

Analysis of Validation of Learning Outcome Test Questions in the Implementation of English Language Learning

Moon Hidayati Otoluwa¹ (✉) Universitas Negeri Gorontalo, Indonesia¹
moon.hidayati@gmail.com¹

Abdul Kadir Ismail² IAIN Sultan Amai Gorontalo, Indonesia²
abdulkadir@iaingorontalo.ac.id²

 <https://doi.org/10.58194/eloquence.v2i2.1453>

Corresponding Author: ✉ Moon Hidayati Otoluwa

Article History	ABSTRACT
Received 16-08-2023 Accepted 18-08-2023 Published: 19-08-2023	<p>Background: The teaching staff is not only responsible for the process of knowledge transactions but more than that, the teaching staff, in this case, lecturers, are also required to make assessment and measurement instruments to find out the extent to which students are able and able to master the English material in an evaluation process or rather a test.</p> <p>Purpose: To find out how the validity analysis of learning outcome tests made by English lecturers at the IAIN Sultan Amai Gorontalo Language Development Center</p> <p>Method: This research is evaluative research that evaluates question items using a quantitative approach.</p> <p>Results and Discussion: IAIN Sultan Amai Gorontalo lecturers tend to make learning outcome test questions in objective form based on practical courses' characteristics. With accurate tests, developing questions that can reveal students' applicative abilities, analysis, synthesis, and evaluation is more accessible than objective tests. In designing learning outcome test questions, not all lecturers (around 76%) make a grid of questions used as a reference for writing question items. Even though the grid is essential in writing good questions, in addition to writing questions, most lecturers use a combination of last year's questions and new ones. Making further questions related to the development of science and technology will be more meaningful.</p> <p>Conclusions and Implications: Not all lecturers of English courses at IAIN Sultan Amai Gorontalo have conducted a theoretical review regarding the quality of the questions that have been asked. However, one thing that is quite good and positive in the review that has been carried out by most lecturers, namely the most re-examined aspect, is the suitability of the question material with the competence of learning outcomes. There are still some weaknesses in several question sets, namely for objective tests. The aspect that is still weak is the non-uniformity of the short length of statements between answer choices and the separation of the main question page from the answer page. Another drawback to this form of the test is that there are still pictures accompanying the questions that are not clear.</p>
Keywords:	<i>Test Questions; Learning Outcomes; English</i>
	ABSTRAK
	<p>Latar Belakang: Para tenaga pengajar tidak hanya bertanggung jawab pada proses transaksi ilmu, namun lebih dari itu tenaga pengajar dalam hal ini dosen dituntut juga untuk membuat instrumen penilaian dan pengukuran untuk mengetahui sejauh mana para mahasiswa mampu dan dapat menguasai materi Bahasa Inggris tersebut dalam sebuah proses evaluasi atau lebih tepatnya tes.</p>

Tujuan: Untuk mengetahui bagaimana analisis validitas ítem tes hasil belajar yang dibuat oleh para dosen Bahasa Inggris di Pusat Pengembangan Bahasa IAIN Sultan Amai Gorontalo

Metode: Penelitian ini merupakan penelitian evaluatif yang mengevaluasi butir soal dengan menggunakan pendekatan kuantitatif.

Hasil dan Pembahasan: Kecenderungan dosen IAIN Sultan Amai Gorontalo membuat soal tes hasil belajar dalam bentuk objektif didasarkan atas karakteristik mata kuliah yang bersifat praktis. Dengan bentuk tes objektif, maka lebih mudah dikembangkan soal- soal yang mampu mengungkap kemampuan aplikatif, analisis, sintesis, dan evaluasi mahasiswa dari pada tes objektif. Dalam pengembangan soal tes hasil belajar belum semua dosen (sekitar 76%) membuat kisi-kisi soal yang dijadikan acuan penulisan butir soal. Padahal kisi-kisi sangat penting artinya pada proses penulisan soal yang baik. Di samping itu, dalam penulisan soal, sebagian besar dosen melakukan kombinasi antara soal tahun lalu dan soal baru. Pembuatan soal baru akan lebih bermakna berkaitan dengan perkembangan sains dan teknologi.

