

Analisis Kualitas Butir Soal dan Hasil Belajar Bahasa Inggris Menggunakan Anates pada Siswa Kelas VII SMPN 1 SOOKO

Vivin Muji Rahayu*, Rofi'i, Sabariah

Universitas PGRI Adi Buana Surabaya, Surabaya, Indonesia

*Corresponding Author: vmujirahayu@gmail.com

Dikirim: 02-06-2026; Direvisi: 16-06-2026; Diterima: 18-06-2026

Abstrak: Penelitian ini bertujuan untuk menganalisis kualitas butir soal dan hasil belajar Bahasa Inggris menggunakan program Anates pada siswa kelas VII SMPN 1 Sooko. Analisis kualitas butir soal menjadi aspek penting dalam evaluasi pembelajaran karena menentukan tingkat validitas dan reliabilitas instrumen yang digunakan untuk mengukur pencapaian kompetensi peserta didik. Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis dokumen terhadap hasil tes Bahasa Inggris siswa kelas VII yang terdiri dari 32 siswa sebagai subjek dan instrumen berupa 25 butir soal pilihan ganda. Analisis dilakukan menggunakan program Anates versi 4.0.2 berbasis Teori Tes Klasik. Hasil menunjukkan bahwa instrumen memiliki reliabilitas sangat tinggi ($r = 0,88$), namun distribusi tingkat kesukaran belum ideal dengan dominasi butir mudah hingga sangat mudah (56%). Daya pembeda 76% butir tergolong baik hingga sangat baik, sementara 64% butir memiliki korelasi skor butir-total yang signifikan. Efektivitas pengecoh menunjukkan 53,3% pengecoh berfungsi baik, sedangkan sisanya perlu diperbaiki. Sebanyak 14 butir layak dipertahankan, 10 butir perlu direvisi, dan 1 butir disarankan diganti. Penelitian ini berkontribusi dengan menyediakan profil psikometrik spesifik soal Bahasa Inggris SMP dan menunjukkan bahwa reliabilitas tinggi tidak selalu beriringan dengan distribusi kesukaran yang ideal, sehingga evaluasi instrumen harus dilakukan secara multidimensi.

Kata Kunci: analisis butir soal; Anates; Bahasa Inggris SMP; evaluasi pembelajaran.

Abstract: This study aims to analyze the quality of English question items and learning outcomes using the Anates program in grade VII students of SMPN 1 Sooko. Analysis of the quality of question items is an important aspect in learning evaluation because it determines the level of validity and reliability of the instruments used to measure the achievement of students' competencies. This study uses a descriptive quantitative approach with a document analysis method on the results of the English test of grade VII students consisting of 32 students as subjects and instruments in the form of 25 multiple-choice questions. The analysis was carried out using the Anates program version 4.0.2 based on Classical Test Theory. The results showed that the instrument had very high reliability ($r = 0.88$), but the distribution of difficulty levels was not ideal with the dominance of easy to very easy items (56%). The differentiating power of 76% of the grains was good to very good, while 64% of the grains had a significant grain-total score correlation. The effectiveness of the trickster showed that 53.3% of the tricksters were functioning well, while the rest needed to be improved. A total of 14 items are worth keeping, 10 items need to be revised, and 1 item is recommended to be replaced. This study contributes by providing a specific psychometric profile of junior high school English and shows that high reliability does not always go hand in hand with an ideal distribution of difficulty, so the evaluation of the instrument must be carried out in a multidimensional manner.

Keywords: question item analysis; Anates; Junior High School English; learning evaluation.

PENDAHULUAN

Evaluasi pembelajaran merupakan komponen integral dalam sistem pendidikan yang tidak dapat dipisahkan dari proses belajar-mengajar. Melalui evaluasi, guru memperoleh informasi yang akurat tentang sejauh mana tujuan pembelajaran telah tercapai, sekaligus mendapatkan umpan balik yang diperlukan untuk memperbaiki kualitas pembelajaran berikutnya. Dalam konteks pendidikan formal di Indonesia, salah satu bentuk evaluasi yang paling sering digunakan adalah tes hasil belajar berbentuk pilihan ganda. Soal pilihan ganda dipilih karena kemampuannya mencakup materi secara luas, objektivitas penilaian yang tinggi, serta kemudahan dalam proses skoring. Meski demikian, kualitas soal pilihan ganda tidak dapat diasumsikan baik hanya karena telah disusun oleh guru yang berpengalaman. Tanpa melalui proses analisis yang sistematis dan terstandar, instrumen tes berisiko menghasilkan data yang tidak sah dan tidak dapat diandalkan sebagai dasar pengambilan keputusan pembelajaran (Fauziah & Raharjo, 2022).

Pentingnya analisis butir soal semakin mendapat perhatian seiring dengan tuntutan sistem penilaian yang lebih akuntabel dan berbasis bukti. Analisis butir soal merupakan proses pengkajian kualitas setiap butir soal secara empiris, mencakup aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Kelima aspek tersebut saling berkaitan dan secara kolektif menentukan apakah suatu instrumen tes layak digunakan sebagai alat ukur kompetensi siswa. Menurut (Fauzi & Lestari, 2022), analisis butir soal tidak hanya mempermudah proses evaluasi, tetapi juga membantu guru mengidentifikasi kelemahan dalam soal yang telah disusun sehingga akurasi hasil pengukuran dapat ditingkatkan. Hal senada dikemukakan oleh (Mustikasari & Akhdinirwanto, 2025) yang menyatakan bahwa evaluasi kualitas instrumen tes melalui analisis item merupakan langkah kritis untuk memastikan bahwa tes benar-benar mengukur apa yang dimaksudkan untuk diukur, bukan sekadar kemampuan siswa menebak jawaban.

Di lapangan, kenyataan menunjukkan bahwa masih banyak soal yang digunakan dalam proses penilaian belum pernah dianalisis secara psikometrik. Penelitian (Agustin et al., 2024) pada soal ujian di SMP Negeri di Kabupaten Rembang menemukan bahwa dari keseluruhan butir yang dianalisis, sebagian besar masuk kategori mudah dengan indeks daya pembeda yang rendah, sehingga tidak efektif dalam membedakan siswa berkemampuan tinggi dan rendah. Temuan serupa dilaporkan dalam penelitian (Prayitasari et al., 2024) yang mengkaji soal pilihan ganda di jenjang SMK, di mana sebagian butir soal terindikasi memiliki pengecoh yang tidak berfungsi dengan baik. Kondisi ini menggambarkan bahwa tanpa analisis yang memadai, kualitas instrumen tes di lapangan cenderung tidak terjaga, yang pada akhirnya berimplikasi pada rendahnya validitas simpulan hasil belajar siswa.

