



Analisis Kelayakan Butir Soal Pilihan Ganda Perpajakan Sebagai Instrumen Evaluasi Pembelajaran SMK Akuntansi melalui Software ANATES

Erina Octaviani Mirantu¹, Rifka Amelia Putri², Ardelia Benita Diva Gristiana³, Fidia Saputri⁴, Nur Fatimah Az Zahra⁵, Ira Yuli Indrianti⁶, Luqman Hakim⁷, Vivi Pratiwi⁸

^{1,2,3,4,5,6,7,8} Universitas Negeri Surabaya, Indonesia

Corresponding Author: ✉ : 24080304036@mhs.unesa.ac.id

ABSTRACT

Penelitian ini bertujuan untuk menganalisis kelayakan butir soal berupa pilihan ganda pada materi Perpajakan sebagai instrumen evaluasi pembelajaran bagi peserta didik SMK Akuntansi dengan memanfaatkan perangkat lunak ANATES. Instrumen penelitian yakni berupa 54 butir soal yang dikembangkan berdasarkan 6 indikator elemen kompetensi perpajakan fase F. Soal tersebut diujikan kepada 20 responden peserta didik SMK, kemudian dianalisis menggunakan ANATES versi 4.09 untuk memperoleh informasi mengenai tingkat kesukaran, daya pembeda, validitas, reliabilitas, serta kualitas pengecoh. Hasil analisis menunjukkan bahwa tes memiliki reliabilitas tinggi dengan koefisien sebesar 0,95, menandakan konsistensi internal yang sangat baik. Berdasarkan tingkat kesukaran, diperoleh 46,3% butir soal tergolong mudah, 31,5% sedang, 5,6% sukar, dan sisanya sangat mudah. Dari aspek daya pembeda, sebagian besar butir soal berada pada kategori baik hingga sangat baik, dengan rata-rata indeks daya pembeda sebesar 0,46. Sementara itu, hasil korelasi butir menunjukkan mayoritas soal memiliki validitas signifikan dan pengecoh yang berfungsi dengan baik. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa instrumen tes pilihan ganda perpajakan yang dikembangkan layak digunakan sebagai alat evaluasi pembelajaran di SMK Akuntansi, meskipun beberapa butir perlu direvisi untuk meningkatkan keseimbangan tingkat kesukaran dan efektivitas pengecoh.

ARTICLE INFO

Article history:
Received
01 November 2025
Revised
05 November 2025
Accepted
20 November 2025

Key Word

Analisis Butir Soal, ANATES, Perpajakan, Evaluasi Pembelajaran, SMK Akuntansi

How to cite

<https://pusdikra-publishing.com/index.php/jsr>



This work is licensed under a
[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

PENDAHULUAN

Pendidikan memiliki peran penting untuk membentuk sumber daya manusia berkualitas serta mampu beradaptasi dengan perkembangan pengetahuan juga teknologi yang semakin pesat. Dalam dunia pendidikan, kualitas pembelajaran tidak hanya diukur dari bagaimana materi disampaikan, tetapi juga dari bagaimana hasil belajar peserta didik dievaluasi (Ikhsaniyah et al., 2024). Beberapa komponen penting dalam proses pendidikan yakni evaluasi pembelajaran, yang berfungsi mengetahui

tingkat pencapaian kompetensi peserta didik serta efektivitas proses pembelajaran yang dilaksanakan. Evaluasi pembelajaran harus dilakukan secara objektif, terukur, dan berkesinambungan, sehingga diperlukan media evaluasi yang memiliki kualitas baik, yaitu validitas, reliabel, tingkat kesukaran, daya pembeda yang tinggi, juga pengecoh yang berfungsi dengan baik (Arikunto, 2019).

Dalam pembelajaran di SMK Akuntansi, instrumen evaluasi yang baik sangat diperlukan untuk mengukur penguasaan peserta didik terhadap kompetensi keahlian, salah satunya pada elemen perpajakan. Kompetensi ini mengharuskan peserta didik bukan hanya paham pada konsep saja, tetapi juga bisa menerapkan prinsip perpajakan dalam situasi kontekstual sesuai dengan dunia kerja. Butir soal yang digunakan dalam evaluasi pembelajaran perlu melalui proses analisis kelayakan agar dapat memberikan hasil penilaian yang akurat dan adil. Perihal ini bukan hanya meningkatkan kualitas dari suatu pembelajaran, akan tetapi mendukung peserta didik untuk bisa terjun ke dunia pekerjaan yang membutuhkan kemampuan berpikir secara kritis (Hidayati & Nisa, 2023). Pada proses ini bisa dilakukan dengan berbagai metode, seperti mengikuti tes soal pilihan ganda. Tes merupakan salah satu alat yang dipakai untuk mengukur sesuatu menurut (Worby, 2022). Evaluasi bisa berupa data kualitatif maupun kuantitatif berdasar kriteria yang sudah ditetapkan. Hasil yang dapat dipakai untuk menentukan keputusan pada bidang pendidikan. Dengan pengukuran bisa membandingkan hasil dengan standar penelitian yang telah ditetapkan (Putri & Hakim, n.d.) .

