OPEN ACCESS

Tadrisia: Education and Islamic Studies | Vol. 1 No. 4 (2026)

Research Article

Analisis Validitas, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Instrumen Tes IPA Materi Lingkungan Kelas VII Berbasis SPSS

Sayyidati Nur Hanna[1*], Hanik Malichatin[2]
Universitas Islam Negeri Sunan Kudus, Indonesia
Email: [1)] sayyidati@mhs.uinsuku.ac.id,[2)] hanikm@uinsuku.ac.id

Received:	Revised:	Accepted:	Online:
November 14 , 2026	November 17, 2026	November 24, 2026	November 25, 2026

Abstract

This study aims to analyze the quality of a Grade VII science test on environmental topics in terms of validity, reliability, item difficulty, and discrimination power. The study employed a descriptive quantitative approach involving 30 Grade VII junior high school students from several schools in Jepara Regency, Indonesia. Data were collected through an online test consisting of 10 multiple-choice questions administered using Google Forms. Students' report-card scores obtained through the Jepara online report-card service were used as supporting information about their learning achievement. The test items were analyzed using SPSS version 27. The results showed that 6 items (60%) were valid and 4 items (40%) were invalid. The Cronbach's Alpha value was 0.351, indicating a low level of reliability. All 10 items (100%) were classified as having moderate difficulty. In terms of discrimination power, 4 items were classified as good, 2 as moderate, 2 required revision, and 2 should be discarded or replaced. These findings indicate that the instrument still requires improvement

Keywords: Validity, Reliability, Item Difficulty, Discrimination, Science Test

Abstrak

Penelitian ini bertujuan untuk menganalisis kualitas instrumen tes IPA materi lingkungan kelas VII berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan melibatkan 30 siswa kelas VII SMP dari beberapa sekolah di Kabupaten Jepara. Data dikumpulkan melalui tes daring menggunakan Google Form yang terdiri atas 10 soal pilihan ganda. Data nilai rapor yang diperoleh melalui layanan rapor online Jepara digunakan sebagai informasi pendukung mengenai capaian belajar siswa. Analisis instrumen dilakukan menggunakan aplikasi SPSS versi 27. Hasil analisis menunjukkan bahwa 6 butir soal (60%) dinyatakan valid dan 4 butir soal (40%) tidak valid. Nilai Cronbach's Alpha sebesar 0,351 menunjukkan reliabilitas instrumen berada pada kategori rendah. Seluruh butir soal (100%) termasuk kategori tingkat kesukaran sedang. Berdasarkan daya pembeda, 4 butir soal termasuk kategori baik, 2 butir soal kategori sedang, 2 butir soal perlu direvisi, dan 2 butir soal perlu dibuang atau diganti. Hasil penelitian menunjukkan bahwa instrumen masih memerlukan perbaikan agar kualitasnya lebih baik.

Kata Kunci: Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, Instrumen Tes IPA

Copyright: © 2026 by the authors.
This is an open access article
distributed under the terms and conditions of the CC BY 4.0.

1. Pendahuluan

Clearly Pendidikan berkualitas merupakan pilar utama dalam mencetak generasi yang kompeten dan responsif terhadap tantangan zaman. Evaluasi memiliki kedudukan yang sangat strategis dalam sistem pembelajaran karena berfungsi sebagai alat ukur untuk mengetahui tingkat ketercapaian tujuan instruksional dan keberhasilan proses belajar mengajar (Saputra et al., 2025). Evaluasi yang objektif tidak hanya memberikan gambaran mengenai tingkat penguasaan materi oleh peserta didik, tetapi juga menjadi umpan balik (*feedback*) bagi pendidik untuk memperbaiki strategi pembelajaran yang telah diterapkan. Pendidik memerlukan alat ukur atau instrumen tes hasil belajar yang berkualitas dan memenuhi syarat-syarat teoretis maupun empiris agar fungsi evaluasi tersebut dapat berjalan optimal.

