



ANALISIS BUTIR TES: DAYA PEMBEDA DAN EFEKTIVITAS PENGECOH PADA SOAL PILIHAN GANDA

Siti Zainab¹, Dina Hermina²

Program Studi Pendidikan Agama Islam, UIN Antasari Banjarmasin^{1,2}

e-mail: sitizainab4598@gmail.com

Diterima: 23/6/2026; Direvisi: 7/8/2026; Diterbitkan: 14/8/2026

ABSTRAK

Analisis butir tes merupakan prosedur ilmiah berbasis data untuk memastikan kualitas instrumen evaluasi, namun dalam praktik di lapangan masih banyak soal pilihan ganda yang disusun tanpa pengujian empiris sehingga kualitasnya tidak terjamin. Artikel ini bertujuan menjelaskan konsep daya pembeda dan efektivitas pengecoh pada soal pilihan ganda serta cara menganalisis keduanya sebagai dasar pengambilan keputusan terhadap kelayakan butir. Kajian disusun dengan metode studi kepustakaan (*library research*) melalui penelusuran, telaah, dan sintesis literatur dari jurnal ilmiah terkini, kemudian dilengkapi ilustrasi perhitungan untuk memperjelas penerapannya. Hasil telaah menunjukkan bahwa daya pembeda berfungsi sebagai indikator sensitivitas butir dalam membedakan peserta didik berkemampuan tinggi dan rendah, sedangkan efektivitas pengecoh menunjukkan sejauh mana alternatif jawaban salah mampu menarik peserta didik yang belum menguasai materi. Berbagai penelitian memperlihatkan bahwa sebagian besar butir soal di lapangan memiliki daya pembeda rendah dan pengecoh yang tidak berfungsi, serta keduanya saling berkaitan: pengecoh yang lemah cenderung disertai daya pembeda yang rendah, demikian pula sebaliknya. Disimpulkan bahwa analisis daya pembeda dan efektivitas pengecoh perlu dijadikan bagian rutin dari siklus evaluasi pembelajaran agar instrumen tes menjadi lebih valid, reliabel, dan adil bagi peserta didik.

Kata Kunci: *Analisis Butir Tes, Daya Pembeda, Efektivitas Pengecoh, Soal Pilihan Ganda, Evaluasi Pembelajaran.*

ABSTRACT

Item analysis is a data-based scientific procedure used to ensure the quality of assessment instruments; however, in practice many multiple-choice items are still developed without empirical testing, so their quality cannot be guaranteed. This article aims to explain the concepts of discriminating power and distractor effectiveness in multiple-choice items and how to analyze both as a basis for deciding on item quality. The study was conducted using a library research method through searching, reviewing, and synthesizing recent scholarly literature, complemented by computational illustrations to clarify its application. The review shows that discriminating power serves as an indicator of an item's sensitivity in distinguishing high- and low-ability students, while distractor effectiveness indicates the extent to which incorrect options attract students who have not mastered the material. Various studies show that most items in the field have low discriminating power and non-functioning distractors, and that the two are interrelated: weak distractors tend to be accompanied by low discriminating power, and vice versa. It is concluded that the analysis of discriminating power and distractor effectiveness should be a routine part of the learning evaluation cycle so that test instruments become more valid, reliable, and fair for students.

Keywords: *Item Analysis, Discriminating Power, Distractor Effectiveness, Multiple-Choice Items, Learning Evaluation*

PENDAHULUAN

Evaluasi pembelajaran merupakan komponen penting dalam sistem pendidikan karena berfungsi untuk mengukur ketercapaian kompetensi peserta didik sekaligus menjadi dasar pengambilan keputusan pedagogis. Salah satu instrumen yang paling banyak digunakan dalam penilaian hasil belajar adalah tes objektif bentuk pilihan ganda. Bentuk tes ini dipilih karena efisien, memiliki cakupan materi yang luas, serta memungkinkan penskoran yang objektif. Namun demikian, kualitas tes pilihan ganda sangat bergantung pada mutu setiap butir soal yang disusun oleh guru. Oleh karena itu, pengujian kualitatif dan kuantitatif secara sistematis mutlak diperlukan untuk menjamin keabsahan penilaian.

