

Analisis Kualitas Butir Soal Tes Kemampuan Akademik (TKA) SMA di Kota Merauke

¹Evani Doana Nababan, ²Andi Nurfahrunnisa, ³Rachmaniar

¹Program Studi Pendidikan Fisika, Universitas Musamus. Jl. Kamizaun Mopah Lama, Rimba Jaya, Kecamatan Merauke, Kabupaten Merauke, Papua 99611, Indonesia

²Program Studi Pendidikan Guru dan Sekolah Dasar, Universitas Musamus. Jl. Kamizaun Mopah Lama, Rimba Jaya, Kecamatan Merauke, Kabupaten Merauke, Papua 99611, Indonesia

³Program Studi Pendidikan Komputer, Universitas Musamus. Jl. Kamizaun Mopah Lama, Rimba Jaya, Kecamatan Merauke, Kabupaten Merauke, Papua 99611, Indonesia

*Corresponding Author e-mail: evaninababan@unmus.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas butir soal Tes Kemampuan Akademik (TKA) SMA di Kota Merauke dengan melihat tingkat kesukaran, daya beda dan efektivitas pengecohnya. Jenis penelitian yang digunakan adalah penelitian deskriptif kuantitatif. Teknik pengumpulan data yang dilakukan adalah menggunakan teknik dokumentasi. Hasil penelitian menunjukkan bahwa, tingkat kesukaran butir soal memiliki sebaran dengan perbandingan sukar 40%: sedang 56%: mudah 4%, daya beda butir soal terdapat 20 % baik, 44 % cukup dan 36 % buruk. Efektivitas pengecoh butir soal secara keseluruhan dikatakan efektif karena dari 75 butir soal yang diujikan terdapat 54 butir soal yang efektivitas pengecoh berfungsi. Kualitas butir soal TKA SMA di kota Merauke secara keseluruhan telah dikatakan cukup baik ditinjau dari teori klasik. Penelitian ini memberikan manfaat praktis bagi pendidik, perancang tes, dan lembaga pendidikan karena hasil analisis dapat dijadikan rujukan untuk memperkuat standar penilaian akademik di sekolah yang ada di Merauke. Selain itu, temuan ini dapat dimanfaatkan untuk mengembangkan bank soal yang lebih representatif serta memperbaiki pola konstruksi pertanyaan pada asesmen berikutnya.

Kata kunci: tingkat kesukaran; daya beda; efektivitas pengecoh

Analysis of the Quality of Test Items in the Senior High School Academic Ability Test (TKA) in Merauke City

Abstract

This study aims to examine the quality of the Academic Ability Test (TKA) items for Senior High School in Merauke City by examining the difficulty level, discrimination power, and the effectiveness of the distractors. The type of research employed is quantitative descriptive research. The data collection technique used is documentation. The results indicate that the distribution of item difficulty levels consists of 40% difficult items, 56% moderate items, and 4% easy items, while the discrimination power of the items comprises 20% good, 44% fair, and 36% poor. The distractor effectiveness of the items is generally considered effective, as 54 out of 75 items have functioning distractors. Overall, the quality of the TKA items for senior high schools in Merauke City is classified as reasonably good when reviewed from the perspective of classical test theory. This research provides practical benefits for teachers, test designers, and educational institutions because the analysis results can be used as a reference to strengthen academic assessment standards in schools in Merauke. Additionally, these findings can be utilized to develop a more representative question bank and improve question construction patterns for subsequent assessments.

Keywords: difficulty level; discrimination power; effectiveness of the distractors

How to Cite: Nababan, E. D., Andi Nurfahrunnisa, & Rachmaniar. (2025). Analisis Kualitas Butir Soal Tes Kemampuan Akademik (TKA) SMA di Kota Merauke. *Journal of Authentic Research*, 4(2), 2075-2083. <https://doi.org/10.36312/ep2nry15>



<https://doi.org/10.36312/jar.v3i2.2025>

Copyright© 2025, Nababan et al.

This is an open-access article under the CC-BY-SA License.



