

Analisis Butir Soal Penilaian Akhir Semester (PAS) Ganjil Mata Pelajaran Bahasa Indonesia Kelas XI Akuntansi SMK Negeri 2 Padang

Nada Salsabilla^{1*}, Zuraida Khairani², Susanti Marisya³

^{1,2,3} Pendidikan Bahasa dan Sastra Indonesia, Universitas Ekasakti-Aai Padang, Indonesia

*Email: salsabillanada96@gmail.com

Received 06/08/2025 ; Revised 14/08/2025 ; Accepted 27/08/2025 ; Published 29/08/2025

Abstrak

Penelitian ini dilatarbelakangi oleh belum dilaksanakannya analisis butir soal Penilaian Akhir Semester (PAS) mata pelajaran Bahasa Indonesia tahun ajaran 2024/2025 di sekolah, sehingga kualitas soal dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh belum diketahui. Penelitian ini bertujuan menganalisis kualitas butir soal PAS serta mengidentifikasi apakah soal terlalu mudah/sulit dan mampu membedakan siswa yang telah menguasai materi dengan yang belum. Penelitian menggunakan metode deskriptif kuantitatif. Data berupa naskah soal dan lembar jawaban peserta didik PAS ganjil 2024/2025, kemudian ditabulasi berdasarkan pilihan jawaban dan dianalisis menggunakan teori pengukuran meliputi uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil penelitian menunjukkan: pada paket A terdapat 7 butir valid (70%) dan 3 tidak valid (30%), reliabilitas sedang (0,620), tingkat kesukaran didominasi kategori sedang 7 butir (70%) dan mudah 3 butir (30%), daya pembeda kategori baik–baik sekali 7 butir (70%) namun masih terdapat 3 butir berkategori jelek (30%), serta pengecoh baik–sangat baik pada seluruh butir (100%). Pada paket B terdapat 8 butir valid (80%) dan 2 tidak valid (20%), reliabilitas tinggi (0,843), tingkat kesukaran didominasi mudah 6 butir (60%) dan sedang 4 butir (40%), daya pembeda baik–baik sekali 7 butir (70%) dengan 3 butir cukup–jelek (30%), serta pengecoh baik–sangat baik 9 butir (90%) dan 1 butir jelek (10%). Rekomendasi penelitian adalah merevisi atau mengganti butir yang tidak valid dan berdaya pembeda rendah, menyeimbangkan tingkat kesukaran terutama pada paket B, serta melakukan analisis butir secara rutin sebelum soal digunakan kembali.

Kata Kunci: Analisis butir soal, penilaian akhir semester, validitas, reliabilitas, efektivitas pengecoh

Abstract

This study was motivated by the absence of item analysis for the Indonesian Language End-of-Semester Assessment (PAS) for the 2024/2025 academic year, so the quality of the test items in terms of validity, reliability, difficulty level, discrimination power, and distractor effectiveness was unknown. The study aimed to analyze the quality of the PAS items and to identify whether the items were too easy or too difficult and whether they could differentiate between students who had mastered the content and those who had not. A descriptive quantitative method was employed. The data consisted of the test papers and students' answer sheets from the odd-semester PAS 2024/2025. Students' responses were tabulated by option choice and analyzed using measurement theory, covering item validity, test reliability, item difficulty, item discrimination, and distractor effectiveness. The results showed that in Package A, 7 items were valid (70%) and 3 were invalid (30%); the reliability was moderate (0.620); difficulty levels were dominated by the medium category (7 items; 70%) and easy category (3 items; 30%); discrimination indices were good to very good for 7 items (70%), but 3 items (30%) were poor; and all distractors were good to very good (100%). In Package B, 8 items were valid (80%) and 2 were invalid (20%); reliability was high (0.843); difficulty levels were dominated by easy items (6 items; 60%) and medium items (4 items; 40%); discrimination indices were good to very good for 7 items (70%), with 3 items (30%) in the fair to poor range; and distractors were good to very good for 9 items (90%), while 1 item (10%) had poor distractors. The study recommends revising or replacing invalid items and items with low discrimination, balancing item difficulty—especially in Package B—and conducting routine item analysis before reusing the test.

Keywords: Item analysis, end-of-semester assessment, validity, reliability, distractor effectiveness

PENDAHULUAN

PENDAHULUAN

Penilaian Akhir Semester (PAS) Bahasa Indonesia berfungsi sebagai evaluasi sumatif untuk menilai capaian belajar peserta didik pada akhir periode pembelajaran sekaligus menjadi dasar pengambilan keputusan akademik. Karena hasil PAS dipakai untuk memotret ketercapaian tujuan pembelajaran, instrumen PAS harus disusun dan dievaluasi berbasis data agar pengukuran capaian siswa tidak bias oleh butir soal yang terlalu mudah, terlalu sulit, atau tidak selaras dengan kompetensi yang diukur (Mertasari, 2022; Himawan et al., 2024). ([Ejournal Undiksha](#))

Kualitas sebuah tes tidak cukup dinilai dari “banyaknya siswa benar” atau nilai rata-rata kelas, tetapi perlu ditopang oleh karakteristik instrumen yang baik. Literatur evaluasi pendidikan menempatkan validitas dan reliabilitas sebagai syarat utama, lalu diikuti analisis karakteristik butir seperti tingkat kesukaran, daya pembeda, dan efektivitas pengecoh untuk memastikan setiap butir berfungsi membedakan siswa yang menguasai materi dan yang belum (Arikunto, 2018; Himawan et al., 2024). ([BintangPusnas](#))

