



Evaluasi Asesmen Sumatif Kelas Akhir (ASKA) pada Mata Pelajaran Projek Kreatif dan Kewirausahaan melalui Analisis Butir Soal Berbasis Anates Kelas XII Pemasaran

Vella Anggresta^{1*}, Fera Shiami²

Universitas Indraprasta PGRI^{1,2}

* Corresponding author, e-mail: vellaanggresta@gmail.com

Abstract: *This study aims to examine the quality of the Final School Examination instrument in the subject of Creative Projects and Entrepreneurship (Projek Kreatif dan Kewirausahaan/PKK). The study employed a descriptive quantitative approach. There were 21 twelfth-grade students from the Marketing specialization program at SMK PGRI 15 Jakarta participated in this study. Data were collected through documentation of students answer containing 45 multiple-choice questions. The data were analyzed by using item analysis of Classical Test Theory using the Anates application including validity, reliability, difficulty level, discrimination power, and distractor quality. The results of the analysis indicate that most of the test items met the validity criteria (68.9%), although a proportion of the items were still invalid (31.1%). The instrument demonstrated very high reliability (0.95); however, the distribution of difficulty levels was dominated by easy and very easy categories (77.7%), making it less capable of comprehensively measuring students' abilities. In addition, the discrimination power of several items was still low, and the distractor quality analysis showed that some answer options did not function optimally. These findings suggest that although the instrument is consistent, the quality of item construction has not yet been evenly developed.*

Keywords: *anates, validity, reliability, discrimination power, difficulty level*



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2021 by author.

PENDAHULUAN

Evaluasi pembelajaran merupakan bagian penting dalam proses pendidikan yang berperan dalam menilai pencapaian hasil belajar peserta didik serta menjadi acuan untuk menentukan keputusan pembelajaran dan meningkatkan mutu proses pendidikan (Sudjana,

2016). Dalam perspektif pendidikan modern, evaluasi tidak lagi diposisikan sekadar sebagai aktivitas administratif di akhir pembelajaran (*assessment of learning*), melainkan berkembang menjadi proses yang berkelanjutan dan berorientasi pada perbaikan pembelajaran (Mardiah & Hamami, 2026). Paradigma ini menuntut agar setiap perangkat evaluasi memiliki mutu yang baik dan terstandar dalam memberikan informasi yang akurat, objektif, dan dapat dipertanggungjawabkan (Firmansyah et al., 2026).

Namun demikian, realitas di lapangan menunjukkan bahwa praktik evaluasi pembelajaran di satuan pendidikan, khususnya pada jenjang Sekolah Menengah Kejuruan (SMK), masih menghadapi berbagai keterbatasan (Syaridawati et al., 2025). Evaluasi pembelajaran cenderung berorientasi pada hasil akhir semata tanpa diimbangi dengan perhatian terhadap kualitas instrumen yang digunakan. Instrumen penilaian, khususnya dalam bentuk tes tertulis pada ujian akhir sekolah, sering kali disusun tanpa melalui proses analisis yang memadai. Kondisi ini mengindikasikan adanya kesenjangan antara tuntutan evaluasi pembelajaran yang ideal dengan praktik yang berlangsung di lapangan, yang pada akhirnya berimplikasi pada kurang optimalnya akurasi hasil penilaian (Kasman, 2017).

Salah satu unsur krusial pada evaluasi pembelajaran adalah kualitas soal yang digunakan sebagai alat penilaian. Secara teoritis, suatu instrumen evaluasi yang baik harus memiliki kriteria validitas, reliabilitas, tingkat kesukaran, serta daya pembeda yang memadai (Augustia et al., 2025). Validitas menunjukkan sejauh mana instrumen mampu mengukur apa yang seharusnya diukur, sementara reliabilitas berkaitan dengan konsistensi hasil pengukuran (Ghozali, 2016). Tingkat kesukaran berfungsi untuk mengidentifikasi apakah suatu soal terlalu mudah atau terlalu sulit, sedangkan daya pembeda menunjukkan kemampuan soal dalam membedakan peserta didik yang berkemampuan tinggi dan rendah (Herdiansyah et al., 2025). Apabila keempat aspek tersebut tidak terpenuhi, maka hasil evaluasi berpotensi tidak mencerminkan kemampuan sebenarnya dari peserta didik, sehingga dapat menimbulkan bias dalam penilaian (Buka et al., 2025).

