	1	1	0	1	1	0	1	1	0	7
1	0	1	1	1	1	1	1	1	0	8
1	1	1	1	0	1	1	1	1	1	9

0	1	1	1	1	1	1	1	1	1	9
1	1	0	0	1	1	1	1	1	1	8
0	0	1	1	0	0	0	0	1	1	4

The table above shows the distribution of answers for each Arabic language test question by each student which does not show a significant spread. Most Arabic language test questions can be answered properly and correctly.

One student can only answer 4 (four) of the 10 (ten) multiple-choice questions. This should be more attention by the teacher concerned to look for aspects of the problems faced by students.

The distribution of data on Arabic language test questions that were not answered correctly was in the following numbered questions: question number 1 (one) with 3 (three) incorrect answers, question number 7 (seven) with 4 (four) incorrect answers. four) people, and question number 10 (ten) with 3 (people) wrong answers.

Table 3. The results of the calculation of the validity

Question No.	rCount	rTable	Valid/Invalid
1	0.604708	0.514	Valid
2	0.64379	0.514	Valid
3	0.008471	0.514	Invalid
4	0.262599	0.514	Invalid
5	0.516727	0.514	Valid
6	0.784987	0.514	Valid
7	0.696747	0.514	Valid
8	0.784987	0.514	Valid
9	-0.11859	0.514	Invalid
10	0.388741	0.514	Invalid

Based on table 3, the results of the calculation of the validity as referred to in Arikunto's opinion, it can be said that the level of validity of the Arabic language test questions from question numbers 1 to 10 which are included in the category of valid and valid questions, there are 6 items (60%) and which include in the category of invalid questions as many as 4 items (40%). This shows that most of the questions made can be used to measure students' knowledge of Arabic language learning and have a level of accuracy and constancy.

The questions with no valid value are in test questions 4, 5, 9, and 10. With the rTable which has been adjusted to the number of participants taking the Arabic language test, there are 10 (ten) students. This analysis shows that most of the questions given are valid and can be used well to evaluate the learning that the teacher has given.

A test has high validity if it performs its measurement function or provides measurement results that follow the intent and purpose of holding the test. On the other hand, a test that provides data that is irrelevant to the measurement objective is called a low-validity test.

Reliability is the consistency of the way used to provide data. Therefore, an item is considered reliable if it provides consistent data even though it is tested repeatedly at different times.

The reliability of the instrument can be tested with several reliability tests. Among them are equivalence and internal consistency. Internal consistency has several different test techniques: split half test, KR 20, KR 21, and Alfa Cronbach. However, each test has criteria for what kind of instrument can be tested with the technique. In this study, researchers focused on the use of the Cronbach Alpha test technique.

Regarding reliability, two queries must be addressed: What differences exist between a dependable instrument and consistent results? What is the reliable instrument's indicator, second? The similarity, uniformity, or lack of noticeably different results between repeated measurements is referred to as consistency (Louangrath & Sutanapong, 2018). The consistency of the survey's results is referred to as the consistency of results. Respondents may consistently give false results even if a poll mismeasures a construct (Louangrath & Sutanapong, 2018).

Furthermore, the reliability of Arabic language questions, to calculate the reliability of the test. In this study, the reliability of each item was measured using Cronbach's Alpha. There are three reasons researchers used Cronbach's Alpha test. First, this technique is most often used in testing the reliability of the questionnaire (Magdalena et al., 2021). Second, Cronbach's Alpha test can detect inconsistent indicators (Magdalena et al., 2021).

It is common practice to calculate either Cronbach Alpha coefficient using an appropriate statistical technique to evaluate the scale's reliability. There is a "0" to "1" dependability coefficient range. By using the scale repeatedly, twice, or the equivalent scale once, the dependability of the scale will be evaluated (Ravinder & Saraswathi, 2020).

