



## Meninjau Kualitas Soal Bahasa Arab di Pendidikan Tinggi: Analisis Empiris dan Rekomendasi Perbaikan Soal

Ulya Nur Alim<sup>1</sup>

<sup>1</sup>Penelitian dan Evaluasi Pendidikan, Universitas Negeri Makassar  
E-mail: [240012301007@student.unm.ac.id](mailto:240012301007@student.unm.ac.id)

---

### Article Info

#### Article history:

Received June 05, 2025

Revised June 15, 2025

Accepted June 21, 2025

---

#### Keywords:

Item Analysis, Difficulty Level, Discrimination Index, Arabic Language, Bloom's Taxonomy

---

### ABSTRACT

*This study aims to analyze the quality of test items in the Mid-Semester Examination (UTS) for the Arabic language course Imla' in the Undergraduate Program of Arabic Language Education at Universitas Negeri Makassar. The primary focus of the research is on the difficulty level and discrimination index of the test items. A descriptive quantitative approach was employed, involving 25 second-semester students as subjects and 44 test items as the objects of analysis. Data were analyzed using Iteman 4.3 software. The results revealed that most items fell into the "easy" and "moderate" difficulty categories, while 45% of the items had poor discrimination indices, indicating their ineffectiveness in distinguishing students' levels of understanding. Five items were identified as problematic due to being too easy and having low discrimination values. Therefore, it is recommended that these items be revised by increasing their complexity in accordance with Bloom's taxonomy to foster higher-order thinking skills. These findings serve as a valuable foundation for designing training programs in item writing and analysis for lecturers and prospective educators, particularly within the context of Arabic language instruction in higher education.*

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Ulya Nur Alim

Penelitian dan Evaluasi Pendidikan, Universitas Negeri Makassar

Email: [240012301007@student.unm.ac.id](mailto:240012301007@student.unm.ac.id)

---

### Article Info

#### Article history:

Received June 05, 2025

Revised June 15, 2025

Accepted June 21, 2025

---

#### Keywords:

Analisis Butir Soal, Tingkat Kesukaran, Daya Beda, Bahasa Arab, Taksonomi Bloom.

---

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Ujian Tengah Semester (UTS) Bahasa Arab (mata kuliah Imla') pada Program Studi S-1 Pendidikan Bahasa Arab di Universitas Negeri Makassar. Fokus utama penelitian adalah pada aspek tingkat kesukaran dan daya beda butir soal. Metode yang digunakan adalah pendekatan deskriptif kuantitatif dengan melibatkan 25 mahasiswa semester II sebagai subjek dan 44 butir soal sebagai objek penelitian. Data dianalisis menggunakan perangkat lunak Iteman 4.3. Hasil analisis menunjukkan bahwa sebagian besar soal tergolong dalam tingkat kesukaran "mudah" dan "sedang", serta 45% soal memiliki daya beda "jelek" atau tidak mampu membedakan kemampuan peserta didik secara efektif. Lima soal diidentifikasi sebagai item bermasalah karena terlalu mudah dan memiliki daya beda rendah. Oleh karena itu, perbaikan soal disarankan dengan meningkatkan kompleksitas soal sesuai dengan taksonomi Bloom agar mendorong keterampilan berpikir tingkat tinggi. Temuan ini menjadi dasar



penting dalam merancang pelatihan penyusunan dan analisis soal bagi dosen dan calon pendidik, khususnya dalam pembelajaran Bahasa Arab di pendidikan tinggi.

*This is an open access article under the [CC BY-SA](#) license.*



**Corresponding Author:**

Ulya Nur Alim  
Penelitian dan Evaluasi Pendidikan, Universitas Negeri Makassar  
Email: [240012301007@student.unm.ac.id](mailto:240012301007@student.unm.ac.id)

## Pendahuluan

Asesmen merupakan komponen penting dalam sistem pendidikan karena menjadi alat untuk mengukur pencapaian belajar peserta didik serta dasar untuk pengambilan keputusan dalam pengajaran (Yulis dkk, 2025). Di tingkat perguruan tinggi, asesmen yang berkualitas merupakan bagian integral dari proses pembelajaran yang sangat diperlukan untuk memastikan bahwa hasil belajar mahasiswa sesuai dengan capaian pembelajaran yang telah ditetapkan dalam kurikulum (Nusaibah, 2023). Salah satu bentuk asesmen yang sering digunakan adalah tes tertulis dengan format pilihan ganda. Namun, kualitas butir-butir soal dalam tes tersebut sangat menentukan validitas dan reliabilitas hasil asesmen (Oc & Hassen, 2025).

