

**PENGUKURAN KUALITAS BUTIR SOAL PILIHAN GANDA UNTUK
MENINGKATKAN EVALUASI PEMBELAJARAN PENGANTAR INDUSTRI
MIGAS**

Eka Megawati¹, Meita Rezki Vegatama², Nusrotus Sa'idah³

^{1,2}Teknik Pengolahan Migas, Sekolah Tinggi Teknologi Migas

³Pendidikan Agama Islam, Universitas Islam Nahdlatul Ulama Jepara

1ekamegawati89@yahoo.com

ABSTRACT

Evaluating the quality of test instruments is an important step in ensuring the acceptability and accuracy of measuring student learning outcomes. This study aims to analyze the quality of 50 multiple-choice questions in the Introduction to the Oil and Gas Industry course in terms of validity, reliability, difficulty level, and discriminatory power. The method used is quantitative item analysis with a descriptive and inferential statistical approach based on the results of validity tests, reliability using Cronbach's Alpha, difficulty level, and discriminatory power. The analysis results show that most of the items are valid. The reliability test obtained a Cronbach's Alpha value of 0.88 which is included in the high category, indicating good internal consistency of the instrument. The level of difficulty of the questions is predominantly in the medium category, while the discriminatory power is mostly in the good category. However, several items were still found to be invalid and have low discriminatory power so that they need to be revised or removed. Overall, the test instrument in the Introduction to the Oil and Gas Industry course can be categorized as good quality and suitable for use with improvements to certain items.

Keywords: item validity, reliability, discrimination analysis

ABSTRAK

Evaluasi kualitas instrumen tes merupakan langkah penting dalam memastikan keberterimaan dan ketepatan pengukuran hasil belajar mahasiswa. Penelitian ini bertujuan untuk menganalisis kualitas 50 butir soal pilihan ganda pada mata kuliah Pengantar Industri Migas ditinjau dari validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Metode yang digunakan adalah analisis butir soal kuantitatif dengan pendekatan statistik deskriptif dan inferensial berdasarkan hasil uji validitas, reliabilitas menggunakan Cronbach's Alpha, tingkat kesukaran, serta daya pembeda. Hasil analisis menunjukkan bahwa sebagian besar butir soal dinyatakan valid. Uji reliabilitas memperoleh nilai Cronbach's Alpha sebesar 0,88 yang termasuk dalam kategori tinggi, menunjukkan konsistensi internal instrumen yang baik. Tingkat kesukaran soal didominasi kategori sedang, sedangkan daya pembeda sebagian besar berada pada kategori baik. Meskipun demikian, masih ditemukan beberapa butir soal yang tidak valid dan memiliki daya pembeda rendah

sehingga perlu dilakukan revisi atau penghapusan. Secara keseluruhan, instrumen tes pada mata kuliah Pengantar Industri Migas dapat dikategorikan berkualitas baik dan layak digunakan dengan perbaikan pada beberapa butir soal tertentu.

Kata Kunci: validitas butir soal, reliabilitas, analisis daya pembeda

A. Pendahuluan

Pada proses pembelajaran di perguruan tinggi, evaluasi hasil belajar merupakan komponen penting untuk mengukur tingkat pencapaian kompetensi mahasiswa. Salah satu instrumen evaluasi yang banyak digunakan adalah tes pilihan ganda karena memiliki kelebihan dalam hal objektivitas, efisiensi waktu, serta kemampuan mencakup materi yang luas. Namun demikian, kualitas butir soal yang digunakan dalam evaluasi sering kali belum dianalisis secara sistematis, sehingga validitas dan reliabilitasnya belum terjamin secara optimal.

Menurut penelitian Abdullah, I., dkk (2025) tes pilihan ganda merupakan instrumen evaluasi yang efektif untuk mengukur kompetensi peserta didik, asalkan setiap butir soal telah melalui proses analisis kualitas seperti validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Hal ini menunjukkan bahwa penggunaan tes pilihan ganda tidak hanya

bergantung pada bentuknya, tetapi juga pada kualitas penyusunan soal.

Sejalan dengan itu, penelitian Laeli, S. (2025) menyatakan bahwa tes pilihan ganda banyak digunakan dalam evaluasi pembelajaran karena mampu memberikan hasil yang objektif dan terstandar, serta dapat dianalisis secara kuantitatif untuk mengetahui kualitas instrumen secara menyeluruh. Dengan demikian, tes pilihan ganda menjadi pilihan utama dalam berbagai konteks pendidikan, baik di sekolah maupun perguruan tinggi.

