
ANALISIS BUTIR SOAL MULTIPLE CHOICE MATERI LINGKUNGAN BERKELANJUTAN XI: PENDEKATAAN EVALUASI PEMBELAJARAN UNTUK PEMBANGUNAN KARAKTER EKOLOGIS

**Mara Bira Koswara¹, Salwa Dhia Tisani², Imelda Triana Putri³,
Mamat Ruhimat⁴, Dina Siti Logayah⁵, Igna Juli Triana**

¹Universitas Pendidikan Indonesia, Bandung, Indonesia

Email Corespondent : Marabira@upi.edu¹

Diterima: 26-05-2025

Direvisi: 20-06-2025

Diterbitkan: 30-10-2025

Abstrak: Menghadapi tantangan abad ke-21, pendidikan Indonesia menekankan penerapan soal Higher Order Thinking Skills (HOTS) untuk melatih kemampuan berpikir kritis, kreatif, dan pemecahan masalah. Soal HOTS mencakup analisis (C4), evaluasi (C5), dan kreasi (C6), mendorong siswa tidak hanya menghafal tetapi juga memahami dan menyelesaikan masalah, termasuk isu lingkungan yang kompleks. Materi lingkungan berkelanjutan di kelas XI SMA penting untuk membangun kesadaran ekologis dan tanggung jawab siswa terhadap pelestarian alam. Penyusunan soal HOTS memerlukan langkah sistematis, mulai dari pemilihan Kompetensi Dasar, penyusunan indikator berbasis ABCD (audiens, perilaku, kondisi, derajat), hingga pembuatan kisi-kisi soal. Penelitian ini bertujuan menganalisis soal pilihan ganda materi lingkungan berkelanjutan bagi siswa kelas XI SMAN 1 Kandanghaur dan SMA Pesantren Unggulan AL-BAYAN. Pelaksanaan dilakukan pada April–Mei 2025 melalui platform Quizizz. Analisis soal mengacu pada prinsip indikator ABCD dan pedoman mutu evaluasi Sunartana (1986). Aspek yang dikaji meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Selain itu, berdasarkan taksonomi Anderson dan Krathwohl (2001), soal dianalisis untuk mengukur pengetahuan, sikap, dan keterampilan secara utuh.

Kata Kunci: HOTS, pendidikan, berpikir kritis.

Abstract: Facing the challenges of the 21st century, Indonesian education emphasizes the application of Higher Order Thinking Skills (HOTS) to train critical thinking, creativity, and problem-solving abilities. HOTS questions encompass analysis (C4), evaluation (C5), and creation (C6), encouraging students not only to memorize but also to understand and solve problems, including complex environmental issues. Sustainable environmental material in the 11th grade of high school is important for building ecological awareness and students' responsibility towards nature conservation. The preparation of HOTS questions requires a systematic approach, starting from the selection of Basic Competencies, the formulation of indicators based on ABCD (audience, behavior, condition, degree), to the creation of question grids. This research aims to analyze multiple-choice questions on sustainable environment material for 11th-grade students at SMAN 1 Kandanghaur and SMA Pesantren Unggulan AL-BAYAN. The implementation was carried out in April–May 2025 through the Quizizz platform. The analysis of the questions refers to the ABCD indicator principles and the quality evaluation guidelines by Sunartana (1986). The aspects examined include validity, reliability, difficulty level, discrimination power, and the effectiveness of distractors. Additionally, based on Anderson and Krathwohl's taxonomy (2001), the questions are analyzed to measure knowledge, attitudes, and skills comprehensively.