Salah satu permasalahan yang kerap muncul dalam praktik penyusunan tes di lingkungan sekolah adalah minimnya pemahaman guru atau kurangnya perhatian terhadap prinsip-prinsip dalam penyusunan dan analisis kualitas butir soal (Subando dkk., 2023; Misda & Mukhlis, 2023; Fietri dkk., 2021). Banyak guru belum mendapatkan pembelajaran dari perguruan tingginya dahulu serta pelatihan yang memadai dalam menganalisis kualitas soal,

khususnya dalam hal tingkat kesukaran dan daya beda. Hal ini berdampak pada penggunaan butir soal yang tidak berkualitas, yang pada akhirnya dapat merugikan peserta didik karena hasil asesmen tidak mencerminkan kemampuan sebenarnya (Iskandar & Rizal, 2018).

Tingkat kesukaran (difficulty index) menunjukkan proporsi peserta didik yang menjawab soal dengan benar. Tingkat kesukaran yang ideal ditandai dengan proporsi butir soal yang sebagian besar berada pada kategori sedang, sementara hanya sedikit yang berada pada kategori mudah maupun sulit (Dia & Syah, 2022).

Sementara itu, daya beda (discrimination index) menunjukkan kemampuan suatu butir soal dalam membedakan antara peserta didik yang memiliki kemampuan tinggi dan rendah. Butir soal yang memiliki daya beda tinggi mampu mengidentifikasi secara tepat peserta didik yang menguasai materi dengan baik (Haladyna & Rodriguez, 2013).

Tanpa analisis terhadap kedua aspek tersebut, soal yang digunakan dalam asesmen dapat menjadi tidak representatif dan tidak adil (Saputri dkk., 2023). Di lingkungan pendidikan tinggi, keberadaan soal-soal berkualitas menjadi semakin



krusial karena menyangkut pengembangan kemampuan berpikir tingkat tinggi dan profesionalisme mahasiswa. Oleh karena itu, pelatihan dan peningkatan pengetahuan dosen dalam menganalisis kualitas butir soal sangat mendesak untuk dilakukan.

Urgensi dari penelitian ini terletak pada pentingnya melakukan analisis butir soal mata kuliah Bahasa Arab di tingkat perguruan tinggi, mengingat peninjauan terhadap kualitas soal masih jarang menjadi fokus dalam pengajaran bahasa. Padahal, validitas asesmen sangat menentukan arah dan efektivitas pembelajaran. Selain itu, kajian ilmiah yang secara khusus membahas analisis kualitas butir soal Bahasa Arab di tingkat perguruan tinggi masih terbatas, sehingga penelitian ini tidak hanya menjawab kebutuhan praktis dalam peningkatan mutu asesmen, tetapi juga berkontribusi dalam memperkaya literatur akademik di bidang tersebut.

Sejalan dengan permasalahan dan urgensi yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk menganalisis kualitas butir soal (tingkat kesukaran dan daya beda) Ujian Tengah Semester (UTS) bahasa Arab mata kuliah Imla' pada mahasiswa Program Studi S-1 Pendidikan Bahasa Arab di Universitas Negeri Makassar.

Hasil dari penelitian ini diharapkan dapat menjadi dasar dalam merancang pelatihan analisis butir soal bagi para guru dan dosen, khususnya dalam konteks pembelajaran Bahasa Arab. Lebih jauh, temuan penelitian ini dapat digunakan untuk mengembangkan modul atau materi ajar tentang analisis kualitas soal yang dapat diajarkan kepada mahasiswa, terutama calon pendidik, sebagai bagian dari upaya sistematis meningkatkan literasi asesmen di lingkungan universitas.

## **Metode**

Penelitian ini menerapkan pendekatan deskriptif kuantitatif dengan fokus pada analisis tingkat kesukaran dan daya beda butir soal dengan pendekatan CTT. Subjek dalam penelitian ini terdiri atas 25 mahasiswa semester 2 Tahun Ajaran 2024/2025 dari Program Studi S-1 Pendidikan Bahasa Arab, Universitas Negeri Makassar. Sementara itu, objek penelitian mencakup 44 butir soal Ujian Tengah Semester (UTS) Bahasa Arab, khususnya materi Imla'.

