

Analisis Butir Soal Penilaian Akhir Semester Matematika Kelas XI SMA

Evan Dari Kristianto Simarmata, Yenita Roza*, Atma Murni

Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Riau, Indonesia

*Email: yenita.roza@lecturer.unri.ac.id

Abstract: *This research is prompted by the unavailability of proper item analysis for the final semester examination. The test scores are used only as final results, ignoring the characteristics of the test items. This research aims to know the item characteristics of Final Semester Examination (PASMAT) Grade 11 in academic year 2024/2025 SMA N 9 Pekanbaru. Data are collected using documentation technique. Data used in this research are students' scores in PASMAT. Data are analysed through descriptive quantitative approach. Test validity and reliability are firstly established, then the item characteristics (difficulty level, discrimination index and distractor efficiency) are calculated. Results of item analysis show that the PASMAT (35 items) has a high reliability (Cronbach's Alpha coefficient = 0.729). Based on difficulty level, 14 items are difficult, 14 items are moderate, and 7 items are easy. Based on discrimination index, 6 items are really bad, 3 items are bad, 14 items are fairly acceptable, 12 items are good. Based on distractor efficiency, 57 of 140 (40.7%) distractors are non-functional. The results of the research can be used as basis of consideration to improve the test quality for the upcoming final semester examinations.*

Keywords: *final semester examination; item analysis; item characteristic; reliability*

Abstrak: Penelitian ini dilatarbelakangi oleh ketidakterediaan analisis butir soal yang memadai pada penilaian akhir semester. Skor tes pada penilaian akhir semester digunakan hanya sebagai hasil akhir tanpa melihat karakteristik butir-butir soal terhadap tes tersebut. Penelitian ini bertujuan untuk mengetahui karakteristik butir soal pada Penilaian Akhir Semester Matematika (PASMAT) Kelas XI tahun ajaran 2024/2025 SMA N 9 Pekanbaru. Pengumpulan data dilakukan dengan metode dokumentasi. Data yang digunakan adalah skor perolehan siswa dari PASMAT. Teknik analisis data menggunakan pendekatan kuantitatif deskriptif. Validitas dan reliabilitas tes dihitung terlebih dahulu, selanjutnya karakteristik butir soal berupa tingkat kesulitan, indeks diskriminasi dan efisiensi pengecoh dihitung. Hasil analisis butir soal PASMAT (35 butir) menunjukkan bahwa sebanyak 80% (28 butir) valid dan reliabilitas dari PASMAT adalah 0,729 (*Cronbach's Alpha*), yang artinya reliabilitas PASMAT tinggi. Berdasarkan tingkat kesulitan, PASMAT terdiri dari 14 butir sulit, 14 butir sedang, dan 7 soal mudah. Berdasarkan daya beda, diperoleh 6 butir buruk sekali, 3 butir buruk, 14 butir cukup baik, 12 butir baik. Berdasarkan efisiensi pengecoh, 57 dari 140 (40,7%) pengecoh tidak berfungsi dengan baik. Manfaat dari hasil penelitian ini dapat digunakan sebagai bahan pertimbangan untuk meningkatkan kualitas butir-butir soal yang akan digunakan pada penilaian-penilaian akhir semester yang akan datang.

Kata kunci: analisis butir; karakteristik butir; penilaian akhir semester; reliabilitas

PENDAHULUAN

Asesmen merupakan bagian penting dalam proses pembelajaran (BSKAP, 2025). Asesmen berfungsi sebagai berfungsi untuk memperoleh informasi mengenai tingkat penguasaan kompetensi peserta didik sekaligus menjadi dasar bagi guru dalam merancang tindak lanjut pembelajaran (Samudra, 2026). Asesmen dalam pembelajaran matematika memiliki peran strategis mengingat mata pelajaran ini menuntut kemampuan penalaran dan pemecahan masalah yang tidak hanya dapat diukur melalui hafalan prosedural semata. Esmael & Rabut (2025) mengemukakan bahwa asesmen dalam pembelajaran matematika bertujuan untuk memberikan bukti yang dapat digunakan sebagai bahan pertimbangan untuk meningkatkan kualitas pembelajaran matematika.