Keywords: HOTS, education, critical thinking

Pendahuluan

Dalam beberapa tahun belakangan, penerapan soal yang berorientasi pada Higher

Order Thinking Skills (HOTS) telah menjadi salah satu tujuan utama dalam sektor pendidikan di Indonesia. Ini sesuai dengan tuntutan untuk memberikan peserta didik keterampilan berpikir kritis, kreativitas, dan kemampuan menyelesaikan masalah secara mandiri, terutama dalam menghadapi tantangan di era abad ke-21. Pertanyaan HOTS dirancang untuk mendorong siswa agar tidak hanya menghafal atau memahami informasi, tetapi juga dapat menganalisis, mengevaluasi, dan mengembangkan solusi untuk masalah yang ada, termasuk isu-isu lingkungan yang semakin rumit. Pembelajaran tentang lingkungan yang berkelanjutan di tingkat Sekolah Menengah Atas, terutama pada kelas XI, sangat penting dalam membangun kesadaran serta karakter ekologis pada para siswa. Dengan mempelajari materi ini, diharapkan siswa tidak hanya mengenal konsep-konsep fundamental mengenai lingkungan, tetapi juga dapat mengembangkan sikap perhatian dan tanggung jawab terhadap pelestarian alam. Salah satu metode yang efektif dalam menilai pemahaman dan karakter siswa adalah dengan melakukan evaluasi menggunakan soal pilihan ganda yang disusun berdasarkan prinsip Higher Order Thinking Skills (HOTS).

Menurut Anderson dan Krathwohl (2001), soal yang termasuk dalam kategori HOTS mengharuskan siswa untuk terlibat dalam kegiatan berpikir yang lebih kompleks, yang terdiri dari analisis (C4), evaluasi (C5), dan kreasi (C6). Proses pembuatan soal jenis ini memerlukan perencanaan yang cermat, dimulai dari pemilihan kompetensi dasar (KD), penetapan indikator soal, sampai penyusunan kisi-kisi yang terstruktur. Setiap pertanyaan dalam ujian sebaiknya dirancang untuk menilai kemampuan siswa secara menyeluruh, mencakup pengetahuan, sikap, dan keterampilan mereka. Selain itu, pengajar juga harus memperhatikan penggabungan faktor-faktor penting seperti audiens, perilaku, kondisi, dan tingkat (ABCD) saat menyusun indikator soal guna memastikan pencapaian tujuan pembelajaran secara maksimal. Signifikansi pemanfaatan soal HOTS dalam penilaian pembelajaran tidak hanya berfokus pada pengukuran pencapaian hasil belajar, tetapi juga berperan dalam membangun karakter siswa. Dengan menghadapi pertanyaan-pertanyaan yang menantang, para siswa digerakkan untuk berpikir secara kritis, membuat pilihan yang bijak, dan bertindak dengan penuh tanggung jawab terhadap lingkungan sekitar. Dengan demikian, evaluasi pembelajaran memiliki dua fungsi, yaitu sebagai instrumen pengukuran akademis dan sebagai sarana untuk membentuk karakter yang peduli lingkungan.

Agar soal HOTS yang diterapkan dapat berfungsi secara optimal, diperlukan analisis mendetail terhadap setiap butir soal. Analisis ini melibatkan uji validitas, Reliabilitas , tingkat kesukaraan, daya pembeda , serta efektivitas pengecoh . Tahapan ini sangat krusial untuk menjamin bahwa setiap soal yang diuji memiliki mutu yang memadai dan dapat secara adil membedakan tingkat kemampuan para siswa. Menurut Sunartana (1986), mutu evaluasi sangat bergantung pada berbagai aspek tersebut, sehingga analisis terhadap item soal menjadi langkah yang tidak bisa diabaikan dalam proses evaluasi pembelajaran. studi ini bertujuan untuk menganalisis item-item soal pilihan ganda mengenai materi lingkungan berkelanjutan untuk siswa kelas XI SMAN 1Kandanghaur & SMA Pesantren Ugulan AL-BAYAN yang dilaksanakan April-Mei dengan menggunakan pendekatan evaluasi atau Quizizz pembelajaran yang fokus pada pengembangan karakter ekologis. Studi ini diharapkan mampu memberikan penjelasan

tentang seberapa akurat dan berkualitasnya soal yang diujikan, serta turut serta dalam pengembangan evaluasi pembelajaran yang lebih efisien dan bermanfaat dalam membentuk karakter siswa yang peduli terhadap lingkungan.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis deskriptif evaluatif. Pendekatan ini dipilih karena data yang dianalisis berupa hasil kuis pilihan ganda yang dikerjakan siswa dan diolah secara statistik sederhana. Tujuan utama penelitian adalah mendeskripsikan kualitas soal berdasarkan indikator teknis seperti tingkat kesukaran, daya pembeda, dan efektivitas distraktor, serta mengevaluasi peran soal dalam membentuk karakter ekologis peserta didik.