Permasalahan kualitas instrumen evaluasi tidak terlepas dari kompetensi guru dalam merancang dan menyusun butir soal. Dalam praktiknya, guru cenderung lebih berfokus pada penyampaian materi pembelajaran, sementara aspek pengembangan instrumen evaluasi belum mendapatkan perhatian yang memadai. Banyak guru yang masih menyusun soal berdasarkan intuisi atau pengalaman semata tanpa didukung oleh analisis empiris (Fajariani & Hanum, 2025). Selain itu, keterbatasan pemahaman terhadap teknik analisis butir soal serta rendahnya pemanfaatan teknologi pendukung menjadi faktor yang memperkuat permasalahan ini. Kondisi tersebut berimplikasi pada rendahnya kualitas soal yang dihasilkan, yang pada akhirnya memengaruhi validitas hasil evaluasi pembelajaran.

Di sisi lain, kesiapan peserta didik dalam menghadapi ujian akhir sekolah juga merupakan faktor penting yang memengaruhi hasil evaluasi. Kesiapan tersebut tidak hanya berkaitan dengan penguasaan materi pembelajaran, tetapi juga mencakup kemampuan memahami karakteristik soal, strategi menjawab, serta kesiapan psikologis dalam menghadapi ujian. Ketika instrumen evaluasi yang digunakan tidak memiliki kualitas yang baik, peserta didik cenderung menghadapi ketidakpastian dalam menginterpretasikan tingkat kesulitan soal

(Buhori, 2024). Hal ini dapat menyebabkan ketidaksesuaian antara kemampuan yang dimiliki dengan hasil yang diperoleh, sehingga berpotensi mengurangi keadilan dalam proses penilaian.

Konteks penelitian ini menjadi semakin relevan ketika dikaitkan dengan karakteristik mata pelajaran Projek Kreatif dan Kewirausahaan (PKK). Mata pelajaran ini menekankan pada pengembangan kreativitas, inovasi, kemampuan berpikir kritis, serta keterampilan kewirausahaan yang bersifat aplikatif. Dengan karakteristik tersebut, instrumen evaluasi yang digunakan seharusnya mampu mengukur kemampuan berpikir tingkat tinggi (*higher order thinking skills*) dan keterampilan kontekstual (Gunawan et al., 2025). Namun dalam praktiknya, soal yang disusun masih cenderung berfokus pada aspek kognitif tingkat rendah, seperti hafalan dan pemahaman sederhana (Savika et al., 2025). Kondisi ini menunjukkan adanya ketidaksesuaian antara tujuan pembelajaran dengan instrumen evaluasi yang digunakan, yang berpotensi menghambat pencapaian kompetensi yang diharapkan.

Dalam upaya meningkatkan kualitas instrumen evaluasi, analisis butir soal berbasis Teori Tes Klasik (*Classical Test Theory*) menjadi pendekatan yang relevan dan banyak digunakan dalam kajian evaluasi pembelajaran (Luthfiyah et al., 2020). Pendekatan ini memungkinkan pengukuran kualitas setiap butir soal secara empiris melalui indikator validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, sehingga dapat memberikan gambaran objektif mengenai kelayakan instrumen yang digunakan. Seiring dengan perkembangan teknologi, pemanfaatan perangkat lunak analisis menjadi semakin penting untuk meningkatkan akurasi dan efisiensi proses evaluasi. Salah satu aplikasi yang banyak digunakan adalah Anates, yang dirancang untuk mengolah data hasil tes secara sistematis dan menghasilkan informasi kuantitatif yang komprehensif terkait kualitas butir soal. Melalui Anates, proses analisis tidak hanya menjadi lebih cepat, tetapi juga mampu meminimalisasi subjektivitas dalam penilaian kualitas instrumen (Alifah et al., 2025). Namun demikian, pemanfaatan aplikasi ini di tingkat satuan pendidikan masih belum optimal karena banyak guru yang belum mampu mengaplikasikannya (Putu et al., 2024), sehingga diperlukan kajian empiris yang dapat menunjukkan efektivitas penggunaannya dalam mendukung perbaikan kualitas instrumen evaluasi pembelajaran secara berkelanjutan.

Sejumlah penelitian sebelumnya telah mengkaji analisis butir soal dalam berbagai konteks mata pelajaran dan jenjang pendidikan (Yustiandi & Saepuzaman, 2024). Namun demikian, kajian yang secara spesifik menyoroti kualitas instrumen evaluasi pada mata pelajaran Projek Kreatif dan Kewirausahaan, khususnya di kelas XII program keahlian Pemasaran, masih relatif terbatas. Selain itu, sebagian penelitian yang ada cenderung bersifat deskriptif tanpa memberikan analisis mendalam yang dapat dijadikan dasar dalam perbaikan instrumen evaluasi. Hal ini menunjukkan adanya kesenjangan penelitian (*research gap*) yang perlu diisi melalui kajian yang lebih komprehensif dan berbasis data empiris.

Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah bagaimana kualitas instrumen penilaian akhir sekolah pada mata pelajaran Projek Kreatif dan Kewirausahaan di kelas XII Pemasaran ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal. Sejalan dengan rumusan masalah tersebut, penelitian ini bertujuan untuk mengevaluasi kualitas butir soal ujian akhir sekolah melalui analisis berbasis Anates, mengidentifikasi butir soal yang memenuhi kriteria kualitas, serta menentukan

butir soal yang perlu direvisi atau dieliminasi. Selain itu, penelitian ini juga bertujuan untuk memberikan rekomendasi berbasis data empiris guna meningkatkan kompetensi guru dalam menyusun instrumen evaluasi yang lebih akurat, objektif, dan sesuai dengan tuntutan pembelajaran berbasis kompetensi.

Dengan demikian, penelitian ini diharapkan tidak hanya memberikan kontribusi praktis dalam perbaikan kualitas instrumen evaluasi di tingkat satuan pendidikan, tetapi juga memberikan kontribusi teoretis dalam pengembangan kajian evaluasi pembelajaran, khususnya yang berkaitan dengan analisis butir soal berbasis Teori Tes Klasik. Lebih lanjut, penelitian ini diharapkan dapat mendorong peningkatan profesionalisme guru dalam merancang instrumen evaluasi yang berkualitas serta memperkuat implementasi evaluasi pembelajaran yang selaras dengan tuntutan pendidikan abad ke-21.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk mengevaluasi kualitas instrumen Asesmen Sumatif Kelas Akhir (ASKA) pada mata pelajaran Projek Kreatif dan Kewirausahaan (PKK) melalui analisis butir soal berbasis Teori Tes Klasik. Subjek penelitian adalah 21 siswa kelas XII program keahlian Pemasaran pada SMK 15 PGRI Jakarta, sedangkan data penelitian berupa 45 butir soal pilihan ganda pada ASKA PKK. Data dikumpulkan melalui teknik dokumentasi berupa naskah soal, kunci jawaban, dan lembar jawaban siswa, kemudian dianalisis menggunakan aplikasi Anates.

Instrumen yang dianalisis disusun berdasarkan kisi-kisi soal yang mengacu pada kompetensi dasar dan indikator pembelajaran. Kisi-kisi tersebut mencakup tujuan pembelajaran, materi, indikator soal, level kognitif, bentuk soal, serta nomor soal. Data yang diperoleh, butir soal tersebar pada beberapa level kognitif, mulai dari C2 hingga C5, yang menunjukkan bahwa instrumen tidak hanya mengukur kemampuan memahami, tetapi juga kemampuan analisis dan evaluasi. Dengan demikian, instrumen yang digunakan memiliki konstruksi yang sistematis serta relevan dengan capaian pembelajaran.

Analisis butir soal dilakukan menggunakan aplikasi Anates yang mengacu pada prinsip Teori Tes Klasik (*Classical Test Theory*), dengan indikator yang meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda fungsi pengecoh (Pratama, 2020). Penggunaan aplikasi ini memungkinkan proses analisis dilakukan secara sistematis, objektif, dan berbasis data empiris hasil tes peserta didik. Prosedur penelitian meliputi pengumpulan data, penginputan data ke dalam aplikasi Anates, pelaksanaan analisis butir soal, serta interpretasi hasil analisis. Kriteria penentuan kategori didasarkan pada terpenuhinya indikator kualitas, yaitu validitas yang memadai, tingkat kesukaran yang proporsional, serta daya pembeda yang baik.

HASIL DAN PEMBAHASAN

Pada tahap awal analisis, dilakukan uji validitas terhadap butir soal untuk mengetahui sejauh mana setiap item mampu mengukur apa yang seharusnya diukur. Hasil uji validitas butir soal pada instrumen Asesmen Sumatif Kelas Akhir (ASKA) disajikan pada Tabel berikut.