Penilaian Akhir Semester (PAS) adalah bentuk asesmen/tes yang diselenggarakan secara rutin di sekolah. PAS dilakukan dua kali dalam satu tahun ajaran, yaitu pada akhir semester ganjil dan akhir semester genap. PAS digunakan sebagai alat untuk mengukur tingkat capaian pembelajaran siswa setelah menyelesaikan satu semester pembelajaran (Asri et al., 2024; Antari et al., 2021). Hasil PAS tidak hanya dimanfaatkan untuk menentukan nilai akademik siswa, tetapi juga menjadi dasar pengambilan keputusan kelulusan, kenaikan kelas atau *placement* (Manasikana et al., 2026).

Asesmen berupa PAS Matematika (PASMAT) secara ideal harus memperhatikan berbagai aspek agar data hasil tes dapat dipertanggungjawabkan. Pada praktiknya, pengembangan PASMAT di SMA masih menghadapi berbagai permasalahan (Syafudin & Vitantri, 2024). Soal-soal yang dikembangkan untuk PASMAT banyak dikembangkan oleh guru mata pelajaran sendiri sehingga perlu usaha untuk memastikan kualitas dari soal itu (Suzana, 2018; Antari et al., 2021; Nisa' & Hidayati, 2023).

Kualitas soal PASMAT dapat dievaluasi melalui analisis butir (*item analysis*). Pada PASMAT dengan bentuk pilihan ganda, analisis butir meliputi analisis validitas, reliabilitas, daya beda, tingkat kesukaran dan fungsi pengecoh (Surbakti, 2025). Analisis butir memperlihatkan kita karakteristik soal yang dapat membantu kita untuk mengidentifikasi butir soal yang mana yang terlalu sukar atau terlalu mudah, butir soal mana yang mampu membedakan siswa berkemampuan tinggi atau rendah, atau pengecoh yang tidak masuk akal (Quaigrain & Arhin, 2017). Analisis butir memungkinkan kita dapat memutuskan apakah suatu butir soal direvisi atau dibuang atau diganti dengan soal baru pada ujian (asesmen) yang akan datang.

Berdasarkan wawancara dengan guru matematika di SMA N 9 Pekanbaru, skor hasil PASMAT masih hanya digunakan sebagai hasil akhir perolehan siswa tanpa melihat karakteristik kualitas butir soalnya. Data perolehan siswa belum digali secara maksimal, yakni skor siswa sekadar berfungsi untuk sebagai komponen pengisian rapor, tanpa dianalisis lebih lanjut untuk mengevaluasi kualitas soal maupun mendiagnosis kemampuan spesifik peserta didik. Sebagaimana tes lainnya, instrumen PASMAT memiliki informasi kritis yang dapat dieksplorasi, seperti tingkat kesulitan, daya pembeda, serta efektivitas pengecoh pada soal pilihan ganda. Jika analisis butir soal diabaikan, guru kehilangan peluang untuk mengidentifikasi materi apa yang benar-benar dikuasai oleh sebagian besar siswa, atau memperlihatkan materi apa yang menjadi kelemahan sebagian siswa yang dimaksudkan untuk pembenahan atau peningkatan kualitas pembelajaran.

Berdasarkan permasalahan yang diuraikan di atas, penelitian ini bertujuan untuk melakukan analisis butir soal PASMAT di SMA N 9 Pekanbaru. Analisis ini memberikan gambaran empiris meliputi analisis validitas, reliabilitas, daya beda, tingkat kesukaran dan efektivitas pengecoh. Analisis ini diharapkan berfungsi sebagai bahan evaluasi alat ukur, tetapi lebih jauh diharapkan dapat menjadi dasar rekomendasi bagi guru dan sekolah untuk meningkatkan kualitas soal dalam penilaian atau asesmen akhir semester.

METODE

Penelitian ini menggunakan metode penelitian kuantitatif deskriptif. Analisis data secara kuantitatif meliputi analisis validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh dengan bantuan *Microsoft Excel*. Analisis data secara deskriptif dalam penelitian ini mencakup interpretasi terhadap analisis butir dan dokumen soal PASMAT. Data dalam pada penelitian ini diperoleh dari dokumen soal (35 butir) dan jawaban PASMAT Semester Ganjil kelas XI (140 siswa) di SMA N 9 Pekanbaru tahun pelajaran 2024/2025.