Mata pelajaran Bahasa Inggris di jenjang Sekolah Menengah Pertama menempati posisi strategis dalam kerangka kurikulum nasional. Sebagai salah satu mata pelajaran yang diujikan secara berkelanjutan, kualitas instrumen penilaian Bahasa Inggris perlu mendapat perhatian serius. Siswa yang memperoleh nilai tinggi seharusnya benar-benar mencerminkan penguasaan kompetensi berbahasa, bukan semata-mata karena soal yang terlalu mudah atau pengecoh yang tidak berfungsi. Dalam konteks ini, analisis butir soal menjadi alat penting bagi guru Bahasa Inggris untuk memastikan bahwa instrumen yang mereka gunakan mampu menghasilkan data yang tepercaya. Penelitian (Yasinta, 2021) yang mengkaji soal Bahasa Inggris



menggunakan aplikasi AnBuSo menemukan bahwa 83% butir soal berada pada kategori kesukaran sedang dan 73% memiliki indeks daya pembeda yang baik, hasil yang mengindikasikan bahwa analisis sistematis mampu mengungkap kualitas instrumen secara komprehensif.

Untuk menunjang efisiensi proses analisis, berbagai perangkat lunak psikometrik telah dikembangkan dan dimanfaatkan secara luas oleh para pendidik di Indonesia. Salah satu yang paling banyak digunakan adalah program Anates, yang dikembangkan oleh Karno To dan Yudi Wibisono dari Universitas Pendidikan Indonesia. Program ini dirancang khusus untuk kebutuhan analisis butir soal di lingkungan pendidikan Indonesia, dengan antarmuka berbahasa Indonesia sehingga relatif mudah dioperasikan oleh guru tanpa latar belakang statistik yang mendalam. (Zukhruffah & Fradana, 2025) mencatat bahwa Anates versi 4.0 mampu menganalisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh secara terintegrasi dalam satu platform, menjadikannya pilihan yang praktis dan efisien dibandingkan perhitungan manual atau penggunaan perangkat umum seperti Microsoft Excel. (Alista & Syahzanani, 2023) juga melaporkan bahwa penggunaan Anates pada soal pilihan ganda kelas VII SMP memberikan hasil analisis yang komprehensif dan mudah diinterpretasikan oleh guru.

Penelitian ini dilaksanakan di kelas VII-H SMPN 1 Sooko dengan melibatkan 32 siswa sebagai subjek. Instrumen yang dianalisis adalah soal pilihan ganda Bahasa Inggris yang terdiri dari 25 butir, dengan empat alternatif pilihan jawaban (A, B, C, D). Data skor mentah diproses menggunakan Anates versi 4.0.2 untuk menghasilkan informasi psikometrik yang mencakup tingkat kesukaran, daya pembeda, reliabilitas tes, korelasi skor butir dengan skor total, serta kualitas pengecoh. Pemilihan lokasi penelitian didasarkan pada pertimbangan bahwa instrumen penilaian yang digunakan di sekolah tersebut belum pernah dianalisis secara formal, sehingga guru belum memiliki informasi yang cukup mengenai kualitas soal yang selama ini digunakan untuk mengukur hasil belajar siswanya.

Berdasarkan pemaparan di atas, terdapat kesenjangan antara kebutuhan akan instrumen penilaian yang berkualitas dengan realitas penggunaan soal yang belum tervalidasi secara empiris di tingkat sekolah menengah pertama, khususnya pada mata pelajaran Bahasa Inggris. Penelitian ini bertujuan untuk: (1) mendeskripsikan tingkat kesukaran butir soal Bahasa Inggris kelas VII-H SMPN 1 Sooko; (2) mendeskripsikan daya pembeda butir soal; (3) menganalisis reliabilitas instrumen; (4) mengkaji korelasi skor butir dengan skor total; serta (5) mengevaluasi efektivitas pengecoh pada setiap butir soal.

METODE PENELITIAN

Penelitian ini merupakan penelitian deskriptif kuantitatif, yaitu suatu pendekatan yang bertujuan mendeskripsikan fenomena secara sistematis berdasarkan data numerik tanpa melakukan manipulasi variabel (Chan et al., 2020). Pendekatan ini dipilih karena penelitian bertujuan menggambarkan kualitas psikometrik butir soal pilihan ganda Bahasa Inggris kelas VII-H SMPN 1 Sooko secara objektif dan terukur, mencakup tingkat kesukaran, daya pembeda, reliabilitas, korelasi skor butir dengan skor total, serta efektivitas pengecoh. Dalam konteks evaluasi pembelajaran, penelitian deskriptif kuantitatif relevan digunakan karena menghasilkan informasi



empiris yang dapat langsung dimanfaatkan sebagai dasar pengambilan keputusan perbaikan instrumen (Alista & Syahzanani, 2023).

Sumber data dalam penelitian ini terbagi menjadi dua, yaitu data primer dan data sekunder. Data primer diperoleh langsung dari lembar jawaban siswa kelas VII-H SMPN 1 Sooko yang mengikuti tes Bahasa Inggris, dengan jumlah subjek sebanyak 32 siswa. Pemilihan kelas VII-H sebagai sumber data didasarkan pada pertimbangan bahwa instrumen penilaian yang digunakan di kelas tersebut belum pernah dianalisis secara psikometrik, sehingga guru belum memiliki informasi yang memadai mengenai kualitas soal yang selama ini digunakan. Instrumen yang dianalisis adalah soal pilihan ganda yang terdiri dari 25 butir dengan empat alternatif pilihan jawaban (A, B, C, D). Data sekunder bersumber dari dokumen kunci jawaban resmi yang ditetapkan oleh guru pengampu mata pelajaran. Jumlah subjek sebanyak 32 siswa dinilai memadai untuk analisis berbasis Teori Tes Klasik (CTT), mengingat pendekatan ini tidak mensyaratkan sampel besar sebagaimana yang dibutuhkan dalam Teori Respons Butir (Ambarwati & Ismiyati, 2021).