PENDAHULUAN

(Care et al., 2018; Hill & Sietz, 2023; Kinuthia et al., 2023; Syatriana et al., 2025) menyatakan penilaian hasil belajar merupakan salah satu pilar penting dalam upaya peningkatan mutu pendidikan. Pada tingkat menengah atas (SMA), Tes Kemampuan Akademik (TKA) diposisikan sebagai alat ukur yang dapat memberi informasi tentang capaian kompetensi peserta didik secara terstandar, mendukung proses pemetaan potensi akademik, seleksi, dan perencanaan intervensi pembelajaran. Kebijakan dan pedoman TKA yang telah dirilis oleh pihak berwenang menegaskan perlunya instrumen penilaian yang valid, andal, serta adil agar keputusan pendidikan yang diambil berdasarkan hasil tes tersebut memiliki dasar yang kuat (Vogl et al., 2018).

(Akbari & Shahrokhi, 2024; Kinuthia et al., 2023) menyatakan kualitas butir soal adalah faktor penentu mutu instrumen asesmen. Parameter dasar yang biasa dianalisis meliputi indeks kesukaran, daya pembeda, dan efektivitas pengecoh (*distractor efficiency*). Bersama-sama parameter ini membantu menentukan apakah sebuah butir mengukur konstruk yang dimaksud dan apakah butir tersebut memberikan kontribusi positif terhadap reliabilitas tes. Praktik analisis butir menggunakan pendekatan seperti Classical Test Theory (CTT) sudah menjadi standar dalam evaluasi instrumen pendidikan karena mampu mengungkap butir-butir bermasalah yang harus direvisi atau dihapus. Beberapa penelitian mutakhir telah mengkaji analisis butir pada konteks asesmen akademik di tingkat SMA. Misalnya, penelitian oleh (Rohmatdi 2024) menggunakan perangkat ITEMAN 4.3 untuk mengevaluasi soal ujian matematika SMA dan menemukan bahwa sekitar 25% butir perlu direvisi karena daya pembeda rendah. (Nisfatulsanah 2024) melaporkan bahwa analisis dengan pendekatan CTT dapat memberikan informasi yang cukup akurat dalam menentukan kualitas soal pada mata pelajaran sains.

Meskipun prosedur analisis butir relatif mapan secara metodologis, penelitian berkelanjutan pada konteks lokal (mis. sekolah, kabupaten/provinsi) terus menunjukkan variasi kualitas butir pada ujian sekolah dan tes sumatif mulai dari butir yang terlalu mudah atau terlalu sulit, daya beda rendah, hingga pengecoh yang tidak berfungsi. Temuan-temuan ini memperlihatkan bahwa tanpa praktik analisis butir yang konsisten dan tindak lanjut revisi, hasil TKA berisiko menyesatkan pemangku kepentingan tentang kemampuan akademik peserta didik. Oleh karena itu, penelitian empiris yang sistematis terhadap butir-butir TKA pada jenjang SMA sangat diperlukan sebagai dasar perbaikan instrumen dan praktik penyusunan soal.

Selain aspek teknis, analisis butir juga memiliki nilai praktis bagi pendidik dan pengembang instrumen. Sebagai contoh, ketika analisis menunjukkan pola tertentu, misalnya banyak butir yang gagal membedakan kelompok peserta didik berkemampuan tinggi dan rendah. Hal ini memberi masukan langsung terkait dengan cakupan kurikulum, kualitas pembelajaran, dan kemampuan kognitif yang belum dicapai oleh peserta didik. Dapat dikatakan bahwa analisis butir bukan hanya tugas psikometrik statis, tetapi juga alat reflektif bagi perbaikan proses pembelajaran dan penyusunan kurikulum. Butir soal dengan kualitas baik sebaiknya disimpan sebagai soal yang siap pakai, butir soal yang kualitasnya belum baik dapat diperbaiki pada bagian mana yang menjadi penyebab butir soal tersebut dikategorikan kurang

baik dan dilakukan pengujian kembali untuk mengetahui kualitasnya terutama dapat dipakai sebagai bank soal di Indonesia bagian Timur (Prayetno et al., 2023).