Analisis Validitas

Validitas butir soal diukur melalui korelasi antara skor butir dengan skor total, yang digunakan untuk mengetahui sejauh mana setiap butir soal mampu mengukur apa yang seharusnya diukur. Rumus yang digunakan untuk menghitung validitas butir soal adalah dengan rumus korelasi *product-moment*, yaitu:

$$r_{xy} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{(n \sum x^2 - x^2)(n \sum y^2 - y^2)}} \dots (1)$$

(Apriatni et al., 2022)

Keterangan:

r_{xy} : r hitung
 n : banyaknya peserta tes
 x : skor butir soal ke- i
 y : skor total

Soal PASMAt dinyatakan valid jika $r_{hitung} > r_{tabel}$. Kategori validitas butir dapat dilihat pada Tabel 1.

Tabel 1. Kategori Validitas Butir

Interval r_{xy}	Validitas
0,800 – 1,000	Sangat tinggi
0,600 – 0,799	Tinggi
0,400 – 0,599	Cukup Tinggi
0,200 – 0,399	Rendah
0,000 – 0,199	Sangat Rendah

Sumber: Apriatni et al., 2022

Analisis Reliabilitas

Reliabilitas berkaitan dengan seberapa konsisten tes yang diujikan. Rumus yang digunakan untuk menghitung reliabilitas PASMAt adalah dengan rumus *Cronbach' Alpha*, yaitu:

$$r_{\alpha} = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum \sigma_t^2}{\sigma_t^2} \right] \dots (2).$$

(Mahkotawati et al., 2025)

Keterangan:

r_{α} : koefisien *Cronbach's Alpha*
 k : banyak butir soal
 $\sum \sigma_t^2$: total varians skor dari setiap butir
 σ_t^2 : varians skor total

PASMAt dinyatakan reliabel jika jika $r_{\alpha} > r_{tabel}$. Kategori reliabilitas berdasarkan koefisien reliabilitas dapat dilihat pada Tabel 2.

Tabel 2 Kategori Reliabilitas

Interval r_{α}	Reliabilitas
$0,90 \leq r_{\alpha} \leq 1,00$	Sangat tinggi
$0,70 \leq r_{\alpha} < 0,90$	Tinggi
$0,40 \leq r_{\alpha} < 0,70$	Sedang
$0,20 \leq r_{\alpha} < 0,40$	Rendah
$r_{\alpha} < 0,20$	Sangat Rendah

Sumber: Apriatni et al., 2022

Analisis Tingkat Kesukaran

Tingkat kesukaran adalah persentase siswa yang menjawab benar dalam tes. Rumus yang digunakan dalam perhitungan tingkat kesukaran adalah:

$$P_i = \frac{\sum B}{N} \dots (3)$$

(Retnawati, 2017)

Keterangan:

 P_i : indeks kesukaran $\sum B$: banyak siswa yang benar pada butir soal N : total siswa

Kategori tingkat kesukaran disajikan dalam Tabel 3.

Tabel 3 Kategori Tingkat Kesukaran

Nilai P_i	Tingkat Kesukaran
$P_i < 0,30$	Sukar
$0,30 \leq P_i \leq 0,70$	Sedang
$P_i > 0,70$	Mudah

Sumber: Iskandar & Rizal (2017)

Analisis Daya Beda

Daya beda adalah kemampuan siswa untuk membedakan siswa dengan kemampuan tinggi (kelompok atas) dan kemampuan rendah (kelompok bawah). Rumus yang digunakan adalah:

$$DI = \frac{BA - BB}{N} \dots (4)$$

(Quaigrain & Arhin, 2017)

Keterangan:

 DI : indeks daya beda BA : banyak siswa pada kelompok atas yang menjawab benar BB : banyak siswa pada kelompok bawah yang menjawab benar N : 27% dari total siswa pada tes

Kategori daya beda disajikan dalam Tabel 4.