Dalam praktik di lapangan, masih banyak ditemukan butir soal yang kurang mampu membedakan peserta didik berkemampuan tinggi dan rendah. Selain itu, pengecoh (*distractor*) pada soal pilihan ganda sering kali tidak berfungsi secara optimal, misalnya tidak dipilih oleh peserta didik atau justru lebih banyak dipilih oleh kelompok berkemampuan tinggi (Kaka et al., 2024; Sene et al., 2022). Fenomena tidak berfungsinya opsi distraktor ini membuktikan bahwa penyusunan instrumen tes masih sering dilakukan tanpa pertimbangan konseptual yang matang. Kondisi ini menunjukkan bahwa penyusunan soal belum sepenuhnya berbasis analisis empiris, sehingga kualitas instrumen penilaian menjadi kurang akurat dan kurang adil bagi peserta didik.

Analisis butir tes merupakan pendekatan ilmiah yang digunakan untuk mengevaluasi kualitas soal berdasarkan data hasil pengerjaan peserta didik. Dalam kerangka *Classical Test Theory*, analisis butir mencakup pengujian validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh sebagai indikator kelayakan instrumen (Hermanto et al., 2023). Dua indikator penting di antaranya adalah daya pembeda dan efektivitas pengecoh. Daya pembeda menunjukkan kemampuan suatu butir soal dalam membedakan peserta didik berdasarkan tingkat penguasaan materi, sedangkan efektivitas pengecoh menunjukkan sejauh mana alternatif jawaban yang salah mampu berfungsi sebagaimana mestinya.

Kedua indikator tersebut saling berkaitan erat dan menjadi tolok ukur utama dalam menentukan kelayakan suatu butir soal. Butir yang memiliki daya pembeda baik biasanya didukung oleh opsi pengecoh yang bekerja secara ideal dalam mengecoh peserta didik yang belum menguasai materi. Sebaliknya, ketika opsi pengecoh tidak berfungsi, kapasitas diskriminasi butir soal tersebut cenderung mengalami penurunan yang signifikan. Oleh sebab itu, pemeriksaan terhadap kedua parameter ini secara bersamaan menjadi kunci utama dalam merevisi atau merekonstruksi instrumen tes pilihan ganda.

METODE PENELITIAN

Penelitian ini menggunakan metode studi kepustakaan (*library research*) dengan pendekatan kualitatif-deskriptif untuk mengkaji secara mendalam konsep daya pembeda dan efektivitas pengecoh pada tes pilihan ganda. Pengumpulan data literatur dilakukan secara elektronik melalui penelusuran sistematis pada empat *database* jurnal ilmiah bereputasi, yaitu Google Scholar, Garuda (Garba Rujukan Digital), Crossref, dan Dimensions. Strategi pencarian literatur memanfaatkan kombinasi kata kunci (*keywords*) spesifik, meliputi: “analisis butir tes”, “daya pembeda soal”, “efektivitas pengecoh”, “*distractor effectiveness*”, “*item discrimination index*”, dan “evaluasi pilihan ganda”. Kriteria inklusi difokuskan pada artikel jurnal ilmiah

terindeks publikasi sepuluh tahun terakhir (2015–2025) yang membahas evaluasi psikometrik tes pilihan ganda dalam bidang pendidikan.