Analisis butir soal menjadi langkah penting untuk memastikan bahwa instrumen evaluasi benar-benar mampu mengukur kemampuan yang seharusnya diukur. Kualitas butir soal dapat dilihat dari beberapa indikator seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian oleh Ambarwati, Y. F., & Ismiyati, I. (2021) menunjukkan bahwa meskipun sebagian besar soal tergolong valid, masih terdapat butir soal yang tidak

memenuhi kriteria sehingga perlu perbaikan. Hal ini menunjukkan bahwa tanpa analisis yang tepat, soal yang digunakan dalam evaluasi berpotensi menghasilkan data yang kurang akurat.

Analisis butir soal merupakan proses sistematis yang dilakukan untuk menilai kualitas setiap item dalam suatu instrumen evaluasi. Proses ini bertujuan untuk memastikan bahwa setiap butir soal benar-benar mampu mengukur kemampuan yang seharusnya diukur sesuai dengan tujuan pembelajaran yang telah ditetapkan. Dengan demikian, analisis butir soal menjadi langkah penting dalam menjamin keakuratan dan keandalan hasil evaluasi.

Pada analisis butir soal, terdapat beberapa aspek utama yang perlu diperhatikan, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Melalui analisis ini, pendidik dapat mengidentifikasi butir soal yang berkualitas baik, kurang baik, maupun yang harus diperbaiki atau dibuang. Tanpa adanya analisis yang tepat, instrumen evaluasi berpotensi menghasilkan data yang tidak akurat

dan tidak mencerminkan kemampuan sebenarnya dari peserta didik.

Menurut Savika, H. I., & Zuhriyah, I. A. (2024), analisis butir soal merupakan bagian penting dalam evaluasi pembelajaran karena dapat memberikan informasi empiris mengenai kualitas instrumen yang digunakan. Hasil analisis tersebut dapat dijadikan dasar dalam pengambilan keputusan untuk memperbaiki dan menyempurnakan soal agar sesuai dengan standar penilaian yang baik.

Selain itu, penelitian oleh Hariati, B. S., dkk (2026) menyatakan bahwa analisis butir soal membantu meningkatkan kualitas instrumen evaluasi dengan cara mengidentifikasi kelemahan pada setiap item soal. Dengan melakukan analisis secara berkala, pendidik dapat memastikan bahwa instrumen yang digunakan memiliki tingkat validitas dan reliabilitas yang memadai, sehingga mampu mengukur kemampuan peserta didik secara objektif dan tepat.

Seiring dengan perkembangan teknologi, analisis butir soal tidak hanya dapat dilakukan secara manual, tetapi juga dengan bantuan perangkat lunak seperti Microsoft Excel. Penggunaan Excel dinilai

efektif karena mudah diakses, fleksibel, serta mampu mengolah data secara cepat dan akurat melalui berbagai fungsi statistik seperti korelasi, rata-rata, dan standar deviasi. Penelitian oleh Augustia, A. D., dkk (2025), menegaskan bahwa pemanfaatan aplikasi pengolah data dapat meningkatkan ketelitian dalam analisis validitas dan reliabilitas soal. Selain itu, Anshari dkk. (2024) menyatakan bahwa penggunaan teknologi dalam evaluasi pembelajaran dapat mempermudah dosen dalam mengidentifikasi kualitas butir soal secara lebih sistematis.

Lebih lanjut, penggunaan Microsoft Excel dalam analisis butir soal memungkinkan penghitungan berbagai parameter seperti indeks kesukaran, daya pembeda, serta uji validitas dengan korelasi product moment secara lebih efisien. Hal ini didukung oleh penelitian Laeli, S. (2025) yang menyatakan bahwa analisis berbasis aplikasi memberikan hasil yang lebih akurat dibandingkan perhitungan manual. Selain itu, Rachmadani, F. C., dkk (2024) menegaskan bahwa analisis butir soal merupakan proses penting untuk menghasilkan instrumen evaluasi yang valid dan reliabel.

