

Evaluasi Kuantitatif Butir Soal Pilihan Ganda sebagai Upaya Peningkatan Kualitas Instrumen Pembelajaran IPA pada Materi Usaha, Energi, dan Pesawat Sederhana

Alifia Calista Salsabilla¹, Salisy Chika Rahmawati², 'Afina Zulfa Nur Fauza³

^{1,2,3}Universitas Negeri Semarang, Indonesia

Alamat: Sekaran, Kec. Gn. Pati, Kota Semarang, Jawa Tengah 50229

Korespondensi penulis: alifiacs@students.unnes.ac.id¹, salisyachk@students.unnes.ac.id²,
afinazulfa13@students.unnes.ac.id³

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas butir soal pilihan ganda pada materi usaha, energi, dan pesawat sederhana dalam pembelajaran IPA melalui pendekatan kuantitatif deskriptif. Instrumen yang dianalisis terdiri dari 40 soal yang telah divalidasi oleh lima validator menggunakan Aiken's V, dengan 25 soal layak diujikan kepada 26 peserta didik kelas VIII SMP. Analisis dilakukan terhadap validitas isi, reliabilitas menggunakan KR-20, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Hasil penelitian menunjukkan bahwa 35 soal valid, nilai reliabilitas KR-20 sebesar 0,78 (reliabel), dan sebagian besar soal berada pada tingkat kesukaran mudah hingga sedang. Daya pembeda mayoritas soal tergolong cukup hingga tinggi, namun beberapa soal masih perlu perbaikan karena memiliki daya pembeda rendah dan distraktor tidak efektif. Evaluasi kuantitatif ini memberikan dasar yang kuat untuk meningkatkan kualitas instrumen evaluasi pembelajaran IPA secara berkelanjutan dan kontekstual.

Kata kunci: analisis butir soal, evaluasi pembelajaran, validitas, reliabilitas, IPA SMP.

Abstract

This study aims to evaluate the quality of multiple-choice items on the topic of work, energy, and simple machines in science learning using a descriptive quantitative approach. The instrument analyzed consisted of 40 items validated by five experts using Aiken's V, with 25 items selected for trial with 26 eighth-grade junior high school students. The analysis covered content validity, reliability using KR-20, item difficulty, discrimination index, and distractor effectiveness. The results show that 35 items were valid, the KR-20 reliability coefficient reached 0.78 (reliable), and most items fell into the easy to moderate difficulty range. The majority of items had sufficient to high discrimination indices, although several items required revision due to low discrimination power and ineffective distractors. This quantitative evaluation provides a solid

Article History

Received: Juni 2025

Reviewed: Juni 2025

Published: Juni 2025

Plagiarism Checker No
234.GT8.,35

Prefix DOI : Prefix DOI :
10.8734/Sindoro.v1i2.365

Copyright : Author

Publish by : Sindoro



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

foundation for improving the quality of science assessment instruments in a sustainable and contextualized manner.

Keywords: *item analysis, learning evaluation, validity, reliability, junior high school science.*

A. LATAR BELAKANG

Evaluasi pembelajaran merupakan bagian penting dalam sistem pendidikan karena berfungsi untuk mengukur sejauh mana tujuan pembelajaran telah tercapai serta menilai efektivitas proses belajar mengajar (Magdalena dkk., 2021). Dalam perkembangan kurikulum saat ini, evaluasi tidak hanya berfokus pada hasil akademik, tetapi juga digunakan untuk menilai aspek karakter, keterampilan berpikir kritis, dan kemampuan pemecahan masalah peserta didik. Oleh karena itu, penyusunan instrumen evaluasi terutama dalam bentuk soal pilihan ganda harus dirancang secara tepat agar hasil penilaian benar-benar mencerminkan kompetensi siswa secara objektif, adil, dan menyeluruh.

Namun pada kenyataannya, banyak instrumen evaluasi yang digunakan di sekolah belum dianalisis secara kuantitatif untuk memastikan kualitasnya. Instrumen tersebut seringkali langsung digunakan tanpa menguji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas opsi jawaban (distraktor). Padahal, tanpa analisis tersebut, soal-soal yang digunakan berisiko memberikan gambaran yang keliru terhadap kemampuan siswa dan berpotensi menghasilkan keputusan pembelajaran yang tidak tepat. Instrumen yang tidak memenuhi kriteria kualitas dapat gagal membedakan siswa yang benar-benar memahami materi dari yang belum, serta menghambat upaya peningkatan mutu pembelajaran.

Khususnya dalam mata pelajaran IPA di tingkat SMP, materi seperti usaha, energi, dan pesawat sederhana menuntut pemahaman konseptual serta kemampuan berpikir kritis siswa. Oleh karena itu, dibutuhkan instrumen evaluasi yang tidak hanya valid dan reliabel, tetapi juga mampu mengukur kemampuan siswa secara menyeluruh. Analisis kuantitatif terhadap butir soal, seperti mengkaji validitas isi, reliabilitas, indeks kesukaran, daya pembeda, dan efektivitas distraktor (Anshari dkk., 2024), merupakan langkah penting untuk meningkatkan kualitas instrumen evaluasi. Dengan mengetahui kelebihan dan kekurangan tiap butir soal, guru dapat memperbaiki dan mengembangkan instrumen yang lebih bermutu sebagai bagian dari upaya peningkatan kualitas pembelajaran IPA di sekolah.

B. KAJIAN TEORITIS

Evaluasi pembelajaran merupakan elemen penting dalam sistem pendidikan karena berfungsi sebagai alat untuk mengukur ketercapaian tujuan pembelajaran, perkembangan kompetensi peserta didik, serta efektivitas proses pembelajaran yang telah berlangsung. Evaluasi juga berfungsi sebagai dasar pengambilan keputusan dalam perencanaan, pelaksanaan, dan perbaikan pembelajaran (Suardipa & Primayana, 2023). Dalam konteks kurikulum modern tidak hanya fokus pada pencapaian nilai akademis siswa, tetapi juga pada pengembangan karakter dan keterampilan siswa. Evaluasi pembelajaran menjadi salah satu instrumen utama dalam menentukan efektivitas kegiatan pembelajaran serta memberikan umpan balik bagi pendidik dan siswa (Sa'edi dkk., 2024).

Evaluasi pembelajaran merupakan komponen krusial dalam proses pendidikan, di mana kualitas instrumen evaluasi, khususnya soal-soal tes, sangat menentukan keakuratan pengukuran kemampuan siswa (Surbakti, 2025). Namun, masih banyak soal yang digunakan dalam proses evaluasi belum melalui analisis statistik yang memadai, seperti uji validitas, reliabilitas, daya pembeda, dan tingkat kesukaran (Setiyawan & Wijayanti, 2020). Ketidadaan analisis ini berpotensi menghasilkan instrumen yang tidak valid dan tidak reliabel, sehingga tidak mampu membedakan kemampuan siswa secara tepat.

Validitas soal menunjukkan sejauh mana sebuah soal benar-benar mengukur aspek yang ingin diuji. Reliabilitas berkaitan dengan sejauh mana hasil dari soal tersebut konsisten jika diuji berulang kali. Daya pembeda mencerminkan kemampuan soal dalam membedakan siswa yang berprestasi tinggi dengan yang berprestasi rendah, sedangkan tingkat kesukaran menggambarkan seberapa mudah atau sulit soal tersebut bagi siswa. Jika tidak dilakukan analisis statistik yang tepat, penggunaan soal tersebut bisa memberikan penilaian yang kurang akurat terhadap kemampuan siswa, sehingga dapat mempengaruhi pengambilan keputusan pendidikan secara negatif.

Beberapa penelitian telah menyoroti pentingnya analisis statistik dalam pengembangan dan evaluasi soal. Misalnya, Qohar dan Fauziyah (2024) dalam studi mereka mengenai soal aljabar untuk siswa SMP menekankan perlunya analisis validitas dan reliabilitas untuk memastikan kualitas soal. Demikian pula, Amelia dan Erita (2024) mengeksplorasi validitas dan reliabilitas soal pemahaman konsep dalam asesmen pembelajaran, menunjukkan bahwa analisis statistik membantu dalam mengidentifikasi kelemahan soal yang perlu diperbaiki. Surbakti (2025) menyoroti pentingnya analisis kualitas butir soal dalam evaluasi pembelajaran, menunjukkan bahwa tanpa analisis yang tepat, soal dapat gagal dalam mengukur kemampuan siswa secara akurat.

Dalam penyusunan dan pelaksanaan evaluasi pembelajaran, analisis kuantitatif butir soal memiliki peran yang sangat penting untuk menjamin mutu dan ketepatan instrumen evaluasi yang digunakan. Analisis kuantitatif butir soal merupakan pendekatan berbasis data empiris yang digunakan untuk mengevaluasi kualitas suatu instrumen tes, khususnya soal pilihan ganda. Melalui analisis ini, pendidik dapat mengetahui sejauh mana sebuah soal mampu mengukur kompetensi siswa secara valid, reliabel, dan objektif. Aspek utama yang dianalisis meliputi validitas isi, reliabilitas instrumen, tingkat kesukaran soal, daya pembeda, serta efektivitas distraktor. Setiap aspek ini berperan penting dalam memastikan bahwa soal-soal yang digunakan dalam proses evaluasi benar-benar mampu mencerminkan capaian pembelajaran yang diharapkan.