Pengumpulan data dilakukan melalui teknik dokumentasi, yaitu dengan menghimpun lembar jawaban siswa yang telah terisi beserta kunci jawaban dari guru. Teknik dokumentasi dipilih karena data yang dibutuhkan dalam analisis butir soal bersifat archival, yakni berupa rekaman respons peserta tes yang sudah ada sebagai bagian dari proses penilaian reguler di sekolah (Ikawati et al., 2024). Seluruh lembar jawaban siswa diinput ke dalam program Anates versi 4.0.2 dengan format entri data mentah (*raw data*), di mana setiap jawaban benar diberi skor 1 dan jawaban salah diberi skor 0. Proses entri data dilakukan secara cermat untuk meminimalkan kesalahan input yang dapat memengaruhi akurasi hasil analisis.

Analisis data dilakukan menggunakan program Anates versi 4.0.2, perangkat lunak analisis butir soal berbasis antarmuka Bahasa Indonesia yang dikembangkan oleh Karno To dan Yudi Wibisono dari Universitas Pendidikan Indonesia. Program ini dipilih karena kemampuannya menghasilkan seluruh parameter psikometrik yang dibutuhkan secara terintegrasi dalam satu proses analisis (Ramli et al., 2021). Terdapat lima parameter yang dianalisis, sebagai berikut.

Pertama, reliabilitas tes dihitung menggunakan teknik belah dua (*split-half*) yang dikoreksi dengan formula Spearman-Brown. Koefisien reliabilitas diinterpretasikan berdasarkan kriteria: sangat tinggi (0,80–1,00), tinggi (0,60–0,79), sedang (0,40–0,59), rendah (0,20–0,39), dan sangat rendah ($< 0,20$) (Marianti, 2025). Kedua, tingkat kesukaran (*difficulty index*) dihitung berdasarkan proporsi jawaban benar dari seluruh peserta tes. Kategori interpretasi mengacu pada kriteria: sangat mudah ($> 85\%$), mudah (70–85%), sedang (40–70%), sukar (15–40%), dan sangat sukar ($< 15\%$) (Suryadin, 2020). Ketiga, daya pembeda (*discrimination index*) dianalisis dengan membandingkan proporsi jawaban benar antara kelompok unggul (27% skor tertinggi) dan kelompok asor (27% skor terendah). Nilai daya pembeda diinterpretasikan dengan kriteria: sangat baik ($> 70\%$), baik (40–70%), cukup (20–40%), dan perlu direvisi ($< 20\%$) (Asyaroh et al., 2025). Keempat, korelasi skor butir dengan skor total dihitung menggunakan koefisien korelasi product-moment Pearson untuk mengukur validitas empiris setiap butir soal. Butir dinyatakan valid apabila nilai korelasi berada pada taraf signifikansi $p < 0,05$ atau $p < 0,01$ (Mudiyantri et al., 2025). Kelima, efektivitas pengecoh dianalisis berdasarkan proporsi peserta yang memilih setiap opsi jawaban yang salah. Pengecoh dinyatakan berfungsi baik apabila dipilih oleh minimal 5% peserta tes dari kelompok yang tidak menjawab benar



(Mustaqim & Sulisti, 2024). Kelima hasil analisis tersebut kemudian diinterpretasikan secara deskriptif untuk menghasilkan profil kualitas instrumen secara menyeluruh.

HASIL DAN PEMBAHASAN

Hasil dan pembahasan penelitian berdasarakan analisis Anates dibagi menjadi beberapa hal yakni sebagai berikut:

A. Reliabilitas Instrumen

Hasil analisis menggunakan program Anates versi 4.0.2 menunjukkan bahwa instrumen tes Bahasa Inggris kelas VII-H SMPN 1 Sooko memiliki koefisien reliabilitas sebesar 0,88 dengan korelasi antarbelahan 0,79, rata-rata skor 16,47, dan simpangan baku 5,16. Berdasarkan kriteria klasifikasi reliabilitas, nilai 0,88 masuk dalam kategori sangat tinggi, yang berarti instrumen ini memiliki konsistensi internal yang sangat baik dan layak digunakan sebagai alat ukur hasil belajar siswa. Temuan ini sejalan dengan pandangan (Avianti et al., 2021) yang menegaskan bahwa instrumen tes dengan koefisien reliabilitas di atas 0,80 menunjukkan bahwa butir-butirnya secara kolektif menghasilkan pengukuran yang stabil dan dapat diandalkan lintas kondisi pengujian. Dalam perspektif Teori Tes Klasik, reliabilitas yang tinggi mengindikasikan bahwa proporsi varians skor sejati (*true score*) terhadap varians skor teramati cukup besar, sehingga kesalahan pengukuran acak relatif kecil (Pradani & Efendi, 2023). Nilai simpangan baku sebesar 5,16 dengan rentang skor 6 hingga 24 mengindikasikan adanya dispersi kemampuan yang memadai di antara siswa, kondisi yang secara statistik mendukung terbentuknya koefisien reliabilitas yang tinggi.

Jika dibandingkan dengan penelitian terdahulu, reliabilitas 0,88 dalam studi ini melampaui temuan (Ramadhani & Pratiwi, 2025) yang melaporkan reliabilitas 0,76 pada soal akuntansi SMK, dan lebih tinggi dari Mulyani et al. (2024) yang mendapatkan reliabilitas 0,81 pada soal menyimak di perguruan tinggi. Perbedaan ini dapat dijelaskan melalui karakteristik kelompok subjek: siswa kelas VII yang baru memasuki jenjang SMP memiliki rentang kemampuan berbahasa Inggris yang lebih lebar dibandingkan mahasiswa yang telah melalui seleksi masuk, sehingga variansi skor lebih besar dan secara matematis menghasilkan koefisien reliabilitas yang lebih tinggi. (Zaini et al., 2025) menguatkan argumen ini dengan menyatakan bahwa reliabilitas tes dipengaruhi secara signifikan oleh heterogenitas kemampuan kelompok yang dites; semakin beragam kemampuan peserta, semakin tinggi potensi koefisien reliabilitas yang dihasilkan. Kontribusi temuan ini bagi praktik evaluasi di SMPN 1 Sooko adalah bahwa guru dapat menggunakan skor dari instrumen ini dengan keyakinan bahwa perbedaan skor antarindividu mencerminkan perbedaan kemampuan nyata, bukan sekadar efek fluktuasi pengukuran. Namun demikian, reliabilitas yang tinggi tidak serta-merta menjamin validitas seluruh butir soal, sebagaimana ditunjukkan oleh hasil analisis korelasi skor butir yang dibahas pada bagian selanjutnya.