Pada tahap awal penelusuran *database*, ditemukan sebanyak 45 artikel relevan. Peneliti kemudian melakukan penyaringan (*screening*) berdasarkan kriteria inklusi dan eksklusi, sehingga sebanyak 23 artikel dieksklusi karena tidak menyajikan data empiris yang relevan atau dipublikasikan di luar rentang waktu penelusuran. Akhirnya, sebanyak 22 artikel jurnal ilmiah terpilih untuk dianalisis secara komprehensif. Pengolahan data kualitatif menerapkan teknik analisis isi (*content analysis*) melalui tiga tahapan sistematis: reduksi data, kategorisasi temuan berdasarkan subtema, serta sintesis konseptual untuk penarikan kesimpulan. Sebagai pelengkap analitis, kajian ini dilengkapi dengan simulasi ilustrasi perhitungan matematis indeks daya pembeda dan pola distribusi jawaban peserta didik guna memperjelas pengaplikasian konseptualnya di lapangan.

HASIL DAN PEMBAHASAN

Hasil

Konsep Analisis Butir, Daya Pembeda, dan Efektivitas Pengecoh

Analisis butir tes merupakan prosedur evaluatif untuk menilai kualitas setiap butir soal berdasarkan respons empiris peserta didik. Prosedur ini membantu guru memastikan bahwa instrumen yang digunakan benar-benar mampu mengukur kemampuan yang ditargetkan secara akurat. Mustaqim & Sulisti (2024) menegaskan bahwa analisis butir adalah proses yang harus dilakukan pendidik untuk mengetahui mutu soal yang digunakan dalam penilaian, sementara Zakiyah & Jamillah (2021) menunjukkan bahwa karakteristik tes seperti kesukaran, daya pembeda, serta fungsi pengecoh hanya dapat diungkap secara tepat melalui pendekatan deskriptif kuantitatif berbasis data respons siswa.

Perkembangan teknologi turut memudahkan pelaksanaan analisis butir. Augustia et al. (2025) serta Ikawati et al. (2024) menunjukkan bahwa penggunaan perangkat lunak seperti Anates memungkinkan guru menganalisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh secara cepat dan sistematis, sehingga keputusan terhadap kelayakan butir dapat diambil berdasarkan data yang lebih akurat.

Daya pembeda merupakan indeks yang menunjukkan kemampuan suatu butir soal dalam membedakan peserta didik berkemampuan tinggi dan rendah. Butir yang baik seharusnya lebih banyak dijawab benar oleh kelompok atas dibanding kelompok bawah. Yuliani (2024) menegaskan bahwa semakin tinggi koefisien daya pembeda, semakin baik kemampuan butir dalam memisahkan peserta didik yang telah menguasai kompetensi dari yang belum. Senada dengan itu, Rachmawati dan Pradana (2025) menemukan bahwa butir dengan daya pembeda yang baik layak digunakan karena mampu merefleksikan variasi kemampuan siswa secara lebih akurat, dan Ratni (2023) menunjukkan bahwa variasi kategori daya pembeda memberikan informasi diagnostik penting bagi guru untuk menentukan butir mana yang perlu direvisi.

Efektivitas pengecoh merujuk pada kemampuan alternatif jawaban yang salah dalam menarik peserta didik yang belum menguasai materi; pengecoh yang baik harus tampak logis sehingga tidak mudah dieliminasi. Qadir et al. (2024) menegaskan bahwa pengecoh dikatakan efektif apabila dipilih oleh minimal sekitar 5% peserta tes, sedangkan opsi yang tidak dipilih menunjukkan distraktor tidak berfungsi dan perlu direvisi. Fitriani (2021) menunjukkan bahwa sebagian besar pengecoh yang kurang baik mengindikasikan perlunya perbaikan butir sebelum digunakan kembali, dan Rahayu & Djazari (2016) menemukan variasi kualitas pengecoh mulai dari sangat baik hingga sangat tidak baik dalam satu paket soal.

Temuan Permasalahan di Lapangan

Berbagai penelitian menunjukkan bahwa banyak soal memiliki daya pembeda rendah bahkan negatif. Fitriani (2021) menemukan bahwa 88,89% butir soal memiliki daya pembeda kategori jelek, dan Pradita et al. (2023) menemukan masih terdapat sejumlah butir yang harus direvisi berdasarkan uji daya pembeda. Sebaliknya, Trianingsih (2023) menunjukkan bahwa ketika analisis butir dilakukan dengan baik, seluruh soal dapat mencapai kategori daya pembeda yang dapat diterima hingga sangat baik. Hal ini memperlihatkan bahwa kualitas daya pembeda sangat bergantung pada proses pengembangan dan evaluasi soal.