Ilmu Pengetahuan Alam (IPA) pada jenjang Sekolah Menengah Pertama (SMP) merupakan salah satu mata pelajaran yang menuntut pemahaman konsep yang mendalam serta aplikatif. Materi lingkungan menjadi salah satu materi krusial dalam kurikulum IPA kelas VII. Materi ini mencakup pemahaman tentang ekosistem, interaksi makhluk hidup dengan lingkungannya, serta fenomena pencemaran dan perubahan lingkungan. Penguasaan konsep materi lingkungan sangat penting bagi siswa untuk pencapaian akademis serta menumbuhkan kesadaran ekologis dan kepedulian terhadap kelestarian alam sejak dini (Tomi et al., 2020). Instrumen tes yang digunakan untuk mengevaluasi hasil belajar siswa harus mampu menyaring kemampuan berpikir logis dan analisis siswa secara tepat mengingat kompleksnya materi ini.

Fakta di lapangan menunjukkan bahwa sebagian besar instrumen tes yang digunakan di sekolah, baik untuk Penilaian Tengah Semester (PTS) maupun Penilaian Akhir Semester (PAS), umumnya disusun oleh guru tanpa melalui tahapan analisis butir soal yang memadai (Fathul Uyun, 2023). Guru sering kali terjebak dalam rutinitas penyusunan soal yang hanya mengandalkan intuisi atau sekadar mengambil dari bank soal yang sudah ada tanpa melakukan uji coba (*try out*) terlebih dahulu. Kualitas soal yang diujikan menjadi tidak diketahui secara pasti (Abdullah et al., 2025). Karakteristik instrumen yang tidak teruji berpotensi menghasilkan bias dalam penilaian, di mana skor yang diperoleh siswa tidak mencerminkan kemampuan mereka yang sebenarnya.

Kualitas instrumen tes yang baik dapat diketahui melalui analisis karakteristik empirik. Analisis tersebut meliputi aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda (Aswat et al., 2024). Validitas memastikan bahwa soal benar-benar mengukur apa yang hendak diukur. Reliabilitas menjamin konsistensi hasil pengukuran. Analisis tingkat kesukaran diperlukan untuk memastikan proporsi soal tidak terlalu mudah atau terlalu sukar agar tidak menurunkan motivasi belajar siswa. Daya pembeda berfungsi untuk membedakan antara siswa yang telah menguasai kompetensi (kelompok atas) dengan siswa yang belum menguasai kompetensi (kelompok bawah).

Proses analisis butir soal secara manual sering kali dianggap rumit dan memakan waktu oleh para pendidik (S. Ramadhan, 2024). Pemanfaatan teknologi berupa software statistik menjadi solusi yang sangat relevan. Perangkat lunak *Statistical Package for the Social Sciences* (SPSS) merupakan salah satu alat yang valid dan banyak digunakan dalam dunia akademis untuk melakukan pengolahan data evaluasi. Penggunaan SPSS versi 27 untuk Windows memberikan kemudahan, kecepatan, serta akurasi tinggi dalam membedah karakteristik empirik setiap butir soal melalui output statistika yang komprehensif, seperti tabel *Item-Total Statistics* dan *Reliability Statistics*.

Beberapa penelitian terdahulu telah banyak membahas analisis butir soal. Penelitian yang secara spesifik membedah instrumen evaluasi hasil belajar IPA materi lingkungan kelas VII dengan pendekatan SPSS masih perlu terus diperbarui guna memberikan referensi empiris bagi pengembangan alat evaluasi di sekolah. Penelitian ini bertujuan menganalisis secara mendalam karakteristik empirik instrumen tes evaluasi hasil belajar materi lingkungan kelas VII SMP di Kabupaten Jepara. Hasil analisis



ini diharapkan dapat menjadi rujukan ilmiah bagi para guru IPA dalam merekonstruksi, merevisi, atau mengembangkan butir-butir soal yang berkualitas, valid, dan reliabel di masa yang akan datang.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan deskriptif kuantitatif. Penelitian deskriptif bertujuan untuk menggambarkan karakteristik suatu variabel, kelompok, atau gejala sosial yang terjadi di masyarakat (Martono, 2016). Pendekatan kuantitatif merupakan penelitian yang menggunakan data berbentuk angka, kemudian diolah dan dianalisis untuk memperoleh informasi ilmiah dari data tersebut (Radha, 2025).