B. Tingkat Kesukaran Butir Soal

Hasil analisis tingkat kesukaran terhadap 25 butir soal disajikan pada Tabel 1.



Tabel 1. Distribusi Tingkat Kesukaran Butir Soal

Kategori	Jumlah Butir	Persentase	Nomor Butir
Sangat Mudah (>85%)	4	16%	1, 3, 5, 10
Mudah (70 – 85%)	10	40%	2, 6, 7, 11, 13, 14, 15, 16, 21, 23
Sedang (40 – 70%)	10	40%	4, 8, 9, 12, 17, 18, 19, 20, 22, 24
Sukar (15 – 40%)	0	0%	—
Sangat Sukar (<15%)	1	4%	25
Total	25	100%	

Distribusi tingkat kesukaran menunjukkan bahwa 56% butir soal berada pada kategori mudah hingga sangat mudah, 40% pada kategori sedang, dan hanya satu butir (nomor 25) yang masuk kategori sangat sukar dengan indeks 9,38%. Tidak terdapat butir pada kategori sukar. Dominasi butir kategori mudah mengindikasikan bahwa instrumen secara keseluruhan relatif tidak menantang bagi mayoritas siswa kelas VII-H. Temuan ini memiliki kemiripan pola dengan (Handayani & Nugroho, 2021) yang melaporkan dominasi butir mudah pada soal ujian SMP, namun berbeda dari (Rahardian et al., 2025) yang menemukan 83% butir pada kategori sedang dalam soal EFL untuk mahasiswa. Perbedaan ini sangat mungkin dipengaruhi oleh faktor jenjang dan pengalaman belajar subjek: siswa kelas VII yang baru beralih dari SD dengan bekal materi Bahasa Inggris dasar cenderung menghadapi soal yang dirancang untuk mengakomodasi transisi tersebut, sehingga soal terkesan lebih mudah. Hal ini diperkuat oleh (Wahyuni & Hartanto, 2023) yang menyatakan bahwa tingkat kesukaran instrumen tidak dapat dilepaskan dari konteks kurikulum dan karakteristik demografis peserta tes. Secara normatif, proporsi ideal distribusi tingkat kesukaran dalam sebuah instrumen tes yang baik adalah sekitar 25% mudah, 50% sedang, dan 25% sukar (Az-Zahra & Rochmawati, 2024). Mengacu pada standar ini, instrumen dalam penelitian ini belum memenuhi distribusi ideal karena terlalu banyak butir mudah dan tidak ada butir sukar. Kondisi tersebut berisiko menurunkan kemampuan tes dalam mendiferensiasi siswa berkemampuan tinggi secara lebih presisi, sebab siswa dengan berbagai level kemampuan cenderung sama-sama menjawab benar soal-soal mudah. Oleh karena itu, guru perlu merevisi atau mengganti butir-butir dengan kategori sangat mudah khususnya butir 1, 3, 5, dan 10 dengan butir yang menguji kompetensi pada level kognitif lebih tinggi, misalnya inferensi makna kontekstual atau analisis teks pendek, agar distribusi kesukaran mendekati proporsi yang lebih ideal. Butir nomor 25 yang masuk kategori sangat sukar (9,38%) perlu mendapat perhatian khusus. Dari keseluruhan 32 siswa, hanya 3 siswa yang menjawab benar butir ini. Jika dikombinasikan dengan hasil analisis daya pembeda dan korelasi skor butir yang juga rendah sebagaimana akan dibahas pada bagian berikutnya maka butir nomor 25 berpotensi mengukur konstruk yang berbeda atau memiliki kekeliruan dalam penulisan stem maupun kunci jawaban, sehingga memerlukan telaah substantif lebih lanjut oleh guru.

C. Daya Pembeda Butir Soal

Analisis daya pembeda dilakukan dengan membandingkan respons kelompok unggul (9 siswa dengan skor tertinggi, rentang 21–24) dan kelompok asor (9 siswa



dengan skor terendah, rentang 6–13). Distribusi daya pembeda disajikan pada Tabel 2.

Tabel 2. Distribusi Daya Pembeda Butir Soal

Kategori	Jumlah Butir	Persentase	Nomor Butir
Sangat Baik ($>70\%$)	5	20%	4, 6, 9, 19, 20
Baik (40–70%)	14	56%	1, 8, 10, 11, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24
Cukup (20–40%)	3	12%	2, 3, 5
Perlu Direvisi ($<20\%$)	3	12%	7, 11, 25
Total	25	100%	

Sebagian besar butir (76%) memiliki daya pembeda pada kategori baik hingga sangat baik, yang berarti instrumen secara keseluruhan cukup efektif membedakan siswa berkemampuan tinggi dari yang berkemampuan rendah. Butir nomor 4, 6, 9, 19, dan 20 tampil paling diskriminatif dengan indeks masing-masing di atas 70%. Sebaliknya, butir nomor 7, 11, dan 25 memiliki daya pembeda yang sangat rendah (sekitar 11,11%) dan termasuk kategori perlu direvisi. Hasil ini lebih baik dibandingkan temuan (Utami et al., 2024) yang melaporkan 42% butir dengan daya pembeda tidak baik pada soal di SMPN 2 Kanatang, dan sebanding dengan (Salsabilla et al., 2025) yang mencatat 67% butir berkategori baik pada soal akuntansi SMK. Tingginya proporsi butir dengan daya pembeda baik dalam penelitian ini dapat dikaitkan dengan reliabilitas instrumen yang tinggi (0,88), karena secara teoretis terdapat hubungan positif antara konsistensi internal instrumen dengan kemampuan diskriminatifnya: semakin homogen butir-butir dalam mengukur konstruk yang sama, semakin besar kemungkinan mereka secara konsisten membedakan kelompok kemampuan tinggi dan rendah (Lestari et al., 2024). Perhatian khusus perlu diberikan pada butir nomor 7 yang meskipun memiliki daya pembeda rendah (11,11%), ternyata memiliki pengecoh yang berfungsi sangat baik opsi B dipilih oleh 7 siswa dari kelompok asor. Pola ini konsisten dengan temuan (Pangastuti et al., 2025) yang menunjukkan bahwa daya pembeda rendah dapat berdampingan dengan pengecoh yang efektif, terutama ketika opsi pengecoh tertentu begitu menarik sehingga juga dipilih oleh sebagian siswa kelompok unggul, yang justru merefleksikan kualitas pengecoh yang terlampaui menjebak. Kondisi ini mengindikasikan bahwa butir nomor 7 perlu direformulasi pada bagian stem-nya agar lebih jelas membedakan antara jawaban benar dan pengecoh, tanpa mengorbankan fungsionalitas pengecoh yang sudah baik.