Analisis butir soal merupakan kegiatan penting dalam evaluasi pembelajaran untuk menilai kualitas instrumen tes agar dapat mengukur kemampuan peserta didik secara tepat. Seiring perkembangan teknologi, proses analisis butir soal tidak hanya dilakukan secara manual, tetapi juga dapat dibantu dengan aplikasi pengolah data seperti Microsoft Excel. Penggunaan aplikasi ini memberikan kemudahan dalam mengolah data hasil tes secara lebih cepat, sistematis, dan akurat.

Microsoft Excel banyak digunakan dalam analisis butir soal karena memiliki fitur-fitur statistik yang dapat membantu menghitung indikator kualitas soal, seperti tingkat kesukaran, daya pembeda, validitas (korelasi), serta reliabilitas sederhana. Fungsi-fungsi seperti AVERAGE, COUNTIF, IF, dan CORREL memungkinkan peneliti untuk melakukan perhitungan tanpa harus menggunakan perangkat lunak statistik yang lebih kompleks. Hal ini menjadikan Excel sebagai alat yang praktis dan mudah diakses oleh pendidik maupun peneliti pendidikan.

Menurut Ihsan, L. N., dkk (2025), penggunaan aplikasi spreadsheet seperti Excel dalam evaluasi

pembelajaran dapat meningkatkan efisiensi pengolahan data dan meminimalkan kesalahan perhitungan manual. Selain itu, penelitian oleh Junaidi, J. K., dkk. (2024) menunjukkan bahwa analisis butir soal berbantuan Excel mampu memberikan hasil yang konsisten dalam menentukan kualitas soal, terutama pada aspek tingkat kesukaran dan daya pembeda.

Pada mata kuliah Pengantar Industri Migas, yang memiliki karakteristik materi kompleks dan mencakup aspek teknis, ekonomi, serta regulasi, diperlukan instrumen evaluasi yang berkualitas tinggi. Soal yang tidak memenuhi kriteria kualitas dapat menyebabkan ketidaktepatan dalam mengukur pemahaman mahasiswa terhadap konsep-konsep dasar industri migas. Oleh karena itu, analisis kualitas butir soal dengan bantuan Microsoft Excel menjadi solusi yang praktis dan efektif untuk meningkatkan kualitas evaluasi pembelajaran.

Berdasarkan uraian tersebut, maka penelitian mengenai analisa kualitas butir soal pilihan ganda pada mata kuliah Pengantar Industri Migas menggunakan Microsoft Excel perlu dilakukan. Penelitian ini diharapkan

dapat memberikan gambaran mengenai kualitas soal yang digunakan serta menjadi dasar dalam perbaikan instrumen evaluasi pembelajaran di masa yang akan datang.

B. Metode Penelitian

Validitas

Validitas merupakan salah satu aspek penting dalam analisis kualitas instrumen evaluasi yang menunjukkan sejauh mana suatu instrumen mampu mengukur apa yang seharusnya diukur. Dalam konteks tes hasil belajar, validitas menggambarkan ketepatan butir soal dalam merepresentasikan kompetensi atau kemampuan yang dimiliki oleh peserta didik.

Uji validitas butir soal umumnya dilakukan menggunakan teknik korelasi, salah satunya adalah korelasi Product Moment Pearson. Teknik ini digunakan untuk mengetahui hubungan antara skor setiap butir soal dengan skor total tes. Suatu butir soal dinyatakan valid apabila nilai korelasi yang diperoleh lebih besar daripada nilai r tabel pada taraf signifikansi tertentu. Sebaliknya, jika nilai korelasi lebih kecil dari r tabel,

maka butir soal tersebut dinyatakan tidak valid dan perlu direvisi atau dibuang.

Validitas sangat penting dalam penyusunan instrumen evaluasi karena berkaitan langsung dengan ketepatan hasil pengukuran. Instrumen yang valid akan mampu memberikan gambaran yang akurat mengenai kemampuan peserta didik, sehingga hasil tes dapat digunakan sebagai dasar pengambilan keputusan dalam proses pembelajaran. Tanpa validitas yang baik, hasil evaluasi berisiko tidak mencerminkan kemampuan sebenarnya dari peserta didik.

Menurut Muhadi, M., dkk (2025), validitas merupakan syarat utama dalam penyusunan instrumen evaluasi yang baik karena menentukan tingkat keakuratan hasil pengukuran. Sementara itu, penelitian oleh Ketaren, M. A., dkk. (2024) menunjukkan bahwa penggunaan korelasi Product Moment dalam uji validitas butir soal efektif dalam mengidentifikasi kualitas item tes, sehingga dapat digunakan untuk memperbaiki instrumen evaluasi agar lebih akurat dan reliabel.