Tabel 4 Kategori Daya Beda

DI	Daya Beda
$DI \geq 0,70$	Baik sekali
$0,40 \leq DI < 0,70$	Baik
$0,20 \leq DI < 0,40$	Cukup
$0,00 < DI < 0,20$	Buruk
$DI \leq 0,00$	Buruk Sekali

Sumber: Pradita et al. (2023)

Analisis Pengecoh (Efektivitas Pengecoh)

Pengecoh yang efektif adalah opsi di mana terdapat minimal 5% dari peserta uji yang memilih opsi itu. Rumus yang digunakan adalah:

$$\text{persentase yang memilih opsi} = \frac{\text{banyak peserta yang memilih opsi}}{\text{jumlah peserta tes}} \times 100\% \dots (5)$$

Sumber: Yani et.al. (Pradita et al., 2023)

HASIL DAN PEMBAHASAN

Hasil

Pada penelitian ini, bentuk soal dari PASMAt adalah pilihan ganda, di mana jawaban benar diberi skor 1 dan jawaban salah diberi skor 0. Berdasarkan perhitungan maka diperoleh hasil untuk validitas, reliabilitas, tingkat kesukaran, daya beda dan tingkat pengecoh (efektivitas pengecoh).

Validitas

Berdasarkan perhitungan validitas butir menggunakan rumus *product-moment* yang dilakukan dengan bantuan *Microsoft Excel*, maka diperoleh hasil validitas butir soal PASMAt berikut ini (Tabel 5).

Tabel 5 Rekapitulasi Validitas Butir Soal

Nomor Butir	Banyak Butir	Validitas	Kategori
18	1		Sangat Rendah
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 15, 20, 21, 22, 24	15	Valid	Rendah
5, 9, 12, 14, 16, 17, 19, 23, 25, 26, 28, 30	12	(80%)	Cukup
-	-		Tinggi
27, 29, 31, 32, 33, 34, 35	7	Tidak Valid	-
		(20%)	
Total	35	100%	

Reliabilitas

Berdasarkan perhitungan menggunakan rumus *Cronbach's Alpha* yang dilakukan dengan bantuan *Microsoft Excel*, maka diperoleh reliabilitas *Cronbach's Alpha* PASMAt sebesar 0,729.

Tingkat Kesukaran

Berdasarkan perhitungan menggunakan rumus tingkat kesukaran yang dilakukan dengan bantuan *Microsoft Excel*, maka diperoleh rekapitulasi tingkat kesukaran soal sebagai berikut (Tabel 6).

Tabel 6 Rekapitulasi Tingkat Kesukaran

Nomor Butir	Banyak Butir	Kategori
2, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 26, 30	14 (40%)	Mudah
1, 3, 4, 11, 17, 18, 19, 20, 21, 22, 23, 24, 25, 28	14 (40%)	Sedang
27, 29, 31, 32, 33, 34, 35	7 (20%)	Sukar
Total	35 (100%)	

Daya Beda

Berdasarkan perhitungan menggunakan rumus daya beda yang dilakukan dengan bantuan *Microsoft Excel*, maka diperoleh rekapitulasi daya beda soal sebagai berikut (Tabel 7).

Tabel 7 Rekapitulasi Daya Beda

Nomor Butir	Banyak Butir	Kategori
27, 29, 31, 33, 34, 35	6 (17%)	Buruk Sekali
2, 18, 32	3 (9%)	Buruk
1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 24	14 (40%)	Cukup
11, 16, 17, 19, 20, 21, 22, 23, 25, 26, 28, 30	12 (34%)	Baik
-	0 (0%)	Baik Sekali
Total	35 (100%)	

Efektivitas Pengecoh

Pada PASMAt Semester Ganjil Kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025, terdapat 35 butir soal dengan satu kunci jawaban dan empat pengecoh. PASMAt terdiri dari 140 pengecoh (4 pengecoh setiap butir). Sebanyak 57 dari 140 pengecoh (40,7%) tidak berfungsi dengan baik atau tidak efektif. Berdasarkan perhitungan menggunakan rumus efektivitas pengecoh yang dilakukan dengan bantuan *Microsoft Excel*, maka diperoleh rekapitulasi efektivitas pengecoh soal sebagai berikut (Tabel 8).