Validitas isi seringkali dianalisis dengan menggunakan pendekatan seperti Aiken's V untuk mengetahui sejauh mana butir soal sesuai dengan indikator pembelajaran. Soal yang valid akan mengukur kompetensi yang seharusnya diukur, bukan kemampuan lain yang tidak relevan. Menurut Pratiwi dan Rufi'i (2023), validitas isi sangat penting karena menjadi pondasi utama dari kualitas soal dimana instrumen yang tidak valid akan menghasilkan informasi yang menyesatkan dalam proses evaluasi. Selain itu, reliabilitas juga perlu diperhatikan, biasanya dianalisis menggunakan rumus KR-20 (untuk soal pilihan ganda), yang mengukur konsistensi hasil tes. Instrumen dengan reliabilitas tinggi menunjukkan bahwa hasil yang diperoleh cenderung stabil meskipun digunakan pada waktu yang berbeda atau dengan kelompok peserta didik yang berbeda.

Selanjutnya, tingkat kesukaran soal perlu diketahui agar pendidik dapat menyesuaikan dengan karakteristik peserta didik. Soal yang terlalu mudah tidak dapat membedakan peserta didik yang memahami materi dengan baik dan yang tidak, sementara soal yang terlalu sulit dapat menurunkan motivasi belajar. Idealnya, soal memiliki indeks kesukaran dalam kategori sedang (antara 0,30–0,70). Hal ini sesuai dengan hasil penelitian oleh Fiska dkk. (2020) yang menunjukkan bahwa distribusi kesukaran soal sangat berpengaruh terhadap efektivitas tes secara keseluruhan.

Selain itu, daya pembeda merupakan indikator penting yang menunjukkan sejauh mana soal mampu membedakan antara siswa berkemampuan tinggi dan rendah. Soal yang baik akan memiliki daya pembeda yang tinggi, yaitu di atas 0,30. Dalam penelitian Mania dkk. (2020), ditemukan bahwa soal dengan daya pembeda rendah cenderung tidak memberikan informasi yang bermakna dalam proses penilaian, sehingga perlu diperbaiki atau diganti. Terakhir, efektivitas distraktor yaitu opsi jawaban yang salah namun logis juga dianalisis untuk memastikan bahwa siswa yang tidak menguasai materi akan cenderung memilih jawaban tersebut. Distraktor yang tidak efektif (misalnya tidak pernah dipilih oleh siswa) harus direvisi karena tidak berfungsi sebagaimana mestinya. Dengan demikian, analisis kuantitatif butir soal bukan hanya membantu dalam memperbaiki instrumen evaluasi yang sudah ada, tetapi juga menjadi acuan penting dalam penyusunan soal yang lebih bermutu di masa depan.

Instrumen yang dianalisis secara kuantitatif akan memberikan hasil evaluasi yang lebih adil, akurat, dan bermanfaat dalam pengambilan keputusan pembelajaran. Soal dalam penelitian ini berasal dari materi Ilmu Pengetahuan Alam (IPA) yang membahas tentang pesawat sederhana. Materi ini tidak hanya menuntut penguasaan konsep secara teoritis, tetapi juga keterampilan berpikir kritis dan logis untuk memahami prinsip kerja alat-alat sederhana yang sering dijumpai dalam kehidupan sehari-hari. Oleh karena itu, penting bagi pendidik untuk menyusun soal yang mampu mengukur kemampuan konseptual dan keterampilan berpikir siswa secara menyeluruh.

Agar instrumen evaluasi benar-benar sesuai dengan tujuan pembelajaran, maka diperlukan analisis soal yang komprehensif. Melalui analisis kuantitatif terhadap aspek-aspek tersebut, kualitas soal dapat ditingkatkan sehingga lebih representatif dalam menggambarkan pencapaian belajar. Hal ini sangat penting, khususnya dalam pembelajaran materi usaha, energi, dan pesawat sederhana yang memiliki aplikasi luas dalam kehidupan nyata, namun tetap membutuhkan pemahaman konseptual yang kuat. Oleh karena itu, pengembangan dan evaluasi soal berbasis analisis kuantitatif merupakan langkah strategis untuk memperkuat peran evaluasi dalam mendukung proses pembelajaran IPA yang bermutu (Ansyah dkk., 2024).

C. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif yang bertujuan untuk menganalisis kualitas butir soal pilihan ganda pada materi usaha, energi, dan pesawat sederhana. Penelitian dilaksanakan di SMP Islam Sultan Agung 1 Semarang dengan subjek penelitian sebanyak 26 peserta didik kelas VIII yang dipilih secara purposive. Variabel dalam penelitian ini meliputi tingkat kesukaran, daya pembeda, efektivitas pengecoh, validitas, dan reliabilitas sebagai kriteria butir soal dalam mata pelajaran IPA Tahun Ajaran 2024/2025. Seperti penelitian yang dilakukan oleh Sopiha (2019) yaitu untuk memastikan kesesuaian isi butir soal dengan indikator pembelajaran, dilakukan analisis validitas isi secara kualitatif. Sebelum penyebaran,

soal melalui telaah ahli oleh lima validator, yang terdiri atas dua guru IPA SMP Islam Sultan Agung 1 Semarang, satu guru IPA SMP N 3 Salatiga, satu guru IPA SMP Al Qolam Muhammadiyah Gemolong, dan satu mahasiswa Pendidikan IPA Universitas Negeri Semarang angkatan 2022. Soal yang digunakan berjumlah 25 butir berbentuk pilihan ganda yang memuat materi usaha, energi, dan pesawat sederhana.

Pengumpulan data dilakukan melalui Google Form yang diberikan kepada siswa dalam bentuk kuis online. Data yang diperoleh berupa hasil respon siswa terhadap soal yang kemudian diunduh dan dianalisis secara manual menggunakan Microsoft Excel dengan rumus-rumus analisis butir soal. Analisis data dilakukan menggunakan Microsoft Excel. Penggunaan Excel dipilih karena program ini memiliki kemampuan pengolahan data numerik yang cukup andal dan fleksibel untuk kebutuhan analisis statistik sederhana (Sarosa, 2021). Excel memudahkan peneliti dalam mengelompokkan data, menghitung nilai proporsi, varians, korelasi, serta membuat tabel dan grafik untuk visualisasi hasil. Selain itu, Excel juga praktis digunakan oleh guru dan tenaga pendidik karena tidak memerlukan instalasi perangkat lunak khusus dan umumnya sudah tersedia di komputer sekolah.

1. Desain Penelitian dan Sumber Data

Penelitian ini menganalisis seperangkat soal evaluasi berjumlah 40 butir, namun hanya 25 butir yang dianalisis lebih lanjut pada aspek reliabilitas dan karakteristik soal (tingkat kesukaran, daya pembeda, dan distraktor). Data diperoleh dari hasil uji coba soal kepada peserta didik kelas VIII SMP dan penilaian ahli untuk validitas isi.

2. Teknik Analisis Butir Soal

Analisis dilakukan dengan beberapa tahapan berikut:

a. Uji Validitas Isi dengan Aiken's V

Menurut Nabil, dkk. (2022) Validitas isi bertujuan untuk menilai sejauh mana suatu butir soal mencerminkan materi atau kompetensi yang ingin diukur. Sebuah tes dapat dikatakan valid apabila mampu mengukur aspek yang dimaksud secara tepat dan sesuai dengan tujuan pengukurannya.

Rumus Aiken's V:

$$V = \frac{\sum s}{n(c-1)}$$

Keterangan:

- s = skor yang diberikan rater dikurangi skor minimum
- n = jumlah rater
- c = jumlah kategori skala

Langkah-langkah:

- Setiap butir soal dinilai oleh beberapa ahli berdasarkan skala tertentu (misalnya 1-4)
- Nilai-nilai yang diperoleh dimasukkan ke rumus Aiken's V
- Hasil dibandingkan dengan nilai minimum signifikan pada Tabel Aiken (misal, $V \geq 0,88$ soal tersebut *valid* dinyatakan untuk 4 rater skala 1-4)
- Butir dengan nilai di atas ambang batas dianggap valid secara isi

Tujuannya adalah memastikan bahwa setiap soal benar-benar merepresentasikan kompetensi dan indikator yang dituju.

b. Uji Reliabilitas Instrumen dengan KR-20

Reliabilitas tes merujuk pada tingkat ketepatan dan konsistensi alat ukur dalam mengukur sesuatu secara berulang (Nabil dkk., 2022). Suatu tes dianggap baik dan layak apabila memiliki tingkat keandalan yang tinggi, yaitu mampu memberikan hasil yang konsisten ketika digunakan dalam kondisi yang serupa. Digunakan untuk mengetahui konsistensi internal dari instrumen pilihan ganda. Reliabilitas mengukur apakah setiap soal saling berkorelasi dengan baik sebagai satu kesatuan instrumen.