Pengumpulan data dilakukan melalui dokumentasi soal UTS dan hasil jawaban mahasiswa. Selanjutnya, data dianalisis menggunakan perangkat lunak Itean versi 4.3 untuk memperoleh informasi kuantitatif mengenai indeks tingkat kesukaran dan daya beda tiap butir soal. Hasil analisis diinterpretasikan berdasarkan kriteria klasifikasi yang telah baku.

## **Hasil**

Soal UTS idealnya mencakup soal dengan variasi tingkat kesukaran—mudah, sedang, dan sulit—agar dapat memberikan gambaran yang proporsional tentang penguasaan materi siswa pada pertengahan semester (Damayanti dkk, 2021). Soal mudah tetap penting untuk menilai pemahaman dasar (Rajagukguk & Naibaho, 2023), soal sedang untuk membedakan siswa berdasarkan tingkat penguasaan (Andini & Mukhlis, 2023), dan soal sulit untuk menantang siswa yang lebih unggul (Taftazani dkk, 2025). Jika soal UTS terlalu didominasi oleh soal mudah, maka hasil tes menjadi kurang informatif dan tidak mencerminkan variasi kemampuan siswa secara akurat. Selain itu, distribusi skor bisa menjadi bias dan reliabilitas tes pun menurun.



Tabel 1 menyajikan analisis tingkat kesukaran butir soal berdasarkan nilai proporsi jawaban benar (nilai P). Tingkat kesukaran butir soal dikelompokkan ke dalam lima kategori: terlalu mudah, mudah, sedang, sulit, dan terlalu sulit.

Dari total 44 butir soal, distribusi tingkat kesukaran menunjukkan bahwa sebagian besar butir soal (16 butir atau 36%) tergolong dalam kategori "sedang", menunjukkan butir-butir ini memiliki tingkat kesulitan yang optimal untuk membedakan kemampuan peserta didik.

Sebanyak 15 butir soal (34%) berada dalam kategori "mudah", dan 12 butir soal (27%) tergolong "terlalu mudah". Ini mengindikasikan adanya sejumlah butir soal yang mungkin kurang menantang bagi sebagian besar peserta didik. Hanya ada 1 butir soal (2%) yang termasuk dalam kategori "sulit" (butir nomor 36) dan tidak ditemukan butir soal yang termasuk dalam kategori "terlalu sulit" (0 butir atau 0%).

Analisis ini menunjukkan bahwa instrumen memiliki kecenderungan butir soal yang lebih banyak berada pada kategori mudah hingga sedang.

Butir soal yang terlalu mudah atau terlalu sulit perlu dievaluasi karena tidak mampu membedakan kemampuan mahasiswa secara akurat. Soal semacam ini cenderung menurunkan kualitas asesmen dan tidak mencerminkan pencapaian belajar yang sebenarnya (Iskandar & Rizal, 2018).

Instrumen tes harus memiliki daya beda yang baik karena daya beda mencerminkan kemampuan suatu butir soal untuk membedakan antara peserta tes yang memiliki kemampuan tinggi dan rendah. Butir yang memiliki daya beda yang baik mampu mengidentifikasi peserta yang

menguasai materi dengan baik, sekaligus membedakan mereka dari peserta yang kurang menguasai. Sebaliknya, butir dengan daya beda rendah atau negatif cenderung tidak memberikan informasi yang valid tentang perbedaan kemampuan peserta (Magdalena dkk., 2021).

Tabel 2 menyajikan distribusi butir soal berdasarkan daya bedanya. Daya beda menunjukkan kemampuan butir soal untuk membedakan antara peserta didik yang berkemampuan tinggi dan rendah. Hasil analisis menunjukkan bahwa 20 butir soal (45%) memiliki daya beda "jelek" (Nilai  $D \leq 0,20$ ) dan "perlu diperbaiki". Butir-butir ini tidak efektif dalam membedakan peserta didik. Sebanyak 9 item (20%) memiliki daya beda "cukup" (Nilai  $D 0,20 - 0,30$ ), 4 item (9%) "baik" (Nilai  $D 0,30 - 0,40$ ), dan 11 item (25%) memiliki daya beda "baik sekali" (Nilai  $D > 0,40$ ).