Kajian 10 tahun terakhir menunjukkan masalah yang berulang pada tes buatan guru: analisis butir belum menjadi kebiasaan, sehingga masih ditemukan butir berkualitas rendah yang tetap digunakan dalam ujian semester. Pada konteks PAS Bahasa Indonesia tingkat SMP, Pohan dan Nurrahmah (2025) menemukan proporsi butir berkategori “jelek” dan bahkan “negatif” pada daya pembeda masih muncul, serta sebagian pengecoh tidak berfungsi sehingga perlu diganti atau dibuang. Temuan serupa pada mata pelajaran lain juga memperlihatkan bahwa evaluasi butir (validitas, reliabilitas, kesukaran, pembeda, pengecoh) diperlukan untuk menentukan butir yang layak masuk bank soal dan butir yang wajib direvisi (Dzahabiyah et al., 2024). Pada konteks tes bahasa Indonesia berbasis ITEMAN, Himawan et al. (2024) menegaskan masih banyak praktik penilaian yang langsung memberi nilai tanpa analisis butir, padahal hasil analisis digunakan untuk memutuskan kualitas instrumen secara objektif. ([Jurnal LLDikti 13](#))

Kondisi lapangan di sekolah menunjukkan persoalan yang sejalan. Wawancara peneliti dengan guru Bahasa Indonesia (Nurhamidah, 16 Desember 2024) menunjukkan soal PAS Bahasa Indonesia tahun ajaran 2024/2025 belum pernah dianalisis dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Situasi ini berisiko menghasilkan keputusan penilaian yang kurang akurat karena kualitas butir tidak diketahui, padahal PAS dipakai sebagai cermin ketercapaian tujuan pembelajaran dan pemetaan kemampuan siswa. ([Ejournal Undiksha](#))

Kesenjangan penelitian terletak pada belum tersedianya bukti empiris berbasis data respons siswa untuk menilai kualitas soal PAS Bahasa Indonesia tahun ajaran 2024/2025 pada konteks sekolah ini, sementara penelitian terdahulu lebih banyak mengambil lokasi dan tahun pelaksanaan berbeda atau hanya menekankan sebagian indikator. Kebaruan penelitian ini adalah evaluasi komprehensif terhadap paket soal PAS Bahasa Indonesia 2024/2025 menggunakan lima komponen pengukuran (validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh) sehingga menghasilkan keputusan berbasis data tentang butir yang dipertahankan, direvisi, atau diganti untuk perbaikan kualitas tes berikutnya (Himawan et al., 2024; Pohan & Nurrahmah, 2025).

Penelitian ini diarahkan untuk: (1) mendeskripsikan validitas butir soal PAS Bahasa Indonesia 2024/2025, (2) menentukan reliabilitas tes, (3) memetakan tingkat kesukaran butir, (4) menilai daya pembeda tiap butir, (5) mengevaluasi efektivitas pengecoh pada setiap opsi jawaban, serta (6) merumuskan rekomendasi perbaikan berupa klasifikasi butir layak pakai,

revisi, dan tidak layak untuk mendukung penguatan bank soal dan peningkatan kualitas evaluasi pembelajaran (Dzahabiyah et al., 2024; Himawan et al., 2024).

METODE PENELITIAN

Pada penelitian ini metode yang digunakan merupakan metode deskriptif kuantitatif. Menurut Rustamana, dkk (2024) penelitian kuantitatif didefinisikan sebagai investigasi sistematis terhadap fenomena dengan mengumpulkan data yang dapat diukur dengan melakukan teknik statistik, matematika atau komputasi. Penelitian deskriptif adalah penelitian yang dilakukan untuk mengetahui nilai variabel, baik satu variabel atau lebih yang sifatnya independen tanpa membuat hubungan maupun perbandingan dengan variabel yang lain. Penelitian deskriptif bertujuan untuk memberikan gambaran atau deskripsi tentang suatu keadaan secara objektif (Sujarweni dalam Purnia dkk, 2020).

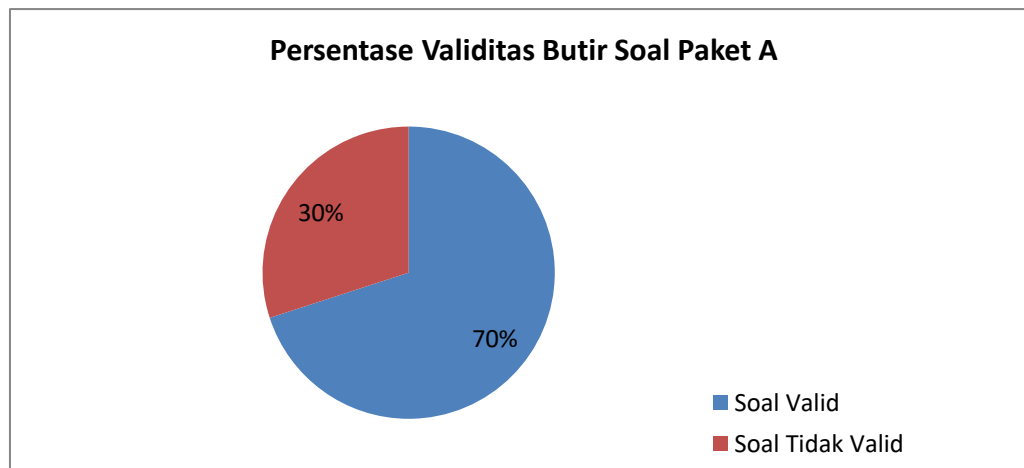
Pada penelitian ini informasi dan data yang diperoleh berupa data kuantitatif. Penelitian deskriptif kuantitatif dapat berupa pengelompokan, peringkasan, dan penyajian data berupa angka-angka yang lebih informatif. Penelitian ini menginformasikan keadaan objek yang diamati sesuai dengan apa yang ada sebenarnya dengan mendeskripsikan butir soal. Peneliti mengambil dan mengumpulkan data, lalu data dianalisis dengan mengacu teori pengukuran yang meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.