$$\text{Rumus KR-2C} \quad r_{11} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum pq}{S_t^2} \right)$$

Keterangan:

- k = jumlah soal
- p = proporsi siswa yang menjawab benar
- q = 1 - p
- $\sum pq$ = jumlah hasil perkalian p dan q untuk semua soal
- S_t^2 = varians total dari skor siswa

Interpretasi:

Jika nilai $r_{11} \geq 0,70$ maka soal dianggap reliabel (Juliani & Erita, 2023). Nilai mendekati 1 menunjukkan bahwa instrumen memiliki konsistensi yang tinggi antar butir.

c. Tingkat Kesukaran (Difficulty Index)

Soal yang berkualitas adalah soal yang memiliki tingkat kesulitan yang seimbang, tidak terlalu mudah maupun terlalu sulit (Juliani & Erita, 2023). Jika soal terlalu mudah, peserta didik tidak terdorong untuk mengembangkan kemampuan berpikirnya dalam menyelesaikan soal tersebut. Sebaliknya, soal yang terlalu sulit dapat membuat peserta didik merasa frustrasi dan enggan untuk mencoba kembali karena berada di luar jangkauan $P = \frac{B}{N}$ puan mereka. Tingkat kesukaran soal (TK) mengacu pada sejauh mana soal sulit untuk dijawab. Oleh karena itu, soal yang ideal adalah soal dengan tingkat kesukaran yang proporsional dan sesuai dengan kemampuan peserta didik. Tujuannya adalah untuk mengetahui sejauh mana soal mudah atau sulit bagi siswa. Rumus:

Keterangan:

- B = jumlah siswa yang menjawab benar
- N = jumlah total siswa

Interpretasi nilai P:

- $P \leq 0,30$ = Soal Sukar
- $0,31 \leq P \leq 0,70$ = Soal Sedang
- $P \geq 0,71$ = Soal Mudah

Analisis ini membantu penyusun soal menyeimbangkan tingkat kesulitan secara proporsional.

d. Daya Pembeda (Discrimination Index)

Daya pembeda soal merujuk pada kemampuan suatu butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Indikator ini digunakan untuk menilai seberapa efektif soal dalam mengidentifikasi

perbedaan tingkat pemahaman antar siswa (Juliani & Erita, 2023). Soal dengan daya pembeda yang baik cenderung dijawab benar oleh siswa dengan kemampuan tinggi dan dijawab salah oleh siswa dengan kemampuan rendah.

Rumus:

$$DP = P_A - P_B = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Keterangan:

- BA dan BB = jumlah siswa kelompok atas dan bawah yang menjawab benar
- JA dan JB = jumlah siswa kelompok atas dan bawah

Interpretasi:

- $DP < 0,20$ = Sangat Rendah
- $0,20 - 0,39$ = Cukup
- $0,40 - 0,69$ = Baik
- $\geq 0,70$ = Sangat Baik

Daya pembeda sangat penting untuk menjaga agar soal memberikan informasi akurat tentang perbedaan kemampuan siswa. Daya pembeda disebut indeks diskriminasi yang nilainya berkisar antara 0,00-1,00.

e. Analisis Efektivitas Distraktor

Distraktor merupakan opsi jawaban yang salah dalam suatu soal pilihan ganda. Distraktor yang efektif adalah yang mampu menarik perhatian peserta didik yang belum memahami materi dengan baik. Efektivitas distraktor menggambarkan pola pemilihan jawaban oleh peserta tes terhadap berbagai pilihan yang tersedia dalam setiap butir soal (Juliani & Erita, 2023). Tujuan utama penggunaan distraktor adalah untuk membedakan antara peserta didik yang memahami materi dengan yang belum. Sebuah distraktor dianggap berfungsi dengan baik jika cukup menarik bagi siswa yang tidak menguasai konsep atau belum memahami materi secara menyeluruh.

Kriteria Distraktor Efektif:

- Dipilih oleh minimal 5% siswa
- Tidak boleh kosong (tidak dipilih sama sekali)

Analisis distraktor dilakukan dengan melihat distribusi pilihan siswa terhadap setiap opsi (Juliani & Erita, 2023). Jika distraktor tidak efektif, maka revisi diperlukan agar jawaban pengecoh benar-benar menilai miskonsepsi siswa.

Metode penelitian ini dirancang secara sistematis untuk menganalisis kualitas butir soal pilihan ganda pada materi usaha, energi, dan pesawat sederhana. Melalui pendekatan kuantitatif deskriptif, penelitian ini menggabungkan analisis validitas isi secara kualitatif dengan penilaian para ahli serta pengujian reliabilitas dan karakteristik butir soal secara kuantitatif, meliputi tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Penggunaan Microsoft Excel sebagai alat bantu analisis dipilih karena kemudahan dan kepraktisannya dalam mengolah data numerik. Dengan demikian, metode yang digunakan diharapkan mampu memberikan gambaran yang komprehensif dan akurat

mengenai kualitas instrumen evaluasi yang dikembangkan, serta dapat dijadikan dasar dalam penyempurnaan soal untuk pembelajaran IPA yang lebih efektif dan tepat sasaran.

D. HASIL DAN PEMBAHASAN

Untuk mengetahui kualitas butir-butir soal yang digunakan dalam suatu evaluasi pembelajaran, dilakukan analisis terhadap berbagai komponen seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Hasil analisis ini bertujuan untuk mengevaluasi sejauh mana soal-soal yang disusun mampu mengukur kemampuan peserta didik secara tepat dan adil. Langkah pertama yang dilakukan dalam analisis butir soal adalah menguji validitas isi menggunakan pendekatan Aiken’s V. Uji validitas ini bertujuan untuk menilai sejauh mana setiap butir soal telah mewakili kompetensi dan indikator pembelajaran yang telah ditetapkan. Penilaian dilakukan oleh lima orang validator yang terdiri dari guru dan mahasiswa pendidikan IPA, dengan menggunakan rentang skala penilaian 1 sampai 4. Hasil penilaian tersebut kemudian dihitung menggunakan rumus Aiken’s V, dan dibandingkan dengan nilai kritis berdasarkan tabel Aiken, yaitu $V \geq 0,80$ dinyatakan valid dan $V < 0,80$ dinyatakan tidak valid. Analisis validitas ini menjadi dasar penting sebelum melanjutkan ke tahapan analisis reliabilitas dan karakteristik butir soal, karena memastikan bahwa soal yang diuji telah sesuai secara konten dengan tujuan pembelajaran. Adapun hasil uji validitas isi disajikan dalam tabel berikut.

Tabel 1. Hasil Uji Validitas

BUTIR SOAL	VALIDATOR					S1	S2	S3	S4	S5	ΣS	n(c-1)	V	HASIL
	I	II	III	IV	V									
Butir-1	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-2	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-3	3	4	3	3	4	2	3	2	2	3	12	15	0,80	Tidak Valid
Butir-4	3	4	4	4	4	2	3	3	3	3	14	15	0,93	Valid
Butir-5	4	4	3	4	4	3	3	2	3	3	14	15	0,93	Valid
Butir-6	4	3	4	3	3	3	2	3	2	2	12	15	0,80	Tidak Valid
Butir-7	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-8	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-9	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-10	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-11	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-12	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-13	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-14	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-15	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-16	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-17	3	4	3	3	4	2	3	2	2	3	12	15	0,80	Tidak Valid
Butir-18	3	4	3	3	4	2	3	2	2	3	12	15	0,80	Tidak Valid
Butir-19	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-20	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-21	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-22	4	3	4	3	4	3	2	3	2	3	13	15	0,9	Valid
Butir-23	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-24	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-25	3	4	3	4	4	2	3	2	3	3	13	15	0,9	Valid
Butir-26	4	4	4	3	4	3	3	3	2	3	14	15	0,93	Valid
Butir-27	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-28	4	4	3	3	4	3	3	2	2	3	13	15	0,9	Valid
Butir-29	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-30	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-31	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-32	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-33	4	4	4	4	4	3	3	3	3	3	15	15	1,00	Valid
Butir-34	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-35	3	4	3	3	4	2	3	2	2	3	12	15	0,80	Tidak Valid
Butir-36	4	4	4	3	4	3	3	3	2	3	14	15	0,93	Valid
Butir-37	4	4	4	3	4	3	3	3	2	3	14	15	0,93	Valid
Butir-38	4	4	4	3	4	3	3	3	2	3	14	15	0,93	Valid
Butir-39	3	4	4	3	4	2	3	3	2	3	13	15	0,9	Valid
Butir-40	4	4	4	3	4	3	3	3	2	3	14	15	0,93	Valid

Hasil uji validitas menunjukkan bahwa dari 40 butir soal yang dianalisis, sebanyak 35 butir (85%) dinyatakan valid berdasarkan perhitungan Aiken’s V ($V \geq 0,80$). Hal ini menunjukkan bahwa mayoritas soal telah sesuai dengan indikator pembelajaran dan mampu mengukur kompetensi yang dituju. Validitas ini mencerminkan bahwa soal-soal tersebut memiliki kesesuaian isi dan struktur dengan kompetensi dasar yang ingin dicapai dalam materi usaha, energi, dan pesawat sederhana. Penilaian dilakukan oleh lima validator dengan rentang skala 1–4, yang merupakan komposisi representatif dari guru dan mahasiswa pendidikan IPA, sehingga dapat memberikan penilaian yang objektif dan akurat terhadap relevansi isi soal.