Temuan serupa dilaporkan Nafs et al. (2023) dan Ikawati et al. (2024) yang menemukan bahwa sebagian besar butir soal pada ujian di sekolah belum berfungsi optimal sehingga lebih banyak yang harus direvisi daripada yang langsung layak digunakan. Bahkan Nurhalimah et al. (2022) menunjukkan adanya keterkaitan antara validitas item, daya pembeda, dan tingkat kesukaran, yang menegaskan bahwa kelemahan pada satu aspek cenderung memengaruhi aspek lainnya.

Pada aspek pengecoh, masalah yang paling sering ditemukan adalah pengecoh yang tidak berfungsi (*non-functional distractor*). Meriyotin et al. (2015) menunjukkan bahwa dari 200 pengecoh yang dianalisis, hanya 46,5% yang efektif sementara 50,5% perlu diperbaiki. Sene et al. (2022) menemukan distribusi kualitas pengecoh yang beragam, sebagian berada pada kategori buruk hingga sangat buruk, dan Taufik & Suryadi (2023) menegaskan bahwa banyak guru belum pernah melakukan analisis distraktor terhadap soal yang mereka susun sehingga mutu soal sebenarnya tidak diketahui.

Ilustrasi Perhitungan Butir

Sebagai ilustrasi hipotetis (bukan data penelitian lapangan), sebuah tes pilihan ganda dimisalkan dikerjakan oleh 40 siswa yang dibagi menjadi kelompok atas (20 siswa) dan kelompok bawah (20 siswa). Hasil analisis salah satu butir disajikan pada Tabel 1, sedangkan distribusi pemilihan opsinya disajikan pada Tabel 2. Butir tersebut dirujuk sebagai Butir A, sementara contoh butir bermasalah dirujuk sebagai Butir B.

Tabel 1. Hasil Analisis Daya Pembeda Dua Butir Soal (Ilustrasi Hipotetis)

Komponen	Butir A	Butir B
Kelompok atas benar	16/20	11/20
Kelompok bawah benar	8/20	9/20
Indeks daya pembeda	0,40 (baik)	0,10 (jelek)

Tabel 2. Distribusi Pemilihan Opsi Jawaban Butir A (Ilustrasi Hipotetis)

Opsi	Status	Proporsi	Keterangan
A	Kunci	60%	Berfungsi
B	Pengecoh	25%	Berfungsi
C	Pengecoh	10%	Berfungsi
D	Pengecoh	5%	Berfungsi (batas)

Berdasarkan Tabel 1 dan Tabel 2, Butir A memiliki daya pembeda yang baik dan seluruh pengecohnya berfungsi sehingga layak dipertahankan. Sebaliknya, Butir B memiliki daya pembeda yang lemah dan salah satu opsinya (opsi D) tidak dipilih oleh seorang pun peserta,

sehingga butir tersebut perlu direvisi karena distraktornya tidak berfungsi dan kemampuan diskriminasinya rendah.

Pembahasan

Hasil telaah menegaskan bahwa daya pembeda berfungsi sebagai indikator utama kualitas butir karena menunjukkan sensitivitas soal terhadap perbedaan kemampuan peserta didik. Temuan rendahnya daya pembeda pada banyak penelitian (Fitriani, 2021; Pradita et al., 2023) dapat dimaknai secara psikometrik: butir dengan daya pembeda rendah tidak sensitif terhadap variasi kemampuan siswa sehingga validitas internal tes melemah. Rendahnya daya pembeda umumnya disebabkan oleh stem soal yang kurang jelas, kunci jawaban yang tidak tegas, tingkat kesukaran yang ekstrem, serta pengecoh yang tidak berfungsi. Oleh karena itu, butir semacam ini tidak seharusnya digunakan kembali tanpa revisi.