Kesimpulan: Belum semua dosen mata kuliah Bahasa Inggris di IAIN Sultan Amai Gorontalo melakukan telaah ulang secara teoretis menyangkut kualitas soal yang sudah dibuat. Meskipun demikian, ada satu hal yang cukup baik dan positif dalam telaah ulang yang sudah dilakukan oleh sebagian besar dosen, yaitu aspek yang paling banyak ditelaah ulang adalah kesesuaian materi soal dengan kompetensi hasil belajar. Masih ditemukan beberapa kelemahan pada sejumlah perangkat soal, yaitu untuk tes objektif, aspek yang masih lemah adalah adanya ketidakseragaman panjang pendeknya pernyataan antar pilihan jawaban, terpisahnya halaman pokok soal dengan halaman jawaban. Kelemahan lain untuk bentuk tes tersebut adalah masih ada gambar-gambar yang menyertai soal yang kurang jelas.

Kata Kunci

Soal Tes; Hasil Belajar; Bahasa Inggris



Copyright: © 2023 by the author(s).

This is open access article under the

[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

INTRODUCTION

English is a compulsory subject that must be taken and studied by students of all faculties at IAIN Sultan Amai Gorontalo from semester I to semester II. English language learning is carried out consecutively at the language development centre. Because this course is compulsory, students have no other choice, regardless of the student's school origin. So, in the learning process, there are many very significant obstacles. However, IAIN Sultan Amai Gorontalo has prepared professional teaching staff in their fields.

The teaching staff is not only responsible for the process of knowledge transactions but more than that, the teaching staff, in this case, lecturers, are also required to make assessment and measurement instruments to find out the extent to which students are able and able to master the English material in an evaluation process or rather a test. Two tests measure students' ability at the language centre: oral and written tests.[1] In this research proposal, researchers focus on written tests because in making written test instruments, lecturers have a significant role; lecturers are required to make questions related to themes that the language centre institution has determined. The form of questions created by lecturers is objective, with 50 questions. Thus, many questions will be collected in the language centre. These questions will then be filtered so that later appropriate or quality question items will be selected to be contained in the result evaluation process.

Question items are said to be of high quality if the question items made have high validity.[2] Validity is a tool for assessing the accuracy of questions so that a lecturer can determine what he wants to score.[3] According to Uno, four kinds of question item validity originate and are the basis

for the division of types, namely (a) logical validity includes content validity and construction validity, (b) empirical validity includes current validity and prediction validity.[4]

In addition to the four types of validity of the questions above, one type of validity is considered to have an important role in determining the quality of the question items, namely the validity of learning outcome test items. The item validity of a test is the accuracy of measuring what an item has (which is an integral part of the test as a totality) in measuring what should be measured through the item.[5], [6], [7]

If we want to pay close attention, then the learning outcome tests made or compiled by teachers, both teachers, lecturers, and other teaching staff, are a collection of many items, with items where the test compilers want to measure and reveal the learning outcomes that each student has achieved after they follow the learning process for a certain period. This statement confirms that each item in the learning outcome test is an inseparable part of the learning outcome test.

The problem of item validity is seen when teachers, lecturers or teaching staff only focus on making question items without considering matters related to the validity of question items. Systematically making question items, after writing the question items, a tryout or trial should be held to find out the extent to which the question items made by teachers and lecturers are in the valid category. Valid question items are worthy of being used as material in implementing learning outcome tests.

As stated earlier, the main task of lecturers is to carry out learning. The lecturer concerned must make measurements by making learning outcomes tests to determine the learning outcomes. Lecturer-made tests are significant concerning the effectiveness of the learning process and improving student learning outcomes. Based on the description above, the problem the researcher will examine is "how to analyze the validity of ítem learning outcome tests made by English lecturers at the IAIN Sultan Amai Gorontalo Language Development Center".[8]

LITERATURE REVIEW

English is a subject that encourages, guides, develops, and fosters abilities and positive attitudes towards English, both receptive and productive.[9], [10], [11] Assertive ability is the ability to understand other people's speech and understand reading.[12], [13], [14] Productive ability is the ability to use language as a means of communication, both orally and in writing.[15] English language skills and a positive attitude towards English are essential in helping understand the source of Islamic teachings,[16] namely the Qur'an and Hadith as English books related to Islam for students.

For this reason, English at the University level is prepared to achieve essential language competencies, which include four language skills taught integrally: listening, speaking, reading, and writing.[17] Even so, at the elementary education level, the emphasis is on listening and speaking skills as a basis for language. At the secondary education level (intermediate), the four language skills are taught in a balanced manner. The advanced education level is concentrated on reading and writing skills, so students are expected to be able to access various English references.