Meskipun sebagian besar soal menunjukkan validitas yang baik, terdapat 5 butir soal (15%) yang tidak valid karena nilai Aiken's V-nya berada di bawah batas minimum yang ditetapkan ($V < 0,80$). Soal-soal ini perlu direvisi atau ditinjau ulang karena tidak menunjukkan hubungan yang signifikan dengan indikator pembelajaran. Ketidakvalidan tersebut bisa disebabkan oleh berbagai faktor, seperti rumusan soal yang ambigu, konteks soal yang tidak sesuai, atau pilihan jawaban (opsi) yang kurang representatif terhadap konsep yang hendak diukur. Dalam beberapa kasus, soal tidak menunjukkan daya ukur yang kuat terhadap kompetensi siswa, sehingga dikhawatirkan hasil penilaiannya tidak dapat mencerminkan kemampuan yang sebenarnya.

Dengan demikian, hasil uji validitas memberikan informasi awal yang penting dalam proses evaluasi kualitas instrumen. Soal-soal yang dinyatakan valid dapat digunakan dalam proses penilaian, sementara soal yang tidak valid perlu dilakukan revisi dengan memperbaiki redaksi, memperjelas konteks, atau menyusun ulang opsi jawaban (Abdullah dkk., 2024). Revisi ini bertujuan agar soal menjadi lebih representatif dan terarah dalam mengukur kompetensi siswa. Selain itu, hasil uji validitas ini juga memberikan gambaran bahwa prosedur telaah ahli dan pendekatan kuantitatif melalui Aiken's V efektif digunakan dalam mengidentifikasi kekuatan dan kelemahan soal dari aspek isi. Evaluasi ini menjadi pijakan penting sebelum melanjutkan pada analisis reliabilitas dan karakteristik statistik lainnya seperti tingkat kesukaran, daya pembeda, dan efektivitas distraktor.

Sebelum melanjutkan ke tahap uji reliabilitas instrumen, langkah penting yang dilakukan adalah pemilihan soal-soal yang akan diujikan kepada peserta didik. Proses ini diawali dengan penyusunan tabel spesifikasi soal (blueprint) yang berfungsi sebagai acuan untuk memetakan keterkaitan antara indikator pembelajaran, level kognitif, dan materi yang diujikan. Melalui tabel ini, peneliti dapat mengevaluasi sejauh mana setiap butir soal telah mencakup kompetensi yang diharapkan secara proporsional. Selanjutnya, dilakukan analisis terhadap hasil validitas isi untuk menentukan soal-soal yang layak digunakan. Soal yang dinyatakan valid langsung dipilih untuk diujikan, sementara soal yang tidak valid dan tidak memungkinkan untuk diperbaiki akan dibuang. Adapun soal yang validitasnya belum memenuhi kriteria namun masih memiliki potensi untuk direvisi akan ditinjau lebih lanjut dan diperbaiki sebelum diikutsertakan dalam uji coba kepada siswa. Proses seleksi ini memastikan bahwa hanya soal-soal yang berkualitas dan relevan yang digunakan dalam pengukuran kemampuan peserta didik.

Berdasarkan hasil validasi dan blueprint soal yang telah disusun, 40 butir soal dalam penelitian ini dirancang secara proporsional dan merata sesuai dengan cakupan materi pokok pada bab usaha, energi, dan pesawat sederhana. Distribusi soal telah dibagi secara adil berdasarkan sub-bab yaitu: Usaha (20%), Energi (30%), dan Pesawat Sederhana (50%), dengan mempertimbangkan kedalaman level kognitif mulai dari C1 (mengingat) hingga C6 (mencipta). Hal ini mencerminkan keseimbangan antara cakupan konten materi dan tuntutan berpikir tingkat tinggi, yang sesuai dengan karakteristik peserta didik jenjang SMP.

Tabel 2. Spesifikasi Soal (Blueprint Soal)

Sub-Bab dan Sub-Sub Bab	Persentase	Level Kognitif						Jumlah
		C1	C2	C3	C4	C5	C6	
A. Usaha	20%							8
A.1. Memindahkan Benda		1	1	1	1	0	0	4
A.2. Daya		0	1	0	1	1	1	4
B. Energi	30%							12
B.1. Energi Kinetik		1	1	1	0	0	1	4
B.2. Energi Potensial		1	1	1	1	0	0	4
B.3. Energi Mekanik		0	1	1	0	1	1	4
C. Pesawat Sederhana	50%							20
C.1. Manusia Membutuhkan Pesawat Sederhana		1	0	1	0	0	1	3
C.1.1. Meningkatkan Daya Angkat/Dorong		0	0	1	0	1	0	2
C.1.2. Meningkatkan Jarak Gaya		0	1	0	0	0	0	1
C.1.3. Mengubah Arah Gaya		0	0	1	0	0	0	1
C.2. Katrol (Tetap, Bebas, Majemuk)		1	1	0	0	1	1	4
C.2.2. Roda		1	0	0	1	0	0	2
C.2.3. Bidang Miring		0	0	1	0	1	0	2
C.2.4. Pengungkit (Jenis 1, 2, 3)		1	1	1	1	1	0	5
Jumlah Soal	100%	8	10	11	6	2	3	40

Dari 40 soal tersebut, sebanyak 25 soal dipilih untuk diujikan kepada siswa dalam tahap uji reliabilitas dan karakteristik butir soal. Pemilihan dilakukan berdasarkan hasil uji validitas Aiken's V ($V \geq 0,80$), serta mempertimbangkan relevansi soal terhadap indikator pembelajaran yang telah ditentukan dalam kisi-kisi. Beberapa soal yang awalnya tidak valid tetapi masih memiliki struktur soal yang kuat telah direvisi agar sesuai dengan kaidah penulisan soal yang baik. Revisi mencakup perbaikan redaksi, penyusunan opsi jawaban yang lebih representatif, dan penyelarasan konteks soal agar lebih sesuai dengan pengalaman siswa di lingkungan sekolah. Upaya ini penting agar instrumen yang digunakan tidak hanya valid secara statistik tetapi juga kontekstual terhadap karakteristik siswa SMP Islam Sultan Agung 1 Semarang sebagai lokasi penelitian. Berikut adalah tabel pemetaan 25 soal yang layak diujikan, yang memuat keterkaitan antara capaian pembelajaran, indikator soal, level kognitif, dan nomor soal yang dipilih:

Tabel 3. Soal yang diujikan

No	Sub Bab Materi	Indikator Soal	Level Kognitif	Nomor Soal
1	Usaha	Mengidentifikasi pengertian usaha dari peristiwa sehari-hari	C1	1
2	Usaha	Menjelaskan hubungan gaya dan perpindahan	C2	2
3	Usaha	Menghitung besar usaha berdasarkan data gaya dan perpindahan	C3	3
4	Usaha	Mengidentifikasi contoh usaha dalam kehidupan	C4	4
5	Usaha dan Daya	Menjelaskan perbedaan daya dari aktivitas berbeda	C2	5
6	Usaha dan Daya	Menghitung energi dari daya dan waktu	C4	6
7	Energi Kinetik	Mengidentifikasi benda yang memiliki energi kinetik	C1	9
8	Energi Kinetik	Menjelaskan hubungan kecepatan dan energi kinetik	C2	10
9	Energi Kinetik	Menerapkan konsep energi kinetik	C3	11
10	Energi Potensial	Menjelaskan faktor yang mempengaruhi energi potensial	C2	14
11	Energi Potensial	Menerapkan konsep energi potensial pegas	C3	15

12	Energi Mekanik	Menjelaskan energi mekanik sebagai gabungan energi	C2	17
13	Energi Mekanik	Menerapkan energi mekanik dalam gerak benda	C3	18
14	Energi Mekanik	Menilai efisiensi penggunaan energi mekanik	C6	20
15	Pesawat Sederhana	Mengidentifikasi alat sebagai pesawat sederhana	C1	21
16	Pesawat Sederhana	Menerapkan prinsip tuas saat membuka botol	C3	22
17	Katrol	Menjelaskan penggunaan katrol tetap dalam kehidupan	C3	24
18	Bidang Miring	Mengevaluasi efisiensi memindahkan benda berat	C5	25
19	Pengungkit Jenis 1	Menjelaskan manfaat pegangan panjang pada gunting	C2	26
20	Katrol Tetap	Menjelaskan penggunaan katrol saat upacara	C3	27
21	Katrol Tetap	Mengidentifikasi ciri katrol tetap dari gambar	C1	28
22	Katrol Bebas	Menilai kebenaran pernyataan tentang katrol bebas	C2	29
23	Katrol Majemuk	Mengevaluasi penggunaan katrol majemuk dalam proyek	C4	30
24	Roda dan Poros	Menyebutkan contoh alat roda berporos	C1	32

25	Pengungkit Jenis 3	Menganalisis penggunaan stapler sebagai pengungkit	C4	39
----	-----------------------	--	----	----

Sebanyak 25 butir soal yang digunakan dalam uji coba kepada peserta didik telah dipilih dengan seksama berdasarkan hasil uji validitas isi menggunakan Aiken's V dan telah mengalami proses revisi sesuai dengan masukan para validator. Soal-soal ini dipastikan telah memenuhi kriteria kelayakan dari aspek isi, konstruksi, dan bahasa. Selain itu, seluruh soal juga telah disesuaikan dengan indikator dalam tabel spesifikasi soal yang mencakup submateri Usaha, Energi, dan Pesawat Sederhana. Dalam proses seleksi dan perbaikan, setiap soal dipastikan memiliki keterkaitan yang kuat dengan tujuan pembelajaran yang telah ditentukan dan menyajikan konteks yang sesuai dengan kehidupannya nyata siswa SMP. Ini penting untuk menjaga agar soal tidak hanya mengukur hafalan, tetapi juga mampu mengevaluasi kemampuan berpikir siswa dalam konteks yang mereka pahami.