Peneliti melakukan analisis terhadap butir soal ujian akhir semester. Data yang diperoleh berupa lembar jawaban peserta didik dan soal dari penilaian akhir semester ganjil tahun ajaran 2024/2025 mata pelajaran Bahasa Indonesia. Data yang didapatkan kemudian ditabulasi untuk setiap pilihan jawaban peserta didik. Bagi peserta didik yang menjawab soal dengan benar maka diberi skor (1) dan peserta didik yang menjawab salah diberi skor (0). Software yang digunakan untuk mengolah dan menganalisis data menggunakan program SPSS Version 22 dan Microsoft Excel.

HASIL DAN PEMBAHASAN

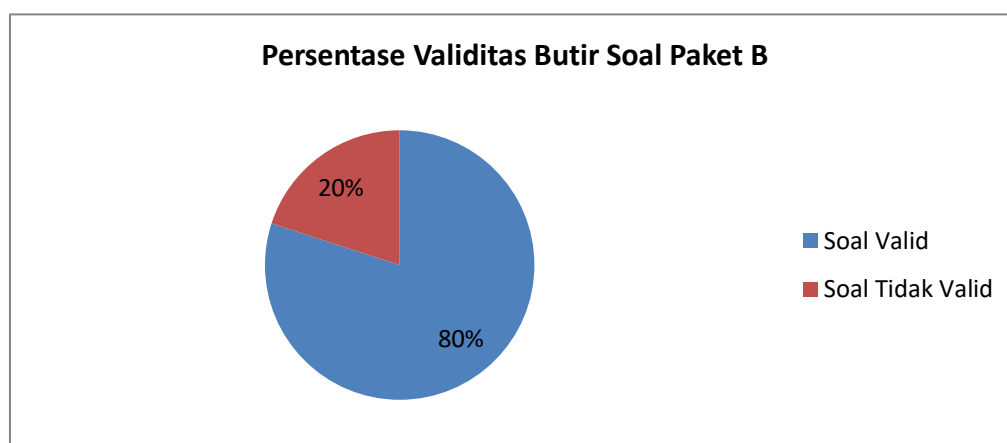
Penelitian yang dilakukan ini untuk mengetahui tingkat kualitas analisis butir soal Ujian Akhir Semester Ganjil pada mata pelajaran Bahasa Indonesia kelas XI program keahlian Akuntansi Keuangan Lembaga SMK Negeri 2 Padang tahun ajaran 2024/2025 yang dilihat dari segi Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh. Data yang digunakan dalam penelitian ini berupa 20 butir soal pilian ganda dengan bentuk soal terbagi atas 10 soal paket A dan 10 soal paket B yang diikuti oleh 34 peserta didik dimana 17 peserta didik untuk soal paket A dan 17 peserta didik untuk soal paket B pada kelas XI program keahlian Akuntansi Keuangan Lembaga.

Instrumen yang digunakan untuk memperoleh data melalui metode dokumentasi berupa soal ujian, jawaban peserta didik, dan kunci jawaban. Selanjutnya, data yang telah ada dianalisis dengan menggunakan software Excel dan SPSS sehingga mendapatkan hasil untuk mengetahui kualitas butir soal baik dari segi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.



Gambar 1. Validitas butir soal paket A

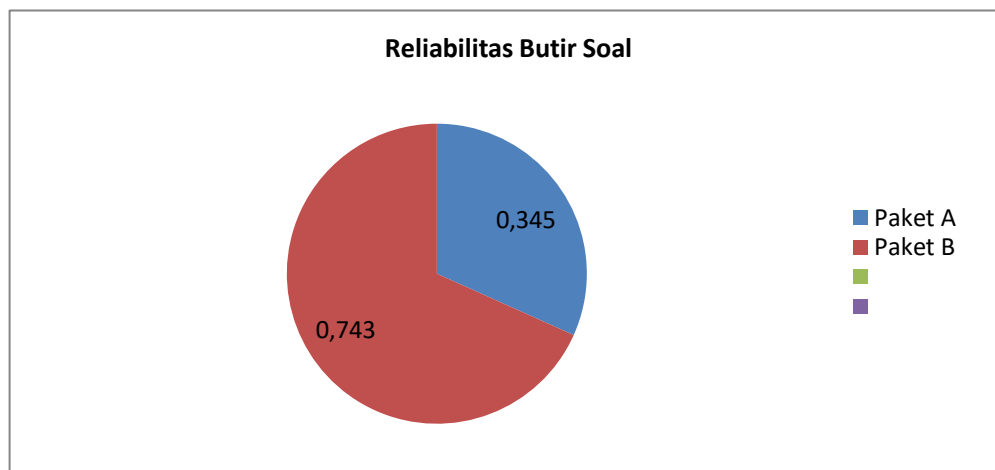
Hasil analisis validitas butir soal Ujian Akhir Semester Ganjil pada mata pelajaran Bahasa Indonesia kelas XI program keahlian Akuntansi Keuangan Lembaga SMK Negeri 2 Padang tahun ajaran 2024/2025 memiliki validitas pada soal paket A yang baik. Hal ini membuktikan bahwa terdapat 7 butir soal (70%) yang dinyatakan valid dan 3 butir soal (30%) dinyatakan tidak valid. Soal yang valid terdiri dari butir soal nomor 1,3,4,6,7,8,10 karena pada perhitungan menggunakan rumus dan SPSS menunjukkan $r_{hitung} \geq r_{tabel}$. Peninjauan data dari bentuk soal pilihan ganda, butir soal yang valid ditemukan sesuai dengan capaian pembelajaran dan indikator pembelajaran serta untuk option yang digunakan relevan dengan soal dan tidak mudah ditebak oleh siswa. Pada soal yang tidak valid terdapat pada butir soal nomor soal 2,5,9 karena pada perhitungan dengan menggunakan rumus dan SPSS menunjukkan $r_{hitung} < r_{tabel}$. Peninjauan data dari bentuk soal pilihan ganda, butir soal yang tidak valid ditemukan ketidaksesuaian dengan capaian pembelajaran dan indikator pembelajaran serta untuk option yang digunakan tidak relevan dengan soal dan mudah ditebak oleh siswa.