Analisis butir-butir soal mempunyai posisi sangat penting guna mendapatkan pertanyaan-pertanyaan yang baik serta bisa digunakan untuk alat tes. Analisis soal juga mampu meningkatkan suatu keterampilan persiapan pertanyaan peserta didik dan pendidik (Yusuf, 2024). Dengan melaksanakan analisis butir soal pendidik dapat mengidentifikasi kekurangan dalam soal tersebut dan memperoleh informasi yang diperlukan untuk melakukan perbaikan, sehingga kualitas soal dapat ditingkatkan (Ayu, 2024). Salah satu cara untuk menilai kelayakan butir soal adalah dengan analisis butir soal (item analysis) menggunakan bantuan perangkat lunak. Pemanfaatan teknologi informasi pada analisis masalah yakni dengan salah satu alternatif metode pemecahan masalah yang dapat diterapkan sehingga memudahkan pendidik untuk mengevaluasi pertanyaan dengan lebih efisien (Febrina et al., 2022). Salah satu program yang sering digunakan adalah software ANATES (Wiguna, 2021). Yang dapat menganalisis berbagai aspek kualitas soal seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (Mardiyah, 2024).

Beberapa penelitian sebelumnya menunjukkan bahwa penerapan ANATES dalam menganalisis soal memberikan hasil yang efektif untuk meningkatkan kualitas instrumen evaluasi pembelajaran (Ikawati et al., 2024). Namun, hasil dari analisis menyatakan masih banyak soal yang tergolong mudah, dengan daya pembeda yang

rendah, dan beberapa pengecoh tidak berfungsi dengan baik. Kondisi ini menunjukkan pentingnya analisis mendalam sebelum soal digunakan dalam ujian atau penilaian akhir.

Berdasarkan latar belakang tersebut, penelitian dengan tujuan untuk menganalisis kelayakan butir soal pada pilihan ganda pada materi perpajakan sebagai instrumen evaluasi pembelajaran SMK Akuntansi dengan bantuan software ANATES. Evaluasi yang baik memberikan pemahaman tentang kualitas pembelajaran dan memudahkan guru merencanakan strategi pengajaran (Magdalena et al., 2023). Melalui penelitian ini diharapkan dapat diketahui kualitas butir soal dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan pengecoh, sehingga hasilnya dapat menjadi acuan dalam penyusunan instrumen evaluasi yang lebih baik dan sesuai dengan tujuan pembelajaran vokasi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif karena berfokus pada penyajian dan analisis data hasil tes secara numerik untuk menggambarkan kelayakan butir soal pilihan ganda. Pendekatan ini dipilih agar peneliti dapat mendeskripsikan secara objektif kualitas instrumen evaluasi yang dikembangkan, baik dari segi validitas, reliabilitas, tingkat kesukaran, daya pembeda, maupun efektivitas pengecohnya. Membantu juga dalam memahami rincian data dengan membandingkan dan menganalisis pola dari sampel data yang relevan (Aziza, 2023). Subjek penelitian ini adalah 20 peserta didik SMK Akuntansi yang telah menerima pembelajaran materi Perpajakan pada fase F. Mereka menjadi responden dalam uji coba instrumen tes yang dikembangkan. Objek penelitian berupa 54 butir soal pilihan ganda, yang disusun berdasarkan enam indikator elemen kompetensi dalam elemen perpajakan fase F. Keenam indikator ini mencakup pemahaman terhadap ketentuan umum perpajakan, identifikasi wajib pajak dan jenis pajak, perhitungan pajak terutang, ketentuan pembayaran dan pelaporan pajak, serta analisis ketentuan dan sanksi perpajakan. Analisis butir soal yaitu proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban peserta didik untuk membuat keputusan (Sedya Santosa, 2022).

Tes dapat didefinisikan dalam bentuk serangkaian pertanyaan yang dijawab dengan maksud untuk menilai sejauh mana kinerja individu atau mengungkapkan aspek-aspek khusus dari orang yang mengikuti tes (Susanto, 2023). Instrumen penelitian berbentuk tes objektif pilihan ganda dengan lima opsi jawaban (A, B, C, D, dan E), di mana satu merupakan jawaban benar dan empat lainnya berfungsi sebagai pengecoh. Butir soal disusun dengan memperhatikan kaidah penulisan soal yang baik agar dapat mengukur kemampuan peserta didik secara proporsional pada ranah pengetahuan (kognitif). Tes diimplementasikan melalui Google Form, kemudian hasil jawaban peserta diunduh dan diolah menggunakan perangkat lunak ANATES versi

4.09. Data yang diperoleh berupa hasil pekerjaan peserta didik untuk setiap butir soal. Jawaban tersebut dimasukkan ke dalam program ANATES untuk dilakukan analisis statistik. ANATES menghasilkan data kuantitatif yang meliputi tingkat kesukaran, daya pembeda, validitas, reliabilitas, serta efektivitas pengecoh pada setiap butir soal. Dalam pembelajaran yang diterapkan peserta didik diharapkan mampu menerapkan proses berfikir tingkat tinggi (Tasrif, 2022).