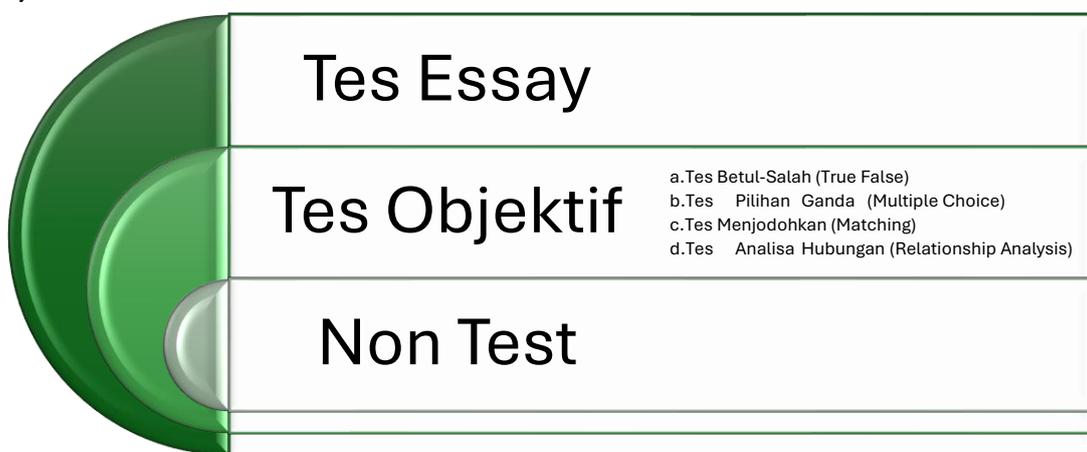
English subjects have the following objectives:

- a. Develop the ability to communicate in oral and written English, which includes four language skills.
- b. Raising awareness about the importance of English as a foreign language to be the primary learning tool, especially in studying the sources of Islamic teachings.
- c. Develop an understanding of the interconnectedness between language and culture, broaden cultural horizons, and engage in cultural diversity.

There are several opinions from some experts about the meaning of the test. According to Anne Anastasi in her paper entitled Psychological Testing,[18] What is meant by a test is a measuring device with an objective standard so that it can be used widely and can be used to measure and compare psychological states or individual behaviour levels.[19] A test is a systematic procedure for comparing the behaviour of two or more people.[20], [21] According to Goodenough, as quoted by Robert Lado, a test is a task or series of tasks given to individuals or groups of individuals intended to compare proficiency.

From the understanding of these experts in the world of education, it can be concluded that tests are methods used or procedures taken in the context of measurement and assessment in the field of education, which provide tasks and a series of tasks given by teachers so that values can be produced that symbolize the level of behaviour or achievement of students. [22]

In terms of the form of questions and possible answers, the test is divided into three parts, namely:



In principle, the objective test form above has its disadvantages and advantages. Still, the accurate form can usually test all the material taught. In contrast, the description form is rather difficult to measure all the material taught because the test form's scope is very narrow. For more details, it is necessary to explain first the weaknesses and virtues of objective form tests. The advantage of the goodness of the accurate form in the evaluation of English learning outcomes for students is the objective form test; [23], [24] (1) appropriate to express learning outcomes that are ordered by knowledge, understanding, application, and analysis, (2) have a learning impact that encourages students to remember, interpret, and analyze opinions, and (3) The answers given can describe the realm of educational goals according to Bloom, primarily the cognitive domain.

The downside is that the test is objective; (1) Students are not required to organize answers because the answers have been provided, (2) Students may be able to guess the answers that are already available, (3) they are unable to unravel thought processes and reasoning, and (4) Measuring only the lowest cognitive domains does not reveal more complex abilities. The characteristics of a good test, according to Arikonto, must have a test requirement, which is to have:

Table 1. Features of a good test

No	Data	Information
1	Validitas	A test is called valid if the test can precisely measure what it wants to measure. For example, student participation in the teaching and learning process

2	Reliability	The test can be reliable if it gives regular results if tested many times. A test is said to be reliable if the test results show determination. If it relates to validity, validity is accuracy, and reliability is permanence.
3	Objectivity	A test is said to have objectivity if, in carrying out the test, no subjective factors influence it. This is especially true of its scoring system. When related to reliability, objectivity emphasizes permanence in the scoring system, while reliability emphasizes permanence in test results.
4	Practicality	A test is said to have high practicability if the test is practical and easy to administer. A good test is easy to do, easy to check, and comes with clear instructions.
5	Economy	It is that the implementation of the test does not require expensive costs or costs, a lot of energy, and a long time.