Proporsi butir soal dengan daya beda "jelek" yang mencapai 45% mengindikasikan bahwa hampir separuh dari instrumen tes tidak efektif dalam membedakan antara peserta didik berkemampuan tinggi dan rendah. Kondisi ini berimplikasi pada validitas serta fungsi diagnostik tes yang rendah. Akibatnya, skor yang diperoleh peserta didik dengan kemampuan berbeda mungkin tidak secara akurat mencerminkan kapabilitas mereka yang sebenarnya. Apabila butir-butir soal ini terus digunakan tanpa perbaikan, akurasi hasil tes dalam mengukur capaian pembelajaran siswa akan berkurang, dan ini berpotensi menyebabkan keputusan penilaian yang tidak adil, karena semua siswa, tanpa memandang tingkat kemampuannya, memiliki peluang yang sama untuk menjawab benar.



Tabel 1. Tingkat Kesukaran Butir (Item) Soal

Nilai P	Kategori	Item	Frekuensi	Persentase
>0,90	Terlalu mudah	2,4,6,13,14,15,17,18, 31,32,33,34	12	27%
0,71-0,90	Mudah	1,8,9,10,12,16,22,24, 25,27,29,37,38,40,41,	15	34%
0,30-0,70	Sedang	3,5,7,11,19,20,21,23, 26,28,30,35,39,42,43,44	16	36%
0,10-0,30	Sulit	36	1	2%
<0,10	Terlalu sulit	-	0	0%
Total			44	100%

Tabel 2. Daya Beda Butir (Item) Soal

Nilai D	Kategori	Item	Frekuensi	Persentase
D>0,40	Baik sekali	2,4,8,9,13,19, 23,25,35,38,42	11	25%
0,30<D≤0,40	Baik	7,29,37,44	4	9%
0,20<D≤0,30	Cukup	5,6,11,14,17,20, 27,28,32	9	20%
D≤0,20	Jelek (perlu diperbaiki)	1,3,10,12,15,16, 18,21,22,24,26,30, 31,33,34,36,39,40, 41,43	20	45%
Total			44	100%

Butir soal dengan daya beda yang lemah juga berarti siswa dengan kemampuan tinggi tidak mendapatkan keuntungan yang sepadan. Jika soal-soal tidak cukup menantang dan tidak mampu mengukur kedalaman pemahaman siswa yang unggul, skor mereka mungkin tidak menunjukkan perbedaan signifikan dibandingkan siswa lain, sehingga hasil tes menjadi kurang sensitif dalam membedakan tingkat penguasaan materi. Kesalahan dalam penilaian dan pengambilan keputusan juga dapat timbul akibat soal dengan daya beda yang tidak baik. Siswa berkemampuan tinggi mungkin tidak tampak unggul, yang dapat menghambat peluang mereka untuk

mendapatkan penghargaan, beasiswa, atau rekomendasi. Sebaliknya, siswa dengan kemampuan rendah bisa saja "lulus" atau melanjutkan ke jenjang yang tidak sesuai dengan kapabilitasnya, dan ini merupakan bentuk ketidakadilan dalam asesmen pendidikan.

Butir-butir soal dengan daya beda "jelek" segera direvisi atau dieliminasi. Revisi dapat dilakukan dengan memperbaiki redaksi, tingkat kesukaran, atau ketepatan pengecoh (distraktor) agar butir mampu membedakan peserta didik secara lebih efektif.

Terdapat Lima butir soal UTS bahasa Arab yang menunjukkan karakteristik



sebagai item bermasalah karena memiliki daya beda rendah ( $D < 0.20$ ) dan tingkat kesukaran yang ekstrem, yaitu terlalu mudah ( $P > 0.90$ ).

Kelima item tersebut yaitu item 15, 18, 31, 33, dan 34, tidak efektif dalam mengukur kemampuan peserta secara optimal. Soal yang terlalu mudah cenderung tidak mampu mendistribusikan peserta berdasarkan kemampuan mereka, dan daya beda yang rendah mengindikasikan bahwa soal tersebut tidak mampu membedakan peserta yang berkemampuan tinggi dan rendah. Disarankan untuk mengeliminasi, merevisi, atau mengganti item-item ini.

Berikut merupakan rekomendasi perbaikan untuk salah satu item yang bermasalah (item 18):

18. أي من الكلمات التالية تبدأ بهمزة قطع؟

Tandai satu oval saja.