Analisis validitas digunakan untuk mengetahui sejauh mana suatu butir soal mampu mengukur kompetensi yang diharapkan. Suatu soal dikatakan valid apabila nilai korelasi antara skor butir dan skor total signifikan pada taraf kepercayaan 5%. Selanjutnya, reliabilitas menunjukkan konsistensi hasil tes secara keseluruhan. Hasil analisis ANATES menunjukkan nilai reliabilitas sebesar 0,95, yang termasuk kategori sangat tinggi, artinya instrumen memiliki konsistensi internal yang baik antarbutir soal. Analisis tingkat kesukaran digunakan untuk mengetahui proporsi peserta didik yang dapat menjawab benar setiap soal. Soal yang baik umumnya memiliki tingkat kesukaran sedang, karena dapat menggambarkan variasi kemampuan peserta didik secara lebih objektif. Hasil penelitian menunjukkan bahwa 46,3% butir soal tergolong mudah, 31,5% tergolong sedang, 5,6% tergolong sukar, dan sisanya sangat mudah.

Selanjutnya, daya pembeda digunakan untuk melihat sejauh mana suatu soal mampu membedakan antara peserta didik yang berkemampuan tinggi dan rendah. Rata-rata daya pembeda pada penelitian ini sebesar 0,46, yang termasuk kategori baik. Nilai tersebut menunjukkan bahwa sebagian besar soal sudah cukup efektif untuk membedakan tingkat kemampuan peserta. Selain itu, dilakukan pula analisis terhadap efektivitas pengecoh. Pengecoh dinilai berfungsi dengan baik apabila dipilih oleh minimal 5% peserta tes. Berdasarkan hasil ANATES, sebagian besar pengecoh pada soal berfungsi dengan baik, walaupun terdapat beberapa butir yang masih perlu direvisi karena pengecohnya tidak menarik bagi peserta.

Berdasarkan hasil analisis tersebut, suatu butir soal dinyatakan layak digunakan apabila memenuhi kriteria valid, reliabel, memiliki tingkat kesukaran sedang hingga mudah, daya pembeda baik, serta pengecoh yang berfungsi. Sebaliknya, butir yang tidak memenuhi kriteria tersebut disarankan untuk direvisi atau diganti agar kualitas instrumen evaluasi pembelajaran meningkat. Metode penelitian ini tidak hanya bertujuan untuk mengetahui hasil pengujian kelayakan butir soal, tetapi juga memberikan gambaran menyeluruh tentang bagaimana perangkat evaluasi dapat dioptimalkan agar sesuai dengan tujuan pembelajaran perpajakan di SMK Akuntansi.

HASIL DAN PEMBAHASAN

Salah satu cara untuk menganalisis sebuah pertanyaan adalah dengan melakukan analisis kualitatif atau kuantitatif. Dalam penelitian ini kami mengumpulkan jawaban butir soal sebagai hasil pengukuran dan melakukan analisis kuantitatif (Mustaqim, M.,

& Sulisti, 2024). Analisis butir soal dalam penelitian ini dilakukan menggunakan software ANATES versi 4.09 terhadap 54 butir soal pilihan ganda pada materi Perpajakan Fase F SMK Akuntansi. Keunggulan dari Anates yakni secara otomatis menganalisis unsur-unsur pertanyaan dan memungkinkan melihat jawaban benar serta salah dengan cepat. Selain itu, anates juga memiliki kelebihan langkah-langkah menjalankan program dalam bahasa Indonesia sehingga mudah dimengerti, dan hasil analisa dapat dipindahkan ke Ms Excel untuk dihitung nilainya (Elviana, 2020). Instrumen diberikan kepada 20 peserta didik, kemudian hasil tes dianalisis untuk mengetahui kelayakan butir soal berdasarkan lima aspek utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Tujuan analisis yakni meneliti serta mempertimbangkan setiap soal yang akan diajukan untuk peserta didik (Garut et al., 2019). Secara umum, hasil analisis menunjukkan bahwa instrumen soal yang disusun memiliki kualitas yang baik dan dapat digunakan untuk mengevaluasi hasil belajar peserta didik, meskipun beberapa butir perlu direvisi agar kualitasnya lebih optimal.

Hasil

Validitas

Validitas menunjukkan sejauh mana suatu butir soal mampu mengukur kompetensi yang seharusnya diukur. Hasil analisis ANATES menunjukkan bahwa dari 54 butir soal, terdapat 40 butir soal valid (74,1%) dan 14 butir soal tidak valid (25,9%). Soal yang valid memiliki korelasi signifikan antara skor butir dan skor total tes, sedangkan soal yang tidak valid umumnya memiliki redaksi kurang jelas, indikator yang tidak sesuai, atau pengecoh yang tidak berfungsi.