An item can be said to have high validity or be declared valid if the scores on the item concerned have conformity or alignment of direction with the total score. Or, with statistical language, a significant positive correlation exists between an item's and complete scores. The total score here is a dependent variable, while the item score is an independent variable. If so, then to conclude that the items that want to know their validity, that is, valid or not, we can use the correlation technique as an analysis technique. An item can be declared valid if its score is proven to have a significant positive correlation with its total score.

The next issue is to choose and determine the type of correlation technique considered appropriate to test the item's validity. In objective tests, there are only two possible answers: right and wrong. Each question item answered correctly is generally given a score of 1 (one), while 0 (zero) is provided for each wrong answer. This data type, namely true-false, yes-no, or something similar, is known in statistics as pure discrete or dichotomous data. While the total score owned by each testee is the sum of each score owned by each item (for example: 0+1+1+0+1+0+1+1+0+0+1=6), it is continuous data.

According to existing theory, if variable I is pure discrete data or dichotomous data, while variable II is continuous data, then the correlation technique between variable I and variable II is a biserial point correlation technique, where the correlation index number given the *rpbi* symbol can be obtained using the formula:

Table 1. Correlation formula between test variables

No	Symbol	Rumus
		$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$
1	<i>rpbi</i>	The biserial point correlation coefficient symbolizes the strength of the correlation between variable I and variable II, which, in this case, is considered an item validity coefficient.
2	<i>M_p</i>	The calculated average score of the testee, for which the item in question has been answered correctly.
3	<i>M_t</i>	Average score from the total score.

4	SDt	The standard deviation of the total score.
5	p	The proportion of testees answered correctly to the item being tested for item validity.
6	q	The proportion of testees answered incorrectly to the item being tested for item validity.

METHOD

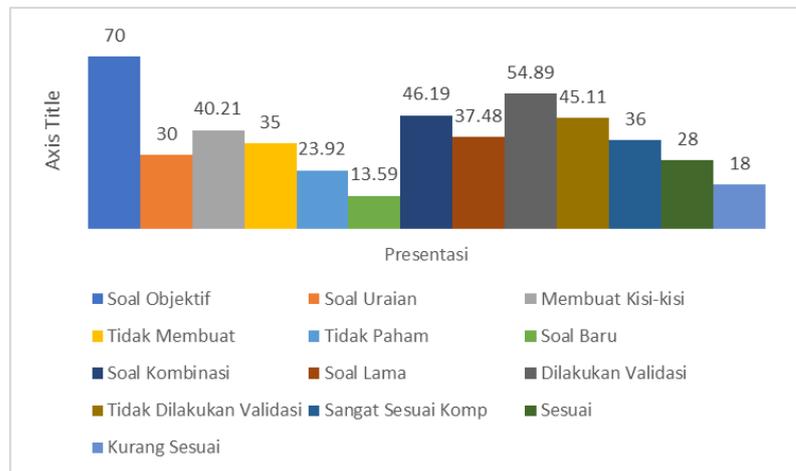
This research is an evaluative research that evaluates question items using a quantitative approach.[25][26][27] Based on certain criteria, evaluative research assesses the success, benefits, usefulness, contribution, and feasibility of an institution's program, product, or activity.[28], [29], [30]

This research is located at the IAIN Sultan Amai Gorontalo Language Development Center. The population of this study was all English lecturers at the Language Center, as many as 192 people, then researchers took a random sample of 92 lecturers.

The data collection technique is by documentation, a collection of questions in the IAIN Sultan Amai Gorontalo Language Development Center. This question will be tested to determine the validity level of the question items. The data analysis technique used to complete this research is quantitative descriptive analysis, namely by exposing and presenting data found in the field with a quantitative display. This is because the data found and completed is in the form of numbers or results of tryouts or trials of English questions at the Language Development Center.[31].

RESULT AND DISCUSSION

Data analysis of research results focused on the form of questions made by lecturers, steps in preparing questions, steps in validating questions, and the quality of questions based on theoretical reviews.



Picture 1. Validation of English Learning Outcomes Test Questions

The description of the results of the analysis is as follows.

1. Based on the analysis results, it appears that more than half of the lecturers of the Language Center make soat learning outcomes tests in objective form, which is as much as 70%, while about 30% make tests in the form of description. The preparation of learning outcome test questions that tend to be objective is carried out considering the nature of

many English courses that are applicable.