Tabel 1: Rekapitulasi Validitas Butir Soal ASKA PKK

Kategori	Item Soal	Jumlah Soal	Proporsi
Valid (korelasi positif & signifikan)	1, 4, 8, 9, 12, 14, 16, 17, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45	31	68,9%
Tidak Valid (negatif / tidak signifikan / NaN)	2, 3, 5, 6, 7, 10, 11, 13, 15, 18, 19, 20, 22, 38	14	31,1%
Total		45	100%

Sumber: Data penelitian diolah dengan bantuan aplikasi Anates, 2026

Hasil analisis validitas menunjukkan bahwa sebagian besar butir soal telah memenuhi kriteria valid, yaitu sebesar 68,9%, sedangkan 31,1% lainnya belum valid. Proporsi butir soal yang tidak valid tergolong cukup tinggi dan menjadi indikator bahwa kualitas instrumen belum sepenuhnya optimal. Meskipun mayoritas soal telah mampu mengukur konstruk yang diharapkan, keberadaan butir tidak valid menunjukkan adanya kelemahan dalam penyusunan instrumen, khususnya dalam aspek kesesuaian indikator, kejelasan redaksi, dan ketepatan materi yang diujikan.

Setelah butir soal dinyatakan valid, langkah selanjutnya adalah melakukan uji reliabilitas. Uji ini digunakan untuk menilai konsistensi butir soal dalam mengukur tingkat pemahaman peserta didik terhadap materi yang diujikan.

Tabel 2: Nilai Koefisien Reliabilitas Instrumen ASKA PKK Kelas

Koefisien Reliabilitas	Kategori
0,95	Sangat Tinggi

Sumber: Data penelitian diolah dengan bantuan aplikasi Anates, 2026

Berdasarkan hasil uji reliabilitas, diperoleh koefisien sebesar 0,95 yang termasuk dalam kategori sangat tinggi, sehingga menunjukkan bahwa butir soal memiliki tingkat konsistensi yang sangat baik dalam mengukur kemampuan peserta didik. Tingginya nilai reliabilitas ini didukung oleh jumlah butir soal yang cukup banyak (31 butir), kesamaan kompetensi yang diukur, serta kualitas soal yang telah tersaring karena hanya melibatkan butir yang valid.

Setelah instrumen dinyatakan memiliki tingkat konsistensi yang baik, analisis dilanjutkan pada aspek daya pembeda butir soal. Analisis ini bertujuan untuk melihat kemampuan setiap butir soal dalam membedakan peserta didik berdasarkan tingkat penguasaan materi yang dimiliki. Hasil perhitungan daya pembeda pada butir soal Asesmen Sumatif Kelas Akhir (ASKA) mata pelajaran Projek Kreatif dan Kewirausahaan (PKK) ditampilkan pada tabel berikut.

Tabel 3: Distribusi Indeks Daya Pembeda Butir Soal ASKA PKK

Rentang Nilai	Kategori	Jumlah Soal	Proporsi
< 0,20	Jelek (<i>Poor</i>)	16	35,6%
0,20 – 0,40	Cukup (<i>Satisfactory</i>)	3	6,7%
0,40 – 0,70	Baik (<i>Good</i>)	17	37,8%
0,70 – 1,00	Baik Sekali (<i>Excellent</i>)	9	20,0%
Total		45	100%

Sumber: Data penelitian diolah dengan bantuan aplikasi Anates, 2026

Hasil analisis daya pembeda menunjukkan bahwa kualitas instrumen belum optimal. Meskipun 37,8% butir soal tergolong baik, tingginya persentase soal berkategori jelek (35,6%) menunjukkan bahwa instrumen belum efektif dalam membedakan kemampuan peserta didik. Secara konsep, daya pembeda penting untuk memastikan akurasi penilaian. Butir dengan daya pembeda rendah umumnya disebabkan oleh distraktor yang tidak efektif, ketidaksesuaian indikator, atau tingkat kesukaran yang tidak seimbang. Di sisi lain, adanya 20% soal berkategori sangat baik menunjukkan potensi kualitas instrumen, meskipun distribusinya belum merata dan kategori cukup hanya 6,7%. Berdasarkan klasifikasi (Anas Sudijono, 2017), kualitas instrumen berada pada kategori cukup dan masih perlu perbaikan, terutama melalui revisi butir soal dengan daya pembeda rendah agar evaluasi pembelajaran lebih akurat.

Selain aspek daya pembeda, evaluasi kualitas butir soal juga diarahkan pada tingkat kesukaran sebagai salah satu indikator penting dalam menilai proporsionalitas instrumen. Analisis ini diperlukan untuk memastikan bahwa setiap butir soal tidak terlalu mudah maupun terlalu sulit, sehingga mampu memberikan gambaran yang lebih akurat terhadap kemampuan peserta didik. Hasil pengolahan data terkait indeks kesukaran butir soal pada Asesmen Sumatif Kelas Akhir (ASKA) mata pelajaran Projek Kreatif dan Kewirausahaan (PKK) dirangkum dalam tabel berikut sebagai dasar interpretasi kualitas instrumen.