Table 1.
Hasil Analisis Validitas

No. Butir Soal	Korelasi	Signifikasi	No. Butir Soal	Korelasi	Signifikasi	No. Butir Soal	Korelasi	Signifikasi
1.	NAN NAN		19.	0,74 1	Sangat Signifikan	37.	0,55 0	Sangat Signifikan
2.	NAN NAN		20.	0,59 1	Sangat Signifikan	38.	0,44 8	Sangat Signifikan
3.	0,851	Sangat Signifikan	21.	0,59 2	Sangat Signifikan	39.	0,52 1	Sangat Signifikan
4.	0,537	Sangat Signifikan	22.	0,46 8	Sangat Signifikan	40.	0,54 4	Sangat Signifikan
5.	NAN NAN		23.	0,70 2	Sangat Signifikan	41.	0,75 8	Sangat Signifikan
6.	0,702	Sangat Signifikan	24.	0,85 1	Sangat Signifikan	42.	0,18 9	
7.	0,851	Sangat Signifikan	25.	0,17 3		43.	0,71 5	Sangat Signifikan
8.	0,627	Sangat Signifikan	26.	0,74 4	Sangat Signifikan	44.	0,70 2	Sangat Signifikan
9.	0,754	Sangat Signifikan	27.	0,77 2	Sangat Signifikan	45.	0,53 7	Sangat Signifikan
10.	0,754	Sangat Signifikan	28.	0,23 3		46.	0,41 5	Sangat Signifikan
11.	0,901	Sangat Signifikan	29.	- 0,05 8		47.	0,48 7	Sangat Signifikan
12.	0,230		30.	0,33 8	Sangat Signifikan	48.	0,79 3	Sangat Signifikan

Hasil ini menunjukkan bahwa sebagian besar soal sudah baik dan sesuai indikator pembelajaran. Namun, soal yang tidak valid sebaiknya direvisi agar lebih mewakili kompetensi pada elemen perpajakan fase F.

Reabilitas

Reliabilitas tes merupakan seberapa jauh alat ukur bisa mengukur secara konsisten dan tepat apa yang sudah diukur. Rangkaian pengujian dianggap baik dan benar jika dapat diandalkan. Jika dilakukan pada subjek yang serupa dan di kondisi yang sama, bakal diperoleh hasil yang sama (Kaka et al., 2024). Reabilitas berfungsi untuk menilai tingkat konsistensi hasil suatu instrumen evaluasi. Tes yang reliabel akan memberikan hasil pengukuran yang relatif stabil ketika digunakan dalam kondisi serupa. Instrumen dengan tingkat reabilitas tinggi dipandang memiliki keajegan yang baik karena mampu menghasilkan data yang dapat dipercaya (Hasanah, 2015) dan (Endah et al., 2023). Temuan penelitian lain juga menekankan bahwa konsistensi internal merupakan salah satu indikator penting dalam menentukan keandalan sebuah alat ukur.

Tingkat Kesukaran

Table 2.
Hasil Analisis Tingkat Kesukaran

No. Butir Soal	Indeks Tingkat Kesukaran	Kategori	No. Butir Soal	Indeks Tingkat Kesukaran	Kategori	No. Butir Soal	No. Butir Soal	Indeks Tingkat Kesukaran
1.	100,00	Sangat Mudah	19.	80,00	Mudah	37.	80,00	Mudah
2.	100,00	Sangat Mudah	20.	75,00	Mudah	38.	80,00	Mudah
3.	90,00	Sangat Mudah	21.	70,00	Sedang	39.	75,00	Mudah
4.	70,00	Sedang	22.	70,00	Sedang	40.	85,00	Mudah
5.	100,00	Sangat Mudah	23.	95,00	Sangat Mudah	41.	85,00	Mudah
6.	95,00	Sangat Mudah	24.	90,00	Sangat Mudah	42.	55,00	Sedang
7.	90,00	Sangat Mudah	25.	85,00	Mudah	43.	85,00	Mudah
8.	80,00	Mudah	26.	75,00	Mudah	44.	95,00	Sangat Mudah
9.	80,00	Mudah	27.	85,00	Mudah	45.	80,00	Mudah
10.	80,00	Mudah	28.	60,00	Sedang	46.	70,00	Sedang
11.	85,00	Mudah	29.	60,00	Sedang	47.	85,00	Mudah

Tingkat kesukaran menggambarkan proporsi peserta tes yang mampu menjawab benar suatu butir soal. Kemampuan menilai topik pada tingkat tertentu yang diwakili oleh indeks biasanya (Akhmadi, 2021). Soal dengan tingkat kesukaran yang seimbang (tidak terlalu mudah atau sukar) dianggap ideal karena dapat mengukur variasi kemampuan peserta didik secara objektif. Berdasarkan hasil analisis ANATES, distribusi tingkat kesukaran soal ditunjukkan dalam tabel berikut:

Dari hasil tersebut terlihat bahwa sebagian besar soal tergolong mudah dan sedang, sehingga tes ini masih lebih banyak mengukur kemampuan dasar. Diagram menunjukkan bahwa sebagian besar butir soal perpajakan berada pada kategori mudah sebanyak 38,9%, diikuti kategori sedang sebesar 31,5%. Sementara itu, 24,1% soal tergolong sangat mudah dan hanya 5,56% yang termasuk kategori sukar. Agar lebih representatif, proporsi soal dengan tingkat kesukaran sedang perlu ditingkatkan. Hasil ini sesuai dengan pendapat yang menyatakan bahwa soal yang baik adalah soal dengan tingkat kesukaran sedang karena mampu menggambarkan perbedaan kemampuan peserta didik secara optimal (Arikunto, 2019).