2. Based on the results of the analysis, it can be stated that around 40.21% of English lecturers at the Language Center pay attention to the grids in the process of preparing learning outcome test questions. In comparison, 35% of lecturers do not make grids, and 23.92% do not respond to whether to make grids because the manufacturing process has paid attention to the competence of the learning outcomes set.
3. Based on the data description, only about 13.59% of lecturers made new questions for the students' final semester exams. In comparison, 46.19% of lecturers made a combination of last year's questions and further questions, and only a number used in the previous year's questions. In addition, there were 37.48% of lecturers did not respond to making questions.
4. Theoretical validity: The results of the analysis showed that only 54.89% of lecturers re-examined the questions prepared to prepare quality learning outcome test questions. The remaining 45.11% of lecturers did not conduct a review because this study had been carried out when the question was made. All aspects related to the quality of objective form questions have been reviewed theoretically by lecturers (85.72%).
5. Quality of questions: based on the theoretical review of the data description, it can be stated that with an academic review of aspects related to the quality of the questions, it can be said that the overall learning outcome test questions made by lecturers are 85% good. However, the results of the analysis conducted by researchers still found that some can be stated that 36% of all objective form questions do not include the weight of the value of each item, 28% of questions give freedom of choice of the questions to do, and 18% of question sets that are considered weak in the images accompanying the questions. In addition, it is still found in 35% of question sets where some of the main question items are not one page with answer choices, and 15% of questions still have pictures accompanying the questions, including unclear. In addition, 34% of question sets only assess the competence of learning outcomes limited to knowledge and understanding.

Based on a brief description of the data description in advance, four main things can be revealed that are problems in this study. First, more than half of the lecturers of English courses at the IAIN Sultan Amai Gorontalo Language Development Center (70.00%) made learning outcome tests in the form of multiple-choice objective tests. This is understandable because the characteristics of English courses are more inclined to application and problem-solving. With accurate tests, students are expected to develop their reasoning skills in solving various language problems practically. Thus, it is unsurprising that only 30.00% of lecturers make learning outcome test questions in descriptions. According to the author, it will be even more meaningful if the learning outcomes test is made with objective tests.

Second, in developing and making questions, only 40.21% of lecturers made a grid of questions, 23.92% did not, and even 35.87% of lecturers were unclear whether to make a grid of questions. Looking at these data, it can also be said that lecturers have not all completed the initial steps of making tests as expected in planning and developing learning outcome test questions. This is certainly not optimal because to be able to make good questions, a grid of questions is needed. With the question grid, the questions made are expected not to deviate from the essential compatibility set. In addition, around 46.19% of the lecturers of English courses at the IAIN Sultan Amai Gorontalo Language Center made questions by combining new and last year's questions. It would be better if all the test questions were made fresh because the previous year's questions may still be easy to remember by students who happen to repeat them unless there is a modification to the year's questions. Based on data analysis, only 13.59% of IAIN Sultan Amai Gorontalo course lecturers asked utterly new questions. This percentage is still considered low because all lecturers should be able to develop further test questions following the demands of competence and the

development of science.

Third, in the process of validating and verifying the quality of the questions, it can be said to be quite good. As many as 54.89% of IAIN Sultan Amai Gorontalo lecturers reviewed the learning outcome test questions before being tested. However, it would be perfect and meaningful if all lecturers re-reviewed the test questions that had been prepared. The review includes three things, namely, from the aspects of matter, construction, and language. Because almost half of the lecturers did not conduct a theoretical review, it is understandable if there are still some sets of learning outcome test questions made by lecturers that have not met the rules as a good test. In addition, a good review is if it is done together with other lecturers in the same field, not done alone. Based on data analysis, only a tiny percentage of re-studies were carried out with other lecturers in the same field, around 17.85%. Suppose all re-studies are carried out with other lecturers in the same field. In that case, the inaccuracy of the language and substance of the test can be overcome because of input and suggestions from various parties involved in the re-review.

In examining the questions, it can be said that all aspects that should be observed have been carried out. However, there are still some aspects that lecturers pay less attention to. For example, for the objective test of multiple-choice form, observing image clarity was only done by 49.50% of lecturers, and keeping language accuracy was carried out by 26.50%. It can be said that examining these two aspects is still incomplete. If the picture accompanying the question item is unclear, it will affect students' accuracy in answering the test questions. Including in terms of language should not cause multiple interpretations or various interpretations.