2.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di Kabupaten Jepara, Jawa Tengah, dengan responden siswa kelas VII dari beberapa SMP. Pengumpulan data dilakukan secara daring melalui Google Form yang dibagikan kepada responden. Penelitian dilaksanakan pada bulan Juni tahun ajaran 2025/2026.

2.2 Subjek dan Objek Penelitian

Penelitian ini dilaksanakan di Kabupaten Jepara, Jawa Tengah, dengan responden siswa kelas VII dari beberapa SMP. Pengumpulan data dilakukan secara daring melalui Google Form yang dibagikan kepada responden. Penelitian dilaksanakan pada bulan Juni tahun ajaran 2025/2026.

2.3 Populasi dan Sampel

Populasi dalam penelitian ini adalah siswa kelas VII SMP di Kabupaten Jepara. Sampel penelitian berjumlah 30 siswa kelas VII yang berasal dari beberapa SMP di Kabupaten Jepara. Penentuan sampel menggunakan teknik convenience sampling, yaitu pemilihan responden berdasarkan kemudahan akses dan kesediaan untuk berpartisipasi dalam penelitian. Pengumpulan data dilakukan secara daring melalui Google Form yang dibagikan kepada siswa melalui WhatsApp atau grup kelas hingga mencapai target 30 responden.

2.4 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan kuesioner digital dan dokumentasi. Kuesioner digital disusun menggunakan platform Google Form untuk memperoleh data berupa respons siswa terhadap soal pilihan ganda yang diberikan secara daring. Dokumentasi merupakan teknik pengumpulan data dengan mencari informasi mengenai hal-hal atau variabel yang berupa catatan, transkrip, buku, surat kabar, majalah, prasasti, notulen rapat, legger, agenda, dan sebagainya (Martono, 2016). Dalam penelitian ini, dokumentasi digunakan untuk memperoleh data pendukung berupa nilai rapor mata pelajaran IPA siswa melalui portal resmi Rapor Online Jepara. Data tersebut digunakan sebagai informasi pendukung untuk melihat capaian hasil belajar siswa (S. Ramadhan, 2024).



2.5 Teknik Analisis Data

Teknik analisis pada penelitian ini menggunakan bantuan software statistik SPSS versi 27 untuk Windows yang mencakup 4 analisis utama:

1. **Validitas Empirik**

Validitas empirik diketahui dengan menghitung validitas butir soal. Perhitungan koefisien validitas dalam program SPSS berbasis pada rumus korelasi *Product Moment Pearson* sebagai berikut (M. F. Ramadhan et al., 2024):

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}}$$

Keterangan:

r_{xy} = Koefisien korelasi antara variabel X dan variabel Y

N = Jumlah subjek penelitian (30 responden)

X = Skor butir soal siswa

Y = Skor total siswa

Nilai korelasi tiap nomor (r_{hitung}) dibandingkan dengan nilai (r_{tabel})pada taraf signifikansi 5% dengan nilai batas sebesar 0,361. Butir soal dinyatakan valid jika nilai (r_{hitung}) ≥ 0,361.

2. **Reliabilitas**

Uji reliabilitas digunakan untuk melihat tingkat konsistensi instrumen tes secara keseluruhan (S. Ramadhan, 2024). Perhitungan indeks keandalan perangkat tes dalam penelitian ini menerapkan rumus *Cronbach's Alpha* pada program SPSS:

$$r_{11} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2}\right)$$

Keterangan:

r_{11} = Reliabilitas instrumen perangkat tes

k = Banyaknya butir soal ujian

$\sum \sigma_b^2$ = Jumlah varians skor tiap butir soal

$\sum \sigma_b^2$ = Varians skor total keseluruhan

Perangkat tes dinilai memenuhi kriteria reliabel minimum jika indeks menunjukkan angka koefisien ≥ 0,60.