D. Korelasi Skor Butir dengan Skor Total

Korelasi skor butir dengan skor total merupakan indikator validitas empiris setiap butir soal. Hasil analisis menunjukkan bahwa dari 25 butir, sebanyak 16 butir (64%) memiliki korelasi yang signifikan pada taraf $p < 0,05$ atau $p < 0,01$, sedangkan 9 butir sisanya (36%) menunjukkan korelasi rendah dan tidak signifikan. Distribusi lengkap disajikan pada Tabel 3.

Tabel 3. Distribusi Korelasi Skor Butir dengan Skor Total

Kategori Signifikansi	Jumlah Butir	Persentase	Nomor Butir
Sangat Signifikan ($p < 0,01$)	12	48%	1, 4, 6, 8, 9, 10, 13, 16, 18, 19, 20, 21



Signifikan ($p < 0,05$)	4	16%	5, 15, 17, 24
Tidak Signifikan	9	36%	2, 3, 7, 11, 12, 22, 23, 25
Total	25	100%	

Butir-butir dengan korelasi sangat signifikan menunjukkan nilai koefisien tertinggi pada nomor 20 ($r = 0,752$), nomor 6 ($r = 0,702$), dan nomor 19 ($r = 0,653$), yang mengindikasikan bahwa ketiga butir ini paling konsisten dalam mengukur konstruk kemampuan Bahasa Inggris yang sama dengan keseluruhan tes. Di sisi lain, butir nomor 7 ($r = 0,140$), nomor 25 ($r = 0,118$), dan nomor 22 ($r = 0,162$) mencatat korelasi terendah, mengonfirmasi bahwa butir-butir ini tidak cukup koheren dengan dimensi kemampuan yang diukur oleh instrumen secara keseluruhan. Temuan ini sejalan dengan (Putra et al., 2021) yang menegaskan bahwa korelasi skor butir dengan skor total merupakan indikator paling langsung dari validitas konstruk internal sebuah instrumen: butir yang berkorelasi tinggi dengan skor total berarti mengukur atribut yang sama dengan keseluruhan tes, sehingga keberadaannya memperkuat konstruk yang ingin diukur. Sebaliknya, butir yang tidak berkorelasi signifikan dengan skor total kemungkinan besar mengukur aspek yang berbeda, mengandung ambiguitas redaksional, atau memiliki kunci jawaban yang tidak tepat. (Rahma et al., 2025) memperkuat pandangan ini dengan menyatakan bahwa dalam konteks tes EFL, butir yang tidak signifikan secara statistik sering kali mencerminkan ketidaksesuaian antara indikator soal dengan kompetensi yang hendak diukur dalam kurikulum. Menarik dicermati bahwa butir nomor 1, meskipun memiliki tingkat kesukaran sangat mudah (87,50%), tetap menunjukkan korelasi sangat signifikan ($r = 0,594$). Hal ini menunjukkan bahwa kemudahan sebuah butir tidak selalu berarti butir tersebut tidak valid; selama sebagian besar siswa yang menjawab benar memang berasal dari kelompok berkemampuan tinggi, validitas empiris butir tersebut tetap terjaga. Pola ini menegaskan pentingnya mengintegrasikan analisis tingkat kesukaran, daya pembeda, dan korelasi skor butir secara bersamaan bukan secara parsial untuk mendapatkan gambaran kualitas instrumen yang komprehensif (Adestya & Supriatna, 2025).

E. Efektivitas Pengecoh

Analisis efektivitas pengecoh dilakukan terhadap tiga opsi pengecoh (non-kunci jawaban) dari setiap butir soal, menghasilkan total 75 pengecoh yang dievaluasi. Ringkasan hasil analisis disajikan pada Tabel 4.

Tabel 4. Distribusi Kualitas Pengecoh

Kategori Pengecoh	Jumlah	Persentase
Sangat Baik (++)	18	24%
Baik (+)	22	29,3%
Kurang Baik (-)	20	26,7%
Buruk (- -)	12	16%
Sangat Buruk (- - -)	3	4%
Total	75	100%

Secara keseluruhan, 53,3% pengecoh berfungsi pada kategori baik hingga sangat baik, sementara 46,7% sisanya termasuk kurang baik, buruk, atau sangat buruk. Pengecoh yang berkategori sangat buruk ditemukan pada butir nomor 7 (opsi B: dipilih 7 siswa secara tidak proporsional dari kelompok asor, namun hanya 1 dari kelompok unggul), butir nomor 15 (opsi C dipilih kurang dari 5% peserta), dan butir nomor 20 (opsi C dengan pemilih sangat sedikit dari kedua kelompok). Hasil ini



melampaui temuan (Mahrawi et al., 2021) yang melaporkan rata-rata efektivitas pengecoh sebesar 48% pada soal akuntansi SMK, dan lebih dekat dengan temuan (Effendi & Andromeda, 2022) yang mencatat efektivitas pengecoh sebesar 88% pada soal SMP meskipun perbedaan metodologi pengukuran perlu dipertimbangkan dalam perbandingan tersebut. (Rahmadhani et al., 2024) mengingatkan bahwa pengecoh yang tidak dipilih oleh siapapun (zero-distractor) bukan hanya tidak berguna, tetapi secara implisit mengurangi jumlah pilihan jawaban efektif dari empat menjadi tiga atau bahkan dua, yang berakibat pada meningkatnya probabilitas menebak jawaban benar secara kebetulan dari 25% menjadi 33% atau 50%.