Likewise, a review of objective tests shows that not half of IAIN Sultan Amai Gorontalo lecturers conduct theoretical studies of all aspects that should be considered. Three aspects are still very poorly considered by lecturers in making objective form tests, namely the homogeneity of answer choices (17.75%), the length of alternative answer choices that are relatively the same (16.75%), and the study of the function of images/graphics (14%). Based on the data, it can also be said that theoretical studies of essay-form test questions are more concerned than studies of objective-form test questions.

Fourth, several weaknesses were found based on the researcher's analysis of the question documents sampled. The most dominant weakness is that test questions are more likely to measure aspects of knowledge and understanding that have not penetrated much on application, analysis, synthesis and evaluation. Measuring learning ability from application to evaluation is the main thing to do for the student level. In addition, unclear images are still found for objective tests, freedom to choose the questions to do and the weight of values for each item. This will all affect the accuracy of the conclusion of the assessment results on the level of learning competence. In addition, objective tests, the inhomogeneity of answer choices (the short length of statements of all answer choices), and the separation of the subject matter page from the answer page indicate that the study has not been conducted in depth. In other words, there is a relationship between the activities carried out in reviewing the quality of the questions and the various weaknesses of the questions used for the exam.

CONCLUSION AND IMPLICATIONS

IAIN Sultan Amai Gorontalo lecturers tend to make learning outcome test questions in objective form based on practical courses' characteristics. With accurate tests, developing questions

that can reveal students' applicative abilities, analysis, synthesis, and evaluation is more accessible than objective tests.

In designing learning outcome test questions, not all lecturers (around 76%) make a grid of questions used as a reference for writing question items. Even though the grid is essential in the process of writing good questions. In addition, in writing questions, most lecturers combine last year's and new questions. Making further questions related to the development of science and technology will be more meaningful.

Not all lecturers of English courses at IAIN Sultan Amai Gorontalo have conducted a theoretical review regarding the quality of the questions that have been asked. However, one thing that is quite good and positive in the review that has been carried out by most lecturers, namely the most re-examined aspect, is the suitability of the question material with the competence of learning outcomes. There are still some weaknesses in some question sets, namely for objective tests. The aspect that is still weak is the non-uniformity of the short length of statements between answer choices and the separation of the main question page from the answer page. Another drawback to this form of the test is that there are still pictures accompanying the questions that are not clear.

BIBLIOGRAPHY

- [1] I. R. N. Hula, M. Z. Papatungan, and Ana Mariana, "Pengembangan Hybrid Learning Berbasis Aplikasi Computer Assited Test (Cat) Pada Program Arabic Proficiency Test," *Tadbir J. Manaj. Pendidik. Islam*, vol. 9, no. 1, pp. 103–125, 2021, <http://doi.org/10.30603/tjmpi.v9i1.2063>
- [2] F. S. Siskawati, A. A. Suci Qurrota A'yun, and T. N. Irawati, "Analisis Kelayakan Butir Soal pada Media INTERMATHLY (Interesting Mathematic Monopoly)," *J. Cendekia J. Pendidik. Mat.*, vol. 6, no. 1, pp. 634–654, 2022, <http://doi.org/10.31004/cendekia.v6i1.1181>
- [3] I. Magdalena, F. Ramadanti, S. Tsania, H. N. Fauzi, N. Rosita, and S. Arif, "Validitas dan Reabilitas Soal Tes Uji Kemampuan Matematika Siswa Kelas V MIT Al Fuad Kota Tangerang," *Masaliq*, vol. 3, no. 1, pp. 28–38, 2023, <http://doi.org/10.58578/masaliq.v3i1.760>
- [4] A. Karim, "Analisis Kualitas Soal Perlombaan Matematika Tingkat SMA," *Titian Ilmu J. Ilm. Multi Sci.*, vol. 10, no. 1, pp. 1–8, 2018, <http://doi.org/10.30599/jti.v10i1.126>
- [5] A. Fauziyah MF, Y. Hala, and A. Asmawati Aziz, "Uji Kevalidan Instrumen Tes Berorientasi Higher Order Thinking Skill (Hots) Pada Materi Sistem Ekskresi Kelas Xi Sma/Ma," *J. Biog.*, vol. 8, no. 1, pp. 329–333, 2023, <http://doi.org/10.30605/biogenerasi.v8i1.2125>
- [6] W. A. Nugroho, R. P. Yudha, S. Sundari, and H. N. Praja, "Analisis Instrumen Asesmen Unjuk Kerja pada Pembelajaran PJOK di Sekolah Dasar Kota Cirebon," *Gelangg. Olabraga J. Pendidik. Jasm. dan Olabraga*, vol. 4, no. 2, pp. 126–141, 2021, <http://doi.org/10.31539/jpjo.v4i2.1795>
- [7] K. Basri, B. Baidowi, J. Junaidi, and M. Turmuzi, "Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP," *Griya J. Math. Educ. Appl.*, vol. 1, no. 4, pp. 682–694, 2021, <http://doi.org/10.29303/griya.v1i4.107>
- [8] Y. Song, "System Design for Online Foreign Language Education Based on Blockchain Technology," vol. 2022, 2022.
- [9] J. Ilmu and P. Islam, "Teori Operant Conditioning Dalam Pembelajaran Bahasa Inggris Dan Bahasa Arab," *J. Ilmu Pendidik. Islam*, vol. 20, no. 1, pp. 139–158, 2022, [doi:](https://doi.org/10.30603/tjmpi.v20i1.139)