Meskipun dalam blueprint awal terdapat beberapa soal yang dirancang hingga level kognitif C5 dan C6 (misalnya menganalisis efisiensi alat atau merancang alat), berdasarkan masukan dari para validator, soal dengan level kognitif C5 dan C6 dianggap kurang sesuai untuk jenjang SMP karena dinilai terlalu tinggi dan tidak representatif dengan perkembangan kognitif peserta didik. Validator merekomendasikan agar level kognitif soal maksimal berada pada level C4 (analyze), yang dianggap sudah cukup menantang namun masih dalam jangkauan kemampuan berpikir siswa SMP. Oleh karena itu, meskipun dalam kisi-kisi terdapat soal berlevel C5–C6, soal-soal tersebut direvisi menjadi lebih sederhana atau diganti dengan soal baru yang berada pada level C1–C4 tanpa mengurangi esensi kompetensi yang diukur.

Pemilihan level kognitif hingga C4 juga sejalan dengan prinsip pembelajaran berdiferensiasi dan penyesuaian terhadap karakteristik peserta didik. Pada jenjang SMP, kemampuan berpikir abstrak siswa masih berkembang, sehingga penilaian yang terlalu kompleks justru bisa menimbulkan kebingungan dan tidak menggambarkan kemampuan yang sesungguhnya. Soal-soal pada level C1 hingga C4 telah mencakup kemampuan mulai dari mengingat konsep dasar (C1), memahami dan menjelaskan hubungan antar konsep (C2), menerapkan pengetahuan dalam konteks nyata (C3), hingga menganalisis perbedaan, perbandingan, atau hubungan sebab-akibat (C4). Dengan demikian, cakupan kognitif ini telah mencerminkan kedalaman berpikir yang memadai untuk mengukur kemampuan konseptual dan keterampilan berpikir kritis siswa secara representatif.

Dengan mempertimbangkan hasil validitas, revisi konstruksi soal, serta saran dari para validator, dapat disimpulkan bahwa 25 butir soal yang digunakan dalam uji coba ini layak untuk digunakan dalam mengukur capaian pembelajaran IPA khususnya pada materi usaha, energi, dan pesawat sederhana. Soal-soal tersebut tidak hanya memenuhi aspek teknis dan substansi, tetapi juga telah mempertimbangkan perkembangan kognitif siswa serta kesesuaian konteks lokal sekolah. Kombinasi antara validitas isi, kesesuaian indikator, cakupan level kognitif, dan revisi berbasis masukan ahli menjadikan 25 soal ini sebagai instrumen evaluasi yang tepat, proporsional, dan mendukung peningkatan kualitas pembelajaran IPA di tingkat SMP.

Setelah melalui tahap seleksi dan revisi berdasarkan hasil uji validitas isi serta saran dari para validator, sebanyak 25 butir soal yang telah dinyatakan layak kemudian diujikan kepada peserta didik untuk dianalisis lebih lanjut dari segi reliabilitasnya. Uji reliabilitas ini bertujuan

untuk mengetahui sejauh mana instrumen tes memiliki konsistensi internal, yaitu apakah setiap butir soal saling berkorelasi dan mampu mengukur kompetensi yang sama secara stabil. Analisis reliabilitas dilakukan menggunakan rumus KR-20 yang sesuai untuk soal pilihan ganda dengan skor dikotomis (benar-salah). Nilai reliabilitas yang diperoleh dari penghitungan akan memberikan gambaran mengenai sejauh mana instrumen tersebut dapat diandalkan dalam mengukur kemampuan siswa secara akurat. Hasil uji reliabilitas disajikan dalam tabel berikut.

Tabel 4. Hasil Uji Reliabilitas

Aspek	Hasil
Jumlah Soal (k)	25
Varian Total Skor (S^2)	15,52
$\Sigma(pq)$	3,90
KR-20	0,78
Interpretasi Reliabilitas	Reliabel

Nilai reliabilitas instrumen yang diperoleh sebesar 0,78 menunjukkan bahwa 25 butir soal yang digunakan tergolong reliabel. Dalam interpretasi standar, nilai KR-20 $\geq 0,70$ menandakan bahwa soal memiliki konsistensi internal yang baik (Juliani & Erita, 2023). Ini berarti bahwa setiap soal dalam instrumen tersebut saling berkorelasi dan mampu mengukur kompetensi yang sama secara konsisten. Dengan kata lain, jika soal yang sama diberikan kepada kelompok siswa serupa dalam kondisi yang sama, maka hasilnya cenderung stabil dan tidak berubah secara signifikan.

Reliabilitas sebesar 0,78 juga mencerminkan bahwa variasi skor yang dihasilkan oleh siswa lebih disebabkan oleh perbedaan kemampuan individu, bukan karena ketidakpastian atau kesalahan dalam penyusunan soal. Hal ini sangat penting dalam konteks evaluasi pembelajaran, karena tujuan utama dari penggunaan instrumen tes adalah untuk mengukur kemampuan siswa secara adil dan akurat. Nilai reliabilitas yang tinggi memberikan kepercayaan kepada guru bahwa instrumen yang digunakan dapat menjadi dasar pengambilan keputusan pembelajaran secara objektif, misalnya dalam penentuan remedial, klasifikasi kemampuan, maupun evaluasi efektivitas pengajaran.

Dengan nilai KR-20 sebesar 0,78, dapat disimpulkan bahwa soal-soal dalam penelitian ini telah memenuhi kriteria kualitas instrumen yang baik dari sisi reliabilitas. Meskipun demikian, masih terdapat ruang untuk penyempurnaan, seperti peningkatan proporsi soal dengan daya pembeda tinggi atau pengayaan variasi konteks soal. Namun secara keseluruhan, hasil ini sudah sangat layak digunakan untuk keperluan evaluasi hasil belajar pada mata pelajaran IPA, khususnya materi usaha, energi, dan pesawat sederhana di tingkat SMP. Reliabilitas ini juga menjadi indikator bahwa proses validasi dan revisi soal yang telah dilakukan sebelumnya telah berhasil meningkatkan kualitas instrumen secara signifikan.

Setelah instrumen dinyatakan reliabel berdasarkan hasil perhitungan KR-20, tahap selanjutnya yang dilakukan adalah analisis butir soal secara lebih mendalam untuk mengevaluasi kualitas tiap-tiap soal dari tiga aspek utama, yaitu tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Analisis ini bertujuan untuk menilai sejauh mana setiap butir soal mampu mengukur kemampuan siswa secara adil, membedakan antara siswa berkemampuan tinggi dan rendah, serta memiliki opsi jawaban yang fungsional dalam mengungkap pemahaman maupun miskonsepsi peserta didik (Alamsyah & Sudrajat, 2021). Ketiga aspek tersebut menjadi indikator penting dalam menentukan apakah suatu soal perlu

dipertahankan, direvisi, atau diganti dalam rangka meningkatkan mutu instrumen evaluasi pembelajaran. Adapun hasil analisis butir soal disajikan dan dibahas pada bagian berikut ini.