Berdasarkan konteks dan alasan yang telah dipaparkan, penelitian ini bertujuan untuk menganalisis butir soal TKA pada jenjang SMA dengan fokus pada parameter CTT: indeks kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian ini diharapkan dapat: (1) menggambarkan profil kualitas butir TKA pada sampel yang diteliti, (2) mengidentifikasi butir-butir bermasalah yang memerlukan revisi atau penghapusan, dan (3) memberikan rekomendasi teknis bagi penyusunan dan pemeriksaan soal TKA di masa mendatang. Hasil penelitian ini akan relevan bagi pengembang instrumen, pendidik, dan pembuat kebijakan yang berkepentingan terhadap mutu asesmen di tingkat menengah.

Berdasarkan latar belakang permasalahan diatas, maka peneliti terdorong untuk mengadakan penelitian lebih lanjut tentang kualitas soal Tes Kemampuan Akademik (TKA) di Kabupaten Merauke. Adapun judul yang diajukan peneliti adalah analisis kualitas butir soal Tes Kemampuan Akademik (TKA) tingkat SMA di Kabupaten Merauke.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif, karena tujuan utamanya adalah menggambarkan kualitas butir soal Tes Kemampuan Akademik (TKA) berdasarkan analisis statistik. Pendekatan ini dipilih untuk memperoleh gambaran objektif mengenai tingkat kesukaran, daya pembeda, dan efektivitas pengecoh setiap butir soal (Sugiyono, 2013). Pendekatan deskriptif memberikan informasi empiris tentang sejauh mana butir soal TKA telah memenuhi kriteria kualitas instrumen yang baik menurut teori tes klasik (*Classical Test Theory*). Dengan demikian, hasil penelitian diharapkan dapat digunakan sebagai dasar perbaikan dan penyusunan soal TKA berikutnya.

Jumlah responden disesuaikan dengan kebutuhan analisis statistik, yaitu minimal 100 peserta untuk menjaga stabilitas estimasi parameter butir (Rohmatdi, 2024). Sampel yang digunakan pada penelitian ini sebanyak 148 peserta. Data diambil dari lembar jawaban peserta didik atau data *response sheet* hasil TKA yang meliputi 75 butir soal pilihan ganda

Data dikumpulkan melalui metode dokumentasi, yaitu dengan mengumpulkan naskah soal, kunci jawaban, dan lembar hasil respon peserta didik. Data sekunder diperoleh dari pihak sekolah atau panitia pelaksana TKA. Analisis data dilakukan dengan menggunakan pendekatan Teori Tes Klasik (*Classical Test Theory/CTT*) dengan menggunakan program komputer *Item And Tes Analysis* (ITEMAN) versi 4.3. Software analisis butir digunakan untuk mengolah data hasil tes secara sistematis sehingga peneliti dapat memperoleh informasi mengenai kualitas setiap butir soal. Peneliti menginput jawaban peserta didik dan kunci jawaban, kemudian software menghitung secara otomatis parameter kualitas butir seperti indeks kesukaran, daya beda, efektivitas distraktor, dan reliabilitas tes. Hasil analisis ditampilkan dalam bentuk tabel dan laporan ringkas, sehingga memudahkan peneliti menilai apakah suatu butir perlu dipertahankan, direvisi, atau dibuang.

Data respons peserta dianalisis menggunakan software analisis butir, dimana mencakup tiga parameter utama:

1. Indeks Kesukaran (*Difficulty Index*, P)

Interpretasi indeks kesukaran yang digunakan dapat dilihat pada Tabel 2.1.

Tabel 1. Interpretasi Indeks Kesukaran

$P < 0,30$	sukar
$0,30 \leq P \leq 0,70$	sedang
$P > 0,70$	mudah

2. Daya Pembeda (*Discrimination Index*, D)

Daya pembeda menunjukkan sejauh mana suatu butir dapat membedakan peserta berkemampuan tinggi dan rendah. Interpretasi daya pembeda yang digunakan dapat dilihat pada Tabel 2.2.

Tabel 2. Interpretasi Daya Pembeda

$D < 0,20$	buruk
$0,20 \leq D < 0,40$	cukup
$0,40 \leq D < 0,70$	baik
$D \geq 0,70$	sangat baik

3. Efektivitas Pengecoh (*Distractor Efficiency*)

Pengecoh yang baik adalah opsi jawaban salah yang dipilih oleh minimal 5% peserta. Apabila pengecoh tidak dipilih sama sekali, maka butir dianggap memiliki pengecoh tidak berfungsi. Distraktor yang tidak dipilih oleh peserta sama sekali menunjukkan ketidakefektifan butir tersebut (Rohmatdi, 2024).