Analisis lebih mendalam terhadap pola pemilihan pengecoh mengungkap insight yang signifikan untuk perbaikan instrumen. Pada butir nomor 25 yang juga merupakan butir paling sukar pengecoh A dipilih oleh 21 siswa (65,6% dari seluruh peserta), jauh melampaui kunci jawaban D yang hanya dipilih 3 siswa. Kondisi ini sangat mengkhawatirkan karena menunjukkan bahwa sebagian besar siswa, termasuk kelompok unggul, justru tertarik pada opsi yang salah. Fenomena ini dapat mengindikasikan dua kemungkinan: (1) kunci jawaban butir nomor 25 perlu ditelaah ulang karena mungkin terdapat kekeliruan penetapan kunci, atau (2) stem soal mengandung ambiguitas yang mengarahkan pembaca pada kesimpulan yang berbeda dari yang dimaksud penulis soal. Temuan serupa dilaporkan oleh (Amiruddin et al., 2020) yang menyebut fenomena ini sebagai *answer key mismatch*, yaitu kondisi di mana pengecoh justru lebih atraktif daripada kunci jawaban karena adanya kelemahan substantif atau redaksional dalam butir soal. Secara keseluruhan, efektivitas pengecoh yang bervariasi mengindikasikan perlunya revisi selektif terhadap pengecoh-pengecoh yang tidak berfungsi. (Azizah et al., 2024) merekomendasikan bahwa pengecoh yang tidak dipilih oleh minimal 5% peserta sebaiknya diganti dengan distractor yang lebih plausibel secara substantif, misalnya dengan memanfaatkan *common errors* atau miskonsepsi yang lazim ditemukan pada siswa kelas VII dalam pembelajaran Bahasa Inggris, seperti kebingungan antara simple present dan simple past, atau kesalahan dalam penggunaan artikel.

F. Profil Integratif Kualitas Instrumen dan Implikasinya

Jika kelima dimensi psikometrik diintegrasikan, tergambar profil instrumen yang menunjukkan kekuatan sekaligus keterbatasan yang saling berkaitan. Instrumen ini unggul dalam hal reliabilitas (0,88) dan daya pembeda (76% butir berkategori baik–sangat baik), tetapi masih memiliki keterbatasan pada distribusi tingkat kesukaran yang condong ke kategori mudah (56%) dan proporsi butir yang tidak valid secara empiris (36% korelasi tidak signifikan). Profil ini membentuk gambaran instrumen yang konsisten namun belum optimal dalam cakupan kesukaran.

Tabel 5. Rekomendasi Tindak Lanjut Berdasarkan Profil Psikometrik

Kategori Tindak Lanjut	Nomor Butir	Alasan
Dipertahankan	4, 6, 8, 9, 10, 13, 14, 15, 16, 18, 19, 20, 21, 24	Daya pembeda baik, korelasi signifikan
Direvisi	1, 2, 3, 5, 7, 11, 12, 17, 22, 23	Tingkat kesukaran tidak ideal atau korelasi rendah
Diganti/ditelaah ulang	25	Sangat sukar, korelasi tidak signifikan, pengecoh A lebih atraktif dari kunci



Temuan integratif ini berkontribusi pada pengisian dua celah literatur yang diidentifikasi sebelumnya. Pertama, penelitian ini menyediakan profil psikometrik spesifik untuk soal Bahasa Inggris jenjang SMP menggunakan Anates v4.0.2 pada populasi siswa di lingkungan semi-urban, yang belum tersedia dalam literatur yang ada. Kedua, dengan mengintegrasikan data skor siswa (rata-rata 16,47, skor tertinggi 24, terendah 6) ke dalam analisis kualitas instrumen, penelitian ini menunjukkan bahwa reliabilitas yang tinggi tidak selalu beriringan dengan distribusi tingkat kesukaran yang ideal sebuah temuan yang secara metodologis penting sebagai pengingat bahwa evaluasi kualitas instrumen tidak boleh hanya bertumpu pada satu parameter psikometrik saja (Sabella et al., 2025). Implikasi praktis temuan ini mencakup tiga hal. Pertama, guru Bahasa Inggris kelas VII disarankan untuk merevisi bank soal dengan menambahkan butir-butir pada tingkat kesukaran sedang hingga sukar, khususnya yang menguji kompetensi membaca kritis dan pemahaman konteks. Kedua, butir yang valid dan diskriminatif (14 butir) layak dijadikan bank soal inti yang dapat digunakan ulang pada penilaian berikutnya. Ketiga, hasil analisis ini sebaiknya dijadikan bahan refleksi dalam forum Musyawarah Guru Mata Pelajaran (MGMP) Bahasa Inggris untuk mendorong budaya analisis butir soal secara rutin sebagai bagian dari siklus penjaminan mutu evaluasi pembelajaran di tingkat sekolah.

KESIMPULAN

Berdasarkan hasil analisis butir soal pilihan ganda Bahasa Inggris kelas VII-H SMPN 1 Sooko menggunakan program Anates versi 4.0.2 terhadap 25 butir soal yang dikerjakan oleh 32 siswa, diperoleh gambaran bahwa kualitas instrumen secara umum sudah cukup baik, meskipun masih terdapat beberapa aspek yang perlu diperbaiki. Tingkat kesukaran soal menunjukkan distribusi yang belum proporsional karena didominasi oleh kategori mudah hingga sangat mudah sebanyak 56% (14 butir), sedangkan kategori sedang sebesar 40% (10 butir), dan hanya satu butir (4%) yang tergolong sangat sukar, yaitu nomor 25 dengan indeks kesukaran 9,38%. Tidak ditemukannya butir pada kategori sukar menunjukkan perlunya pengembangan soal yang mampu mengukur kemampuan berpikir pada tingkat yang lebih tinggi. Dari aspek daya pembeda, instrumen tergolong baik karena 76% butir berada pada kategori baik hingga sangat baik. Namun, butir nomor 7, 11, dan 25 memiliki daya pembeda rendah sehingga perlu direvisi agar lebih mampu membedakan siswa berkemampuan tinggi dan rendah.