Penelitian dilaksanakan di SMA Negeri 1 Kandanghaur dan SMA Pesantren Unggul Al Bayan pada April–Mei 2025. Sampel penelitian berjumlah 34 siswa kelas XI yang dipilih secara *purposive* karena telah menyelesaikan materi lingkungan berkelanjutan. Instrumen yang digunakan adalah tes pilihan ganda sebanyak 20 soal yang diunggah ke platform Quizziz. Soal dirancang tidak hanya untuk mengukur aspek kognitif, tetapi juga menanamkan nilai karakter ekologis. Pengumpulan data dilakukan melalui pelaksanaan kuis daring, kemudian hasilnya diekspor ke Excel untuk dianalisis. Teknik analisis meliputi perhitungan tingkat kesukaran (kategori mudah, sedang, sukar), daya pembeda (sangat baik hingga buruk), serta efektivitas distraktor atau daya kecoh (dilihat dari persentase pemilihannya).

Hasil dan Pembahasan

A. Hasil

Penelitian ini bertujuan untuk menganalisis butir soal yang terdapat dalam alat penilaian berbasis Quizizz mengenai topik lingkungan berkelanjutan. instrumen ini dibagikan kepada murid-murid dari dua lembaga pendidikan, yaitu SMAN 1 Kandanghaur dari Indramayu dan SMA PESANTREN UNGGULAN AL-BAYAN berlokasi di Anyer. Analisis dilakukan terhadap 20 butir soal yang mencakup pengujian validitas, reliabilitas, tingkat kesukaran, kemampuan pembeda, dan daya kecoh. Ujian validitas dilakukan melalui metode korelasi Pearson Product Moment dan hasilnya menunjukkan bahwa mayoritas soal memiliki validitas yang tinggi (nilai korelasi $> 0,30$). Di sisi lain, keandalan instrumen yang dihitung menggunakan rumus KR-20 mendapatkan nilai lebih dari 0,7, yang menunjukkan bahwa instrumen tersebut dapat dianggap reliabel. Analisis mengenai tingkat kesulitan menunjukkan bahwa soal-soal memiliki variasi yang cukup baik. Ada soal yang dibagi ke dalam tiga kategori: mudah ($P > 0,70$), sedang ($0,31 \leq P \leq 0,70$), dan sulit ($P < 0,30$). Sebaran ini menunjukkan bahwa pertanyaan yang diajukan mencakup berbagai tingkat kesulitan. Namun, beberapa soal yang dinilai terlalu mudah atau tingkat kesukaran sebaiknya direvisi untuk mencapai keseimbangan yang lebih baik. Keseimbangan dalam tingkat kesukaran sangat krusial agar setiap siswa memiliki kesempatan yang sama untuk menunjukkan kemampuan mereka dengan adil.

Daya pembeda diukur dengan cara membandingkan nilai kelompok yang berprestasi tinggi dengan kelompok yang berprestasi rendah. Sebagian besar pertanyaan menunjukkan

kemampuan diferensiasi yang baik (nilai $> 0,30$), yang mengindikasikan bahwa pertanyaan tersebut dapat secara efektif membedakan antara siswa yang memiliki pemahaman tinggi dan rendah. Namun, terdapat beberapa soal yang diketahui memiliki kemampuan untuk membedakan yang rendah, bahkan ada yang negatif, sehingga tidak efektif dalam menilai kemampuan siswa. Oleh karena itu, sebaiknya soal-soal tersebut diperbaiki atau dihapus dari alat ukur. dayah kecoh dianalisis berdasarkan seberapa efektif opsi jawaban lain yang dipilih oleh siswa selain jawaban yang benar. Pilihan strategi yang efektif adalah yang dapat menarik perhatian siswa yang belum memahami materi dengan baik. Dalam kajian ini, sebagian besar pilihan pengecoh terbukti berhasil, namun terdapat beberapa opsi yang tidak dipilih sama sekali. Oleh karena itu, perlu ada perbaikan untuk meningkatkan kualitas soal. Secara umum, alat ini menunjukkan standar yang memuaskan dan dapat diterapkan dalam penilaian proses belajar dengan beberapa perbaikan kecil.