Kombinasi dari ketiga parameter ini membantu peneliti dan pendidik dalam memutuskan butir mana yang dipertahankan, direvisi, atau dihapus.

HASIL DAN PEMBAHASAN

1. Tingkat Kesukaran Butir Soal

Tingkat kesukaran merupakan ukuran seberapa mudah atau sulit suatu soal bagi peserta didik. Sesuai dengan hasil analisis tingkat kesukaran bahwa dari 75 butir soal TKA SMA di kota Merauke dengan bantuan program *Iteman versi 4.3* yang dapat dilihat pada nilai P pada tiap butir soal, lebih jelasnya dapat dilihat pada Tabel 3.1 berikut.

Tabel 3. Hasil klasifikasi tingkat kesukaran TKA SMA di kota Merauke

Kriteria	Nomor Butir Soal	Jumlah	Persentase (%)
Sukar	11, 12, 15, 16, 20, 26, 27, 29, 32, 39, 42, 44, 45, 46, 47, 48, 51, 52, 54, 57, 59, 60, 61, 64, 67, 68, 69, 71, 73, dan 74	33	40

Kriteria	Nomor Butir Soal	Jumlah	Persentase (%)
Sedang	1, 2, 5, 6, 8, 9, 10, 13, 14, 17, 18, 19, 21, 22, 23, 24, 25, 28, 30, 31, 33, 34, 35, 36, 37, 38, 40, 41, 43, 49, 50, 53, 55, 56, 58, 62, 63, 65, 66, 70, 72, dan 75	42	56
Mudah	3, 4, 7	3	4

Berdasarkan Tabel 3.1 tentang hasil klasifikasi tingkat kesukaran soal TKA SMA di kota Merauke, dari 75 butir soal diperoleh bahwa 33 (40%) butir berkategori sukar, 42 (56%) butir berkategori sedang, dan 4% butir soal di kategori mudah. Berdasarkan data tersebut diperoleh perbandingan tingkat kesukaran soal pada kategori (sukar: sedang: mudah) adalah (40%: 56%: 4%). Soal yang baik apabila memiliki sebaran tingkat kesukaran dengan perbandingan antara soal mudah: sedang: sukar yaitu 30%: 40%: 30% atau 30%: 50%: 20%. Hasil analisis tingkat kesukaran menunjukkan bahwa soal TKA SMA di kota Merauke memiliki sebaran tingkat kesukaran kategori soal yang sukar. Butir soal yang termasuk dalam kategori sukar memiliki tiga kemungkinan yaitu, butir soal tersebut dibuang, diteliti ulang, dan butir soal tersebut tetap dipertahankan (Fitriani, 2021).

2. Daya Beda Butir Soal

Soal yang baik adalah butir soal yang dapat membedakan peserta didik berkemampuan tinggi dan rendah. Untuk menentukan daya beda butir soal dapat dilihat dari nilai rpbis pada hasil analisis melalui program *Iteman* versi 4.3. Hasil analisis butir soal berdasarkan kategori indeks daya beda dapat dilihat pada tabel berikut.

Tabel 4. Hasil Klasifikasi Daya Beda TKA SMA di kota Merauke

Kriteria	Nomor Butir Soal	Jumlah	Persentase (%)
Buruk	3, 4, 6, 11, 14, 15, 20, 21, 33, 38, 43, 44, 46, 47, 48, 51, 53, 56, 62, 67, 68, 69, 70, 71, 72, 73, 74	27	36
Cukup	1, 7, 8, 9, 12, 16, 18, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 36, 37, 39, 41, 42, 45, 49, 50, 54, 55, 57, 59, 60, 61, 64, 66	33	44
Baik	2, 5, 10, 13, 17, 19, 29, 34, 35, 40, 52, 58, 63, 65, 75	15	20
Sangat Baik	-	0	0