Realibilitas

Reliabilitas merupakan salah satu karakteristik penting dalam suatu instrumen evaluasi yang menunjukkan tingkat konsistensi hasil pengukuran. Instrumen dikatakan reliabel apabila memberikan hasil yang relatif sama ketika digunakan pada kondisi yang sama dan pada waktu yang berbeda. Dengan kata lain, reliabilitas berkaitan dengan keajegan atau kestabilan suatu alat ukur dalam mengukur suatu variabel.

Pada konteks evaluasi pembelajaran, reliabilitas sangat penting karena menentukan sejauh mana hasil tes dapat dipercaya. Instrumen yang memiliki reliabilitas tinggi akan menghasilkan data yang konsisten sehingga dapat digunakan sebagai dasar pengambilan keputusan yang tepat dalam proses pembelajaran. Sebaliknya, instrumen yang memiliki reliabilitas rendah cenderung menghasilkan data yang tidak stabil dan kurang dapat diandalkan.

Pengukuran reliabilitas pada umumnya dilakukan dengan beberapa metode, salah satunya adalah Cronbach's Alpha. Metode ini

digunakan untuk mengukur konsistensi internal antarbutir soal dalam suatu instrumen tes. Nilai Cronbach's Alpha berkisar antara 0 hingga 1, di mana semakin mendekati 1 menunjukkan tingkat reliabilitas yang semakin tinggi. Instrumen dianggap reliabel apabila nilai Cronbach's Alpha $\geq 0,70$, meskipun dalam beberapa penelitian pendidikan nilai minimal dapat bervariasi tergantung konteks pengukuran.

Menurut Saputra, A. (2025), reliabilitas merupakan indikator penting dalam menentukan kualitas instrumen evaluasi karena berkaitan langsung dengan konsistensi hasil pengukuran. Sementara itu, penelitian oleh Lestari, T. W. M., & Maulida, S. (2025) menyatakan bahwa penggunaan Cronbach's Alpha merupakan metode yang efektif dalam menguji reliabilitas instrumen tes pilihan ganda karena mampu memberikan gambaran konsistensi internal antarbutir soal secara akurat.

Tingkat Kesukaran

Tingkat kesukaran merupakan salah satu indikator penting dalam analisis butir soal yang menunjukkan seberapa sulit atau mudah suatu soal dikerjakan oleh peserta tes. Tingkat kesukaran digunakan untuk mengidentifikasi kualitas butir soal dalam membedakan kemampuan peserta didik, sehingga dapat diketahui apakah suatu soal tergolong mudah, sedang, atau sukar.

Secara umum, tingkat kesukaran dihitung berdasarkan proporsi peserta didik yang menjawab benar suatu butir soal. Semakin banyak peserta yang menjawab benar, maka soal tersebut cenderung dianggap mudah. Sebaliknya, semakin sedikit peserta yang menjawab benar, maka soal tersebut dianggap sulit. Dalam praktik analisis, indeks tingkat kesukaran biasanya dinyatakan dalam rentang 0 sampai 1, di mana nilai mendekati 1 menunjukkan soal mudah, sedangkan nilai mendekati 0 menunjukkan soal sulit.

Kriteria interpretasi tingkat kesukaran umumnya dibagi menjadi tiga kategori, yaitu soal mudah, sedang, dan sukar. Soal yang baik biasanya berada pada kategori sedang karena mampu memberikan

informasi yang lebih akurat dalam membedakan kemampuan peserta didik. Soal yang terlalu mudah atau terlalu sulit kurang efektif dalam mengukur variasi kemampuan peserta didik.

Menurut Valen, A., & Satria, T. G. (2021), analisis tingkat kesukaran sangat penting dalam evaluasi pembelajaran karena dapat membantu pendidik dalam menyeimbangkan kualitas instrumen tes agar sesuai dengan tujuan pengukuran. Sementara itu, penelitian oleh Junaidi, J. K., dkk. (2024) menunjukkan bahwa distribusi tingkat kesukaran yang baik dalam suatu tes akan meningkatkan kualitas pengukuran hasil belajar karena mampu memberikan variasi respons yang lebih representatif dari peserta didik.