B. Pembahasan

Penelitian ini bertujuan untuk menganalisis kualitas butir soal pilihan ganda pada mata pelajaran Ekologi tentang amdal Kelas XI SMAN 1 Kandanghaur sendiri berlokasi di indramayu sedangkan SMA PESANTREN UNGGULAN AL-BAYAN berlokasikan dikawasan anyer instrumen yang digunakan yaitu berupa dua soal paket yaitu pada materi amdal dan lingkungan berkelanjutan,masing masing berjumlah 10 butir soal dan diujikan kepada siswa sebagai sempel penelitian.analisis ini dilakukan melalui uji validitas, reliabilitas,tingkat kesukaraan,daya pembeda,dan daya pengecoh.

1. Uji validitas

Berdasarkan hasil uji validitas terhadap 20 butir soal:

- Sebanyak 19 butir soal dinyatakan valid (nilai korelasi $> r$ tabel 0,388 untuk $n = 60$).
- Hanya 1 soal yang tidak valid,yaitu pada nomor 10 dari materi lingkungan berkelanjutan .

Tabel 1. Rincian Validitas 20 soal Lingkungan berkelanjutan

No Soal	Nilai Korelasi (r -Hitung)	r -tabel ($n = 34, \alpha = 0,05$)	Keterangan
1	0,6651	0,388	Valid
2	0,6397	0,388	Valid
3	0,6035	0,388	Valid
4	0,5612	0,388	Valid
5	0,6847	0,388	Valid
6	0,7456	0,388	Valid
7	0,7221	0,388	Valid
8	0,6691	0,388	Valid
9	0,6244	0,388	Valid
10	0,2917	0,388	Tidak Valid
11	0,5311	0,388	Valid

12	0,3473	0,388	Valid
13	0,5244	0,388	Valid
14	0,7089	0,388	Valid
15	0,5418	0,388	Valid
16	0,6510	0,388	Valid
17	0,5262	0,388	Valid
18	0,6027	0,388	Valid
19	0,7691	0,388	Valid
20	0,7497	0,388	Valid

Berdasarkan evaluasi dari tabel 1 ini terlihat bahwa mayoritas butir soal memiliki tingkat validitas yang memuaskan. Dari 20 soal yang dianalisis, sebanyak 19 butir soal (95%) memenuhi kriteria validitas dengan nilai r -hitung yang lebih besar r -tabel sebesar 0,388. Analisis menunjukkan bahwa sebagian besar butir soal memiliki korelasi yang signifikan dengan total skor instrumen, yang mengidentifikasi yang baik dalam mengukur konstruk yang diinginkan. Metodologi yang diterapkan mencakup penghitungan koefisien korelasi antara nilai setiap butir soal dengan jumlah total nilai, selanjutnya membandingkan nilai r -hitung dengan r -tabel guna menetapkan status validitas masing-masing item. Tahapan ini memastikan bahwa setiap item soal yang dianggap valid memberikan kontribusi yang signifikan terhadap keseluruhan pengukuran konstruk, sehingga dapat diandalkan untuk memberikan data yang tepat mengenai kemampuan atau karakteristik yang dimiliki oleh responden yang diukur. Hasil utama dari analisis ini mengindikasikan bahwa terdapat satu item soal yang membutuhkan perhatian khusus karena tidak sesuai dengan kriteria validitas yang telah ditentukan. Nilai korelasinya hanya 0,20, yang berada jauh di bawah ambang batas kritis yaitu 0,632. Nilai korelasi yang rendah pada soal tersebut menunjukkan bahwa item ini tidak berkorelasi yang secara signifikan dengan total skor instrumen. Hal ini mungkin disebabkan oleh beberapa faktor, seperti ketidakjelasan dalam penulisan soal, tingkat kesulitan yang tidak sejalan dengan kemampuan responden, atau ketidaksesuaian dengan konstruk yang ingin diukur.