Berdasarkan Tabel 3.2 tentang hasil klasifikasi daya beda soal TKA SMA di kota Merauke dari 75 butir soal, 27 butir (36 %) termasuk dalam kriteria buruk, 33 butir (44 %) termasuk kriteria cukup, 15 butir (20 %) termasuk dalam kriteria baik, dan tidak ada butir yang termasuk dalam kriteria sangat baik, sehingga dapat dikatakan bahwa soal TKA SMA di kota Merauke belum dapat membedakan peserta didik yang berkemampuan tinggi dan berkemampuan rendah. Terdapat beberapa alasan di balik rendahnya daya beda pada beberapa soal, diantaranya: beberapa butir soal terlalu sulit, beberapa butir soal pilihan jawaban (distraktor) tidak efektif, dan efek tebakan yang tinggi, dimana soal dengan peluang benar tinggi karena menebak sehingga mengurangi kemampuan membedakan.

3. Efektifitas Pengecoh Butir Soal

Keefektifan pengecoh dikatakan berfungsi apabila pengecoh sekurang-kurangnya dipilih oleh 5% peserta tes, dan pengecoh lebih banyak dipilih oleh kelompok peserta didik yang belum memahami materi yang diujikan. Untuk menentukan efektivitas pengecoh pada tiap butir soal dapat dilihat dari nilai *proporsi* (Prop.) tiap pilihan jawaban pada hasil analisis melalui program *Iteman versi 4.3*. Berikut ini pengelompokan butir soal berdasarkan efektifitas pengecoh dari soal TKA SMA di kota Merauke.

Tabel 5. Hasil Klasifikasi efektifitas Pengecoh TKA SMA di kota Merauke

Kriteria	Nomor Butir Soal	Jumlah	Persentase
a		h	(%)
Efektif	1, 3, 4, 6, 7, 8, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 56, 57, 59, 60, 61, 62, 64, 67, 68, 69, 70, 71, 72, 73, 74	54	72
Tidak Efektif	2, 5, 9, 10, 17, 19, 21, 24, 25, 26, 36, 40, 41, 49, 50, 55, 58, 63, 65, 66, 75	21	28

Berdasarkan tabel 3.3 tentang hasil klasifikasi efektifitas pengecoh soal TKA SMA di kota Merauke dari 75 butir, diperoleh bahwa sebanyak 54 (72%) butir soal berada pada kriteria efektif dan 21 (28%) butir soal berada pada kriteria tidak efektif. Hal ini menunjukkan bahwa, berdasarkan hasil persentase efektifitas pengecoh soal TKA SMA di kota Merauke, secara keseluruhan dapat dikatakan efektif, karena sebagian besar soal yang digunakan sudah efektif. Beberapa penelitian mutakhir telah mengkaji analisis butir pada konteks asesmen akademik di tingkat SMA. Misalnya, penelitian oleh Rohmatdi (2024) menggunakan perangkat ITEMAN 4.3 untuk mengevaluasi soal ujian matematika SMA dan menemukan bahwa sekitar 25% butir perlu direvisi karena daya pembeda rendah. Penelitian serupa oleh Nisfatulsanah (2024) menunjukkan bahwa analisis dengan pendekatan CTT dapat memberikan informasi yang cukup akurat dalam menentukan kualitas soal pada mata pelajaran sains.

4. Hasil Analisis Kualitas Butir Soal TKA SMA di kota Merauke

Hasil analisis yang diperoleh dengan menggunakan pendekatan teori klasik terdiri atas tingkat kesukaran, daya pembeda, dan efektifitas pengecoh butir soal TKA SMA di kota Merauke. Menurut statistik butir, kualitas butir soal secara keseluruhan dapat dikategorikan sebagai berikut.

- 1) Butir soal dikatakan memiliki kualitas yang baik, apabila soal tersebut memenuhi tiga kriteria yaitu tingkat kesukaran (sedang), daya pembeda (sangat baik, baik dan cukup), keefektifan pengecoh (efektif).
- 2) Butir soal dikatakan memiliki kualitas yang cukup baik, apabila soal tersebut hanya memenuhi dua dari tiga kriteria.
- 3) Butir soal dikatakan memiliki kualitas yang tidak baik, apabila soal tersebut tidak memenuhi satu atau lebih kriteria butir soal yang baik.