Gambar 2. Validitas butir soal paket B

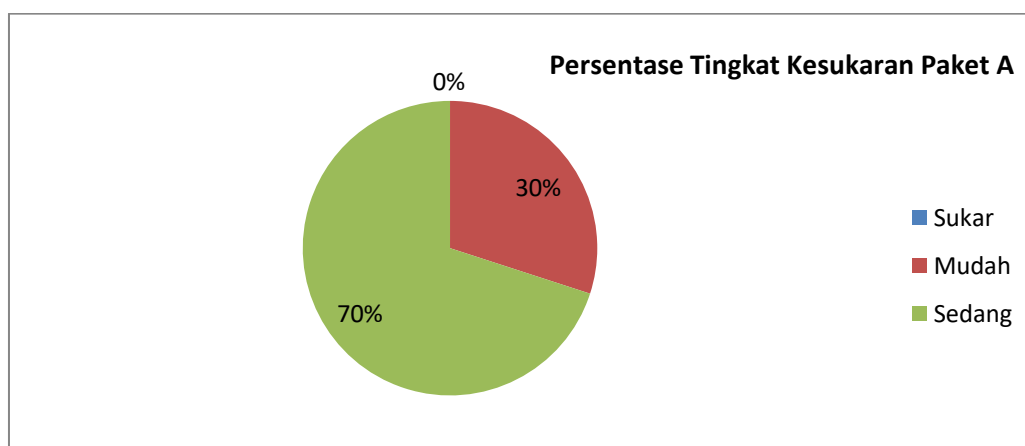
Validitas soal paket B memiliki validitas soal yang sangat baik. Hal ini membuktikan bahwa terdapat 8 butir soal (80%) yang dinyatakan valid dan 2 butir soal (20%) dinyatakan tidak valid. Soal yang valid terdiri dari butir soal nomor 1,2,3,6,7,8,9,10 karena pada perhitungan menggunakan rumus dan SPSS menunjukkan $r_{hitung} \geq r_{tabel}$. Peninjauan data dari bentuk soal pilihan ganda, butir soal yang valid ditemukan sesuai dengan capaian pembelajaran dan indikator pembelajaran serta untuk option yang digunakan relevan dengan

soal dan tidak mudah ditebak oleh siswa. Pada soal yang tidak valid terdapat pada butir soal nomor soal 4 dan 5 karena pada perhitungan dengan menggunakan rumus dan SPSS menunjukkan $r_{hitung} < r_{tabel}$. Peninjauan data dari bentuk soal pilihan ganda, butir soal yang tidak valid ditemukan ketidaksesuaian dengan capaian pembelajaran dan indikator pembelajaran serta untuk option yang digunakan tidak relevan dengan soal dan mudah ditebak oleh siswa.



Gambar 3. Reabilitas butir soal

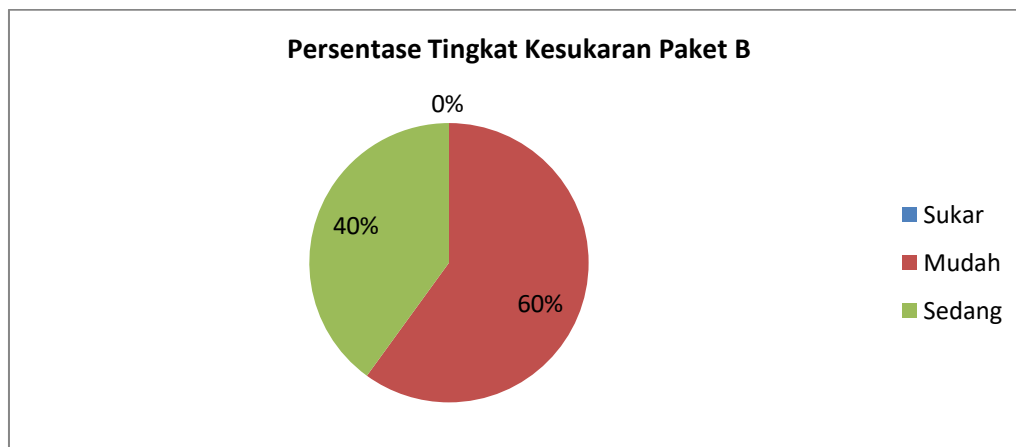
Hasil reliabilitas soal Ujian Akhir Semester Ganjil pada mata pelajaran Bahasa Indonesia kelas XI program keahlian Akuntansi Keuangan Lembaga SMK Negeri 2 Padang tahun ajaran 2024/2025 yang berbentuk pilihan ganda dengan nilai r_{11} 0,345 pada soal paket A yang termasuk dalam kategori reliabilitas rendah. Hal ini peneliti telah melakukan pengujian ulang terhadap soal paket A pada waktu yang berbeda dengan responden yang sama yaitu kelas XI AK 1 ternyata hasil penelitian tidak dapat diandalkan dan tidak dapat memberi informasi yang akurat dengan hasil penelitian yang tidak tetap atau konsisten. Pada soal paket B dengan nilai r_{11} 0,743 yang termasuk dalam kategori reliabilitas tinggi. Hal ini peneliti telah melakukan pengujian ulang terhadap soal paket B pada waktu yang berbeda dengan responden yang sama yaitu kelas XI AK 1 memberikan hasil yang konsisten dan stabil yang dapat memberikan informasi yang akurat.