Pada aspek pengecoh, distraktor yang tidak efektif biasanya terlalu jelas salah, tidak homogen dengan kunci, panjang opsinya tidak seimbang, atau tidak berbasis miskonsepsi nyata siswa. Dari sudut pandang teori pengukuran, distraktor yang tidak dipilih akan meningkatkan peluang menebak (*guessing*) sehingga skor tes kurang mencerminkan kemampuan sebenarnya. Bahkan, Ahmad & Nipaah (2017) menunjukkan bahwa efektivitas distraktor berhubungan dengan prestasi belajar siswa, sehingga kualitas opsi jawaban tidak boleh diabaikan.

Hal yang sering luput disadari adalah bahwa daya pembeda dan efektivitas pengecoh saling berkaitan. Kasanova & Sulistiyono (2023) menegaskan bahwa evaluasi butir pilihan ganda harus melihat secara simultan validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh untuk memperoleh gambaran mutu soal yang utuh. Secara teoretis, distraktor yang tidak berfungsi mengurangi jumlah opsi yang benar-benar bersaing sehingga peluang menebak meningkat, dan hal ini melemahkan kemampuan butir dalam membedakan kelompok atas dan bawah. Secara empiris, butir dengan pengecoh yang tidak berfungsi cenderung memiliki daya pembeda yang rendah, sebaliknya, butir berdaya pembeda rendah kerap disertai distraktor yang lemah, dan apabila keduanya buruk maka butir tersebut sangat bermasalah. Dengan demikian, analisis yang parsial berpotensi menghasilkan keputusan yang kurang tepat.

Implikasinya bagi praktik evaluasi cukup nyata: masih banyak instrumen tes di sekolah yang belum dianalisis secara empiris, memiliki daya pembeda rendah, dan memuat distraktor yang tidak berfungsi, sehingga belum layak masuk ke bank soal berkualitas. Padahal analisis butir berfungsi untuk menentukan apakah soal dipertahankan, direvisi, atau dibuang. Jika proses ini diabaikan, keputusan penilaian berisiko kurang akurat dan kurang adil bagi peserta didik.

Secara operasional, implementasi analisis butir di kelas dapat dilakukan setelah tes dilaksanakan dengan memanfaatkan data respons peserta didik. Nafs et al. (2023) menunjukkan bahwa analisis butir berbasis jawaban nyata siswa mampu memberikan informasi diagnostik yang rinci mengenai validitas, daya pembeda, dan efektivitas pengecoh, sehingga guru dapat menentukan butir yang layak dipertahankan, direvisi, atau dibuang secara lebih objektif. Langkah-langkahnya meliputi pengumpulan data respons siswa, pengelompokan peserta didik ke dalam kelompok atas dan bawah, penghitungan indeks daya pembeda, serta pemeriksaan distribusi pemilihan setiap opsi sebagaimana diilustrasikan pada Tabel 1 dan Tabel 2. Sejalan dengan itu, Masullah et al. (2024) menegaskan bahwa kualitas hasil analisis sangat bergantung pada kelengkapan dan kecukupan data respons peserta didik, karena tanpa data yang memadai interpretasi indeks butir menjadi kurang akurat. Dengan menjadikan analisis ini sebagai bagian dari siklus evaluasi rutin, guru dapat secara bertahap membangun bank soal yang lebih bermutu.

Dalam praktiknya, pemanfaatan perangkat lunak analisis butir dapat menjadi solusi atas keterbatasan waktu dan keterampilan guru. Kaka et al. (2024) menegaskan bahwa aplikasi seperti Anates terbukti efektif membantu pendidik mengidentifikasi butir berkualitas, butir yang perlu direvisi, dan butir yang sebaiknya dibuang secara lebih efisien, sehingga analisis empiris menjadi lebih mudah diintegrasikan ke dalam rutinitas evaluasi di sekolah.