Selain itu, reliabilitas instrumen mencapai koefisien 0,88 yang termasuk kategori sangat tinggi, sehingga menunjukkan konsistensi internal yang sangat baik dan layak digunakan sebagai alat evaluasi pembelajaran. Analisis korelasi skor butir dengan skor total menunjukkan bahwa 64% butir (16 butir) memiliki korelasi signifikan hingga sangat signifikan, sedangkan 36% butir (9 butir) memiliki korelasi rendah dan tidak signifikan sehingga perlu ditinjau kembali dari segi isi maupun redaksi. Dari aspek efektivitas pengecoh, sebanyak 53,3% pengecoh berfungsi dengan baik hingga sangat baik, sementara 46,7% lainnya masih kurang efektif karena jarang atau tidak dipilih siswa, terutama pada butir nomor 15, 20, dan 25. Secara keseluruhan, instrumen tes Bahasa Inggris kelas VII-H SMPN 1 Sooko memiliki reliabilitas yang sangat tinggi dan daya pembeda yang baik, namun masih memerlukan perbaikan pada distribusi tingkat kesukaran, validitas empiris beberapa



butir, serta kualitas pengecohnya. Sebanyak 14 butir dapat dipertahankan sebagai bank soal inti, 10 butir perlu direvisi, dan 1 butir disarankan untuk diganti atau ditelaah ulang kunci jawabannya. Hasil penelitian ini diharapkan dapat menjadi acuan bagi guru dalam meningkatkan kualitas penyusunan instrumen evaluasi serta mendorong pelaksanaan analisis butir soal secara berkelanjutan.

DAFTAR PUSTAKA

- Adestya, A. I., & Supriatna, E. (2025). *Analisis Kualitas Butir Soal Penilaian Tengah Semester Berbasis HOTS Menggunakan Program Anates pada Mata Pelajaran PKn di Sekolah Dasar*. researchgate.net. https://www.researchgate.net/profile/Alvin-Adestya/publication/397731048_Analisis_Kualitas_Butir_Soal_Penilaian_Tengah_Semester_Berbasis_HOTS_Menggunakan_Program_Anates_pada_Mata_Pelajaran_PKn_di_Sekolah_Dasar/links/691cfd711555db2ebd5fa84f/Analisis-Kualitas-Butir-Soal-Penilaian-Tengah-Semester-Berbasis-HOTS-Menggunakan-Program-Anates-pada-Mata-Pelajaran-PKn-di-Sekolah-Dasar.pdf
- Agustin, B. I., Nuriani, R. I., & ... (2024). ... of TOAFL Questions at KH. Abdul Wahab Hasbullah University Jombang: Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal TOAFL di Universitas KH In *Edulab: Majalah Ilmiah* journal.uin-suka.ac.id. <https://journal.uin-suka.ac.id/tarbiyah/edulab/article/view/8644>
- Alista, Y. F., & Syahzanani, R. A. (2023). Analisis butir soal ulangan harian fisika dengan pendekatan teori tes klasik menggunakan program Anates. *SIPTEK: Seminar Nasional* <https://proceeding.unesa.ac.id/index.php/siptek/article/view/173>
- Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis butir soal pilihan ganda ulangan akhir semester genap mata pelajaran kearsipan. In *Measurement In Educational Research*.
- Amiruddin, K., Mania, S., Ichiana, N. N., Majid, A. F., & ... (2020). Analysis of Final School Examination Questions for Mathematics Subjects. In *Alauddin Journal of* [academia.edu](https://www.academia.edu/download/88922728/pdf.pdf). <https://www.academia.edu/download/88922728/pdf.pdf>
- Asyaroh, S. P., Febriyanti, A. C., & ... (2025). Analisis butir soal HOTS pada elemen perpajakan fase F SMK akuntansi menggunakan software Anates. *PESHUM: Jurnal* <https://al-haramjournal.com/index.php/PESHUM/article/view/7176>
- Avianti, M. N., Mahrawari, M., & Usman, U. (2021). Pengembangan Instrumen Asesmen Critical Thinking Skill pada Materi Sistem Ekskresi. *Jurnal Siliwangi: Seri* <https://jurnal.unsil.ac.id/index.php/JSSP/article/view/4211>
- Az-Zahra, I. B., & Rochmawati, R. (2024). PENGEMBANGAN INSTRUMEN ASESMEN BERBASIS KURIKULUM MERDEKA DENGAN BERBANTUAN LIVEWORKSHEET PADA MATA PELAJARAN DASAR *Promosi: Jurnal Program Studi* <https://ojs.fkip.ummetro.ac.id/index.php/ekonomi/article/view/10872>



- Azizah, N. L., Arliani, E., & ... (2024). Literasi matematika materi sistem persamaan dan pertidaksamaan linier: pengembangan butir soal. In ... *Matematika*. pdfs.semanticscholar.org.
<https://pdfs.semanticscholar.org/4907/e958744b95285ee92c857fbfc58941ebe990.pdf>
- Chan, K. W., Bard, A. M., Adam, K. E., Rees, G. M., Morgans, L., Cresswell, L., Hinchliffe, S., Barrett, D. C., Reyher, K. K., & Buller, H. (2020). Diagnostics and the challenge of antimicrobial resistance: A survey of UK livestock veterinarians' perceptions and practices. *Veterinary Record*, 187(12), E125–E125. <https://doi.org/10.1136/vr.105822>
- Effendi, Z., & Andromeda, A. (2022). Development of a Two-Tier Higher Order Thinking Skill (HOTS) Test Instrument Based on IBT on Chemical Equilibrium Material for SMA/MA Students. In *Entalpi Pendidikan Kimia*. entalpipendidikan.ppj.unp.ac.id.
<http://entalpipendidikan.ppj.unp.ac.id/index.php/epk/article/download/264/145>
- Fauzi, A., & Lestari, D. (2022). Analisis Daya Pembeda Butir Soal pada Evaluasi Pembelajaran Matematika Sekolah Dasar. *Jurnal Pendidikan Matematika Indonesia*, 7(2), 115–123.
- Fauziah, S., & Raharjo, T. (2022). Reliabilitas instrumen tes hasil belajar menggunakan pendekatan KR-20. *Jurnal Pendidikan Dan Evaluasi Pembelajaran*, 4(1), 33–41.
- Handayani, L., & Nugroho, P. (2021). Analisis hasil belajar matematika siswa sekolah dasar berdasarkan statistik deskriptif. *Jurnal Pendidikan Guru Sekolah Dasar*, 6(3), 210–218.
- Lestari, R., Surayanah, S., & ... (2024). Pengembangan Instrumen Tes IPA Berbasis Kearifan Lokal Blitar pada Siswa Sekolah Dasar. *Proceedings Series of*
<https://conference.um.ac.id/index.php/pses/article/view/9677>
- Mahrawi, M., Usman, U., & Avianti, M. N. (2021). Pengembangan Instrumen Asesmen Critical Thinking Skill pada Materi Sistem Ekskresi. *Indonesian Journal of* <https://mass.uinkhas.ac.id/index.php/mass/article/view/72>
- MUDIYANTRI, N. S., INDRIANI, N. S., & ... (2025). PENGUJIAN KUALITAS SOAL ETIKA PROFESI KEUANGAN SEBAGAI ALAT EVALUASI PEMBELAJARAN KELAS X MELALUI UJI ANATES VERSI 4.0. *Jurnal Pendidikan*
<https://jurnal.isdikkieraha.ac.id/index.php/jupek/article/view/1143>
- Mustaqim, M., & Sulisti, H. (2024). *Analisis butir soal pas matematika peminatan: daya pembeda, tingkat kesukaran, dan kualitas pengecoh.*
- Mustikasari, W. P., & Akhdinirwanto, R. W. (2025). Analisis Kualitas Instrumen Tes Fisika Kelas X Dengan Memanfaatkan Aplikasi AnBuso. ... : *Jurnal Pendidikan Fisika Dan* <https://ejurnalunsam.id/index.php/JPFS/article/view/13932>
- Pangastuti, A. P. W., Hakim, L., Pratiwi, V., & ... (2025). Analisis Butir Soal K3LH Pada Tingkat SMK Menggunakan Aplikasi ANATES. *Multiplier: Jurnal*
<https://multiplier.upstegal.ac.id/index.php/MLT/article/view/277>