Tabel 4. Hasil Analisis Indeks Kesukaran Butir Soal ASKA PKK

Kategori Pada Anates	Item Soal	Jumlah Soal	Proporsi
Sangat Mudah	1, 4, 5, 6, 9, 11, 12, 13, 17, 18, 21, 32, 33	13	28,9%
Mudah	7, 8, 10, 14, 16, 20, 23, 25, 26, 27, 28, 29, 30, 31, 34, 35, 39, 41, 42, 44,	21	46,7%

Kategori Pada Anates	Item Soal	Jumlah Soal	Proporsi
	45		
	15, 24, 36, 37, 38, 40, 43		15,6%
Sedang		7	
Sukar	2	1	2,2%
Sangat Sukar	3, 19, 22	3	6,7%
Total		45	100%

Sumber: Data penelitian diolah dengan bantuan aplikasi Anates, 2026

Hasil analisis tingkat kesukaran menunjukkan bahwa distribusi butir soal belum proporsional. Dari 45 butir soal yang dianalisis, sebanyak 28,9% tergolong sangat mudah dan 46,7% termasuk kategori mudah. Sementara itu, hanya 15,6% berada pada kategori sedang, 2,2% kategori sukar, dan 6,7% kategori sangat sukar. Temuan ini mengindikasikan bahwa sebagian besar butir soal berada pada tingkat kesukaran rendah. Dominasi soal pada kategori mudah dan sangat mudah menunjukkan bahwa instrumen cenderung kurang memberikan tantangan bagi peserta didik. Kondisi ini berpotensi menyebabkan instrumen kurang mampu mengukur variasi kemampuan secara optimal, terutama dalam membedakan peserta didik dengan kemampuan tinggi dan rendah. Di sisi lain, proporsi soal kategori sedang yang relatif kecil menunjukkan belum tercapainya keseimbangan tingkat kesukaran yang ideal.

Setelah mengkaji tingkat kesukaran, evaluasi kualitas butir soal selanjutnya diarahkan pada efektivitas pengecoh sebagai komponen penting dalam struktur soal pilihan ganda. Analisis ini difokuskan pada kemampuan alternatif jawaban yang tidak benar dalam berfungsi sebagai distraktor, yaitu mampu menarik peserta didik yang belum menguasai materi sehingga setiap butir soal dapat bekerja secara optimal. Hasil pengolahan data terkait efektivitas pengecoh pada butir soal Asesmen Sumatif Kelas Akhir (ASKA) mata pelajaran Projek Kreatif dan Kewirausahaan (PKK) dirangkum dalam tabel berikut sebagai dasar penilaian kualitas distraktor.

Tabel 5. Distribusi Efektivitas Distraktor Butir Soal ASKA PKK

Kualitas Pengecoh	Item Soal	Jumlah Soal	Proporsi
Sangat Baik (++)	3, 7, 8, 14, 23, 24, 27, 37, 38, 40, 43	11	24,4%
Baik (+)	2, 10, 14, 25, 28, 33, 35, 36, 41, 42, 44, 45	12	26,7%
Kurang (-)	1, 4, 6, 9, 12, 13, 15, 16, 17, 20, 21, 22, 26, 29, 30, 31, 32, 34	18	40,0%
Buruk (--/---)	Hampir seluruh soal (dominan tidak berfungsi)	4	8,9%

Sumber: Data penelitian diolah dengan bantuan aplikasi Anates, 2026

Hasil analisis kualitas pengecoh menunjukkan bahwa distribusi efektivitas distraktor belum optimal. Dari keseluruhan butir soal, sebagian besar pengecoh berada pada kategori kurang, yaitu sebesar 40,0%, diikuti kategori baik sebesar 26,7% dan sangat baik sebesar 24,4%. Sementara itu, masih terdapat pengecoh yang tergolong buruk sebesar 8,9%, yang menunjukkan bahwa alternatif jawaban tidak berfungsi sebagaimana mestinya. Dominasi pengecoh pada kategori kurang mengindikasikan bahwa sebagian besar alternatif jawaban belum mampu berperan efektif dalam mengelabui peserta didik yang memiliki kemampuan rendah. Pengecoh yang baik seharusnya dipilih oleh kelompok peserta didik berkemampuan rendah dan tidak dipilih oleh kelompok berkemampuan tinggi. Namun, temuan ini menunjukkan bahwa banyak distraktor yang tidak memenuhi kriteria tersebut, bahkan beberapa di antaranya tidak dipilih sama sekali.