3. Tingkat Kesukaran

Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Angka indeks kesukaran soal (ρ) ini diperoleh melalui nilai rata-rata (*mean*) dari setiap butir pada tabel *Item Statistics* di software SPSS (Saputra et al., 2025). Kolom nilai *maximum* bernilai 1 digunakan sebagai parameter penanda bahwa data skor dikotomi (skor 1 dan 0) telah terinput dan terproses secara akurat (Afriani, 2025). Interpretasi tingkat kesukaran dibagi menjadi tiga kriteria seperti pada tabel berikut:

Tabel 1. Interpretasi Tingkat Kesukaran

Besarnya P	Interpretasi
Kurang dari 0,30	Sukar
0,30-0,70	Sedang
Lebih dari 0,70	Mudah

4. Daya Pembeda

Daya pembeda mengukur kemampuan suatu butir soal dalam membedakan kelompok siswa berkemampuan tinggi dengan siswa berkemampuan rendah. Indeks daya pembeda (*ID*) pada output SPSS diambil berdasarkan nilai kolom *Corrected Item-Total Correlation* pada tabel *Item-Total Statistics* (Nurhalimah et al., 2022). Klasifikasi indeks daya pembeda butir tes dirinci berdasarkan rentang interpretasi berikut:

Tabel 2. Indeks Daya Pembeda Butir Soal

Indeks Daya Pembeda	Interpretasi
$ID \geq 0{,}40$	Sangat Baik
$0{,}30 \leq ID \leq 0{,}39$	Baik
$0{,}20 \leq ID \leq 0{,}29$	Sedang
$0{,}00 \leq ID \leq 0{,}19$	Direvisi
$ID < 0{,}00$	Dibuang / Diganti

3. Results and Discussion/ Hasil dan Diskusi

Sebelum masuk ke tahap analisis kualitas butir soal, tahap awal yang dilakukan adalah melaksanakan telaah soal sesuai dengan lembar telaah yang telah disediakan. Kedua, melakukan pengoreksian lembar jawaban yang telah diperoleh dari pengisian kuesioner digital melalui Google Form oleh siswa kelas VII di Kabupaten Jepara untuk dicocokkan dengan kunci jawaban yang telah dibuat. Tahap akhir barulah dilakukan proses analisis kualitas butir soal menggunakan program SPSS versi 27 untuk Windows.

Plaksanaan evaluasi tes hasil belajar yang diolah menggunakan program SPSS versi 27 diperoleh hasil analisis butir soal sebagai berikut:

3.1 Validitas Tes

Hasil uji validitas menunjukkan bahwa dari 10 butir soal, terdapat 6 butir soal (60%) yang dinyatakan valid, yaitu nomor 1, 2, 5, 8, 9, dan 10. Sementara itu, 4 butir soal (40%), yaitu nomor 3, 4, 6, dan 7, dinyatakan tidak valid. Dimana validitas sebuah tes akan dipengaruhi oleh validitas butir-butir soal yang mendukungnya. Jika butir-butir soal sebuah tes validitasnya tinggi, maka validitas tes itu secara keseluruhan akan tinggi pula. Demikian pula sebaliknya, sebuah tes yang sudah diketahui tingkat validitasnya tinggi, validitas butir-butir soalnya biasanya akan tinggi pula (M. F. Ramadhan et al., 2024). Distribusi frekuensi tingkat validitas dari 10 butir soal tersebut dapat dilihat pada Tabel 3:

Tabel 3. Distribusi Frekuensi dan Persentase Validitas Soal Ujian Materi Lingkungan Kelas VII

Indeks Validitas	Nomor Soal	Jumlah Soal
Valid ($r_{hitung} \geq 0{,}361$)	1, 2, 5, 8, 9, 10	6
Tidak Valid ($r_{hitung} < 0{,}361$)	3, 4, 6, 7	4
Total		**10**