- اذهب  
 ابتسم  
 أكرم  
 اجتهذ

Gambar 1. Butir Soal Nomor 18

Sebagian besar mahasiswa menjawab benar pada butir soal ini, yang menunjukkan bahwa soal tersebut tidak efektif dalam membedakan mahasiswa dengan kemampuan tinggi dan rendah. Kemungkinan besar, soal ini tergolong terlalu mudah karena sebagian besar mahasiswa telah menguasai materi yang hanya menuntut kemampuan berpikir tingkat rendah menurut taksonomi Bloom.

Butir soal nomor 18, misalnya, hanya meminta mahasiswa untuk memilih pilihan yang memuat kata dengan awalan *hamzah qath'i*, yang berada pada level C1 (mengingat). Mahasiswa cukup mengingat ciri-ciri *hamzah qath'i* untuk dapat

menjawab soal tersebut dengan mudah. Selain itu, tingkat kemudahan soal ini juga diduga dipengaruhi oleh penyusunan opsi jawaban yang terlalu mencolok. Sebagai contoh, opsi C diawali dengan huruf *hamzah qath'i* (أ) yang secara visual berbeda mencolok dari opsi lainnya yang menggunakan *hamzah washal* (ا), sehingga peserta tes dapat menebaknya dengan cepat tanpa perlu melakukan analisis mendalam.

Revisi terhadap butir soal ini sebaiknya dilakukan dengan menaikkan tingkat kognitif, antara lain melalui penambahan konteks kalimat agar menuntut pemahaman yang lebih mendalam dari peserta tes. Contoh:

Perhatikan kalimat berikut!

- (1) أَكَلَ خَالِدٌ التَّفَاحَةَ بَعْدَ الْغَدَاءِ.
- (2) اسْتَعْفَرَ الرَّجُلُ رَبَّهُ فِي مِنتَصَفِ اللَّيْلِ.
- (3) أَحَبَّ الْوَلَدُ الْقِصَّةَ كَثِيرًا.
- (4) اعْتَمَدَ الطَّالِبُ عَلَى الْمَعْجَمِ لِفَهْمِ الْمَعْنَى.

Kalimat nomor berapa yang mengandung *hamzah qath'i* pada kata yang ditebali?

- a. 1 dan 2
- b. 2 dan 3
- c. 1 dan 3\*
- d. 3 dan 4

Melalui penyusunan soal dengan tingkat kompleksitas yang lebih tinggi, mahasiswa tidak hanya mengandalkan kemampuan kognitif dasar seperti mengingat, tetapi diarahkan untuk mengembangkan kemampuan analisis. Pendekatan ini berpotensi meningkatkan mutu asesmen serta menstimulasi keterampilan berpikir kritis.

Temuan dalam penelitian ini dapat dijadikan pijakan dalam merancang pelatihan penyusunan dan analisis soal bagi dosen maupun calon pendidik, khususnya dalam konteks pengajaran Bahasa Arab di tingkat perguruan tinggi.



## Kesimpulan

Penelitian ini menyimpulkan bahwa kualitas butir soal Ujian Tengah Semester (UTS) Bahasa Arab pada mata kuliah *Imla'* di lingkungan pendidikan tinggi masih perlu ditingkatkan. Dari 44 butir soal yang dianalisis, sebagian besar tergolong pada tingkat kesukaran “mudah” (34%) dan “sedang” (36%), sedangkan butir soal dengan tingkat kesukaran “sulit” sangat sedikit (2%), dan tidak ada yang termasuk kategori “terlalu sulit”. Hal ini mengindikasikan bahwa instrumen belum optimal dalam mencerminkan variasi kemampuan mahasiswa secara proporsional.

Analisis daya beda menunjukkan bahwa hampir setengah dari butir soal (45%) tergolong “jelek”, artinya tidak mampu membedakan mahasiswa dengan kemampuan tinggi dan rendah. Kondisi ini menurunkan validitas asesmen serta berpotensi menimbulkan keputusan evaluasi yang tidak adil.

Sebanyak lima butir soal (nomor 15, 18, 31, 33, dan 34) diidentifikasi sebagai item bermasalah karena memiliki daya beda rendah dan tingkat kesukaran yang terlalu mudah. Oleh karena itu, soal-soal tersebut disarankan untuk direvisi, diganti, atau dieliminasi.