Daya Pembeda

Menurut Iskandar dan Rizal dalam (Mania et al., 2020) bahwa daya beda suatu butir bertujuan untuk membedakan antara peserta didik yang berkemampuan tinggi dan rendah. Daya pembeda juga menunjukkan sejauh mana suatu soal mampu membedakan antara peserta didik yang memiliki kemampuan tinggi dan rendah. Hasil analisis ANATES menunjukkan bahwa rata-rata daya pembeda soal sebesar 0,46, yang termasuk kategori baik.

Table 3.
Hasil Analisis Daya Pembeda

No. Butir Soal	Indeks DP (%)	Kategori	No. Butir Soal	Indeks DP (%)	Kategori	No. Butir Soal	Indeks DP (%)	Kategori
1.	0,00	Jelek	19.	60,00	Baik	37.	60,00	Baik
2.	0,00	Jelek	20.	60,00	Baik	38.	40,00	Cukup
3.	40,00	Cukup	21.	80,00	Sangat Baik	39.	80,00	Sangat Baik
4.	60,00	Baik	22.	80,00	Sangat Baik	40.	40,00	Cukup
5.	0	Jelek	23.	95,00	Sangat Baik	41.	40,00	Cukup
6.	10,00	Jelek	24.	80,00	Sangat Baik	42.	80,00	Sangat Baik
7.	40,00	Cukup	25.	60,00	Baik	43.	70,00	Sangat Baik
8.	60,00	Baik	26.	20,00	Jelek	44.	80,00	Sangat Baik
9.	60,00	Baik	27.	20,00	Jelek	45.	100,0	Sangat Baik
10.	40,00	Cukup	28.	60,00	Baik	46.	60,00	Baik
11.	40,00	Cukup	29.	-20,00	Jelek	47.	50,00	Baik
12.	40,00	Cukup	30.	20,00	Cukup	48.	100,0	Sangat Baik
13.	20,00	Jelek	31.	20,00	Jelek	49.	40,00	Cukup
14.	40,00	Cukup	32.	-60,00	Jelek	50.	60,00	Baik
15.	40,00	Cukup	33.	-20,00	Jelek	51.	40,00	Cukup
16.	40,00	Cukup	34.	60,00	Baik	52.	100,0	Sangat Baik
17.	20,00	Jelek	35.	0,00	Jelek	53.	60,00	Baik
18.	40,00	Cukup	36.	20,00	Jelek	54.	40,00	Cukup

Hasil analisis terhadap kategori daya pembeda menunjukkan bahwa sebagian besar butir soal berada pada kategori Cukup dan Jelek, yaitu masing-masing sebesar 32,1% dan 30,2%. Kondisi ini menggambarkan bahwa masih banyak soal yang belum mampu memisahkan secara optimal antara peserta didik berkemampuan tinggi dan rendah. Sementara itu, kategori Baik mencakup 24,5% dari keseluruhan soal, yang berarti terdapat sejumlah butir yang sudah efektif dalam membedakan kemampuan peserta tes. Adapun kategori Sangat Baik hanya berjumlah 13,2%, sehingga butir-butir dengan kualitas sangat tinggi masih relatif sedikit. Secara keseluruhan, distribusi ini menunjukkan bahwa instrumen evaluasi memerlukan perbaikan pada sebagian besar soal, terutama yang berada pada kategori Jelek dan Cukup, agar kualitas pengukuran kemampuan peserta didik dapat meningkat dan hasil evaluasi menjadi lebih akurat.

Kualitas Pengecoh

Distractor quality adalah Bahasa Inggris dari kualitas pengecoh, yaitu ukuran seberapa efektif pilihan jawaban yang salah dalam sebuah butir tes (mis-multiple choice) dalam “mengecoh” peserta didik yang kurang paham dengan materi. Distraktor berfungsi untuk membingungkan peserta didik jika mereka tidak memilih jawaban yang benar (Radja et al., 2023) . Distraktor yang berkualitas baik harus cukup meyakinkan agar peserta yang tidak menguasai materi bisa tertarik memilihnya, tetapi tidak begitu menarik sehingga pesaing kunci jawaban menjadi tidak dipilih (Uswatun et al., 2024). Faktor kualitas pengecoh meliputi ketidakjelasan dan ambiguitas (distraktor tidak terlalu mirip dengan jawaban benar), gaya bahasa & format, distribusi kesalahan (keseimbangan pilihan), konteks & relevansi serta keseragaman kesulitan antara pengecoh. Sesuai dengan hasil butir soal terhadap kualitas pengecoh dalam bentuk pilihan ganda disajikan dengan diagram batang seperti dibawah:

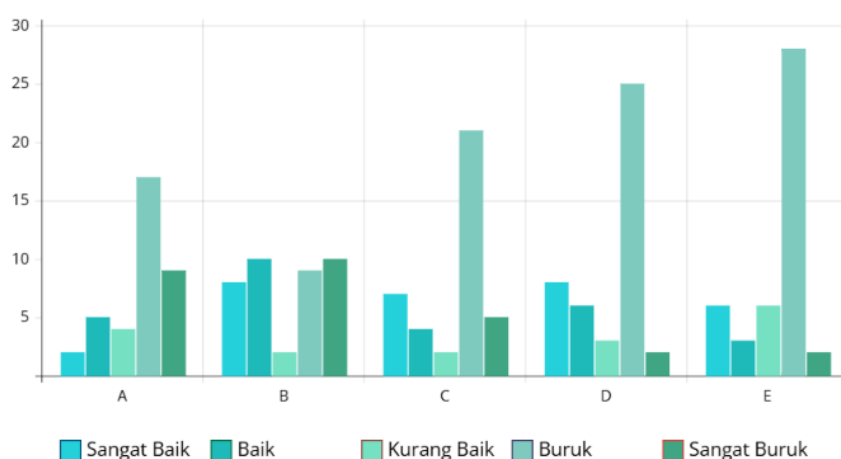


Diagram 1.
Hasil Analisis Daya Pengecoh

Pembahasan

1. Validitas

Berdasarkan hasil perhitungan menggunakan aplikasi ANATES, maka penulis dapat menyimpulkan bahwa analisis validitas butir soal dilakukan dengan membandingkan nilai koefisien korelasi butir soal dengan nilai r-tabel pada tingkat signifikansi yang ditetapkan (Issn et al., 2018). Butir soal dengan kategori sangat signifikan sebanyak 38 butir yakni pada nomor 3, 4, 6, 7, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 30, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 50, 52, 53, 54. Butir-butir ini dikategorikan sebagai soal sangat baik serta dapat diterima untuk digunakan pada tes selanjutnya. Lalu soal dengan kriteria signifikan berjumlah 3 soal pada nomor 13, 31, dan 51. Butir-butir ini memiliki validitas kategori rendah, yang berarti masih dapat diterima tetapi disarankan untuk ditinjau kembali atau diperbaiki. Lalu Soal dengan Kriteria Tidak Valid atau yang ditandai dengan (-) berjumlah 10 soal pada nomor 12, 25, 28, 29, 32, 33, 35, 36, 42, 49. Butir-butir ini Tidak diterima atau seharusnya dibuang karena tidak memenuhi kriteria validitas minimum atau memiliki koefisien korelasi di bawah r-tabel. Lalu soal dengan kriteria NAN atau tidak terhitung berjumlah 3 soal pada nomor 1, 2, dan 5. Butir-butir ini terjadi ketika semua responden memberikan jawaban yang sama sehingga tidak ada variasi skor. Butir-butir ini perlu dipertimbangkan untuk dibuang atau diubah.

2. Reabilitas

Hasil dari perhitungan reabilitas menggunakan aplikasi ANATES pada instrumen tes mata pelajaran perpajakan kelas 11 menunjukkan koefisien sebesar 0,95. Berdasarkan kriteria penilaian reliabilitas, nilai tersebut termasuk dalam kategori sangat tinggi karena berada pada rentang $\geq 0,90$ (Arini & Dewi, 2016). Temuan ini mengindikasikan bahwa instrumen soal memiliki tingkat konsistensi internal yang sangat baik dalam mengukur kemampuan peserta didik. Nilai reabilitas yang tinggi tersebut memperkuat bahwa instrumen pada penelitian berjudul "Analisis Kelayakan Butir Soal Pilihan Ganda Perpajakan sebagai Instrumen Evaluasi Pembelajaran melalui Software ANATES" layak digunakan sebagai alat evaluasi, karena mampu memberikan hasil pengukuran yang stabil dan dapat dipercaya (Arini & Dewi, 2016). Penelitian lain juga menyatakan bahwa koefisien reliabilitas di atas 0,90 menjadi indikator bahwa suatu instrumen memiliki tingkat keandalan yang sangat baik.

3. Tingkat Kesukaran

Berdasarkan hasil analisis tingkat kesukaran terhadap 54 butir soal perpajakan, diperoleh gambaran bahwa sebaran tingkat kesukaran cenderung mengarah pada kategori mudah. Hal ini terlihat dari diagram yang menunjukkan bahwa 38,9% butir berada pada kategori mudah, diikuti 31,5% kategori sedang, 24,1% kategori