Tabel 8 Rekapitulasi Efektivitas Pengecoh

Banyak Pengecoh Efektif	Nomor Butir	Banyak Butir Soal
0	5, 7, 12	3 (9%)
1	2, 4, 6, 8, 13, 14, 26	7 (20%)
2	1, 9, 15, 20, 30, 35	6 (17%)
3	10, 11, 16, 18, 19, 21, 22, 23, 25, 27, 28, 33	12 (34%)
4	3, 17, 24, 29, 31, 32, 34	7 (20%)
Total		100%

Pembahasan

PASMAt Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025 terdiri dari 35 butir soal di mana setiap butir soal terdiri dari lima (5) pilihan: satu kunci jawaban dan empat pengecoh.

Validitas

Validitas soal PASMAt Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025, berdasarkan hasil pada Tabel 5 sudah didominasi oleh soal-soal yang valid (80%), artinya PASMAt sudah didominasi oleh soal yang mengukur apa yang hendak diukur sebanyak 28 butir. Namun, perlu diperhatikan bahwa kategori maksimal validitas butir soal masih berada pada kategori cukup. Butir soal nomor 18 memiliki validitas yang sangat rendah.

Sebanyak 20% soal masih dalam kategori tidak valid. Hal ini disebabkan karena koefisien *product-moment*nya sangat rendah tetapi masih di atas nol, yaitu butir no. 27. Butir soal lain yang tidak valid memiliki koefisien *product-moment* dengan nilai negatif (butir no. 29, 31, 32, 33, 34, 35).

Reliabilitas

Reliabilitas PASMAT Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025 memiliki koefisien *Cronbach's Alpha* sebesar 0,729 yang berada pada kategori tinggi. Reliabilitas yang tinggi menunjukkan bahwa PASMAT memiliki konsistensi internal yang dapat diterima atau memiliki kelayakan dapat dipercaya untuk mengukur kemampuan matematika siswa. Berdasarkan analisis validitas dan reliabilitas, maka ada kesesuaian dengan teori, di mana tes yang valid sudah pasti reliabel (Arbeni et al., 2025).

Reliabilitas dapat ditingkatkan dengan cara menghapus butir-butir yang tidak valid (Zijlmans et al., 2019). Misalnya, dengan menghapus butir yang tidak valid (butir no. 27, 29, 31, 31, 32, 33, 34, 35), maka akan didapatkan koefisien reliabilitas *Cronbach's Alpha* sebesar 0,779. Akan tetapi, keputusan menghapus butir tersebut juga harus dengan penuh pertimbangan karena butir-butir tidak valid adalah butir yang memiliki tingkat kesukaran kategori sukar.

Tingkat Kesukaran

Hasil pada Tabel 6 menunjukkan bahwa PASMAT Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025 memiliki distribusi soal dengan tingkat kesukaran kategori mudah (40%), sedang (40%) dan sukar (20%). Sudjana (Fauzie et al., 2021) mengemukakan bahwa proporsi tingkat kesukaran yang berimbang adalah dengan perbandingan 3-4-3 (30% mudah, 40% sedang, 20% sukar). PASMAT belum memiliki proporsi yang berimbang secara teoritis, tetapi sudah memiliki kategori yang tepat untuk butir dengan kategori sedang.

Soal-soal sukar memiliki peran dalam proporsi sebuah tes secara teoritis. Akan tetapi, soal-soal sukar pada PASMAT ini memiliki kategori soal tidak valid. Secara statistik, seperti pada pembahasan reliabilitas, menghapus butir-butir yang tidak valid dapat meningkatkan koefisien reliabilitasnya. Akan tetapi, menghapus butir-butir tidak valid akan mengubah tes dalam kategori mudah dan sedang saja.

Daya Beda

Hasil pada Tabel 7 menunjukkan bahwa PASMAT Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025 menunjukkan bahwa terdapat enam (6) butir soal yang memiliki daya beda soal yang buruk sekali, yaitu di bawah nol. Seperti pada pembahasan sebelumnya, soal-soal tersebut dalam kategori tidak valid. PASMAT juga memiliki tiga butir soal yang memiliki daya beda yang buruk, yaitu butir no. 2, 18 dan 32.

Butir no. 2 valid akan tetapi memiliki tingkat kesukaran yang termasuk terlalu mudah (88,5% menjawab benar) sehingga soal dapat dengan mudah dijawab oleh siswa dengan kemampuan tinggi dan kemampuan rendah. Butir no. 18 memiliki tingkat validitas sangat rendah walaupun memiliki tingkat kesukaran sedang. Butir no. 32 juga dikategorikan tidak valid.