2. Relabilitas

Tabel 2. Interpretasi Reliabilitas

Reliabilitas	Nilai	Keterangan
varians	364,934	Sangat tinggi
KR-20	0,9156	Sangat tinggi
KR-21	0,9044	Sangat tinggi
CA	0,8989	Sangat tinggi

Berdasarkan data pada tabel Reliabilitas ini menunjukkan tingkat keandalan yang sangat baik dengan berbagai indikator psikometrik yang beragam. Data ini menunjukkan bahwa total varians mencapai 364,934 dan berada dalam kategori sangat tinggi. Hal ini menandakan bahwa distribusi skor yang diperoleh cukup baik dan alat ukur tersebut efektif dalam membedakan tingkat kemampuan siswa. Nilai Kuder-Richardson 20 (KR-20) yang mencapai 0,9156 menunjukkan adanya konsistensi internal yang sangat tinggi. Di sisi lain, KR-21 dengan nilai 0,9044 juga termasuk dalam kategori yang sama, yang menunjukkan bahwa

semua item pertanyaan memberikan kontribusi yang konsisten terhadap pengukuran kemampuan yang diinginkan. Nilai Cronbach's Alpha (CA) sebesar 0,8989 semakin memperkuat bahwa alat ukur ini memiliki tingkat reliabilitas yang sangat baik dalam menilai konstruk yang serupa. Tingginya skor reliabilitas di semua parameter menunjukkan bahwa instrumen evaluasi ini telah mencapai standar kualitas psikometrik yang diperlukan untuk pengukuran yang tepat dan dapat diandalkan. Semua indikator reliabilitas yang melebihi 0,89 menunjukkan bahwa instrumen ini dapat dipercaya untuk menghasilkan pengukuran yang stabil dan konsisten saat digunakan berulang kali dalam situasi yang sama. Keunggulan ini memiliki peranan yang krusial dalam evaluasi pembelajaran, karena menjamin bahwa penilaian yang dilakukan secara akurat mencerminkan kompetensi siswa, bukan dipengaruhi oleh fluktuasi alat ukur.

3. Uji Tingkat Kesukaran

Tabel 3. Rincian Tingkat Kesukaran 20 soal Lingkungan Berkelanjutan

Butir soal	Tingkat kesukaran	Kategori
1	0.26	sukar
2	0.58	Sedang
3	0.67	sedang
4	0.75	sedang
5	0.52	Sedang
6	0.64	Sedang
7	0.75	sedang
8	0.74	mudah
9	0.64	Sedang
10	0.79	Mudah
11	0.73	Mudah
12	0.29	Sukar
13	0.85	mudah
14	0.55	sedang
15	0.44	sedang
16	0.76	mudah
17	0.5	Sedang
18	0.73	mudah
19	0.70	Sedang
20	0.67	sedang

Analisis tingkat kesukaran soal merupakan salah satu aspek penting. Tabel yang disajikan di atas menampilkan data hasil jawaban 34 siswa terhadap 20 butir soal. Baris pada tabel awal menggambarkan jawaban setiap siswa terhadap masing-masing soal menghitung proporsi terhadap total soal. Proporsi ini menjadi dasar penentu tingkat kesukaran butir soal.

Tingkat kesukaran butir soal dibagi menjadi 3 kategori utama: sukar (0,00 sampai 0,30), sedang, (0,31-0,70), dan mudah (0,71-100). Klasifikasi ini membantu guru atau pengembang tes untuk mengetahui sejauh mana soal tersebut mampu dijawab oleh siswa secara tepat. Dari hasil perhitungan, terlihat bahwa soal nomor 1 dan 12 termasuk ke dalam kategori sukar,

Dengan proporsi jawaban masing masing sekitar 0,26 dan 0,29. Soal soal ini hanya mampu dijawab oleh kurang dari 1/3 siswa ,Sehingga dapat dikatakan bahwa soal ini memang cukup menantang bagi sebagian besar peserta didik.

Mayoritas soal justru berada pada kategori sedang, Mencakup soal nomor 2,3,4,5,6,7,9,13,15,17,19,dan 20. Proporsi jawaban benar pada soal soal ini berkisar Antara 0,31 hingga 0,70. Soal soal ini dapat dikatakan cukup seimbang,Tidak terlalu sulit, namun tidak terlalu mudah bagi siswa. Kategori sedang ini penting karena soal yang mudah atau sulit kurang efektif dalam membedakan kemampuan siswa soal dihentikan kesukaran sudah justru lebih mampu memberikan informasi tentang kemampuan siswa.