Hasil analisis secara komprehensif menunjukkan bahwa kualitas instrumen masih belum optimal dan memperlihatkan pola kelemahan yang bersifat sistemik antarindikator. Meskipun koefisien reliabilitas tergolong sangat tinggi (0,95), temuan ini tidak serta-merta mencerminkan kualitas instrumen yang baik. Reliabilitas yang tinggi dalam kondisi distribusi soal yang tidak proporsional cenderung dipengaruhi oleh homogenitas tingkat kesukaran, sehingga menghasilkan konsistensi skor yang tinggi tetapi kurang mampu merepresentasikan variasi kemampuan peserta didik secara akurat. Temuan ini sejalan dengan penelitian yang menyatakan bahwa reliabilitas yang tinggi tidak selalu diikuti oleh validitas yang baik apabila konstruksi butir soal tidak optimal (Saputri et al., 2023)

Permasalahan mendasar terlihat dari masih adanya 31,1% butir soal yang tidak valid, yang menunjukkan adanya ketidaksesuaian antara indikator pembelajaran dengan konstruksi soal. Hal ini mengindikasikan bahwa sebagian butir soal belum mampu mengukur kompetensi yang seharusnya diukur secara tepat. Kondisi tersebut diperkuat oleh distribusi tingkat kesukaran yang didominasi kategori mudah dan sangat mudah (77,7%), sehingga instrumen cenderung hanya mengukur kemampuan dasar dan kurang mampu mengeksplorasi kemampuan berpikir tingkat tinggi. Penelitian sebelumnya juga menunjukkan bahwa dominasi soal mudah akan menurunkan kualitas pengukuran karena tidak mampu menggambarkan kemampuan peserta didik secara komprehensif (Fitri et al., 2024; Magdalena et al., 2021)

Secara teoretis, distribusi tingkat kesukaran yang ideal seharusnya didominasi oleh kategori sedang, karena pada tingkat ini butir soal memiliki daya diskriminasi yang lebih baik (Komarudin & Sarkadi, 2017). Namun, hasil penelitian ini menunjukkan deviasi yang cukup signifikan dari kondisi ideal tersebut. Ketidakseimbangan ini berdampak langsung pada rendahnya daya pembeda pada sebagian butir soal, sehingga instrumen menjadi kurang sensitif dalam membedakan peserta didik berkemampuan tinggi dan rendah. Hal ini sejalan dengan temuan penelitian yang menyatakan bahwa ketidakseimbangan tingkat kesukaran berimplikasi pada rendahnya daya pembeda dan akurasi pengukuran (Pradita & Megawanti, 2023)

Lebih lanjut, analisis kualitas pengecoh menunjukkan bahwa hampir separuh butir soal memiliki distraktor yang tidak berfungsi secara efektif. Pengecoh yang tidak optimal menyebabkan peserta didik cenderung mudah mengidentifikasi jawaban yang benar, sehingga meningkatkan kemungkinan menjawab benar bukan karena penguasaan materi, melainkan

karena kelemahan desain soal (Arikunto, S. 2018). Kondisi ini memperkuat kecenderungan soal menjadi terlalu mudah dan menurunkan daya diskriminatifnya. Penelitian lain juga menegaskan bahwa kualitas distraktor yang rendah merupakan salah satu faktor utama yang memengaruhi validitas dan daya pembeda instrumen (Sari et al., 2022)

Dengan demikian, temuan ini menunjukkan bahwa kelemahan utama instrumen terletak pada aspek konstruksi butir soal, khususnya dalam penyusunan pengecoh dan distribusi tingkat kesukaran. Kedua aspek tersebut saling berkaitan dan secara simultan memengaruhi validitas serta daya pembeda. Oleh karena itu, kualitas instrumen tidak dapat dinilai secara parsial, melainkan harus dipahami sebagai suatu sistem yang terintegrasi (Anike, 2022; Zahiroh & Ritonga, 2021). Perbaikan instrumen perlu dilakukan secara menyeluruh melalui revisi konstruksi soal berbasis analisis empiris agar instrumen mampu berfungsi secara optimal dalam mengukur kemampuan peserta didik secara komprehensif

SIMPULAN

Berdasarkan hasil analisis butir soal kualitas instrumen Asesmen Sumatif Kelas Akhir (ASKA) pada mata pelajaran Projek Kreatif dan Kewirausahaan (PKK) secara umum masih berada pada kategori cukup dan belum sepenuhnya optimal. Meskipun instrumen menunjukkan reliabilitas yang sangat tinggi, kualitas butir soal secara individual masih belum merata. Hal ini ditunjukkan oleh masih adanya butir soal yang tidak valid, dominasi soal pada kategori mudah, serta daya pembeda yang belum optimal. Selain itu, kelemahan pada kualitas pengecoh menjadi faktor utama yang memengaruhi rendahnya kemampuan soal dalam membedakan tingkat kemampuan peserta didik. Dengan demikian, kualitas instrumen tidak hanya ditentukan oleh konsistensi hasil pengukuran, tetapi juga oleh keterpaduan antara validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.