- <https://doi.org/10.36835/jipi.v20i1.3938>.
- [10] R. vahrotun Nisa, “Peranan Madrasah Dalam Meningkatkan Kemampuan Bahasa Arab Sebagai Bahasa Internasional,” *J. An Nabighoh*, vol. 19, no. 02, pp. 225–248, 2017, doi: <https://doi.org/10.32332/an-nabighoh.v19i2.1001>.
- [11] R. D. Rahmawati and A. Shofiyani, “Strategi Pembelajaran Menulis Bahasa Arab Untuk Mahasiswa Program Studi Bahasa Inggris,” *J. Educ. Dev.*, vol. 8, no. 3, pp. 298–303, 2020, doi: <https://doi.org/10.37081/ed.v8i3.1902>.
- [12] S. Khotiah, “Peningkatan Kompetensi Membaca Teks Berbahasa Arab melalui Metode Qiraah Mata Pelajaran Bahasa Arab di MTs Negeri 8 Karangmojo,” *J. Pendidik. Madrasah*, vol. 5, no. 2, pp. 237–246, 2020, <http://doi.org/10.14421/jpm.2020.52-09>
- [13] Hidayatul Khoiriyah, “Metode Qirā’ah dalam Pembelajaran Keterampilan Reseptif Berbahasa Arab untuk Pendidikan Tingkat Menengah,” *Lisanuna*, vol. 10, no. 1, pp. 32–44, 2020, doi: <http://dx.doi.org/10.22373/ls.v10i1.7804>.
- [14] A. Husna and D. Eliza, “Strategi Perkembangan dan Indikator Pencapaian Bahasa Reseptif dan Bahasa Ekspresif pada Anak Usia Dini,” *J. Fam. Educ.*, vol. 1, no. 4, pp. 38–46, 2021, <http://doi.org/10.24036/jfe.v1i4.21>
- [15] N. H. Amalia and N. Hidayat, “Penggunaan Media Teka-Teki Silang (Crossword Puzzle) dalam Meningkatkan Kemampuan Baca Tulis Kosakata Bahasa Arab Peserta Didik Kelas III MI Ma’arif Giriloyo 1 Bantul,” *Al-Bidayah J. Pendidik. Dasar Islam*, vol. 10, no. 1, pp. 119–134, 2018, <http://doi.org/10.14421/al-bidayah.v10i1.133>
- [16] D. C. Sondakh and M. F. Sya, “Kesulitan Pembelajaran Kosakata Bahasa Inggris Tingkat Sekolah Dasar,” *Karimah Taubid*, vol. 1, pp. 9–10, 2022, doi: <https://doi.org/10.30997/karimahtauhid.v1i3.7818>.
- [17] A. M. Albantani, “Pembelajaran Bahasa Arab di Madrasah Ibtidaiyah : Sebuah Ide Terobosan,” *Attadib J. Elem. Educ.*, vol. 3, no. 2, pp. 160–173, 2018, doi: <https://doi.org/10.32507/attadib.v2i2.417>.
- [18] E. Mayasari, “Instrumen Tes Sebagai Alat Evaluasi (Analisis Soal, Indek Kesukaran, Daya Pembeda dan Fungsi Distraktor),” *J. Pendidik. dan Pengajaran*, vol. 2, no. 1, pp. 56–66, 2023.
- [19] L. Dachliyani, “Instrumen yang Sahih : Sebagai Alat Ukur Keberhasilan Suatu Evaluasi Program Diklat (evaluasi pembelajaran),” *Med. Media Inf. dan Komun. Diklat Kepustakawanan*, vol. 5, no. 1, pp. 57–65, 2019, [Online]. Available: <https://ejournal.perpusnas.go.id/md/article/view/721/0>
- [20] H. A. Multin, W. Munawar, and A. A. M. Noor, “Penyusunan Dan Analisis Tes Kinerja (Performance Test) Pada Kompetensi Praktik Memasang Sistem Penerangan Dan Wiring Kelistrikan Di Smk,” *J. Mech. Eng. Educ.*, vol. 5, no. 2, p. 176, 2019, <http://doi.org/10.17509/jmcc.v5i2.15185>
- [21] D. Paramartha, “Pengembangan Asesmen Minat-Bakat Berbasis Computer Based Test,” *J. Bimbing. Konseling Indones.*, vol. 1, no. 1, p. 19, 2016, doi: https://doi.org/10.23887/jurnal_bk.v6i2.763.
- [22] I. F. Ahmad, “Asesmen Alternatif Dalam Pembelajaran Jarak Jauh Pada Masa Darurat Penyebaran Coronavirus Disease (Covid-19) Di Indonesia,” *Pedagog. J. Pendidik.*, vol. 7, no. 1, pp. 195–222, 2020, <http://doi.org/10.33650/pjp.v7i1.1136>
- [23] Muhammad Nashrullah, “Penggunaan Media Google Form Dalam Evaluasi Pembelajaran