If the Cronbach's Alpha reliability coefficient has been calculated (ri), this value is then compared with the Cronbach's Alfa reliability coefficient criteria for a reliable instrument. According to Nunnally in Magdalena (Magdalena et al., 2021) states that the instrument is said to be reliable if the reliability coefficient of Cronbach's Alpha is more than 0.70 ($ri > 0.70$) and Streiner himself in Yusup (2018) states that the reliability coefficient of Cronbach's Alpha, cannot be more than 0,90 ($ri < 0.9$).

If the reliability coefficient of Cronbach's Alfa is less than 0.70 ($ri < 0.70$), Tavakol & Dennick in Yusup (2018) suggest revisiting or deleting items that have less correlation. An easy way to determine the item in question is with the help of a program on a computer, in this case, the researcher uses a formula in the Microsoft Excel program. Suppose the reliability coefficient of Cronbach's Alpha is more than 0.90 ($ri > 0.90$). In that case, it is directed to reduce the number of questions with the same question criteria even though in different sentence forms.

While Pallant in Daud (2018) claims that an index with a Cronbach's Alpha value of 0.6 or higher is considered sufficient dependability. A Cronbach's Alpha value of less than 0.6 is considered low. Values of 0.60 to 0.80 on the Cronbach's Alpha scale are regarded as moderate yet acceptable. While Cronbach's Alpha between 0.8 and 1.00 is excellent (Daud et al., 2018). Therefore, the level of dependability of the constructed instrument is assessed through this article using the Cronbach's Alpha value measurement.

The value of the reliability level of Cronbach's Alpha can be shown in the following table (Table 4).

Table 4. The value of the reliability level of Cronbach's Alpha

Cronbach's Alpha. value	Reliability Level
0.0 – 0.20	Less Reliable
>0.20 – 0.40	Somewhat Reliable
>0.40 – 0.60	Reliable enough
>0.60 – 0.80	Reliable
>0.80 – 1.00	Very Reliable

The Cronbach Alpha value above becomes the main reference in determining each item's reliability level. the range of values from 0.0-0.02 is less reliable to 0.80-1.00 is very reliable. After calculating using the Cronbach Alpha test technique, it was found the following,

Table 5. The reliability of the Arabic language test questions

K	10
K/K-1	1.111111
Total Variance Per-Item	0.261194
Sigma	0.738806
Cronbach Alpha Reliability Coefficient	0.820896

Researchers used the Microsoft Excel 2019 software program to help assess each item. From the Microsoft Excel 2019 software program output, the number 0.82 (table 4) is obtained for the reliability of the Arabic language test questions. Thus, it can be concluded that the Arabic language test questions have a high level of reliability or are very reliable because they are in the 0.80-1.00 interval so that this test score can be trusted to use.

From the description above, it can be seen that the Arabic language test items tested on students in the multiple-choice form have good quality questions, in addition to meeting validity and reliability, there is a balance between the level of difficulty of the questions and conformity with the grid provided. The balance in question is the number of questions classified as easy, medium, and difficult to balance. The difficulty level of the questions needs to be seen from the student's ability to answer the questions given, not from the point of view of the teacher who made the questions. researchers must test both the validity and reliability of the measuring instrument they want to use. A teacher must continue to measure items with these two test tools.

Researcher reiterate the incorrect use of the Cronbach's alpha as a gauge of reliability; this metric gauges the consistency of the responses generated by a survey. It is not a tool to assess the instrument's resilience, though. In order to directly and empirically evaluate the effectiveness of the instrument, a way of measuring the robustness of the quantitative scale was suggested in this study. This method can be used as a useful tool for instrument calibration. Any inference made from data produced by a flawed tool will fall short of scientific empiricism. On the other side, a well-calibrated device makes the data error-free; a conclusion drawn from such data is more reliable.

Heri Susanto, et. al., in Ina Magdalena (2021) states that items included in the good category, namely those with high validity and reliability values, should be inventoried in a question bank book. Furthermore, the items can be reused in the test of learning outcomes in the future. Three follow-up actions can be carried out for difficult items: (1) these items are no longer used and will not be issued again in the next learning outcome test. (2) a review is carried out, researched, and examined more deeply so that it can be found the factors that cause students to be unable to answer the question. (3) items that are difficult should have other benefits in the form of a loose selection, since most of the students will be declared to have passed the selection test. In this condition, giving easy questions will allow many students to pass the selection test or exam.