Berdasarkan temuan tersebut, perbaikan instrumen perlu dilakukan secara sistematis dan berkelanjutan. Guru disarankan untuk merevisi butir soal yang tidak valid dan memiliki daya pembeda rendah dengan memperbaiki kesesuaian indikator, memperjelas redaksi soal, serta merancang pengecoh yang lebih fungsional dan logis. Selain itu, penyusunan soal perlu memperhatikan keseimbangan tingkat kesukaran dengan meningkatkan proporsi soal kategori sedang dan sukar agar instrumen mampu mengukur variasi kemampuan peserta didik secara lebih optimal. Secara praktis, guru dapat melakukan uji coba soal sebelum digunakan, menganalisis hasilnya menggunakan aplikasi Anates, serta menyusun bank soal yang telah teruji kualitasnya sebagai referensi untuk evaluasi berikutnya. Dengan langkah tersebut, kualitas instrumen evaluasi diharapkan dapat meningkat secara bertahap dan berkontribusi pada penilaian pembelajaran yang lebih akurat dan objektif.

DAFTAR PUSTAKA

- Alifah, F., Septianingsih, R., Rohmah, A. T. N., Arif, A., & Hakim, L. (2025). Analisis butir soal melalui aplikasi anates oleh siswa smk. *Analisis Butir Soal Melalui Aplikasi Anates Oleh Siswa Smk*, 2(2), 235–242.
- Anas Sudijono. (2017). *Pengantar Evaluasi Pendidikan (cetakan terbaru)*. RajaGrafindo Persada.
- Anike, A. (2022). Analisis Kualitas Butir Soal Ujian Tengah Semester (UTS) Ganjil Biologi Kelas XI SMAN 1 Pancung Soal Pendahuluan Metode. *Jurnal Metaedukasi: Jurnal Ilmiah Pendidikan*, 4(2), 61–69. <https://doi.org/https://doi.org/10.37058/metaedukasi.v4i2.4087>
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan (Edisi 3)*. Bumi Aksara.
- Augustia, A. D., Agustia, C. N., Azzahra, D., Hakim, L., & Pratiwi, V. (2025). Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan. *Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250–265. <https://doi.org/https://doi.org/10.37481/jmneb.v5i1.1165>
- Buhori. (2024). Kualitas Instrumen Evaluasi Pembelajaran Oleh: *Jurnal Ilmiah Keislaman*, 03(01).
- Buka, N., Jayamin, M. A., Kholis, N., & Rasyid, N. A. (2025). Peran Validitas dan Rehabilitas Dalam Evaluasi Pembelajaran Nurhan. *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial*, 02(June), 224–227. <https://doi.org/https://doi.org/10.5281/zenodo.15557188>
- Fajariani, K., & Hanum, E. (2025). Evaluasi Kualitas Instrumen Tes Pilihan Ganda pada Mata Pelajaran IPA Kelas V Sekolah Dasar Berdasarkan Analisis Butir Soal. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 4(2), 9224–9229.
- Firmansyah, Ilham, Syarifuddin, & Luthfiah. (2026). Pengembangan Instrumen Evaluasi Pembelajaran. *Jurnal Ilmu Pendidikan (SOKO GURU)*, 5(3), 489–504. <https://doi.org/https://doi.org/10.55606/sokoguru.v5i3.67112>
- Fitri, A., Pangestuti, N., Ana, S., Fahira, E., Syahra, A., & Sari, S. N. (2024). Analisis Distribusi Soal Higher Order Thinking Skills (HOTS) dan Lower Order Thinking Skills (LOTS) dalam Buku Ajar Bahasa Indonesia. 26 *IJAR: Indonesian Journal of Action Research*, 3(1). <https://doi.org/https://doi.org/10.14421/ijar.2024.31-03> Analisis
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate SPSS 23*. Universitas Diponegoro.
- Gunawan, S. T., Elida, Gusnita, W., & Insan, R. R. (2025). Implementasi Pembelajaran Projek Kreatif dan Kewirausahaan di Teaching Factory SMK Negeri 9 Padang. *Inovasi: Jurnal Sosial Humaniora Dan Pendidikan*, 4(1). <https://doi.org/https://doi.org/10.55606/inovasi.v4i1.4192>
- Herdiansyah, A. A., Nurfauzi, D., Khoirunnisa, N. A., Putri, H. E., & Rahayu, T. G. (2025). Pengembangan instrumen untuk mengukur kemampuan berpikir kritis siswa melalui konsep operasi hitung perkalian dan pembagian pecahan pada siswa kelas vi sd 3 babakan sari. *Jurnal Review Pendidikan Dan Pengajaran*, 8(1), 191–200.
- Kasman, T. (2017). *Evaluasi Tata Kelola Kelembagaan dan Pembelajaran SMK 3 Tahun*. Direktorat Pembinaan Sekolah Menengah Kejuruan Direktorat Jenderal Pendidikan Dasar dan Menengah Kementerian Pendidikan dan Kebudayaan RI.
- Komarudin, & Sarkadi. (2017). *Evaluasi pembelajaran*. Laboratorium Sosial Politik Press.
- Luthfiah, F., Arthur, R., & Maulana, A. (2020). Penggunaan teori tes klasik pada asesmen pembelajaran di smk dan peluang perbaikannya. *Educational Technology Journal*, 5(1), 1–9.
- Magdalena, I., Fauziah, S. N., Faziha, S. N., & Nupus, F. S. (2021). Analisis validitas, reliabilitas, tingkat kesulitan dan daya beda butir soal ujian akhir semester tema 7 kelas iii sdn karet 1 sepatan. *Bintang: Jurnal Pendidikan Dan Sains*, 3(2), 198–214.