Tabel 5. Hasil Perhitungan Analisis Butir Soal (25 soal)

NOMOR SOAL	EFEKTIVITAS DISTRAKTOR				T. KESULITAN	KATEGORI	DAYA PEMBEDA	
	A	B	C	D			INDEKS	KATEGORI
1	4%	0%	96%	0%	0,96	MUDAH	0,00	RENDAH
2	92%	8%	0%	0%	0,92	MUDAH	0,29	CUKUP
3	12%	58%	15%	15%	0,54	SEDANG	0,86	SANGAT TINGGI
4	12%	35%	42%	12%	0,35	SEDANG	0,86	SANGAT TINGGI
5	50%	31%	12%	8%	0,50	SEDANG	0,43	TINGGI
6	4%	73%	19%	4%	0,73	MUDAH	0,14	RENDAH
7	31%	23%	35%	12%	0,35	SEDANG	0,14	RENDAH
8	0%	4%	8%	88%	0,88	MUDAH	0,14	RENDAH
9	54%	23%	19%	4%	0,54	SEDANG	1,00	SANGAT TINGGI
10	4%	0%	96%	0%	0,96	MUDAH	0,14	RENDAH
11	73%	15%	12%	0%	0,73	MUDAH	0,14	RENDAH
12	0%	4%	92%	4%	0,96	MUDAH	0,00	RENDAH
13	4%	0%	4%	92%	0,92	MUDAH	0,29	CUKUP
14	73%	15%	4%	8%	0,73	MUDAH	0,71	SANGAT TINGGI
15	46%	23%	15%	15%	0,23	SULIT	0,14	RENDAH
16	4%	23%	4%	69%	0,69	SEDANG	0,57	TINGGI
17	27%	50%	23%	0%	0,58	SEDANG	0,57	TINGGI
18	88%	4%	8%	0%	0,88	MUDAH	0,29	CUKUP
19	100%	0%	0%	0%	1,00	MUDAH	0,00	RENDAH
20	92%	0%	8%	0%	0,92	MUDAH	0,29	CUKUP
21	50%	31%	12%	8%	0,50	SEDANG	0,57	TINGGI
22	0%	0%	85%	15%	0,85	MUDAH	0,43	TINGGI
23	15%	23%	19%	42%	0,38	SEDANG	0,43	TINGGI
24	65%	19%	8%	8%	0,65	SEDANG	0,71	SANGAT TINGGI
25	4%	8%	88%	0%	0,88	MUDAH	0,29	CUKUP

Tabel 6. Hasil Analisis Tingkat Kesulitan, Daya Pembeda, dan Efektivitas Distraktor

No	Tingkat Kesukaran	Daya Pembeda	Efektivitas Distraktor
1	Mudah (0,96)	Rendah (0,00)	Tidak efektif (A, B, D < 5%)
2	Mudah (0,92)	Cukup (0,29)	Efektif (A dan B ≥ 5%)
3	Sedang (0,54)	Sangat Tinggi (0,86)	Efektif (semua opsi ≥ 5%)
4	Sedang (0,35)	Sangat Tinggi (0,86)	Efektif (semua opsi ≥ 5%)
5	Sedang (0,50)	Tinggi (0,43)	Efektif (semua opsi ≥ 5%)
6	Mudah (0,73)	Rendah (0,14)	Efektif (semua opsi ≥ 5%)
7	Sedang (0,35)	Rendah (0,14)	Efektif (semua opsi ≥ 5%)
8	Mudah (0,88)	Rendah (0,14)	Tidak efektif (A dan B < 5%)
9	Sedang (0,54)	Sangat Tinggi (1,00)	Efektif (semua opsi ≥ 5%)
10	Mudah (0,96)	Rendah (0,14)	Tidak efektif (A, B, D < 5%)
11	Mudah (0,73)	Rendah (0,14)	Efektif (semua opsi ≥ 5%)
12	Mudah (0,96)	Rendah (0,00)	Tidak efektif (A, B, D < 5%)
13	Mudah (0,92)	Cukup (0,29)	Tidak efektif (A, B, D < 5%)
14	Mudah (0,73)	Sangat Tinggi (0,71)	Efektif (semua opsi ≥ 5%)
15	Sukar (0,23)	Rendah (0,14)	Efektif (semua opsi ≥ 5%)
16	Sedang (0,69)	Tinggi (0,57)	Efektif (semua opsi ≥ 5%)
17	Sedang (0,58)	Tinggi (0,57)	Efektif (semua opsi ≥ 5%)
18	Mudah (0,88)	Cukup (0,29)	Tidak efektif (B, C, D < 5%)
19	Mudah (1,00)	Rendah (0,00)	Tidak efektif (B, C, D < 5%)
20	Mudah (0,92)	Cukup (0,29)	Tidak efektif (B, C, D < 5%)
21	Sedang (0,50)	Tinggi (0,57)	Efektif (semua opsi ≥ 5%)
22	Mudah (0,85)	Tinggi (0,43)	Efektif (semua opsi ≥ 5%)
23	Sedang (0,38)	Tinggi (0,43)	Efektif (semua opsi ≥ 5%)
24	Sedang (0,65)	Sangat Tinggi (0,71)	Efektif (semua opsi ≥ 5%)
25	Mudah (0,88)	Cukup (0,29)	Tidak efektif (A dan C < 5%)

Hasil analisis menunjukkan bahwa tingkat kesukaran 25 soal tersebar cukup merata dengan dominasi pada kategori mudah dan sedang. Sebanyak 13 soal termasuk kategori mudah, 11 soal sedang, dan hanya 1 soal tergolong sukar. Soal-soal kategori mudah mendominasi, yang menandakan bahwa sebagian besar peserta didik dapat menjawab benar. Ketidakseimbangan ini mengindikasikan bahwa sebagian besar soal belum mampu menantang kemampuan berpikir kritis peserta didik, karena soal yang terlalu mudah cenderung hanya mengukur kemampuan dasar dan tidak efektif untuk menilai pemahaman mendalam atau kemampuan berpikir tingkat tinggi. Distribusi ini sebenarnya cukup sesuai untuk penilaian formatif maupun sumatif awal, yang bertujuan untuk mengetahui pemahaman dasar peserta didik. Namun, untuk evaluasi pembelajaran yang lebih mendalam dan mengukur kemampuan berpikir tingkat tinggi (Higher Order Thinking Skills atau HOTS), proporsi soal dengan tingkat kesukaran sedang hingga sukar perlu ditingkatkan. Soal-soal yang menantang mendorong siswa berpikir lebih kritis, dan membantu guru memetakan kemampuan secara lebih akurat. Idealnya, distribusi soal sebaiknya mencakup sekitar 30% mudah, 40% sedang, dan 30% sulit agar lebih representatif dan adil bagi seluruh jenjang kemampuan siswa (Zuhriyah dkk., 2025). Ketidakseimbangan dalam distribusi tingkat kesukaran ini menjadi masukan penting bagi guru untuk menyusun soal evaluasi berikutnya agar lebih representatif dan mampu mengukur pencapaian kompetensi siswa secara komprehensif.

Secara umum, daya pembeda soal dalam penelitian ini menunjukkan kualitas yang bervariasi. Dari 25 soal yang dianalisis, 18 soal memiliki daya pembeda yang tergolong cukup hingga sangat tinggi. Hal ini berarti bahwa mayoritas soal memiliki kemampuan yang baik dalam membedakan siswa berkemampuan tinggi dan rendah. Daya pembeda tinggi menunjukkan bahwa soal berhasil mengidentifikasi perbedaan tingkat pemahaman siswa terhadap materi yang diujikan. Dengan demikian, sebagian besar soal dapat dikategorikan sebagai efektif dalam konteks asesmen diagnostik, formatif, maupun sumatif, karena memberikan informasi yang bermanfaat bagi guru dalam mengambil keputusan lanjutan seperti remedial atau pengayaan.

Berdasarkan hasil analisis, sebanyak 10 soal dari 25 soal memiliki daya pembeda rendah (< 0,20). Untuk meningkatkan kualitas evaluasi pembelajaran, guru perlu meninjau kembali soal-soal dengan daya pembeda rendah. Revisi dapat dilakukan dengan memperbaiki rumusan soal agar lebih jelas dan menantang, memperkuat pengecoh agar mampu menarik siswa yang belum memahami materi, atau menyusun soal baru yang lebih seimbang tingkat kesukarannya (Harahap, 2024). Soal yang memiliki daya pembeda tinggi sangat penting dalam evaluasi karena memberikan informasi yang tajam mengenai tingkat penguasaan siswa terhadap materi ajar. Soal-soal seperti ini dapat membantu guru dalam mengidentifikasi kelompok siswa yang membutuhkan remedial maupun pengayaan, serta menjadi dasar pengambilan keputusan pembelajaran yang lebih tepat sasaran. Dengan demikian, perbaikan terhadap aspek daya pembeda merupakan langkah penting dalam menjamin keadilan dan efektivitas proses evaluasi.

Meskipun sebagian besar soal memiliki daya pembeda yang baik, terdapat 9 soal (28%) yang daya pembeda-nya rendah (misalnya soal nomor 1, 6, 8, 10, 11, 12, dan 19). Soal-soal ini tidak dapat secara efektif membedakan siswa yang menguasai materi dengan yang belum. Hal ini dapat disebabkan oleh beberapa faktor, antara lain: (1) soal terlalu mudah atau terlalu sulit, sehingga semua siswa menjawab benar atau salah; (2) kunci jawaban rancu atau membingungkan; dan (3) materi dalam soal mungkin belum sepenuhnya dipahami oleh peserta

didik. Soal dengan daya pembeda rendah ini perlu dievaluasi lebih lanjut dan direvisi agar fungsinya sebagai alat ukur kemampuan siswa menjadi lebih optimal.