- Pradani, R. A., & Efendi, A. (2023). Analisis Butir Soal Ujian Sekolah Menggunakan Program Itean (Analysis of School Exam Questions Using the Itean Program). *Indonesian Language Education and* <https://journal1.uinssc.ac.id/index.php/jeill/article/view/11002>
- Prayitasari, S. D., Nadjidah, S. J., & ... (2024). Analisis Butir Soal Pilihan Ganda Pada Elemen Job Profile Dan Peluang Usaha Di Bidang Akuntansi Dan Keuangan Lembaga Fase E Kelas X SMK Melalui *Journal of Learning* <https://ejournal.reysha.org/index.php/jlt/article/view/32>
- Putra, E. N., Syamsurizal, S., Lufri, L., & ... (2021). The Teacher-Made Multiple-Choice Tests Quality for Natural Science Subject. *Journal of Education* <https://ejournal.undiksha.ac.id/index.php/JERE/article/view/33219>
- Rahardian, D., Rahayu, A. P., & ... (2025). STRENGTHENING TEACHER COMPETENCE IN ANALYSING QUESTION ITEMS AT SDIT DARUL ABROR, GARUT REGENCY. *DEVELOPMENT* <http://jurnal.staim-probolingo.ac.id/index.php/DJCE/article/view/2547>
- Rahma, A., Meivia, A., Hakim, L., & ... (2025). Analisis Butir Soal K3LH pada Tingkat SMK Menggunakan Aplikasi ANATES. *Journal of Learning and* <https://ejournal.reysha.org/index.php/jlt/article/view/39>
- Rahmadhani, A., Ma'ruf, N., Hakim, L., & ... (2024). Penerapan Aplikasi Anates Dalam Menganalisis Kelayakan Butir Soal Pilihan Ganda Pada Elemen Etika Profesi Akuntansi. *Jurnal Nirta: Studi* <https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/100>
- Ramadhani, N. T., & Pratiwi, V. (2025). Pengembangan Alat Evaluasi Pembelajaran Berbasis Website Wordwall pada Mata Pelajaran Accurate di SMKN 10 Surabaya. *Jurnal Studi Guru Dan Pembelajaran.* <https://ejournal.my.id/jsgp/article/view/6619>
- Ramli, T., Mania, S., Abrar, A. I. P., Nur, F., & ... (2021). Analysis of USBN Items in Mathematics at MTS Madani Alauddin. In ... *Pendidikan Matematika*). academia.edu. <https://www.academia.edu/download/75055834/3393.pdf>
- Sabella, A. N., Puspitasari, L., & ... (2025). Analisis Butir Soal HOTS Pilihan Ganda Pada Mata Pelajaran Akuntansi Fase F Melalui Program Komputer ANATES V4. *Didaktik: Jurnal* <https://www.journal.stkipsubang.ac.id/index.php/didaktik/article/view/9795>
- Salsabilla, A. C., Rahmawati, S. C., & Fauza, A. Z. N. (2025). Evaluasi Kuantitatif Butir Soal Pilihan Ganda sebagai Upaya Peningkatan Kualitas Instrumen Pembelajaran IPA pada Materi Usaha, Energi, dan Pesawat In *Sindoro: Cendikia Pendidikan*.
- Suryadin, A. (2020). *Hubungan antara motivasi berprestasi dan pengetahuan pembuatan tes dengan kualitas soal buatan guru SMA dalam mata pelajaran bahasa Indonesia dan* academia.edu. <https://www.academia.edu/download/119035870/14755.pdf>



- Utami, S. C. P., Subekti, C. P. H., Mukarromah, L., Hakim, L., & ... (2024). Analisi Butir Soal Pilihan Ganda Pada Elemen Aplikasi Pengolah Angka (Spreadsheet) Fase E Melalui Penggunaan Program Komputer Anates. In *Widya Balina*.
- Wahyuni, S., & Hartanto, D. (2023). Pemanfaatan software ANATES dalam analisis butir soal berbasis digital. *Jurnal Teknologi Pendidikan*, 15(2), 101–110.
- Yasinta, L. (2021). ANALISIS BUTIR SOAL MATA KULIAH ALGORITMA PEMROGRAMAN. In *JUTECH: Journal Education and repository.persadakhatulistiwa.ac.id*.
<https://repository.persadakhatulistiwa.ac.id/id/eprint/36/2/1234-3946-1-PB.pdf>
- Zaini, Z., Hardika, R. C., Abrori, M. A. M., & Anggara, F. (2025). Pelatihan Analisis Tingkat Kesukaran Soal dan Daya Pembeda Soal untuk Identifikasi Kualitas Butir Soal Bagi Guru Di SMP Bahrul Ulum. In *Journal of Empowerment Community*.
- Zukhruffah, N., & Fradana, A. N. (2025). Multiple Choice Item Analysis in Indonesian Language Under Merdeka Curriculum: Analisis Soal Pilihan Ganda dalam Bahasa Indonesia Berdasarkan In *Academia Open*.
acopen.umsida.ac.id.
<https://acopen.umsida.ac.id/index.php/acopen/article/view/12130>