sangat mudah, dan hanya 5,56% yang termasuk kategori sukar. Data pada dokumen analisis juga memperlihatkan banyaknya butir dengan indeks kesukaran berada pada rentang 75%-100%, menegaskan bahwa sebagian besar soal dapat dijawab dengan benar oleh mayoritas peserta tes. Dominasi soal kategori mudah dan sangat mudah menunjukkan bahwa instrumen evaluasi cenderung memberikan tingkat tantangan yang rendah, sehingga belum optimal dalam membedakan kemampuan peserta didik yang beragam. Rendahnya proporsi soal kategori sukar (hanya tiga butir) mengindikasikan bahwa peserta didik jarang diuji pada level kemampuan tinggi seperti analisis, evaluasi, atau penerapan konsep kompleks. Kondisi tersebut berpotensi membuat instrumen kurang komprehensif dalam mengukur penguasaan materi secara mendalam. Jika pada soal sangat sulit atau sangat mudah, peserta didik bisa saja merasa tidak mampu dalam menyelesaikan dan berhenti mengerjakan (Alista & Syahzanani, 2023). Oleh karena itu, penyusunan soal di masa mendatang disarankan untuk menambah variasi tingkat kesulitan, terutama pada kategori sedang dan sukar, agar evaluasi pembelajaran mampu mencerminkan kemampuan peserta didik secara lebih akurat, seimbang, dan proporsional.

4. Daya Pembeda

Berdasarkan hasil analisis kualitas pengecoh pada 54 butir soal, terlihat bahwa sebagian besar distraktor tidak berfungsi sebagaimana mestinya. Banyak opsi salah tidak dipilih oleh peserta didik, sehingga dikategorikan sebagai distraktor buruk hingga sangat buruk. Kondisi ini menunjukkan bahwa sebagian besar soal cenderung terlalu mudah, karena peserta didik langsung mengenali kunci jawaban tanpa mempertimbangkan opsi lainnya. Pola ini juga sejalan dengan hasil analisis tingkat kesukaran, di mana banyak butir berada pada kategori mudah hingga sangat mudah, serta didukung oleh data daya pembeda yang menunjukkan masih banyak soal berada pada kategori jelek dan cukup, yang berarti kemampuannya dalam membedakan peserta didik berkemampuan tinggi dan rendah masih rendah. Hanya sebagian kecil butir yang memiliki pengecoh berkualitas baik hingga sangat baik, yaitu pada soal-soal yang distraktornya benar-benar dipilih oleh kelompok peserta didik berkemampuan rendah dan dihindari oleh kelompok peserta didik berkemampuan tinggi. Hal ini menunjukkan bahwa hanya sedikit soal yang dirancang dengan distraktor yang efektif dan mendukung fungsi evaluasi secara optimal. Secara keseluruhan, kualitas pengecoh yang kurang baik ikut mempengaruhi tingkat kesukaran yang cenderung mudah dan menurunkan daya pembeda butir soal. Oleh karena itu, pengecoh perlu direvisi menjadi lebih logis, lebih menyerupai kunci jawaban, dan lebih mampu menarik peserta didik yang belum menguasai materi, sehingga setiap butir soal dapat berfungsi optimal dalam mengukur kemampuan peserta didik secara akurat.

5. Kualitas Pengecoh

Berdasarkan hasil analisis kualitas pengecoh pada 54 butir soal, terlihat bahwa sebagian besar pengecoh tidak berfungsi dengan baik karena banyak opsi salah yang sama sekali tidak dipilih peserta didik, sehingga dikategorikan buruk hingga sangat buruk dan mengindikasikan bahwa soal-soal tersebut terlalu mudah. Kondisi ini menyebabkan mayoritas butir soal memiliki pengecoh yang tidak mampu menjalankan fungsinya untuk membedakan kemampuan peserta didik, karena peserta didik hampir selalu langsung memilih kunci jawaban tanpa mempertimbangkan opsi lain. Hanya beberapa butir yang memiliki pengecoh berkualitas baik hingga sangat baik, terlihat dari adanya distraktor yang dipilih oleh kelompok peserta didik dengan kemampuan rendah dan dihindari oleh kelompok peserta didik berkemampuan tinggi, hal ini menandakan bahwa hanya sebagian kecil soal yang benar-benar dirancang dengan distraktor yang efektif. Banyaknya pilihan adalah 3 sampai 5 (Azzahroh et al., 2022). Secara keseluruhan kualitas pengecoh menyebabkan tingkat kesukaran menjadi terlalu mudah dan daya pembeda butir menurun. Oleh karena itu, pengecoh perlu direvisi agar lebih masuk akal, lebih mirip dengan kunci jawaban, dan lebih mampu menarik peserta didik yang belum menguasai materi.

KESIMPULAN

Berdasarkan hasil analisis menggunakan ANATES terhadap keseluruhan butir soal pada mata pelajaran Perpajakan kelas XI, dapat disimpulkan bahwa sebanyak 40 dari 54 soal dinyatakan valid, sedangkan 14 soal lainnya tidak memenuhi kriteria validitas sehingga perlu dilakukan perbaikan. Daya pembeda menunjukkan bahwa sebagian butir soal masih memerlukan revisi, dengan 30,2% soal berada pada kategori kurang baik, 32,1% kategori cukup, 24,5% kategori baik, dan hanya 13,2% yang termasuk kategori sangat baik. Berdasarkan tingkat kesukaran, sebanyak 24,1% soal tergolong sangat mudah, 38,9% termasuk kategori mudah, 31,5% kategori sedang, dan 5,56% masuk kategori sukar. Komposisi ini menunjukkan bahwa mayoritas soal masih berada pada level kesulitan rendah. Hasil perhitungan reliabilitas menunjukkan koefisien sebesar 0,95, yang berada pada kategori sangat tinggi dan menandakan konsistensi internal instrumen yang sangat baik. Dari aspek kualitas pengecoh, sebagian besar distraktor belum bekerja secara efektif karena banyak pilihan jawaban yang tidak dipilih peserta, meskipun terdapat sejumlah kecil butir soal yang memiliki pengecoh yang berfungsi dengan baik.