Efektivitas distraktor juga menjadi aspek penting yang dianalisis. Dari hasil analisis, ditemukan bahwa 88% opsi jawaban (distraktor) efektif, yaitu dipilih oleh minimal 5% peserta didik, menandakan bahwa sebagian besar pengecoh bekerja dengan baik dalam menggiring siswa yang belum memahami konsep. Artinya sebagian besar pengecoh dalam soal berhasil menarik perhatian peserta didik yang belum menguasai materi. Distraktor yang efektif sangat penting dalam soal pilihan ganda karena menunjukkan bahwa opsi jawaban salah disusun secara logis dan mampu mengecoh siswa yang ragu atau tidak memahami materi (Jannah, 2019). Ini juga menjadi indikator bahwa soal memiliki kualitas yang baik karena dapat mengungkap kesalahan pemahaman siswa secara lebih mendalam. Efektivitas distraktor turut berkontribusi pada validitas dan daya pembeda soal, karena semakin fungsional distraktor, semakin besar kemungkinan soal membedakan dengan baik antara siswa yang menguasai dan yang tidak menguasai materi (Jannah, 2019). Namun demikian, terdapat beberapa soal dengan distraktor yang tidak efektif, seperti pada soal nomor 1, 8, 10, 12, 13, 18, 19, 20, dan 25. Distraktor yang tidak efektif umumnya terlalu mudah dikenali sebagai salah, tidak relevan dengan materi, atau memiliki redaksi yang membingungkan. Distraktor yang seperti ini perlu direvisi agar lebih logis dan menarik, serta dapat berfungsi maksimal untuk mengungkap pemahaman siswa.

Dari 12% distraktor yang tidak efektif tersebut, yaitu pilihan jawaban yang tidak dipilih oleh siswa sama sekali atau dipilih oleh kurang dari 5% peserta. Distraktor yang tidak efektif biasanya disebabkan oleh redaksi jawaban yang tidak logis, terlalu mudah dikenali sebagai salah, atau terlalu mirip dengan kunci jawaban (Jannah, 2019). Distraktor semacam ini perlu direvisi atau diganti agar seluruh opsi jawaban memiliki peluang yang seimbang untuk dipilih, terutama oleh siswa yang belum memahami materi dengan baik. Revisi ini penting untuk meningkatkan kualitas soal secara keseluruhan, karena distraktor yang lemah dapat membuat soal menjadi terlalu mudah ditebak dan menurunkan daya ukur dari instrumen evaluasi.

Distraktor yang efektif tidak hanya meningkatkan daya pembeda, tetapi juga memperkuat validitas soal. Oleh karena itu, penting bagi guru dan pembuat soal untuk melakukan perbaikan pada opsi jawaban yang tidak bekerja sesuai harapan. Distraktor yang baik harus memiliki kemiripan struktur dengan kunci jawaban, tidak terlalu mencolok sebagai jawaban salah, dan berbasis pada miskonsepsi umum peserta didik. Dengan merevisi distraktor yang lemah, kualitas soal secara keseluruhan akan meningkat, sehingga evaluasi menjadi lebih akurat, adil, dan dapat dipercaya sebagai cerminan kemampuan siswa yang sesungguhnya.

Distraktor yang efektif menunjukkan bahwa pilihan jawaban yang salah berhasil mengecoh peserta yang tidak memahami materi, sehingga membantu membedakan antara siswa yang paham dan tidak paham. Distraktor yang tidak efektif perlu diperbaiki karena tidak memberikan daya tarik, sehingga peserta didik mudah menebak jawaban yang benar. Distraktor yang baik harus logis dan serupa dalam struktur dengan kunci jawaban. Efektivitas pengecoh meningkatkan kualitas instrumen soal, membuat evaluasi lebih adil dan menantang. Hal ini penting dalam asesmen formatif dan sumatif untuk memperoleh data hasil belajar yang akurat.

Pada hasil analisis butir soal menunjukkan bahwa sebagian besar instrumen yang digunakan dalam penelitian ini sudah berkualitas baik, khususnya dalam aspek validitas, reliabilitas ($KR-20 = 0,78$), dan efektivitas distraktor. Namun, masih ditemukan beberapa kelemahan pada aspek daya pembeda dan distribusi tingkat kesukaran. Hal ini menjadi catatan

penting untuk penyusunan instrumen evaluasi berikutnya agar mencakup lebih banyak soal tingkat kesukaran sedang hingga sulit, serta memperbaiki butir-butir dengan daya pembeda rendah dan distraktor tidak efektif. Dengan melakukan evaluasi dan revisi secara berkelanjutan, kualitas evaluasi pembelajaran IPA dapat ditingkatkan sehingga benar-benar mendukung pengembangan kemampuan siswa secara menyeluruh.

Soal yang valid mampu mengukur kompetensi yang ditargetkan, sehingga memberikan informasi akurat mengenai pencapaian belajar siswa (Arbiatin & Mulabbiyah, 2020). Soal yang reliabel menunjukkan konsistensi hasil evaluasi dan dapat menjadi dasar penilaian yang objektif (Fatayah dkk., 2022). Selain itu, distribusi tingkat kesukaran yang seimbang penting untuk mengakomodasi seluruh level kemampuan peserta didik, yang merupakan prinsip dasar dalam penyusunan soal berkualitas (Masulili et al., 2021).

Begitu pula soal dengan daya pembeda tinggi mampu membedakan peserta didik yang sudah menguasai materi dengan yang belum, sehingga sangat membantu dalam pengambilan keputusan tindak lanjut pembelajaran (Hapsari dkk., 2023). Distraktor yang efektif pun memiliki peran penting dalam mencegah jawaban acak dan meningkatkan kualitas interpretasi hasil tes (Halik dkk., 2019). Dengan demikian, hasil evaluasi yang berbasis pada analisis butir soal tidak hanya memperbaiki kualitas tes itu sendiri, tetapi juga berkontribusi terhadap peningkatan mutu pembelajaran melalui asesmen yang valid, reliabel, dan adil (Mawardi dkk., 2023).

Pelaksanaan analisis butir soal secara kuantitatif dalam evaluasi pembelajaran IPA, khususnya pada materi usaha, energi, dan pesawat sederhana ini, memberikan implikasi signifikan terhadap peningkatan kualitas instrumen yang digunakan. Melalui pengujian validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor, guru dapat memperoleh informasi objektif dan mendalam mengenai sejauh mana soal yang digunakan benar-benar mengukur kompetensi siswa secara tepat. Proses ini membantu memastikan bahwa setiap butir soal tidak hanya sesuai dengan indikator pembelajaran, tetapi juga memiliki struktur dan opsi jawaban yang mampu menggambarkan kemampuan berpikir siswa secara akurat. Dengan demikian, evaluasi pembelajaran tidak hanya bersifat administratif, tetapi menjadi bagian penting dari upaya perbaikan dan pengembangan pembelajaran yang berkelanjutan.

Implikasi lain dari analisis ini adalah meningkatnya kesadaran guru terhadap pentingnya menyusun soal berdasarkan data dan bukti empiris, bukan sekadar intuisi atau kebiasaan. Dengan hasil analisis, guru dapat mengidentifikasi soal mana yang terlalu mudah, terlalu sulit, atau tidak mampu membedakan antara siswa yang paham dan yang tidak. Misalnya, soal dengan daya pembeda rendah dan distraktor tidak efektif dapat direvisi agar lebih menantang dan fungsional. Hal ini sangat penting dalam konteks pembelajaran IPA, yang tidak hanya menekankan penguasaan konsep, tetapi juga mendorong pengembangan kemampuan berpikir kritis dan pemecahan masalah. Analisis kuantitatif membantu guru untuk menghindari penilaian yang bias dan memberikan umpan balik yang lebih bermakna bagi siswa, sehingga hasil evaluasi benar-benar mencerminkan capaian belajar.

Secara keseluruhan, analisis butir soal memberikan kontribusi nyata dalam meningkatkan mutu evaluasi pembelajaran IPA. Instrumen evaluasi yang berkualitas tidak hanya meningkatkan keakuratan pengukuran hasil belajar, tetapi juga mendukung perencanaan pembelajaran yang lebih efektif dan responsif terhadap kebutuhan siswa. Dalam konteks penelitian ini, penerapan evaluasi kuantitatif pada butir soal pilihan ganda terbukti menjadi strategi yang relevan dan aplikatif untuk meningkatkan kualitas instrumen pembelajaran IPA

pada materi usaha, energi, dan pesawat sederhana. Langkah ini sejalan dengan tujuan pendidikan modern yang menekankan pentingnya asesmen sebagai bagian integral dari proses pembelajaran, bukan hanya sebagai alat pengukur akhir. Dengan analisis yang sistematis dan berkelanjutan, guru dapat mengembangkan instrumen evaluasi yang valid, reliabel, dan adil, yang pada akhirnya akan berdampak positif terhadap kualitas pembelajaran secara keseluruhan.