Berdasarkan hasil analisis soal yang berkualitas baik berjumlah 12 butir soal atau 16%, soal yang berkualitas cukup baik berjumlah 44 butir soal atau 58,70 % dan soal yang berkualitas tidak baik berjumlah 19 butir soal atau 25,30 %. Hasil analisis tersebut dituangkan ke dalam Tabel 3.4 berikut.

Tabel 6. Distribusi Soal berdasarkan Kualitas Butir Soal

Kriteria	Nomor Butir Soal	Jumlah	Persentase (%)
Baik	1, 8, 13, 18, 22, 23, 28, 30,31, 33, 35, 37	12	16
Cukup Baik	2, 5,6,7, 9, 10, 12, 14, 16,17, 19, 24, 25, 27, 29, 32, 34, 36, 38, 39, 40, 41, 42, 45, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 70, 72, 75	44	58,70
Tidak Baik	3, 4, 11, 15, 20, 21, 26, 43, 44, 46, 47, 48, 51,67, 68, 69, 71, 73, 74	19	25,30

Hasil analisis butir memberikan dasar kuat bagi peningkatan kualitas pembelajaran, penyesuaian kurikulum, dan perbaikan sistem penyusunan soal di Merauke. Analisis butir tidak hanya bermanfaat bagi pengembangan tes, tetapi juga memberi masukan terhadap proses pembelajaran di kelas, merancang kurikulum yang relevan dengan konteks lokal, serta menghasilkan soal yang lebih representatif terhadap kemampuan akademik peserta didik. Hasil analisis dapat menunjukkan topik yang paling sulit dikuasai peserta didik, materi yang perlu penekanan ulang, atau pola miskonsepsi yang sering terjadi. Oleh karena itu, analisis butir berfungsi sebagai umpan balik (*feedback mechanism*) bagi pendidik dan pengembang kurikulum di kota Merauke (Savika, 2024).

Temuan mengenai hasil analisis butir soal bukan hanya berfungsi sebagai evaluasi teknis, tetapi juga menjadi pijakan strategis untuk memperkuat proses belajar-mengajar, merancang kurikulum yang relevan dengan konteks lokal, serta

menghasilkan soal yang lebih representatif terhadap kemampuan akademik peserta didik.

Studi terbaru menunjukkan bahwa implementasi analisis butir secara berkala dapat meningkatkan mutu asesmen sekolah dan memperbaiki kualitas instrumen ujian berbasis kompetensi (Mariana et al., 2023). Pada tingkat SMA, analisis butir juga dapat digunakan untuk menyesuaikan soal dengan karakteristik peserta didik dan konteks sekolah masing-masing, sejalan dengan semangat Kurikulum Merdeka yang menekankan pembelajaran diferensiatif (Kemdikbudristek, 2024).

KESIMPULAN

Hasil penelitian menunjukkan bahwa, tingkat kesukaran butir soal memiliki sebaran dengan perbandingan sukar 40%: sedang 56%: mudah 4%, daya beda butir soal terdapat 20 % baik, 44 % cukup dan 36 % buruk. Efektivitas pengecoh butir soal secara keseluruhan dikatakan efektif karena dari 75 butir soal yang diujikan terdapat 54 butir soal yang efektivitas pengecoh berfungsi. Kualitas butir soal Tes Kemampuan Akademik (TKA) SMA di kota Merauke secara keseluruhan telah dikatakan cukup baik ditinjau dari teori klasik. Selain itu juga kegiatan simulasi serupa juga diharapkan dapat dilakukan secara berkala dan terus memberikan manfaat yang optimal bagi peserta didik dalam mempersiapkan diri. Terutama untuk peserta didik di daerah-daerah Timur dan Papua khusus Merauke (Jannoke et al., 2023).