Sementara itu, soal masuk ke dalam kategori mudah, yaitu soal nomor 8,10,11,14,16 dan 18, dengan proporsi jawaban benar di atas 0,71. Soal soal ini mampu dijawab Sebagian besar siswa, sehingga dapat dikatakan bahwa materi yang diujikan pada soal soal ini sudah dikuasai dengan baik oleh siswa. namun, soal yang terlalu mudah kurang mampu membedakan siswa yang benar-benar paham dengan hanya sekedar menebak atau mengandalkan ingatan.

4. Daya Pembeda

Tabel 4. Rincian Daya Pembeda 20 Soal Lingkungan Berkelanjutan

Butir soal	Rata-Rata Atas	Rata-Rata bawah	Daya Pembeda	Kategori
1	0.94	0.47	0.47	Sangat baik
2	0.88	0.29	0.58	Sangat baik
3	0.88	0.47	0.41	Sangat baik
4	0.88	0.52	0.35	baik
5	0.82	0.23	0.58	Sangat baik
6	0.94	0.35	0.58	Sangat baik
7	1	0.41	0.58	Sangat baik
8	0.94	0.64	0.29	cukup
9	0.82	0.47	0.35	baik
10	0.88	0.70	0.17	Kurang baik
11	0.88	0.58	0.29	cukup
12	0.35	0.23	0.11	Kurang baik
13	1	0.70	0.29	cukup
14	0.88	0.23	0.64	Sangat baik
15	0.64	0.23	0.41	Sangat baik
16	0.94	0.58	0.35	baik
17	0.64	0.35	0.29	cukup
18	0.94	0.52	0.41	Sangat baik
19	0.94	0.47	0.47	Sangat baik
20	1	0.35	0.64	Sangat baik

Analisis daya pembeda butir soal merupakan salah satu indikator penting dalam evaluasi kualitas soal tes, yang digunakan untuk mengetahui seberapa efektif suatu soal

mampu membedakan antara siswa yang memiliki kemampuan tinggi dan siswa yang memiliki kemampuan rendah. Pada tabel yang diberikan, terlihat data rata-rata kelompok atas (rata-rata atas) dan rata-rata kelompok bawah (rata-rata bawah) untuk setiap soal, serta nilai daya pembeda yang dihasilkan dari selisih antara kedua rata-rata tersebut. Selanjutnya, daya pembeda tersebut diklasifikasikan ke dalam beberapa kategori, yaitu sangat baik ($\geq 0,40$), baik (0,30–0,39), cukup/perlu diperbaiki (0,20–0,29), dan kurang baik ($\leq 0,19$).

Dari hasil analisis, sebagian besar soal memiliki daya pembeda yang sangat baik, yaitu soal nomor 2, 3, 5, 6, 7, 9, 14, 15, 18, dan 19. Soal-soal ini mampu membedakan secara signifikan antara siswa yang memiliki kemampuan tinggi dan rendah, sehingga sangat efektif untuk digunakan dalam evaluasi. Selain itu, terdapat beberapa soal dengan daya pembeda kategori baik, seperti soal nomor 4 dan 16, yang masih cukup efektif namun belum mencapai kriteria sangat baik. Ada pula soal dengan daya pembeda cukup, yaitu soal nomor 8, 11, 13, 17, dan 20. Soal-soal ini masih mampu membedakan siswa, namun efektivitasnya relatif rendah sehingga perlu diperbaiki agar bisa menjadi lebih baik di masa mendatang. Sedangkan soal dengan daya pembeda kurang baik, seperti soal nomor 10 dan 12, menunjukkan bahwa soal tersebut kurang efektif dalam membedakan kemampuan siswa, sehingga perlu dilakukan revisi atau bahkan dihapus dari instrumen tes.