Daya Pembeda

Daya pembeda merupakan salah satu indikator penting dalam analisis butir soal yang menunjukkan kemampuan suatu soal dalam membedakan antara peserta didik yang memiliki kemampuan tinggi dan peserta didik yang memiliki kemampuan rendah. Semakin baik daya pembeda suatu butir soal, maka semakin efektif soal tersebut dalam

mengelompokkan peserta berdasarkan tingkat penguasaannya terhadap materi.

Secara umum, daya pembeda dihitung dengan membandingkan proporsi jawaban benar antara kelompok atas (high group) dan kelompok bawah (low group). Selisih proporsi tersebut kemudian digunakan untuk menentukan indeks daya pembeda. Nilai daya pembeda berkisar antara -1 hingga 1, di mana nilai positif yang tinggi menunjukkan bahwa soal tersebut mampu membedakan dengan baik antara peserta didik berkemampuan tinggi dan rendah. Sebaliknya, nilai rendah atau negatif menunjukkan bahwa butir soal kurang baik dan perlu direvisi atau dibuang karena tidak berfungsi sebagaimana mestinya.

Interpretasi daya pembeda biasanya dikategorikan menjadi sangat baik, baik, cukup, jelek, dan sangat jelek. Butir soal yang baik umumnya memiliki daya pembeda minimal pada kategori cukup hingga sangat baik. Soal dengan daya pembeda yang rendah tidak efektif digunakan dalam evaluasi karena tidak mampu memberikan informasi yang jelas mengenai perbedaan kemampuan peserta didik.

Menurut Muslimah, M., & Widiyanti, A. (2023), daya pembeda merupakan komponen penting dalam analisis butir soal karena menentukan kualitas soal dalam mengidentifikasi perbedaan kemampuan peserta didik secara akurat. Sementara itu, penelitian oleh Nurrisa, F., dkk (2025) menyatakan bahwa analisis daya pembeda membantu pendidik dalam memperbaiki kualitas instrumen evaluasi sehingga soal yang digunakan benar-benar mampu mengukur kemampuan peserta didik secara objektif.

C. Hasil Penelitian dan Pembahasan

Hasil Uji Validitas Butir Soal

Uji validitas dilakukan untuk mengetahui ketepatan setiap butir soal dalam mengukur kemampuan yang seharusnya diukur. Analisis validitas dilakukan menggunakan korelasi Product Moment dengan bantuan Microsoft Excel.

Berdasarkan hasil analisis terhadap 50 butir soal, diperoleh bahwa 38 butir soal dinyatakan valid karena nilai r hitung $>$ r tabel. 12 butir soal dinyatakan tidak valid karena nilai r hitung $<$ r tabel. Dengan demikian, sebagian besar butir soal telah mampu mengukur kompetensi

mahasiswa pada mata kuliah Pengantar Industri Migas, namun masih terdapat beberapa soal yang perlu direvisi atau dibuang.

Hasil Uji Reliabilitas

Uji reliabilitas dilakukan untuk mengetahui konsistensi internal instrumen tes secara keseluruhan. Perhitungan dilakukan menggunakan Cronbach's Alpha dengan bantuan Microsoft Excel.

Hasil analisis menunjukkan nilai Cronbach's Alpha sebesar 0,88, yang berarti lebih besar dari 0,70. Dengan demikian, instrumen tes termasuk dalam kategori reliabilitas tinggi. Hal ini menunjukkan bahwa 50 butir soal memiliki konsistensi internal yang baik dalam mengukur kemampuan mahasiswa secara stabil.

Hasil Uji Tingkat Kesukaran

Uji tingkat kesukaran dilakukan untuk mengetahui kategori kesulitan setiap butir soal berdasarkan proporsi jawaban benar.

Hasil analisis menunjukkan bahwa 10 soal termasuk kategori mudah, 32 soal termasuk kategori sedang, 8 soal termasuk kategori sukar. Secara umum, distribusi tingkat kesukaran didominasi oleh soal kategori sedang. Hal ini menunjukkan bahwa instrumen tes sudah cukup

baik karena sebagian besar soal berada pada tingkat kesulitan yang ideal untuk membedakan kemampuan mahasiswa.

Hasil Uji Daya Pembeda

Uji daya pembeda dilakukan untuk mengetahui kemampuan soal dalam membedakan mahasiswa berkemampuan tinggi dan rendah.