KESIMPULAN

Daya pembeda merupakan indeks yang menunjukkan kemampuan suatu butir soal dalam membedakan peserta didik berkemampuan tinggi dan rendah, sedangkan efektivitas pengecoh adalah kemampuan alternatif jawaban yang salah dalam menarik peserta didik yang belum menguasai materi. Kedua indikator ini saling berkaitan dan harus dianalisis secara simultan agar gambaran mutu butir soal menjadi utuh. Analisis keduanya dapat dilakukan melalui langkah sederhana, yaitu mengelompokkan siswa ke dalam kelompok atas dan bawah, menghitung indeks daya pembeda, serta memeriksa distribusi pemilihan setiap opsi, yang hasilnya digunakan sebagai dasar untuk mempertahankan, merevisi, atau membuang butir soal.

Penerapan analisis butir berbasis data empiris sangat penting untuk meningkatkan kualitas instrumen evaluasi. Oleh karena itu, guru disarankan menjadikan analisis daya pembeda dan efektivitas pengecoh sebagai bagian rutin setelah pelaksanaan tes, meningkatkan keterampilan penyusunan distraktor yang plausibel berdasarkan miskonsepsi umum siswa, serta memanfaatkan perangkat lunak analisis butir. Sekolah perlu memfasilitasi pelatihan dan mendorong terbentuknya bank soal yang teruji kualitasnya. Pengembangan selanjutnya dapat diarahkan pada pemanfaatan Teori Respons Butir (*Item Response Theory/IRT*) dan analisis berbantuan komputer agar proses evaluasi menjadi lebih efisien, valid, reliabel, dan adil.

DAFTAR PUSTAKA

- Ahmad, A., & Nipaah, N. (2017). Pengaruh efektivitas distraktor pada item tes pilihan ganda terhadap prestasi belajar matematika siswa. *Jurnal Varian*, 1(1), 76–82. <https://doi.org/10.30812/varian.v1i1.52>
- 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. <https://doi.org/10.37481/jmeh.v5i1.1165>
- Fitriani, N. (2021). Analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal pelatihan kewaspadaan kegawatdaruratan maternal dan neonatal. *Paedagoria: Jurnal Kajian Penelitian dan Pengembangan Kependidikan*, 12(2), 199–205. <https://doi.org/10.31764/paedagoria.v12i2.4956>
- Hermanto, F. Y., Anggraeni, A. D., & Sutirman, S. (2023). Analisis butir soal dengan classical test theory untuk mengukur kemampuan kognitif siswa. *Efisiensi: Kajian Ilmu Administrasi*, 20(2), 159–170. <https://doi.org/10.21831/efisiensi.v20i2.80225>
- Ikawati, D. S., Agustin, E. W., Hakim, L., & Pratiwi, V. (2024). Analisis butir soal pilihan ganda pada siswa SMK menggunakan Anates. *Edukatif: Jurnal Ilmu Pendidikan*, 6(6), 6773–6781. <https://doi.org/10.31004/edukatif.v6i6.7780>
- Kaka, L., Bano, V. O., & Njoeroemana, Y. (2024). Efektivitas analisis butir soal pilihan ganda menggunakan aplikasi Anates di SMPN 2 Kanatang. *Jurnal Inovasi Penelitian*, 4(9), 1441–1450. <https://doi.org/10.47492/jip.v4i9.3102>