Gambar 4. Persentase tingkat kesukaran paket A

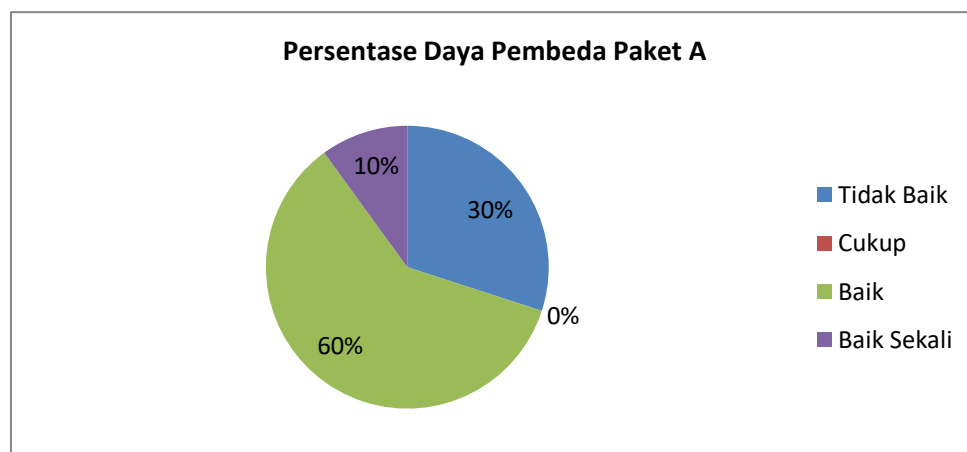
Berdasarkan hasil analisis pada soal paket A tingkat kesukaran didapatkan soal yang tergolong sukar berjumlah 0 butir (0%) atau tidak adanya ditemukan butir soal yang sukar pada soal paket A. Hal ini dibuktikan dengan sebagian besar peserta didik dapat menjawab soal dengan benar. Pada soal yang tergolong sedang berjumlah 7 butir (70%), yang terdapat

pada butir soal nomor 2,3,4,5,6,9,10. Peninjauan data pada butir soal yang tergolong sedang dapat dilihat dari bentuk butir soal yang lebih berfokus pada keterampilan berfikir menggunakan logika dengan mengasosiasi, mengurutkan, mengelompokkan, menghubungkan dan menyimpulkan yang bertujuan dengan pengetahuan dan pemahaman peserta didik. Soal yang tergolong mudah berjumlah 3 butir (30%) yang terdapat pada butir soal nomor 1,7,8. Peninjauan data pada butir soal yang tergolong mudah dapat dilihat dari bentuk butir soal yang tidak jauh beda dari materi yang sudah diajarkan dengan bentuk butir soal yang peserta didik mampu meniru, mengikuti, mengingat, dan mengidentifikasi.



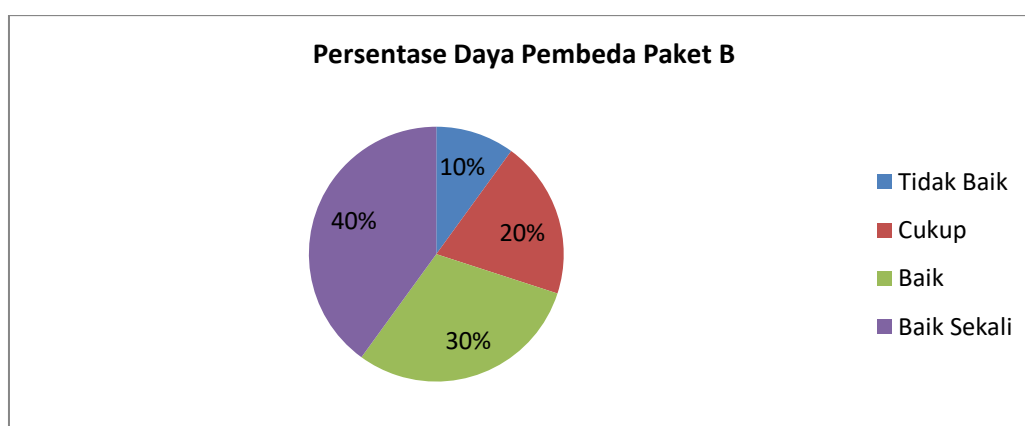
Gambar 5. Persentase tingkat kesukaran paket B