3.2 Reliabilitas Tes

Hasil uji reliabilitas memperoleh nilai Cronbach's Alpha sebesar 0,351. Berdasarkan interpretasi yang digunakan dalam artikel, nilai tersebut menunjukkan bahwa instrumen memiliki tingkat reliabilitas rendah. Dengan hasil ini maka dapat dikatakan bahwa reliabilitas tes soal pilihan ganda ini memiliki interpretasi yang rendah karena harga koefisien *Cronbach's Alpha* statistik ($r_{11} = 0{,}351$) lebih kecil dari harga standar minimal indeks keandalan instrumen pendidikan sebesar 0,60. Dalam buku evaluasi pembelajaran dikatakan juga tes yang memiliki konsistensi reliabilitas tinggi adalah akurat, *reproducible*, dan *generalized* terhadap kesempatan testing dan instrumen tes lainnya (Aswat et al., 2024). Rendahnya nilai keandalan pada perangkat tes ini dipengaruhi oleh keberadaan beberapa butir soal yang tidak valid pada analisis sebelumnya.



3.3 Tingkat Kesukaran

Hasil uji reliabilitas memperoleh nilai Cronbach's Alpha sebesar 0,351. Berdasarkan interpretasi yang digunakan dalam artikel, nilai tersebut menunjukkan bahwa instrumen memiliki tingkat reliabilitas rendah. Perbandingan antara soal mudah, sedang, sukar juga bisa dibuat dengan proporsi tertentu dari teori. Dalam penelitian seharusnya soal dalam kriteria sedang harus lebih diunggulkan atau lebih banyak jumlahnya jika dibandingkan dengan soal dalam kriteria mudah atau sukar, karena soal dalam kriteria sedang sangat cocok digunakan dalam keperluan ujian semester (Saputra et al., 2025). Sebagaimana diterangkan bahwa soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya, soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba karena merasa tidak mampu atau diluar jangkauannya.

Karakteristik tingkat kesukaran pada program SPSS diperoleh dengan memperhatikan parameter nilai rata-rata (*mean*) dan nilai maksimum (*maximum*) pada tabel *Statistics*. Nilai *maximum* untuk seluruh butir soal (soal 1 sampai soal 10) secara konsisten menunjukkan angka 1, yang membuktikan bahwa data respons jawaban siswa telah terinput secara tepat sebagai data dikotomi. Nilai *mean* dari setiap butir soal mencerminkan indeks tingkat kesukaran (*P*). Adapun distribusi tingkat kesukaran dari 10 soal tersebut dapat dilihat pada Tabel 4:

Tabel 4. Distribusi Frekuensi dan Persentase Tingkat Kesukaran Soal Materi Lingkungan

Kriteria	Nomor Soal	Jumlah
Mudah ($P > 0{,}70$)	-	0
Sedang ($0{,}30 \leq P \leq 0{,}70$)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10
Sukar ($P < 0{,}30$)	-	0
Total		**10**

3.4 Daya Pembeda

Jika dianalisis berdasarkan daya pembeda soal ujian akhir semester (UAS) genap mata pelajaran IPA materi lingkungan kelas VII di Kabupaten Jepara termasuk dalam kategori soal jelek (direvisi atau dibuang). Hal ini dikarenakan ada 40% soal yang perlu dilakukan tindakan revisi atau pembuangan secara total. Dasar dan acuan pertimbangannya adalah logika bahwa peserta tes dari kelompok atas seharusnya dapat menjawab dengan benar yang lebih banyak daripada kelompok bawah (Nurhalimah et al., 2022).

Indeks daya pembeda (*ID*) pada output SPSS diambil berdasarkan nilai kolom *Corrected Item-Total Correlation* di dalam tabel *Item-Total Statistics*. Adapun distribusi frekuensi tingkat daya pembeda dari 10 soal tersebut dapat dilihat pada Tabel 5:

Tabel 5. Distribusi Frekuensi dan Persentase Daya Pembeda Soal Materi Lingkungan

Daya Beda	Nomor Soal	Jumlah

Sangat baik ($ID \geq 0{,}40$)	-	0
Baik ($0{,}30 \leq ID \leq 0{,}39$)	1, 2, 5, 10	4
Sedang ($0{,}20 \leq ID \leq 0{,}29$)	8, 9	2
Direvisi ($0{,}00 \leq ID \leq 0{,}19$)	3, 6	2
Dibuang/diganti($ID < 0{,}00$)	4, 7	2
Total		**10**