Quaigrain & Arhin (2017) menyatakan bahwa: *“Items with negative discrimination indices must be deleted or replaced. Classroom teachers must rewrite all items with zero discrimination indices. Also, teachers must replace or rewrite all items with low positive discrimination indices. When all these are done the reliability of the test will be increased.”*

Berdasarkan pernyataan di atas, keenam butir soal dengan daya beda soal yang buruk sekali harus dihapus. Butir yang memiliki indeks daya beda nol, atau daya beda rendah (buruk) harus ditulis ulang (revisi). Dengan menghapus butir-butir itu, maka reliabilitas PASMAT dapat ditingkatkan.

Efektivitas Pengecoh

Sebanyak 57 dari 140 pengecoh (40,7%) tidak berfungsi dengan baik atau tidak efektif. Hasil pada Tabel 8 menunjukkan bahwa 46% dari butir soal dalam tes memiliki paling banyak 2 pengecoh yang efektif. Butir no.5, 7, dan 12 tidak memiliki pengecoh yang efektif. Butir no. 2, 4, 6, 8, 13, 14, dan 26 hanya memiliki 1 pengecoh yang efektif, sementara butir no. 1, 9, 15, 20, 30, 35 memiliki 2 pengecoh yang efektif.

Tarrant et. al. (Quaigrain & Arhin, 2017) menyatakan bahwa efektivitas pengecoh adalah sebuah alat yang mampu memperlihatkan apakah butir soal itu disusun dengan baik atau gagal memenuhi tujuannya. Butir yang tidak memiliki pengecoh yang efektif masih memiliki ruang untuk perbaikan yaitu dengan mengganti pengecoh yang tidak berfungsi dengan pilihan jawaban yang lebih masuk akal. Berdasarkan dari hasil analisis, walaupun secara perhitungan ketiga butir ini berada pada kategori validitas cukup, pengecoh butir-butir no.5, 7 dan 12 perlu direvisi karena dapat dikategorikan dalam kategori *sangat* mudah (masing-masing 92%, 91%, dan 90% siswa dapat menjawab benar).

SIMPULAN

Analisis butir soal PASMAt Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025 menunjukkan bahwa PASMAt didominasi oleh soal-soal yang valid (sebanyak 80% atau 28 butir) dan memiliki reliabilitas tinggi (*Cronbach's Alpha* = 0,729). Berdasarkan tingkat kesulitan, PASMAt terdiri dari 14 butir sulit, 14 butir sedang, dan 7 soal mudah. Berdasarkan daya beda, diperoleh 6 butir buruk sekali, 3 butir buruk, 14 butir cukup baik, 12 butir baik. Berdasarkan efisiensi pengecoh, 57 dari 140 (40,7%) pengecoh tidak berfungsi dengan baik.

Berdasarkan analisis butir soal PASMAt Semester Ganjil kelas XI di SMA N 9 Pekanbaru tahun pelajaran 2024/2025, terdapat saran-saran untuk penelitian yang akan datang, yaitu: 1) soal-soal yang valid dapat disimpan dan digunakan untuk PASMAt yang akan datang sesuai dengan tujuannya dan dapat diteliti ulang validitas dan reliabilitasnya, 2) validasi internal (*content validity*) berdasarkan *expert review* terhadap PASMAt perlu dilakukan sehingga soal dapat direvisi dari segi isi, konstruk, dan bahasa, terutama pada soal-soal yang tidak valid, memiliki daya beda buruk sekali atau yang tidak memiliki pengecoh yang berfungsi dengan baik.

DAFTAR PUSTAKA

- Antari, B., Susanta, A., & Siagian, T. A. (2021). Analisis Soal Penilaian Akhir Semester Matematika SMA Negeri 8 Kota Bengkulu Tahun Ajaran 2020 / 2021. *JEMS Jurnal Edukasi Matematika dan Sains*, 9(2), 299–312. <https://doi.org/10.25273/jems.v9i2.10421>
- Apriatni, S., Yuhana, Y., & Sukirwan, S. (2022). Pengembangan Instrumen Literasi Numerasi Materi Trigonometri Kelas X SMA. *EDU-MAT: Jurnal Pendidikan Matematika*, 10(2), 185. <https://doi.org/10.20527/edumat.v10i2.13720>
- Arbeni, W., Windiani, A., Sihotang, D. S. B., Anggraini, N., Wulandari, S., & Nugroho, A. (2025). Test Reliability Analysis in Educational Evaluation: A Quantitative Approach to Consistency and Validity. *Jurnal Nasional Holistic Science*, 5(1), 59–64.
- Asri, S. A., Syahrul, & Ishak, S. (2024). Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Matematika Kelas VIII. *Pedagogy*, 10(1), 120–133.