Secara keseluruhan, hasil analisis daya pembeda menunjukkan bahwa mayoritas soal sudah memenuhi kriteria sangat baik dan baik, sehingga instrumen tes ini sudah cukup efektif dalam membedakan kemampuan siswa. Namun, masih ada beberapa soal yang perlu diperbaiki agar kualitas tes menjadi lebih optimal dan mampu memberikan gambaran yang lebih akurat tentang kemampuan siswa. Dengan demikian, guru atau pengembang tes dapat menggunakan hasil analisis ini sebagai bahan evaluasi dan perbaikan instrumen evaluasi di masa mendatang.

5. Daya Kecoh

Analisis terakhir yang dilakukan adalah menilai efektivitas pengecoh. Penilaian ini dilakukan dengan menghitung jumlah siswa yang memilih masing-masing opsi jawaban a, b, c, d, dan e, termasuk yang tidak menjawab sama sekali pada tiap butir soal. Suatu pengecoh dianggap efektif jika dipilih oleh minimal 5% dari total siswa. Proses pengukuran efektivitas pengecoh dilakukan menggunakan program Excel dengan kriteria penilaian berupa diterima, perlu direvisi, atau ditolak.

Tabel 5. Distribusi 20 soal Lingkungan berkelanjutan berdasarkan efektivitas pengecoh

No	$\leq 0\%$ (ditolak)	$0,1\text{--}4,9\%$ (direvisi)	$\geq 5\%$ (diterima)	Jumlah Ditolak	Persentase Ditolak (%)
1	D, E	-	A, B, C	2	10
2	E	-	A, B, C, D	1	5
3	E	-	A, B, C, D	1	5
4	E	-	A, B, C, D	1	5
5	A	-	B, C, D, E	1	5

6	A, E	-	B, C, D	2	10
7	D	-	A, B, C, E	1	5
8	B, D	-	A, C, E	2	10
9	C, D	-	A, B, E	2	10
10	A	-	B, C, D, E	1	5
11	A, E	-	B, C, D	2	10
12	E	-	A, B, C, D	1	5
13	D, E	-	A, B, C	2	10
14	A, E	-	B, C, D	2	10
15	-	-	A, B, C, D, E	0	0
16	C, E	-	A, B, D	2	10
17	-	-	A, B, C, D, E	0	0
18	B, D	-	A, C, E	2	10
19	E	-	A, B, C, D	1	5
20	C	-	A, B, D, E	1	5
Jumlah			27	27%	

Tabel di atas menunjukkan bahwa terdapat 27 opsi jawaban (27%) yang tidak efektif atau ditolak dari total 100 pengecoh pada 20 butir soal. Opsi-opsi ini tidak berhasil berfungsi sebagai pengecoh karena tidak dipilih sama sekali oleh siswa (efektivitas $\leq 0\%$). Dengan demikian, terdapat 73 opsi pengecoh (73%) yang efektif atau diterima karena dipilih oleh siswa dengan proporsi $\geq 5\%$, sehingga berfungsi baik dalam mengelabui peserta yang belum memahami materi secara benar.

Persentase pengecoh yang ditolak menunjukkan bahwa sebagian soal masih perlu diperbaiki dari sisi daya pengecohnya agar dapat meningkatkan kualitas instrumen evaluasi. Revisi dapat dilakukan dengan merancang pengecoh yang lebih menarik atau menyerupai jawaban benar untuk menilai kemampuan berpikir kritis siswa secara lebih akurat.

Kesimpulan

Studi ini bertujuan untuk menilai butir soal pilihan ganda yang dihasilkan melalui aplikasi Quizizz, yang digunakan untuk mengukur pemahaman siswa mengenai materi tentang lingkungan berkelanjutan. Studi ini dilaksanakan di dua sekolah menengah atas, yaitu SMAN 1 Kandanghaur yang terletak di Indramayu dan SMA Al-Bayan di Anyer, dengan melibatkan 60 siswa sebagai sampel dan menyediakan total 20 soal. Analisis dilakukan pada berbagai aspek, uji validitas, Reliabilitas, uji kesukaran, tingkat perbedaan, daya pengecoh. Hasil pengujian validitas mengindikasikan bahwa 19 dari 20 pertanyaan (95%) memenuhi standar validitas (nilai r -hitung $> r$ -tabel = 0,388), yang menunjukkan bahwa pertanyaan-pertanyaan tersebut efektif dalam mengukur konstruk yang dimaksud dengan tepat. Hanya satu pertanyaan yang dianggap tidak valid dan harus direvisi atau dihilangkan. Pengujian reliabilitas yang dilakukan dengan menggunakan KR-20 dan Alpha Cronbach menunjukkan hasil yang sangat baik, dengan KR-20