Sebagai upaya perbaikan, disarankan agar penyusunan soal tidak hanya mempertimbangkan aspek teknis, tetapi juga diarahkan untuk meningkatkan level kognitif mahasiswa. Salah satunya adalah dengan mengubah bentuk soal menjadi lebih kompleks sesuai dengan taksonomi Bloom, misalnya dari level mengingat (C1) menjadi menganalisis (C4), dengan memberikan konteks yang lebih luas dan menuntut pemahaman mendalam. Pendekatan ini diharapkan dapat meningkatkan kualitas asesmen sekaligus

mendorong kemampuan berpikir kritis mahasiswa.

Hasil penelitian ini penting untuk dijadikan dasar dalam merancang pelatihan analisis dan penyusunan soal bagi dosen dan calon guru, khususnya dalam konteks pembelajaran Bahasa Arab di perguruan tinggi.

## Daftar Pustaka

- Andini, D. P., & Mukhlis, M. (2023). Analisis butir soal pada instrumen asesmen kompetensi minimum literasi membaca di SMP IT Insan Utama Pekanbaru. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 6(2), 401-412.
- Damayanti, W. W., Halidjah, S., & Pranata, R. (2021). Analisis Tingkat Kesukaran Butir Soal Pilihan Gandapada Penilaian Tengah Semester Kelas IV. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa (JPPK)*, 10(11).
- Dia, E. E., & Syah, A. S. (2022, September). Tingkat Kesukaran Soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Di Madrasah. In *Prosiding Conference on Research and Community Services* (Vol. 4, No. 1, pp. 14-23).
- Fietri, W. A., Zulyusri, Z., & Violita, V. (2021). Analisis butir soal biologi kelas XI Madrasah Aliyah Sakinah Kerinci menggunakan program komputer Anates 4.0 for Windows. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, 7(1), 28-35.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal*



- Penelitian Dan Evaluasi Pendidikan*, 22(1), 12-23.
- Magdalena, I., Fauziah, S. N., Fiazah, S. N., & Nopus, F. S. (2021). Analisis validitas, reliabilitas, tingkat kesulitan dan daya beda butir soal ujian akhir semester tema 7 kelas III SDN Karet 1 Sepatan.
- Misda, S., & Mukhlis, M. (2023). Analisis Butir Soal Literasi Membaca pada Instrumen Asesmen Kompetensi Minimum Siswa SMK: Analisis Butir Soal Literasi Membaca pada Instrumen Asesmen Kompetensi Minimum Siswa SMK Pekanbaru. *Literasi: Jurnal Ilmiah Pendidikan Bahasa, Sastra Indonesia dan Daerah*, 13(2), 467-481.
- Nusaibah, A. W. (2023). Analisis rumusan capaian pembelajaran pada kurikulum program studi Pendidikan Agama Islam. *Journal of Islamic Education and Innovation*, 82-91.
- Oc, Y., & Hassen, H. (2025). Comparing the effectiveness of multiple-answer and single-answer multiple-choice questions in assessing student learning. *Marketing Education Review*, 35(1), 44-57.
- Rajagukguk, M. J. T., & Naibaho, D. (2023). Mampu Memilih Soal Berdasarkan Tingkat Kesukaran. *Jurnal Pendidikan Sosial Dan Humaniora*, 2(4), 12736-12747. Retrieved from <https://publisherqu.com/index.php/pediaqu/article/view/701>
- Saputri, H. A. S., Zulhijrah, Nabila Joti Larasati, & Shaleh. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986 - 2995. <https://doi.org/10.36989/didaktik.v9i5.2268>
- Subando, J., Nashir, M. J., Rohimah, S., & Amin, H.L. (2023). Pelatihan Analisis Butir Soal dan Analisis Kemampuan Peserta Ujian Bagi Guru Madrasah di Pondok Pesantren Budi Utomo Surakarta. *MASPUL JOURNAL OF COMMUNITY EMPOWERMENT*, 5(2), 313-318. <https://doi.org/10.33487/pengabdian.v5i2.7533>
- Taftazani, T. M. T., Haifa, N. H., Safira, N., Putri, H. E., & Nurussama, A. (2025). Analisis Kualitas Butir Soal Pemahaman Siswa Kelas 1 SD Materi Penjumlahan dan Pengurangan Menggunakan Tes Buatan Guru Berbasis Program Anates. *Jurnal Kolaboratif Sains*, 8(1), 150-159.
- Yulis, S., Mutia, N. B., & Husna, R. (2025). Analisis Keterlaksanaan Asesmen Hasil Belajar Siswa pada Kurikulum Merdeka Belajar. *Education & Learning*, 5(1), 41-46.