Data pada Tabel 5 memperlihatkan kualitas daya pembeda instrumen yang bervariasi. Terdapat 4 butir soal (40%) berkategori baik (soal 1, 2, 5, 10) dan 2 butir soal (20%) berkategori sedang (soal 8, 9). Soal-soal tersebut sudah mampu membedakan tingkat pemahaman antar siswa dengan memadai. Terdapat 2 butir soal (20%) berkategori jelek atau direvisi karena memiliki indeks daya pembeda lemah, yaitu soal 3 (0,13) dan soal 6 (0,094). Terdapat 2 butir soal (20%) berkategori sangat jelek atau harus dibuang karena memiliki nilai indeks korelasi negatif, yaitu soal 4 (-0,011) dan soal 7 (-0,050). Nilai korelasi negatif menandakan adanya cacat konstruk pada soal, di mana siswa kelompok bawah justru lebih banyak menjawab benar dibandingkan siswa kelompok atas.

Jadi, dari 10 butir soal pilihan ganda setelah dilakukan analisis kualitas butir soal tentunya hanya mendapatkan 6 butir soal valid, 10 butir soal dengan tingkat kesukaran sedang, serta 6 butir soal dengan daya pembeda sangat baik, baik, dan sedang. Sehingga dapat diambil simpulan bahwa kualitas butir soal IPA materi lingkungan dalam kategori cukup. Hal ini bisa jadi dipengaruhi oleh beberapa faktor baik dari guru dan peserta didik itu sendiri, dan juga disebabkan karena tidak memenuhi kriteria yang telah ditetapkan.

Sebagaimana yang diterangkan bahwa butir soal yang dapat disimpan di dalam bank soal atau digunakan kembali apabila memenuhi kriteria yaitu ditinjau dari Validitas butir soal yang termasuk butir soal valid, ditinjau dari Tingkat Kesukaran termasuk dalam kategori sedang, ditinjau dari Daya Pembeda termasuk kategori sedang, baik dan baik sekali. Butir soal tersebut perlu dijaga kerahasiaannya untuk dapat dikeluarkan kembali pada ujian berikutnya. Dan butir soal yang hanya memenuhi tiga dari keempat kriteria, sebaiknya diperbaiki. Butir soal yang sudah diperbaiki dapat diujikan kembali untuk mengetahui apakah butir soal tersebut sudah menjalankan fungsinya atau tidak. Sedangkan butir soal yang hanya memenuhi dua kriteria sebaiknya tidak digunakan kembali, karena apabila hendak digunakan kembali memerlukan perbaikan yang signifikan. Butir soal inilah yang termasuk kedalam butir soal jelek.

4. Conclusion/ Kesimpulan

The primary Berdasarkan hasil penelitian dan pembahasan mengenai analisis kualitas butir soal pilihan ganda mata pelajaran IPA materi lingkungan kelas VII SMP di Kabupaten Jepara menggunakan program SPSS versi 27 untuk Windows, diperoleh beberapa simpulan sebagai berikut:

1. Analisis validitas empirik menunjukkan bahwa dari 10 butir soal yang diujikan, terdapat 6 butir soal (60%) berkategori valid (soal 1, 2, 5, 8, 9, 10) dan 4 butir soal (40%) berkategori tidak valid (soal 3, 4, 6, 7). Karakteristik tingkat kevalidan perangkat tes ini berada dalam kategori cukup baik.
2. Analisis reliabilitas tes menghasilkan koefisien *Cronbach's Alpha* sebesar 0,351. Angka koefisien tersebut lebih kecil dari standar minimal indeks keandalan instrumen pendidikan sebesar 0,60, sehingga perangkat tes ini dikategorikan memiliki tingkat reliabilitas atau keandalan yang rendah.
3. Analisis tingkat kesukaran dengan memperhatikan parameter nilai rata-rata (*mean*) dan nilai maksimum (*maximum*) menunjukkan bahwa seluruh butir soal (100%) memiliki nilai *maximum* 1 dan masuk dalam karakteristik kategori sedang ($0{,}30 < P < 0{,}70$). Proporsi tingkat kesukaran paket soal ini dinilai sangat baik dan ideal untuk mengukur kemampuan siswa kelas VII karena terpusat secara merata pada kriteria sedang.
4. Analisis daya pembeda menunjukkan kualitas instrumen yang bervariasi, meliputi 4 butir soal (40%) berkategori baik (soal 1, 2, 5, 10), 2 butir soal (20%) berkategori sedang (soal 8, 9), 2 butir soal (20%) berkategori jelek atau direvisi (soal 3, 6), serta 2 butir soal (20%) berkategori sangat jelek karena bernilai negatif sehingga harus dibuang atau diganti (soal 4, 7).