Secara umum, penelitian ini mengungkapkan dua hal penting: (1) Penyusun soal mata pelajaran Perpajakan perlu memperhatikan kembali kualitas butir soal, terutama sebelum soal digunakan dalam evaluasi berskala luas, sehingga uji coba dan revisi

- <https://ojs.fkip.ummetro.ac.id/index.php/matematika/article/view/7575>
- Ikawati, D. S., Agustin, E. W., Hakim, L., Pratiwi, V., & Surabaya, U. N. (2024). Edukatif: Jurnal Ilmu Pendidikan Analisis Butir Soal Pilihan Ganda pada Peserta didik SMK Menggunakan Anates. 6(6), 6773–6781.
- Ikhsaniyah, S. N., Kurnia, A. D., Zuroida, M., Pratiwi, V., & Hakim, L. (2024). maharani.23089. 12(2).
- Issn, P. E., Des, J., & Agustin, D. (2018). Media Akuntansi Perpajakan Publikasi oleh Fakultas Ekonomi dan Bisnis Universitas 17 Agustus 1945 Jakarta Media Akuntansi Perpajakan ISSN (P): 2355-9993 (E): 2527-953X Publikasi oleh Fakultas Ekonomi dan Bisnis Universitas 17 Agustus 1945 Jakarta. 3(2), 1–9.
- Kaka, L., Bano, V. O., & N., Y. (2024). Efektifitas Analisis Butir Soal Pilihan Ganda Menggunakan Aplikasi Anates Di SMPN 2 Kantang. *Jurnal In Inovasi Penelitian*, 4 (9), 1441–1450.
- Magdalena, I., Hidayati, N., D, R. H., S, S. W., & & M, Z. (2023). Pentingnya Evaluasi Dalam Proses Pembelajaran Dan Akibat Memanipulasinya. *Masaliq. Jurnal Pendidikan Dan Sains*, 3(5), 810–823.
- Mania, S., Fitriani, Majid, A. F., N., I. N., & & Abrar, A. I. (2020). Analisis Butir Soal Ujian Akhir Sekolah. *Journal of Islamic Education*, 274–284. [https://doi.org/10.24252/as%0A ma.v2i2.16569](https://doi.org/10.24252/as%0A%20ma.v2i2.16569)
- Mardiyah, H. (2024). Penggunaan Aplikasi Anates Pada Analisis. 11(2).
- Mustaqim, M., & Sulisti, H. (2024). Analisis Butir Soal Pas Matematika Peminatan: Daya Pembeda, Tingkat Kesukaran, Dan Kualitas Pengecoh. *Jurnal Tadris Matematika*, 44–56. <https://doi.org/10.24260/add.v3i1.3011>
- Putri, D. R., & Hakim, L. (n.d.). SMK Menggunakan Aplikasi Anates. 2047–2060.
- Radja, S. P., Bano, V. O., & (& A. T. I. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda dan Efektivitas Pengecoh di SMAN 1 Pandawai. *Jurnal Edusavana*, 30–41.
- Sedya Santosa, J. A. B. (2022). *Jurnal Basicedu*, 6(2), 1678–1686.
- Susanto, S. (2023). Pengembangan Alat dan Teknik Evaluasi Tes dalam Pendidikan. *Jurnal Tarbiyah Jamiat Kheir*, 1(1), 51–60.
- Tasrif, T. (2022). Higher Order Thinking Skills (HOTS) dalam pembelajaran social studies di sekolah menengah atas. <https://journal.uny.ac.id/index.php/jppfa/article/view/29490>
- Uswatun, Novikasari, I., Islam, U., Prof, N., & Zuhri, K. H. S. (2024). Analisis Butir Soal Akhir Semester I Mata Pelajaran Bahasa Indonesia Kelas II Madrasah Ibtidaiyah. 7(1).
- Wiguna, S. (2021). *Aplikasi Anates dalam Evaluasi Pembelajaran. Banyumas: Pena Persada*.
- Worbay, N. (2022). Analisis Kualitas Butir Soal Ujian Akhir Semester (UAS) Genap Mata Pelajaran Kimia Kelas X SMA N 1 Bobotsari Tahun Ajaran 2019/2020. <https://jurnal.uns.ac.id/jpkim>

Yusuf, F. W. (2024). Analisis Butir Soal Asesmen Sumatif Biologi Materi Perubahan Lingkungan Dengan Menggunakan Anates Pada Kelas X SMA. *Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 4(1), 126-135.