sebesar 0,9156 dan Alpha Cronbach sebesar 0,8989. Ini menunjukkan bahwa instrumen soal memiliki konsistensi internal yang sangat baik dan dapat dipercaya sebagai alat untuk mengukur kemampuan siswa. Tingkat kesulitan soal dibagi menjadi tiga kategori, yaitu sulit, sedang, dan mudah. Soal dengan tingkat kesulitan sedang mendominasi dengan jumlah 13 soal, sementara ada 6 soal yang dikategorikan sebagai mudah dan 2 soal lainnya yang dianggap sulit. Distribusi ini menggambarkan bahwa mayoritas pertanyaan sudah berada dalam tingkat kesulitan yang tepat. Namun, ada beberapa pertanyaan yang terlalu mudah atau terlalu sulit yang sebaiknya diperbaiki agar tidak mengganggu pengaturan nilai. Dari segi daya pembeda, terdapat 10 soal yang sangat baik, 2 soal yang baik, 5 soal yang cukup (memerlukan perbaikan), dan 2 soal yang kurang baik. Kemampuan pembelian yang tinggi mencerminkan seberapa efektif suatu soal dalam membedakan antara siswa yang telah memahami materi dan yang belum. Namun, ada beberapa pertanyaan yang perlu direvisi karena tidak efisien dalam mengidentifikasi kemampuan siswa, bahkan ada yang menunjukkan daya pembeda yang buruk.

Analisis mengenai efektivitas pengecoh mengungkapkan bahwa hanya beberapa pilihan jawaban yang salah berfungsi dengan efektif. Terdapat 6 soal yang menggunakan pengecoh yang efektif, 8 soal dengan pengecoh yang berfungsi sebagian, dan 6 soal yang memiliki pengecoh yang tidak efektif. Beberapa opsi jawaban sama sekali tidak dipilih oleh para siswa, yang menunjukkan bahwa opsi tersebut tidak berperan sebagai pengalih perhatian. Ini menunjukkan pentingnya untuk memperbaiki cara menyusun alternatif jawaban yang dapat menyegarkan siswa yang belum sepenuhnya memahami materi. Secara keseluruhan, dapat disimpulkan bahwa alat evaluasi ini memiliki standar yang memadai dan cocok untuk diterapkan dalam proses pembelajaran, khususnya dalam mengukur kemampuan kognitif siswa terkait materi lingkungan yang berkelanjutan. Namun, tetap diperlukan perbaikan pada beberapa item soal, terutama yang memiliki kemampuan membedakan yang rendah serta pilihan jawaban (distraktor) yang kurang efektif.

Referensi

- Arikunto, S. (2019). *Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.
- Asri, B. N., & Mulyaningsih, R. (2020). Analisis kualitas butir soal pilihan ganda pada mata pelajaran biologi kelas XI. *Jurnal Evaluasi Pendidikan*, 11(2), 125-134.
- Hamzah, B. U. (Jurnal Teknologi Pendidikan). Evaluasi pembelajaran berbasis teknologi: Studi pada penggunaan aplikasi Quizziz. 9(1), 45-46.
- Latip., A. E. (2020). *Evaluasi Pembelajaran*. Puslitpen LP2M UIN Syarif Hidayatullah Jakarta.
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. . Yogyakarta: Parama Publishing.

- Nugraheni, A. D., & Rohana, R. (2022). Penerapan evaluasi formatif dalam pembelajaran lingkungan berkelanjutan. *Jurnal Ilmu Pendidikan. Jurnal Ilmu Pendidikan*, 18(3), 211-219.
- Zainal, A., & Yuliana, D. (2020). Analisis daya pembeda dan tingkat kesukaran soal dalam evaluasi hasil belajar. *Jurnal Penelitian Pendidikan*, 27(1), 88-97.