- <https://doi.org/https://doi.org/10.36088/bintang.v3i2.1291>
- Mardiah, D., & Hamami, T. (2026). Pengembangan Assessment as , for , dan of Learning dalam Pembelajaran Pendidikan Agama Islam : Studi Literatur. *Edu Cendikia: Jurnal Ilmiah Kependidikan*, 6(April), 381–390. <https://doi.org/10.47709/educendikia.v6i01.8123>
- Pradita, E., & Megawanti, P. (2023). Analisis Tingkat Kesukaran , Daya Pembeda , dan Fungsi Distraktor. *Himpunan: Jurnal Ilmiah Pendidika*, 3(80), 109–118.
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Jurnal Pendidikan Islam*, 7(1).
- Putu, N., Camarini, I., Riastini, P. N., & Suarjana, I. M. (2024). Permasalahan Penggunaan Aplikasi Digital : Studi Masalah Guru Sekolah Dasar. *Jurnal Media Dan Teknologi Pendidikan*, 4(2), 158–165. <https://doi.org/https://doi.org/10.23887/jmt.v4i2.62701>
- Saputri, H. A., Zuhijrah, Larasati, N. J., & Shaleh. (2023). Analisis Instrumen Assesmen : Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal Hera. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09(09), 2986–2995. <https://doi.org/https://doi.org/10.36989/didaktik.v9i5.2268>
- Sari, V. N. I., Utomo, A. P. Y., & Sumarwati. (2022). Kualitas Soal Bahasa Indonesia di SMP Muhammadiyah 1 Pontianak: Analisis Butir Soal Vina. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 11(2).
- Savika, H. I., Zuhriyah, I. A., & Susilawati, S. (2025). Peran Guru dalam Analisis Butir Soal di Sekolah Dasar. *Jurnal Ilmiah Ilmu Pendidikan*, 8(3), 3313–3319.
- Sudjana. (2016). *Penilaian Hasil Proses Belajar Mengajar*. Remaja Rosdakarya.
- Syaridawati, Darnely, Yuspiani, & Subhan, A. (2025). Evaluasi dalam proses pembelajaran: analisis kritis terhadap pengertian, tujuan, syarat, dan teknik evaluasi. *Jurnal Pengembangan Dan Penelitian Pendidikan Https://Journalversa.Com/s/Index.Php/Jppp*, 07(3), 171–188.
- Yustiandi, & Saepuzaman, D. (2024). Analisis Karakteristik Butir Soal Seleksi Penerimaan Siswa Baru Berdasarkan Pendekatan Teori Tes Klasik. *Jurnal Basicedu*, 8(6), 4700–4706.
- Zahiroh, U., & Ritonga, P. S. (2021). Analisis kualitas butir soal pilihan ganda mata pelajaran kimia pada ujian akhir semester (uas) kelas xi man 2 kepulauan meranti. *Jedchem (Journal Education and Chemistry)*, 3(1), 11–20.