E. KESIMPULAN

Hasil penelitian menunjukkan bahwa evaluasi kuantitatif terhadap butir soal pilihan ganda memberikan kontribusi penting dalam meningkatkan kualitas instrumen evaluasi pembelajaran IPA. Dari 40 soal yang disusun, 35 soal dinyatakan valid berdasarkan uji Aiken's V, dan 25 soal yang telah direvisi dan disesuaikan dengan karakteristik siswa dinyatakan layak untuk diujikan. Selanjutnya, uji reliabilitas menghasilkan nilai KR-20 sebesar 0,78, yang menunjukkan bahwa instrumen memiliki konsistensi internal yang baik dan dapat diandalkan dalam mengukur kemampuan siswa secara objektif dan akurat. Hal ini membuktikan bahwa prosedur validasi dan revisi yang dilakukan telah berhasil meningkatkan kualitas soal secara signifikan.

Analisis karakteristik soal menunjukkan bahwa sebagian besar soal memiliki tingkat kesukaran yang dominan pada kategori mudah dan sedang, serta daya pembeda yang cukup hingga tinggi. Meskipun demikian, terdapat beberapa soal dengan daya pembeda rendah dan distraktor tidak efektif yang perlu diperbaiki. Analisis ini menjadi penting karena memberikan informasi yang mendalam tentang kekuatan dan kelemahan instrumen secara menyeluruh. Soal-soal yang memiliki distraktor tidak efektif dan daya pembeda rendah dapat direvisi untuk lebih mengukur pemahaman konseptual dan keterampilan berpikir kritis siswa. Dengan demikian, evaluasi kuantitatif butir soal tidak hanya membantu dalam penyaringan soal yang layak, tetapi juga mendorong peningkatan mutu evaluasi secara berkelanjutan.

Berdasarkan temuan penelitian ini, disarankan agar guru dan pengembang instrumen evaluasi melakukan analisis butir soal secara berkala sebelum soal digunakan dalam asesmen pembelajaran. Proses validasi dan analisis kuantitatif seperti uji validitas isi, reliabilitas, indeks kesukaran, daya pembeda, dan efektivitas distraktor sebaiknya dijadikan bagian dari standar penyusunan instrumen evaluasi. Selain itu, penting bagi guru untuk menyusun soal dengan variasi tingkat kesukaran dan level kognitif yang sesuai dengan kemampuan peserta didik, agar instrumen mampu mencerminkan capaian pembelajaran secara lebih adil dan menyeluruh. Dengan menerapkan analisis kuantitatif secara berkelanjutan, mutu pembelajaran IPA, khususnya pada materi usaha, energi, dan pesawat sederhana, dapat terus ditingkatkan.

DAFTAR PUSTAKA

- Abdullah, G., Apriyanto, A., Patahuddin, A., Janah, R., Dia, E. E., Retnoningsih, R., ... & Setyaningrum, V. (2024). *Buku Ajar Evaluasi Pembelajaran*. PT. Sonpedia Publishing Indonesia.
- Alamsyah, H. S., & Sudrajat, H. (2021). *Belajar Pembelajaran Di Sekolah Dasar*. Deepublish.
- Amelia, N., & Erita, S. (2024). Eksplorasi Validitas dan Reliabilitas Soal Pemahaman Konsep Dalam Asesmen Pembelajaran. *Jurnal Bima: Pusat Publikasi Ilmu Pendidikan bahasa dan Sastra*, 2(1), 222-231.

- Ansyah, Y. A. U., Alfianita, A., Syahkira, H.P., & Syahrial, S. (2024). Peran Evaluasi Pembelajaran pada Mata Pelajaran Matematika Kelas V Sekolah Dasar. *Indictika: Jurnal Inovasi Pendidikan Matematika*, 6 (2), 173-184.
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 968–976.
- Arbiatin, E., & Mulabbiyah. (2020). Analisis Kelayakan Butir Soal Tes Penilaian Akhir Semester Mata Pelajaran Matematika Kelas VI di SDN 19 Ampenan Tahun Pelajaran 2019/2020. *El-Midad: Jurnal PGMI*, 12(2), 146–171.
- Fatayah, F., Yuliana, I. F., & Muf'idah, L. (2022). Analisis Validitas dan Reliabilitas dalam Mendukung Ketuntasan Belajar Model STEM. *Buana Pendidikan*, 18(1), 49–60.
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2020). Analisis Butir Soal Ulangan Harian IPA Menggunakan Software Anates pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76.
- Halik, A. S., Mania, S., & Nur, F. (2019). Analisis Butir Soal Ujian Akhir Sekolah (UAS) Mata Pelajaran Matematika SMP Negeri 36 Makassar. *Al-Asma: Journal of Islamic Education*, 1(1), 11–17.
- Hapsari, T. P. R. N., Krissandi, A. D. S., & Sumarwati. (2023). Analisis Butir Soal Model Klasik Pada Penilaian Harian Bersama (PHB) Mata Pelajaran Bahasa Indonesia. *Indonesian Journal of Education and Learning*, 7(1), 31–46.
- Harahap, A. (2024). *Evaluasi Pembelajaran Berbasis Hots Dalam Kurikulum Merdeka*. Penerbit Adab.
- Jannah, M. (2019). *Analisis Butir Soal Ujian Akhir Semester (UAS) Mata Pelajaran Fisika pada Ujian Semester Genap Kelas XI Tahun Pelajaran 2017/2018 di SMAN 16 Banda Aceh* (Doctoral dissertation, UIN Ar-Raniry Banda Aceh).
- Juliani, R. P., & Erita, S. (2023). Analisis Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Kritis dalam Konteks Sekolah Menengah. *JEID: Journal of Educational Integration and Development*, 3(3), 169-179.
- Magdalena, I., Anggraini, I. A., & Khoiriah, S. (2021). Analisis Daya Pembeda, dan Taraf Kesukaran pada Soal Bilangan Romawi Kelas 4 SDN Tobat 1 Balaraja. *Jurnal Pendidikan Dan Ilmu Sosial*, 3(1), 151–158.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis Butir Soal Ujian Akhir Sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274–284.
- Masulili, R. R., Dama, L., & Abdul, A. (2021). Analisis Butir Soal Semester Genap Mata Pelajaran Biologi Kelas XI. *Jambura Edu Biosfer Journal*, 3(2), 58–67.
- Mawardi, M. S., Fuady, A., & Sunismi. (2023). Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum. *Wahana: Tridarma Perguruan Tinggi*, 75(1), 31–41.
- Nabil, N. R. A., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis indeks Aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *Jurnal Penelitian Pendidikan*, 25(2), 184-191.
- Pratiwi, L. E., & Rofi'i, H. (2023). Analisis Soal Sumatif IPA Mengenai Gaya pada Peserta Didik Kelas IV di SD YP Nasional. *Didaktika: Jurnal Kependidikan*, 13(1), 45–55.
- Sa'edi, M., Gaffar, A., & Fajriyah, F. (2024). Menerapkan Evaluasi Pembelajaran Holistik Di Tengah Perubahan Kurikulum (Studi Kasus Di Madrasah Tsanawiyah Al-Falah

Pasongsongan Sumenep: Implementing Holistic Learning Evaluation In The Midst Of Curriculum Changes (Case Study At Madrasah Tsanawiyah Al-Falah Pasongsongan Sumenep. *Journal Al-Ilmu*, 2(8), 24-36.

- Safitri, I., Lestarani, D., Imtikhanah, R. D. N. W., Akbarini, N. R., Sari, M. W., Fitrah, M., & Hapsan, A. (2024). *Teori Pengukuran dan Evaluasi*. CV. Ruang Tentor.
- Sarosa, S. (2021). *Analisis data penelitian kualitatif*. Pt Kanisius.
- Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis kualitas instrumen untuk mengukur kemampuan pemecahan masalah siswa selama pembelajaran daring di masa pandemi. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 130-139.
- Sopiah, A., Sidauruk, S., & Asi, N. B. (2019). Kualitas Soal Penilaian Akhir Semester (PAS) Buatan Guru Mata Pelajaran Kimia Kelas X IPA SMA Negeri Di Kabupaten Seruyan Pada Semester Ganjil Tahun Ajaran 2018/2019. *Jurnal Ilmiah Kanderang Tingang*, 10(2), 110–126.
- Suardipa, I. P., & Primayana, K. H. (2023). Peran desain evaluasi pembelajaran untuk meningkatkan kualitas pembelajaran. *Widyacarya: Jurnal Pendidikan, Agama dan Budaya*, 4(2), 88-100.
- Surbakti, I. W. (2025). Analisis Kualitas Butir Soal Pada Uji Coba Evaluasi Pembelajaran Matematika. *CENDEKIA: Jurnal Ilmu Pengetahuan*, 5(1), 116-125.
- Qohar, M. A., & Fauziyah, F. (2024). Analisis Validitas dan Reliabilitas Soal Aljabar untuk Siswa Sekolah Menengah Pertama: Studi Kuantitatif. *Jurnal Tarbiyatuna: Jurnal Kajian Pendidikan, Pemikiran dan Pengembangan Pendidikan Islam*, 5(2), 128-137.
- Zuhriyah, I. A., Wahyuni, H., Al Aluf, W., Nasir, M., Bukhori, I., Firdaus, Y. I., ... & Nazihah, N. (2025). *Pengembangan Evaluasi Pembelajaran Di Madrasah Ibtidaiyah (MI)*. PT. Penerbit Qriset Indonesia.