- BSKAP. (2025). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar dan Jenjang Pendidikan Menengah*. BSKAP.
- Esmael, N. T., & Rabut, J. F. (2025). Evaluating Teacher-Made Mathematics Test and Teachers' Competency in Assessment. *International Journal of Research and Innovation in Social Sciences (IJRISS)*, 9(4), 1532–1547.
<https://doi.org/10.47772/IJRISS>
- Fauzie, M., Pada, A. U. T., & Supriatno. (2021). Analysis of the difficulty index of item bank according to cognitive aspects during the Covid-19 pandemic. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 25(2), 162–169.
- Iskandar, A., & Rizal, M. (2017). Analisis Kualitas Soal Di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 12–23.
- Mahkotawati, R., Rijanto, T., & Rusimamto, P. W. (2025). Uji Validitas dan Reliabilitas Instrumen Penelitian : Angket Pengalaman Praktik Kerja Lapangan (PKL) pada Siswa SMK. *JIP Jurnal Ilmiah Ilmu Pendidikan*, 8(2), 1830–1835.
- Manasikana, A., Harani, D. A., Nabila, M., Azyati, N., Mumtaza, U., Sukma, Z., Islam, U., Sultan, N., Kasim, S., Islam, U., Sultan, N., Kasim, S., Islam, U., Sultan, N., Kasim, S., Islam, U., Sultan, N., Riau, K., Islam, U., ... Riau, S. K. (2026). Telaah Literatur tentang Efektivitas Evaluasi Sumatif dalam Menilai Pencapaian Kompetensi Peserta Didik. *An-Nahdloh Journal of Education and Islamic Studies*, 1(2), 514–530.
- Nisa', Z. E. K., & Hidayati, K. (2023). Analisis Butir Soal Penilaian Akhir Semester Matematika. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(3), 3516–3529.
- Pradita, E., Megawanti, P., & Yulianingsih. (2023). Analisis Tingkat Kesukaran, Daya Pembeda, dan Fungsi Distraktor. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, 3(80), 109–118.
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 12(1), 1–11. <https://doi.org/10.1080/2331186X.2017.1301013>
- Retnawati, H. (2017). *Validitas, Reliabilitas dan Karakteristik Butir (Panduan untuk Peneliti, Mahasiswa dan Psikometrian) (2nd ed.)*. Parama Publishing.
- Samudra, M. B. (2026). Optimalisasi Asesmen Capaian Belajar Matematika dengan Metode Trial by Error untuk Meningkatkan Hasil Belajar Peserta Didik. *Interdisciplinary Explorations in Research (IERJ)*, 4(1), 89–106.
- Surbakti, I. W. (2025). Analisis Kualitas Butir Soal pada Uji Coba Evaluasi Pembelajaran Matematika Tingkat SMA. *Cendikia: Jurnal Ilmu Pengetahuan*, 5(1), 117–125.
- Suzana, A. (2018). Analisis Tingkat Kesukaran dan Daya Beda Butir-Butir Soal Penilaian Akhir Tahun Matematika Kelas X di SMA Negeri 1 Purbalingga. *Jurnal MathGram Matematika*, 2(2), 1–8.
- Syafrudin, T., & Vitantri, C. A. (2024). Hambatan Guru dalam Pengajuan Soal Literasi Matematika Bertema Word Problem. *Jurnal Tadris Matematika*, 7(2), 207–220.
- Zijlmans, E. A. O., Tijmstra, J., Ark, L. A. Van Der, & Sijtsma, K. (2019). Item-Score Reliability as a Selection Tool in Test Construction. *Frontiers in Psychology*, 9, 1–12.
<https://doi.org/10.3389/fpsyg.2018.02298>