Hasil analisis menunjukkan 6 soal memiliki daya pembeda sangat baik, 22 soal memiliki daya pembeda baik, 14 soal memiliki daya pembeda cukup, 8 soal memiliki daya pembeda rendah/jelek. Hal ini menunjukkan bahwa sebagian besar soal sudah mampu membedakan kemampuan mahasiswa dengan cukup baik, namun masih terdapat beberapa butir soal yang perlu direvisi karena kurang mampu membedakan kemampuan peserta didik.

Pembahasan

Hasil analisis butir soal menunjukkan bahwa secara umum instrumen evaluasi berupa 50 soal pilihan ganda pada mata kuliah Pengantar Industri Migas memiliki kualitas yang cukup baik, meskipun masih terdapat beberapa aspek yang perlu diperbaiki.

Dari segi validitas, sebagian besar soal (38 dari 50 butir) telah

memenuhi kriteria valid. Hal ini menunjukkan bahwa instrumen sudah cukup tepat dalam mengukur kemampuan mahasiswa. Namun, adanya 12 soal yang tidak valid menunjukkan bahwa masih terdapat beberapa butir soal yang tidak sesuai dengan indikator pembelajaran sehingga perlu direvisi atau dihapus.

Hasil uji reliabilitas menunjukkan nilai Cronbach's Alpha sebesar 0,88 yang termasuk kategori tinggi. Hal ini mengindikasikan bahwa instrumen memiliki konsistensi internal yang baik. Dengan kata lain, jika tes diberikan kembali pada kondisi yang sama, hasilnya cenderung stabil.

Pada uji tingkat kesukaran, dominasi soal kategori sedang menunjukkan bahwa instrumen sudah cukup ideal. Soal dengan tingkat kesukaran sedang sangat penting karena mampu memberikan informasi yang lebih akurat dalam membedakan kemampuan mahasiswa dibandingkan soal yang terlalu mudah atau terlalu sulit.

Sementara itu, hasil daya pembeda menunjukkan bahwa sebagian besar soal berada pada kategori baik hingga cukup. Namun, masih terdapat beberapa soal dengan

daya pembeda rendah yang mengindikasikan bahwa soal tersebut kurang efektif dalam membedakan mahasiswa berkemampuan tinggi dan rendah. Soal-soal tersebut perlu diperbaiki dari segi redaksi, opsi jawaban, atau tingkat kognitif yang diukur.

Secara keseluruhan, hasil analisis menunjukkan bahwa instrumen tes sudah layak digunakan, namun tetap diperlukan revisi pada beberapa butir soal, khususnya yang tidak valid dan memiliki daya pembeda rendah, untuk meningkatkan kualitas instrumen secara keseluruhan.

D. Kesimpulan

Berdasarkan hasil analisis 50 butir soal pilihan ganda pada mata kuliah Pengantar Industri Migas, dapat disimpulkan bahwa instrumen tes secara umum berkualitas baik. Hasil uji validitas menunjukkan sebagian besar soal valid, uji reliabilitas menunjukkan nilai Cronbach's Alpha sebesar 0,88 (kategori tinggi), serta tingkat kesukaran didominasi kategori sedang dan daya pembeda mayoritas baik. Meskipun demikian, masih terdapat beberapa butir soal yang tidak valid dan memiliki daya

pembeda rendah sehingga perlu direvisi atau dihapus.

DAFTAR PUSTAKA

- Abdullah, I., Listyotami, M. K., Ferita, R. A., Lisa, L., Maulida, D., Yanti, W., & Hidayat, F. (2025). Analisis Kualitas Butir Soal Asesmen Sumatif Matematika Pilihan Ganda Semester Genap Kelas VIII MTs Baitul Ma'mur Tahun Ajaran 2023/2024: Analysis of the Quality of Multiple-Choice Mathematics Summative Assessment Items for the Even Semester of Grade VIII at MTs Baitul Ma'mur in the 2023/2024 Academic Year. *Nusantara Journal of Education and Social Science*, 2(1), 27-32.
- Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis butir soal pilihan ganda ulangan akhir semester genap mata pelajaran kearsipan. *Measurement In Educational Research*, 1(2), 64-75.
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964-975.
- Augustia, A. D., Agustia, C. N., Azzahra, D., Hakim, L., & Pratiwi, V. (2025). Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software Anates pada mata pelajaran perpajakan. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250-265.