Pada soal paket B tingkat kesukaran soal yang tergolong sukar berjumlah 0 butir (0%) atau tidak adanya ditemukan butir soal yang sukar pada soal paket B. Hal ini dibuktikan dengan sebagian besar peserta didik dapat menjawab soal dengan benar. Soal yang tergolong sedang berjumlah 4 butir (40%) yang terdapat pada butir soal nomor 2,7,9,10. Peninjauan data pada butir soal yang tergolong sedang dapat dilihat dari bentuk butir soal yang lebih berfokus pada keterampilan berfikir menggunakan logika dengan mengasosiasi, mengurutkan, mengelompokkan, menghubungkan dan menyimpulkan yang bertujuan dengan pengetahuan dan pemahaman peserta didik. Soal yang tergolong mudah berjumlah 6 butir (60%) yang terdapat pada butir soal nomor 1,3,4, 5,6,8. Peninjauan data pada butir soal yang tergolong mudah dapat dilihat dari bentuk butir soal yang tidak jauh beda dari materi yang sudah diajarkan dengan bentuk butir soal yang peserta didik mampu meniru, mengikuti, mengingat, dan mengidentifikasi.



Gambar 6. Persentase daya pembeda paket A

Pada soal paket A ada 3 butir soal (30%) yang kategori tidak baik pada soal paket A yang terdapat pada butir soal 2,5,9. Hal ini ditinjau dari data soal yang peneliti miliki yaitu pada butir soal tersebut tergolong ada yang sangat mudah dan ada yang sangat sulit dan kurangnya relevansi soal dengan materi serta kesalahan dalam penyusunan soal yang menyebabkan butir soal tidak efektifnya dalam membedakan kemampuan peserta didik. Soal dengan kategori cukup terdapat 0 butir soal (0%) atau tidak ditemukan soal dengan kategori cukup. Untuk soal dengan kategori baik terdapat 6 butir soal (60%) yaitu pada butir soal 1,3,4,6,7,8. Butir soal tersebut dikatakan baik karena relevansinya soal dengan materi serta penyusunan soal dengan benar yang dapat membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. Pada kategori soal baik sekali ada 1 butir soal (10%) yaitu pada butir soal nomor 10. Dapat dilihat pada butir soal nomor 10 menunjukkan indeks daya pembeda yang tinggi yakni 0,71. Berikut ini persentase daya pembeda paket B.



Gambar 7. Persentase daya pembeda paket B

Pada soal paket B ada 1 butir soal (10%) yang terdapat pada butir soal nomor 4 termasuk kategori tidak baik. Hal ini ditinjau dari data soal yang peneliti miliki yaitu kurangnya relevansi soal dengan materi serta kesalahan dalam penyusunan soal yang menyebabkan butir soal tidak efektif dalam membedakan kemampuan peserta didik. Ada 2 butir soal (20%) termasuk kategori cukup yang terdapat pada butir soal nomor 5 dan 10. Hal ini ditinjau dari data soal yang peneliti miliki yaitu soal masih membutuhkan perbaikan yang lebih baik lagi seperti peserta didik ambigu dalam pertanyaan atau pilihan jawaban pada butir soal tersebut. Ada 3 butir soal (30%) yaitu butir soal nomor 1,9,3 termasuk kategori baik. Butir soal tersebut dikatakan baik karena relevansinya soal dengan materi serta penyusunan soal dengan benar. Soal dengan kategori baik sekali ada 4 butir soal (40%) yaitu terdapat pada butir soal nomor 2,6,7,8, dapat dikatakan baik sekali karena butir soal tersebut sudah mampu membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. Berikut ini adalah hasil uji efektifitas pengecoh soal paket A.

Tabel 1. Efektivitas Pengecoh Soal Dalam Bentuk Data Paket A

ITEMS	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
A	-	200%	-	200%	702%	-	0%	160%	0%	1228%
B	175%	-	67%	200%	351%	150%	400%	-	67%	0%
C	702%	133%	133%	0%	-	50%	0%	80%	267%	0%
D	175%	67%	0%	0%	175%	150%	-	80%	0%	-
E	175%	0%	200%	-	0%	50%	0%	80%	-	0%

Hasil analisis menunjukkan efektivitas pengecoh Paket A berada pada kategori tinggi karena seluruh butir sudah memiliki distraktor yang berfungsi: 1 butir (10%) berkategori sangat baik (nomor 7) dan 9 butir (90%) berkategori baik (nomor 1, 2, 3, 4, 5, 6, 8, 9, 10), tanpa adanya distraktor kurang baik, jelek, atau sangat jelek. Secara psikometrik, distraktor yang berfungsi menekan peluang menebak, memaksa peserta didik melakukan proses pemilihan berbasis pemahaman materi, serta memperkuat kualitas interpretasi skor karena opsi salah benar-benar “bekerja” sebagai pengalih jawaban yang masuk akal bagi peserta didik yang belum menguasai materi (Rezigalla, 2024; Sahoo & Singh, 2017).

Dominasi kategori “baik” (90%) menunjukkan sebagian besar butir sudah memenuhi prinsip konstruksi distraktor: opsi salah relevan dengan pokok soal, memiliki kedekatan makna dengan kunci, dan cukup logis sehingga dipilih oleh sebagian peserta tes. Kondisi seperti ini penting karena efisiensi distraktor berkaitan dengan perilaku butir lainnya; penelitian Rezigalla (2024) menunjukkan efisiensi distraktor berdampak pada indeks kesukaran dan daya pembeda, sehingga distraktor yang fungsional cenderung membuat butir lebih bermakna untuk membedakan kemampuan peserta didik (Rezigalla, 2024).