Conclusion

Evaluation in general can be interpreted as a series of activities designed to measure the system's effectiveness, planning, and implementation of learning as a whole. Furthermore, implemented in the form of identification to see whether the program that has been proclaimed has been achieved or not, valuable or not, efficient or not. Evaluation can be understood as considering values related to students' abilities in mastering learning materials. Evaluation is also defined as an assessment of students' success level in achieving the goals set in a program. This activity is very important to measure the results of students' ability to understand science. Even evaluation is also a place to find weaknesses in a program so that it becomes a reference for correcting errors that already exist and occur. In this regard, evaluation is very important to be carried out based on existing procedures.

The validity and reliability of the items are important things that the teacher must pay attention to in making instruments in the evaluation. Testing quality begins with determining the quality of the test preparation process, which is analyzed qualitatively. In addition, we conducted quantitative analysis on two aspects: the validity aspect and the test reliability aspect. The results of both analyzes are then interpreted. The interpretation results become a benchmark for the quality of the test. Based on the results of research and discussion, it can be concluded that there are 6 valid items (60%) and 4 items (40%) invalid from 10 questions on the Arabic language test in the form of multiple choice. While the level of reliability based on Cronbach's Alpha testing technique is 0.82 which indicates very reliable. Because the results of the validity and reliability tests are very, very important to be used as a reference in making the items in the evaluation instrument, the teachers are expected to continue to pay attention to this, so that the type of test items that are made can be used reliably.[]

References

- Ardiansyah. 2022. "Analisis Butir Soal Tes Uraian pada Mata Pelajaran SKI Kelas IX MTs TI Candung Tahun Ajaran 2020/2021", *Didaktika: Jurnal Kependidikan*, Vol. 11, No. 1.
- Aryobimo, B. 2021. "Analisis Butir Soal Latihan pada Buku *al-Naḥw al-Wāḍiḥ* untuk Siswa Madrasah Aliyah", *Al-Ma'rifah*, Vol. 18, No. 2.
- Cerelia, J. J., Sitepu, A. A., N, F. A. L., Pratiwi, I. R., Almadevi, M., Farras, M. N., Azzahra, T. S., & Toharudin, T. 2021. "Learning Loss Akibat Pembelajaran Jarak Jauh Selama Pandemi Covid-19 di Indonesia", *Seminar Nasional Statistik*.
- Choiroh, M. 2021. "Evaluasi Pembelajaran Bahasa Arab Berbasis Media E-Learning", *Bahasa Arab NASKHI Jurnal Kajian Pendidikan dan Bahasa Arab*, Vol. 3, No. 1.
- Daud, K. A. M., Khidzir, N. Z., Ismail, A. R., & Abdullah, F. A. 2018. "Validity and reliability of instrument to measure social media skills among small and medium entrepreneurs at Pengkalan Datu River", *International Journal of Development and Sustainability*, Vol. 7, No. 3.
- Fahmi, B., Rizqi, S., & Harmeilinda, N. E. 2022. "Analisis Butir Soal Bahasa Arab MAS Pondok Pesantren Assalam Kampar Riau", *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, Vol. 6, No. 1.
- Faizin, K. 2020. "Evaluasi Kurikulum Pembelajaran Bahasa Arab di STAI Attanwir Bojonegoro", *Al-Idaroh: Jurnal Studi Manajemen Pendidikan Islam*, Vol. 4, No. 1.
- al-Fathiyah, S. F. 2019. "Analisis Butir Soal Pelajaran Bahasa Arab", *Tarbiyatuna: Jurnal Pendidikan Ilmiah*, Vol. 4, No. 1.
- Fitriani, V. 2017. "Analisis Soal Ulangan Harian pada Materi Perubahan Lingkungan di Kelas X SMA di Kota Padang", *BioCONCETTA: Jurnal Biologi dan Pendidikan Biologi*, Vol. 3, No. 1.
- Hajaroh, S. 2016. "Perbandingan Reliabilitas Butir Soal Bahasa Arab Antara Pilihan Ganda dengan Menjodohkan pada Tes Buatan Guru", *EL-HIKMAH: Jurnal Kajian dan Penelitian Pendidikan Islam*, Vol. 10, No. 2.
- Indriana, D. 2018. "Evaluasi Pembelajaran dan Penilaian Autentik dalam Pembelajaran Bahasa Arab", *Al-Ittijah : Jurnal Keilmuan Dan Kependidikan Bahasa Arab*, Vol. 10, No. 2.
- Khafidin, Z. 2014. "Analisis Validitas Dan Reliabilitas Tes Mata Pelajaran Pendidikan Agama Islam Tingkat SMA", *Edukasia : Jurnal Penelitian Pendidikan Islam*, Vol. 9, No. 2.
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. 2021. "Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN

- Karet 1 Sepatan”, *BINTANG: Jurnal Pendidikan dan Sains*, Vol. 3, No. 2.
- Maulana, Deni, & Sanusi, A. 2020. “Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (Uambd) Madrasah Ibtidaiyah Tahun 2017-2018”, *Ta’lim Al-’Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, Vol. 4, No. 1.
- Maulayda, M. A., Erfan, M., & Hidayati, V. R. 2022. “Analisis Situasi Pembelajaran Selama Pandemi Covid-19 di SDN Senurus: Kemungkinan Terjadinya Learning Loss”, *Collase (Creative of Learning Students Elementary Education)*, Vol. 4, No. 3.
- Miladya, J. 2018. “Evaluasi dalam Pembelajaran Bahasa Arab”, *An Nabighoh: Jurnal Pendidikan dan Pembelajaran Bahasa Arab*, Vol. 20, No. 1.
- Louangrath, P. I., & Sutanapong, C. 2018. “Validity and Reliability of Survey Scales”, *International Journal of Research & Methodology in Social Science*, Vol. 4, No. 1.
- Qomariyah, L., & Niswah, I. 2019. “Analisis Kualitas Butir Soal Toafl UNHASY”, *Sainsteknopak Ke-5*, Vol. 3, No. 2.
- Ravinder, E. B., & Saraswathi, A. B. 2020. “Literature Review Of Cronbach alpha coefficient (A) and Mcdonald’s Omega Coefficient (Ω)”, *European Journal of Molecular & Clinical Medicine*, Vol. 7, No. 6.
- Rizal, M., Rifky, A., Dharmawan, A., Kusuma, W., Fatiehurrizqie, M., Arifin, M. Z., Pasani, C. F., & Budiarti, I. 2019. “Analisis Validitas Dan Reliabilitas Soal Hots Matematika Untuk Siswa Sekolah Menengah Kejuruan Kelas X”, *SENPIKA II (Seminar Nasional Pendidikan Matematika)*.
- Rosyad, Y. S. 2019. “Validitas dan Reliabilitas Kusiner Pengetahuan, Sikap, dan Tindakan Terhadap Covid-19”, *Jurnal Media Keperawatan: Politeknik Kesehatan Makassar*, Vol. 10, No. 2.
- Söylemez, Y. 2018. “The Development of the Text Evaluation Scale for Child Rights: A Study of Validity and Reliability”, *World Journal of Education*, Vol. 8, No. 5.
- Surucu, L., & Maslakci, A. 2020. “Business & Management Studies”, *Business & Management Studies: An International Journal*, Vol. 8, No. 3.
- Susanty, F. D. 2016. “Analisis Validasi Soal Tes Hasil Belajar pada Pelaksanaan Pembelajaran Bahasa Arab di Pusat Pengembangan Bahasa (P3B) UIN Suska Riau”, *Kutubkhanah: Jurnal Penelitian Sosial Keagamaan*, Vol. 19, No. 2.
- Suswati, I. 2017. “Analisis Validitas Materi Uji Multiple Choise Question dengan Learning Objectives Blok Endokrin”, *Saintika Medika*, Vol. 7, No. 2.
- Yusup, F. 2018. “Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif”, *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, Vol. 7, No. 1.