Secara keseluruhan, kualitas instrumen tes hasil belajar IPA materi lingkungan kelas VII di Kabupaten Jepara ini masuk dalam kategori cukup. Butir soal yang memenuhi kriteria valid, berindeks kesukaran sedang, serta berdaya pembeda sedang atau baik dapat dimasukkan ke dalam bank soal untuk digunakan kembali pada ujian berikutnya. Butir soal yang hanya memenuhi beberapa kriteria disarankan untuk direvisi terlebih dahulu, sedangkan butir soal yang masuk kategori sangat jelek direkomendasikan untuk dibuang secara total dari paket perangkat tes.

5. References/ Daftar Pustaka

Abdullah, I., Listyotami, M. K., Ferita, R. A., Maulida, D., Yanti, W., & Hidayat, F. (2025). *Analisis Kualitas Butir Soal Asesmen Sumatif Matematika Pilihan Ganda Semester Genap Kelas VIII MTs Baitul Ma ' mur Tahun Ajaran 2023 / 2024*. *2*(1), 27–32.

Afriani, K. (2025). *Analisis Kebutuhan Penilaian Kreativitas Ilmiah pada Materi Pengaruh Manusia terhadap Ekosistem di SMP Swasta Muhammadiyah 1 Medan*. *15*(2019), 491–497.

Aswat, H., Manaf, A., & Aminu, N. (2024). *Kualitas Instrumen Penilaian Remodeling Penguatan Pendidikan Karakter Pelajar Pancasila Berdasarkan Validitas Isi*. *6*(6), 6570–6583.

Fathul Uyun, M. (2023). *Pengembangan Instrumen Penilaian Pendidikan Karakter Profil Pelajar Pancasila SIswa SD Berbasis Andoird*. *17*(3), 1781–1804.

Martono, N. (2016). *Metode Penelitian Kuantitatif* (Revisi 2). Rajawali Pers.

Nurhalimah, S., Hidayati, Y., Rosidi, I., Madura, U. T., & Soal, K. (2022). *Hubungan Antara Validitas Item Dengan Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda PAS*. *4*(3), 249–257.

Radha, L. (2025). *Bio-Pedagogi : Jurnal Pembelajaran Biologi Analisis butir soal pengukuran kemampuan berpikir kritis materi pencemaran lingkungan melalui teori tes klasik dan Rasch model*. *14*(1), 36–46.

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). *Validitas Dan Reliabilitas*. *06*(02), 10967–10975.

Ramadhan, S. (2024). *Analisis Butir Soal Posttest Pada Materi Lingkungan Dengan Kemampuan Pemecahab Masalah Dan Kemandirian Belajar*.

Saputra, E., Wulan, R., & Ali, N. (2025). *Peningkatan kualitas pendidikan pada satuan pendidikan melalui perencaan berbasis data*. *10*(2), 286–292.

Tomi, L., Wijaya, T., Jamaluddin, J., Kognitif, D. P., & Pengetahuan, D. (2020). *Penguasaan Konsep Sains Peserta Didik Berdasarkan Dimensi Proses Kognitif Dan Dimensi Pengetahuan*. *15*(4), 357–361. https://doi.org/10.29303/jpm.v15i4.1912