Temuan Paket A ini juga lebih baik dibanding pola masalah yang sering muncul pada evaluasi tes sekolah, khususnya pada tes bahasa, yaitu banyak distraktor yang tidak berfungsi karena jarang dipilih atau mudah dieliminasi. Himawan dkk. (2024) pada studi analisis tes bahasa di konteks sekolah Indonesia melaporkan proporsi distraktor yang tidak berfungsi masih tinggi pada banyak butir, dan menegaskan perlunya analisis butir untuk memutuskan butir yang dipertahankan, direvisi, atau dibuang (Himawan et al., 2024). Dengan hasil Paket A yang seluruhnya minimal “baik”, kualitas distraktor relatif terkendali; fokus perbaikan instrumen (jika ada) lebih tepat diarahkan ke komponen lain seperti reliabilitas, kesukaran, atau daya pembeda, bukan pada distraktor.

Konsistensi hasil yang diperoleh melalui Excel juga sejalan dengan literatur analisis butir berbasis CTT yang menunjukkan berbagai perangkat (termasuk aplikasi dan tabulasi) umumnya menghasilkan informasi yang sepadan untuk parameter kesukaran, daya pembeda, dan efektivitas pengecoh, selama data respons ditabulasi dengan benar (Nurjanah et al., 2024). Berikut ini adalah hasil uji efektifitas pengecoh soal paket B.

Tabel 2. Efektivitas Pengecoh Soal Dalam Bentuk Data Paket B

ITEMS	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
A	240%	0%	-	-	0%	100%	400%	133%	0%	526%
B	80%	0%	200%	160%	0%	-	0%	133%	0%	-
C	-	160%	0%	80%	25%	100%	-	133%	400%	351%
D	80%	-	100%	80%	0%	0%	0%	-	-	0%
E	80%	880%	100%	80%	-	0%	0%	1867%	450%	0%

Berdasarkan hasil analisis efektivitas pengecoh pada soal Paket B menunjukkan 9 dari 10 butir (90%) berada pada kategori baik–sangat baik, dengan 3 butir (30%) berkategori sangat baik (nomor 3, 6, 8) dan 6 butir (60%) berkategori baik (nomor 1, 2, 4, 7, 9, 10). Secara teoritis, pengecoh dinilai efektif apabila opsi salah tampak masuk akal, relevan dengan pokok soal, dan benar-benar dipilih oleh sebagian peserta tes (terutama kelompok berkemampuan

lebih rendah), sehingga menekan peluang menebak dan memperkuat fungsi pengukuran butir (Rezigalla, 2024; Sahoo & Singh, 2017).

Tiga butir dengan pengecoh sangat baik mengindikasikan bahwa opsi-opsi salah pada nomor 3, 6, dan 8 bekerja optimal sebagai distraktor: alternatif salah cukup kredibel untuk “bersaing” dengan kunci, sehingga respons peserta didik tersebar dan butir menjadi lebih bermakna secara psikometrik. Temuan seperti ini penting karena kualitas pengecoh berkaitan dengan kualitas butir secara keseluruhan; distraktor yang fungsional cenderung mendukung tingkat kesukaran yang proporsional dan daya pembeda yang lebih stabil dibanding butir yang memiliki distraktor tidak berfungsi (Rezigalla, 2024).

Enam butir dengan kategori baik menandakan distraktor pada nomor 1, 2, 4, 7, 9, dan 10 pada umumnya sudah logis dan menarik, tetapi biasanya masih ada satu distraktor yang kurang “menggoda” dibanding distraktor lain (misalnya jarang dipilih atau terlalu lemah dibanding opsi salah lainnya). Pola dominasi kategori “baik” juga sering muncul pada analisis butir di konteks asesmen sekolah, di mana sebagian besar distraktor bekerja tetapi tetap ada butir yang perlu penyempurnaan untuk meningkatkan kualitas pilihan jawaban (Firdaus et al., 2025; Himawan et al., 2024).

Satu butir (nomor 5) berkategori pengecoh jelek (10%) menunjukkan adanya distraktor tidak berfungsi, yaitu opsi salah tidak dipilih atau tidak menarik sehingga peserta didik mudah menyingkirkan opsi tersebut. Distraktor yang tidak berfungsi membuat butir cenderung menjadi lebih mudah, meningkatkan peluang menjawab benar dengan eliminasi, dan pada banyak kasus menurunkan kemampuan butir membedakan siswa berkemampuan tinggi dan rendah (Rezigalla, 2024; Himawan et al., 2024). Temuan serupa dilaporkan pada analisis PAS Bahasa Indonesia yang menunjukkan sebagian distraktor tidak berfungsi dan direkomendasikan untuk direvisi, diganti, atau dibuang (Pohan & Nurrahmah, 2025).

Implikasinya, Paket B sudah cukup kuat pada aspek distraktor (90% baik–sangat baik), tetapi perbaikan harus dipusatkan pada butir nomor 5. Revisi butir 5 dilakukan dengan menyusun distraktor yang berbasis miskonsepsi umum materi, menjaga homogenitas bentuk opsi (panjang, struktur gramatikal, dan kategori jawaban), menghindari petunjuk jawaban (kata kunci yang menonjol, opsi paling panjang, atau pola tata bahasa yang berbeda), lalu melakukan uji coba ulang dan analisis ulang agar fungsi distraktor membaik sebelum butir dimasukkan ke bank soal (Sahoo & Singh, 2017; Nurjanah et al., 2024).