-
- Bahasa Arab (Pilihan Ganda),” *J. Naskhi J. Kaji. Pendidik. dan Bhs. Arab*, vol. 3, no. 1, pp. 30–40, 2021, <http://doi.org/10.47435/naskhi.v3i1.553>
- [24] L. Fitrianti, “Prinsip Kontinuitas Dalam Evaluasi Proses Pembelajaran,” *J. Pendidik.*, vol. 10, no. 1, pp. 89–102, 2018, doi: <https://doi.org/10.35445/alishlah.v10i1.68>.
- [25] R. Meisya, R. Jannah, and S. Ramadhan, “Analisis Kualitas Butir Soal Tematik Madrasah Ibtidaiyah Menggunakan Model Rasch,” *Al-Madrasah J. Pendidik. Madrasah Ibtidaiyah*, vol. 7, no. 4, p. 1764, 2023, <http://doi.org/10.35931/am.v7i4.2712>
- [26] A. Fauziana and A. Dessy Wulansari, “Analisis Kualitas Butir Soal Ulangan Harian di Sekolah Dasar dengan Model Rasch,” *Ibriez J. Kependidikan Dasar Islam Berbas. Sains*, vol. 6, pp. 10–19, 2021, <http://doi.org/10.21154/ibriez.v6i1.112>
- [27] S. Ramadhani, “Analisis Butir Soal HOTS (Higher Order Thinking Skills) Mata Pelajaran PAI di SMA Negeri 1 Peukan Bada Aceh Besar,” vol. 13, no. 1, pp. 1–23, 2016, doi: <https://doi.org/10.54604/tdb.v13i1.236>.
- [28] S. Sitepu, “Analisis Kesesuaian Materi Ajar Dengan Tes Hasil Belajar Mahasiswa Di Lingkungan Fkip Uhn Medan,” *Sepren*, vol. 2, no. 1, p. 1, 2020, <http://doi.org/10.36655/sepren.v2i1.276>
- [29] Muhimmatul Choirah, “Evaluasi Pembelajaran Bahasa Arab Berbasis Media E-Learning,” *J. Naskhi J. Kaji. Pendidik. dan Bhs. Arab*, vol. 3, no. 1, pp. 41–47, 2021, <http://doi.org/10.47435/naskhi.v3i1.554>
- [30] M. Arsyam and M. Yusuf Tahir, “Ragam Jenis Penelitian dan Perspektif,” *Al-Ubudiyah J. Pendidik. dan Stud. Islam*, vol. 2, no. 1, pp. 37–47, 2021, <http://doi.org/10.55623/au.v2i1.17>
- [31] J. Fleckenstein, S. Keller, M. Krüger, R. J. Tannenbaum, and O. Köller, “Linking TOEFL iBT® writing rubrics to CEFR levels: Cut scores and validity evidence from a standard setting study,” *Assess. Writ.*, vol. 43, no. July 2018, p. 100420, 2020, <http://doi.org/10.1016/j.asw.2019.100420>