- Hariati, B. S., Hakim, L., Pratiwi, V., Zakia, A., Azizah, R. C. A., Maheswari, A. D., ... & Rahmawati, T. D. (2026). Analisis Kualitas Butir Soal Pilihan Ganda Elemen Dokumen Berbasis Digital Menggunakan ANATES. *Jurnal Jendela Pendidikan*, 6(01), 46-58.
- Ihsan, L. N., Melati, P., & Jumatiah, S. (2025). Perkembangan aplikasi spreadsheet dan pengaruhnya terhadap inovasi pembelajaran berbasis teknologi di dunia pendidikan. *Jurnal Madako Education*, 11(1), 29-34.
- Junaidi, J. K., Husnita, L., Meldawati, M., & Kaksim, K. (2024). Peningkatan Aktifitas Belajar Mahasiswa dengan Memanfaatkan Microsoft Excel dalam Analisis Kualitas Butir Soal. *Journal of Education Research*, 5(1), 875-881.
- Ketaren, M. A., Girsang, K., Manurung, M., & Ginting, E. R. B. (2024). Uji validitas dan uji daya beda soal buatan pilihan ganda dengan tes sumatif siswa kelas IV UPT SD Negeri 065013 Medan. *Jurnal Pengabdian Masyarakat Bangsa*, 1(11), 3278-3283.
- Laeli, S. (2025). Analisis Butir Soal Pilihan Ganda dalam Asesmen Sumatif Akhir Semester (ASAS) Gasal Mata Pelajaran Bahasa Arab Kelas VIII MTs Muhammadiyah Srumbung Kabupaten Magelang Tahun Ajaran 2024/2025 (Doctoral dissertation, UIN Sunan Kalijaga Yogyakarta).
- Lestari, T. W. M., & Maulida, S. (2025). Analisis Reabilitas Instrumen Asesmen Pada Keterampilan Berpikir Kritis dalam Pembelajaran PAI di SDN Koroulon 1. *Al-Muaddib: Jurnal Kajian Ilmu Kependidikan*, 7(3), 235-244.
- Muhadi, M., Jarir, J., Khairina, K., Rajuna, R., & Prasetyo, E. (2025). Evaluasi perencanaan desain pembelajaran, pelaksanaan proses kegiatan pembelajaran, dan evaluasi instrumen hasil pembelajaran. *Edu Society: Jurnal Pendidikan, Ilmu Sosial dan Pengabdian Kepada Masyarakat*, 5(2), 156-165.
- Muslimah, M., & Widiyanti, A. (2023). Analisis Daya Beda Tes Hasil Belajar Bahasa Arab Siswa SMA Mamba'ul Hikmah Paron Ngawi. *Al-Muarrib Journal of Arabic Education*, 3(2), 67-77.
- Nurrisa, F., Sabirin, M., & Huda, N. (2025). Evaluasi Validitas, Reliabilitas, dan Daya Pembeda Instrumen Ujian PAI Kelas XII Melalui Analisis Butir Soal: Evaluation of the Validity, Reliability, and Discriminating Power of Islamic Education Exam Items for Grade XII through Item Analysis. *TA'LIMUNA: Jurnal Pendidikan Islam*, 14(2), 272-285.
- Rachmadani, F. C., Lestari, C. D., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal Pilihan Ganda pada Elemen Akuntansi Keuangan Guna Mengoptimasi Evaluasi Menggunakan Anates V4. 0. *Jurnal Nirta: Studi Inovasi*, 4(1), 71-88.
- Saputra, A. (2025). Pengembangan Instrumen Evaluasi. *Ar-raudah: Jurnal Pendidikan Dan Keagamaan*, 2(4), 1-14.

- Savika, H. I., & Zuhriyah, I. A. (2024). Peran analisis butir soal terhadap kualitas soal, kompetensi guru, dan prestasi belajar peserta didik di sekolah dasar. *PANDU: Jurnal Pendidikan Anak dan Pendidikan Umum*, 2(2), 43-51.
- Valen, A., & Satria, T. G. (2021). Analisis tingkat kesulitan soal PAS (Penilaian Akhir Semester) mata pelajaran IPS di sekolah dasar. *Jurnal Basicedu*, 5(4), 2199-2208.