REFERENSI

- Akbari, A., & Shahrokhi, M. (2024). Unveiling Fairness in Scoring: A Thorough Method for Precise Cutoff Score Calculation in Education Assessment. *Quality Assurance in Education: An International Perspective*, 32(3), 493–509. <https://doi.org/10.1108/QAE-12-2023-0208>
- Care, E., Kim, H., Vista, A., & Anderson, K. (2018). *Education System Alignment for 21st Century Skills: Focus on Assessment*. Center for Universal Education at The Brookings Institution. <https://eric.ed.gov/?id=ED592779>
- Fitriani, N. (2021). Analisis Tingkat Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh Soal Pelatihan Kewaspadaan Kegawatdaruratan Maternal dan Neonatal. *Pedagogia: Jurnal Kajian, Penelitian dan Pengembangan Kependidikan*, 12(2), 199-205. <https://doi.org/10.31764/paedagogia.v12i2.4956>
- Hill, S. L., & Sietz, P. (2023). Culturally Responsive Policy Development: Co-Constructing Assessment and Reporting Practices with First Nation Educators in Alberta. *Journal of Educational Leadership, Policy and Practice*, 37, 1–19.
- Jannoke, L., Susanty, M., Astuti, W., Mayangsari, T. R., & Nugraha, H. D. (2023). Simulasi Ujian Seleksi Nasional Berbasis Tes (SNBT) Bagi Siswa dari 347 SMA. *Jurnal Abdi Insani*, 10(3), 1767-1775. <https://doi.org/10.29303/abdiinsani.v10i3.1052>
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2025). *Pedoman Pelaksanaan Tes Kemampuan Akademik (TKA) SMA*. Jakarta: Pusat Asesmen Pendidikan

- Kinuthia, W., Abiri, A., Hamilton, J. B., & Char, A. (2023). Aligning SDOH Pillars to Learning Outcomes and Assessments. In J. B. Hamilton, B. A. Swan, & L. McCauley (Eds.), *Integrating a Social Determinants of Health Framework into Nursing Education* (pp. 197–210). Springer International Publishing. https://doi.org/10.1007/978-3-031-21347-2_7
- Meguellati, S., Samia, A., Ferhat, A., Djelloul, A., & Khalifa, Z. A. (2024). A Critical Analysis of the Use of Classical Test Theory (CTT) in Psychological Testing: A Comparison with Item Response Theory (IRT). *Pakistan Journal of Life and Social Sciences*, 22(2), 9442–9449. <https://doi.org/10.57239/PJLSS-2024-22.2.00715>
- Mariana., Lessy, D., Riaddin, D., Hardiansyah, M. R., & Pary, C. (2023). Analysis of Item Characteristics of Natural Sciences National Examinations for Junior High School Based on the Classical Test Theory Approach. *Jurnal Penelitian Pendidikan IPA*, 9(10), 7837–7844. <https://doi.org/10.29303/jppipa.v9i10.3698>
- Nisfatulsanah, A., Sugiharto, B. (2024). Comparing Classical Test Theory and Item Response Theory for Respiratory System Question Instruments. *Jurnal Pionir Pendidikan*, 13(3), 70–86. <https://doi.org/10.22373/pjp.v13i3.25435>
- Prayetno, E., Yusuf, I., & Mujasam, M. (2023). Analysis of Items in the Mid-Semester Assessment of Science Class VII SMP Yapis Manokwari. *Kasuari: Physics Education Journal (KPEJ)*, 5(2), 75–87. <https://doi.org/10.37891/kpej.v5i2.221>
- Rohmatdi, A., Martono, M., Zafrullah, Z., & Safitri, R. (2024). Optimizing Question Quality in Junior High School Exams: Classical Test Theory Evaluation with ITEMAN 4.3. *Jurnal Al-Ishlah Pendidikan*, 16(4), 4363–4375. <https://doi.org/10.35445/alishlah.v16i4.5671>
- Sugiyono. (2013). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D (Edisi Revisi)*. Bandung: Alfabeta
- 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. <https://doi.org/10.59966/pandu.v2i2.856>
- Syatriana, E., Saiful, Ariana, & Dzilarsy, A. (2025). The Use of E-Portfolio as Innovative Creation of EFL Students in Language Assessment. *International Journal of Language Education*, 9(1), 101–113.
- Vogl, K., Schmidt, I., & Preckel, F. (2018). The Role of Academic Ability Indicators in Big-Fish-Little-Pond Effect Research: A Comparison Study. *Journal of Educational Research*, 111(4), 429–438. <https://doi.org/10.1080/00220671.2017.1291485>