KESIMPULAN

Berdasarkan hasil penelitian diatas dapat disimpulkan:

1. Dari 10 butir soal, paket A memiliki 7 butir valid (70%) dan paket B 8 butir valid (80%), sehingga validitas kedua paket tergolong baik.
2. Reliabilitas paket A rendah/tidak reliabel ($0,345 < 0,60$), sedangkan paket B reliabel tinggi ($0,743 > 0,60$).
3. Tingkat kesukaran paket A didominasi kategori sedang (7 butir/70%) sehingga tergolong baik, sedangkan paket B didominasi butir mudah (6 butir/60%) sehingga kurang seimbang dari sisi kesukaran.
4. Daya pembeda tergolong baik pada sebagian besar butir: paket A 6 butir (60%) dan paket B 7 butir (70%) sudah mampu membedakan siswa yang menguasai materi dan yang belum.
5. Efektivitas pengecoh umumnya baik: paket A seluruh butir baik–sangat baik (10% sangat baik; 90% baik), sedangkan paket B mayoritas baik–sangat baik (30% sangat baik; 60% baik) namun terdapat 1 butir dengan pengecoh jelek (10%).

DAFTAR PUSTAKA (APA)

- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (Edisi 3). Bumi Aksara.
- Dzahabiyah, S., Ghozali, I., & Putra, Z. A. W. (2024). Analisis kualitas butir soal ulangan akhir semester mata pelajaran Seni Budaya di SMA. *Edukatif: Jurnal Ilmu Pendidikan*, 6(3), 2351–2359. <https://doi.org/10.31004/edukatif.v6i3.6797>
- Himawan, R., Hermanto, Nurgiyantoro, B., Suyono, Widyartono, D., Purwanti, A., Yujing, L., & Pogadaev, V. A. (2024). ITEMAN-based evaluation of end-of-semester assessment items: A case study of language test in Indonesian school context. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 6(3), 387–402. <https://doi.org/10.23917/ijolae.v6i3.23254>
- Hariato, E. (2020). Keterampilan membaca dalam pembelajaran bahasa. *Didaktika: Jurnal Kependidikan*, 9(1), 1–8. <https://doi.org/10.58230/27454312.2>
- Mertasari, N. M. S. (2022). Summative evaluation of ICT-based learning media. *Journal of Education Research and Evaluation*, 6(4), 688–695. <https://doi.org/10.23887/jere.v6i4.54695>
- Pohan, S. N. A., & Nurrahmah. (2025). Analisis butir soal penilaian akhir semester (PAS) Bahasa Indonesia kelas VIII SMP Negeri 18 Banda Aceh 2023/2024. *Jurnal Inovasi Pendidikan Nusantara*, 1(2).
- Firdaus, W. F., Delawanti, D., & Yulianti. (2025). Analisis butir soal pilihan ganda penilaian akhir semester pada pembelajaran tema 8 Praja Muda Karana kelas III. *Jurnal Sadewa: Publikasi Ilmu Pendidikan, Pembelajaran dan Ilmu Sosial*, 3(2), 129–151. <https://doi.org/10.61132/sadewa.v3i2.1727>
- Himawan, R., Hermanto, Nurgiyantoro, B., Suyono, Widyartono, D., Purwanti, A., Yujing, L., & Pogadaev, V. A. (2024). ITEMAN-based evaluation of end-of-semester assessment items: A case study of language test in Indonesian school context. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 6(3), 387–402. <https://doi.org/10.23917/ijolae.v6i3.23254>
- Nurjanah, S., Iqbal, M., Zafrullah, Z., Mahmud, M. N., Seran, D. S. F., Suardi, I. K., & Arriza, L. (2024). Psychometric quality of multiple-choice tests under classical test theory (CTT): AnBuso, Iteman, and R. *Jurnal Penelitian dan Evaluasi Pendidikan*, 28(2), 161–172. <https://doi.org/10.21831/pep.v28i2.71542>
- Pohan, S. N. A., & Nurrahmah. (2025). Analisis butir soal penilaian akhir semester (PAS) Bahasa Indonesia kelas VIII SMP Negeri 18 Banda Aceh 2023/2024. *Jurnal Inovasi Pendidikan Nusantara*, 1(2).
- Rezigalla, A. A. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24, 445. <https://doi.org/10.1186/s12909-024-05433-y>
- Sahoo, D. P., & Singh, R. (2017). Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. *International Journal of Research in Medical Sciences*, 5(12), 5351–5355. <https://doi.org/10.18203/2320-6012.ijrms20175453>
- Himawan, R., Hermanto, Nurgiyantoro, B., Suyono, Widyartono, D., Purwanti, A., Yujing, L., & Pogadaev, V. A. (2024). ITEMAN-Based Evaluation of End-of-Semester Assessment Items: A Case Study of Language Test in Indonesian School Context. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 6(3), 387–402. <https://doi.org/10.23917/ijolae.v6i3.23254>

- Nurjanah, S., Iqbal, M., Zafrullah, Z., Mahmud, M. N., Seran, D. S. F., Suardi, I. K., & Arriza, L. (2024). Psychometric quality of multiple-choice tests under classical test theory (CTT): AnBuso, Iteman, and R. *Jurnal Penelitian dan Evaluasi Pendidikan*, 28(2), 161–172. <https://doi.org/10.21831/pep.v28i2.71542>
- Rezigalla, A. A. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24, 445. <https://doi.org/10.1186/s12909-024-05433-y>
- Sahoo, D. P., & Singh, R. (2017). Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. *International Journal of Research in Medical Sciences*, 5(12), 5351–5355. <https://doi.org/10.18203/2320-6012.ijrms20175453>