- Kasanova, R., & Sulistiyono, R. (2023). Evaluasi butir soal pilihan ganda penilaian tengah semester dalam pembelajaran tematik untuk kelas V di SDN Gladak Anyar 4 Pamekasan. *Journal on Education*, 6(1), 5820–5834. <https://doi.org/10.31004/joe.v6i1.3773>
- Masullah, B. D., Zuhry, L. H., Usman, L. H., & Maulana, L. M. G. (2024). Analisis butir soal ujian akhir semester mata pelajaran matematika SMP Negeri 6 Praya Timur. *ELIPS: Jurnal Pendidikan Matematika*, 5(2), 152–161. <https://doi.org/10.47650/elips.v5i2.1219>
- Meriyotin, M., Martono, M., & Syam, C. (2015). Analisis butir soal ulangan umum semester genap mata pelajaran bahasa Indonesia kelas XI. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 4(3), 1–13. <https://doi.org/10.26418/jppk.v4i3.9319>
- Mustaqim, M., & Sulisti, H. (2024). Analisis butir soal PAS matematika peminatan: Daya pembeda, tingkat kesukaran, dan kualitas pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, 3(1), 44–56. <https://doi.org/10.24260/add.v3i1.3011>
- Nafs, H., Sridana, N., Hikmah, N., & Soeprianto, H. (2023). Analisis kualitas butir soal ulangan akhir semester genap mata pelajaran matematika kelas VII SMPN 6 Mataram tahun ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2324–2331. <https://doi.org/10.29303/jipp.v8i4.1785>
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan antara validitas item dengan daya pembeda dan tingkat kesukaran soal pilihan ganda PAS. *Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>
- Pradita, E., Megawanti, P., & Yulaningsih, Y. (2023). Analisis tingkat kesukaran, daya pembeda, dan fungsi distraktor PTS matematika SMPN Jakarta. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, 3(1), 109–118. <https://doi.org/10.1234/himpunan.v3i1.2023>
- Qadir, A., Huda, N., & Hermina, D. (2024). Analisis butir tes: Tingkat kesukaran, daya pembeda dan efektivitas pengecoh. *Al-Furqan: Jurnal Agama, Sosial, dan Budaya*, 3(3), 1450–1467. <https://doi.org/10.1234/alfurqan.v3i3.2024>
- Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi*, 13(3), 273–284. <https://doi.org/10.26740/jupe.v13n3.p273-284>
- Rahayu, R., & Djazari, M. (2016). Analisis kualitas soal pra ujian nasional mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1), 84–94. <https://doi.org/10.21831/jpai.v14i1.11370>
- Ratni, R. (2023). Analisis tingkat kesukaran, daya pembeda dan efektivitas pengecoh soal ulangan bahasa Indonesia di sekolah dasar. *Prosa: Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, 1(1), 117–123. <https://doi.org/10.1234/prosa.v1i1.2023>
- Sene, C. V. C., Tefa, A., & Lodo, R. Y. (2022). An analysis of multiple choices distractors made by teachers of SMP Negeri Malaka Barat. *Jurnal Pendidikan Tambusai*, 6(1), 31534–31539. <https://doi.org/10.31004/jptam.v6i1.12147>
- Taufik, A., & Suryadi, H. (2023). Analisis distraktor instrumen ujian akhir semester mata pelajaran matematika kelas VII MTs se-kecamatan kediri dengan menggunakan program iteman. *Nusra: Jurnal Penelitian dan Ilmu Pendidikan*, 4(4), 938–946. <https://doi.org/10.55681/nusra.v4i4.1719>



- Trianingsih, L. (2023). Analisis kualitas tes dan butir soal integral pada evaluasi formatif matematika teknik. *IJCEE: Indonesian Journal of Civil Engineering Education*, 9(2), 44–57. <https://doi.org/10.20961/ijcee.v9i2.84711>
- Yuliani, R. D., & Ismiyati, I. (2024). Perbandingan karakteristik butir soal mata pelajaran kearsipan berbantuan ITEMAN. *Efisiensi: Kajian Ilmu Administrasi*, 21(1), 1–15. <https://doi.org/10.21831/efisiensi.v22i1.57989>
- Zakiyah, Z., & Jamillah, J. (2021). Analisis butir, pilihan, dan reliabilitas tes prestasi bahasa Inggris selama covid. *Jurnal Pendidikan LPPM*, 30(3), 547–556. <https://doi.org/10.32585/jp.v30i3.1911>