



Biodik: Jurnal Ilmiah Pendidikan Biologi
ISSN 2580-0922 (*online*), ISSN 2460-2612 (*print*)
Volume 10, Nomor 04, Tahun 2024, Hal. 691-702
Available online at:
<https://online-journal.unja.ac.id/biodik>



Research Article



Pengembangan Instrumen Penilaian *Higher Order Thinking Skills* (HOTS) Pada Materi Ekologi di SMA

(Development of Higher Order Thinking Skills (HOTS) Assessment Instruments on Ecology Material in High School)

Dias Rizki*, Fathan Abiyyu Pakarti Almay, Shefa Dwijayanti Ramadani

Program Studi Pendidikan Biologi, Universitas Tidar

Jl. Kapten Suparman No.39, Potrobangsari, Kec. Magelang Utara, Kota Magelang, Jawa Tengah 56116

*Corresponding Authors: dias.rizki@students.untidar.ac.id

Informasi Artikel	ABSTRACT
Submit: 19 – 08 – 2024 Diterima: 28 – 11 – 2024 Dipublikasikan: 30 – 12 – 2024	<p><i>Evaluation results from national and international education show that students' critical and analytical thinking skills still need to be improved. One way to improve these skills is to develop HOTS questions that can stimulate and assess students' higher-order thinking abilities. However, many schools in Indonesia have not developed enough in creating HOTS instruments that meet good validity standards. Therefore, this research aims to develop a HOTS assessment instrument that meets the criteria for good content and construction validity in ecological material. This research uses a development method (Research & Development) with a 4D approach, which includes the define, design, develop and disseminate stages. The research results showed that the questions developed were valid and reliable. As many as 20% of the questions fall into the very good differential power category. The difficulty level of the questions developed consists of 15% easy questions, 65% medium questions and 20% difficult questions. In addition, construct validity shows that the instrument developed meets the very good criteria based on its aspects.</i></p> <p>Key words: HOTS, R&D, Assessment, Validity</p>
Penerbit	ABSTRAK
Program Studi Pendidikan Biologi FKIP Universitas Jambi, Jambi- Indonesia	<p>Hasil evaluasi dari pendidikan nasional dan internasional menunjukkan bahwa keterampilan berpikir kritis dan analitis siswa masih perlu ditingkatkan. Salah satu cara untuk meningkatkan keterampilan ini adalah dengan mengembangkan soal-soal HOTS yang dapat merangsang dan menilai kemampuan berpikir tingkat tinggi siswa. Namun, banyak sekolah di Indonesia belum cukup berkembang dalam menciptakan instrumen HOTS yang memenuhi standar validitas yang baik. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen penilaian HOTS yang memenuhi kriteria validitas isi dan konstruksi yang baik pada materi ekologi. Penelitian ini menggunakan metode pengembangan (<i>Research & Development</i>) dengan pendekatan 4D, yang mencakup tahap <i>define, design, develop, dan disseminate</i>. Hasil penelitian menunjukkan bahwa soal yang dikembangkan valid dan reliabel. Sebanyak 20% dari soal termasuk dalam kategori daya beda sangat baik. Tingkat kesulitan soal yang dikembangkan terdiri dari 15% soal mudah, 65% soal sedang, dan 20% soal sulit. Selain itu, validitas konstruksi menunjukkan bahwa instrumen yang dikembangkan memenuhi kriteria sangat baik berdasarkan aspeknya.</p> <p>Kata kunci: HOTS, R&D, Penilaian, Validitas</p>



This Biodik : Jurnal Ilmiah Pendidikan Biologi is licensed under a [CC BY-NC-SA \(Creative Commons Attribution-ShareAlike 4.0 International License\)](https://creativecommons.org/licenses/by-nc-sa/4.0/)

PENDAHULUAN

Pendidikan merupakan elemen yang erat kaitannya dalam kehidupan modern. Memasuki era digital, pendidikan diarahkan untuk mencetak peserta didik sebagai insan yang dapat bersaing secara global. Dalam menghadapi persaingan tersebut, diperlukan sumber daya yang berkualitas dan memiliki kemampuan berpikir kritis. Hal ini menjadi tantangan tersendiri bagi penyelenggara pendidikan (Hamidah & Wulandari, 2021). Kemampuan berpikir kritis menjadi hal yang amat penting bagi peserta didik (Živković, 2016, Siburian et al., 2019). Kemampuan tersebut penting dikuasai karena melibatkan kemampuan penyelesaian masalah yang relevan dengan tantangan pendidikan abad 21 (Afrita & Rahmawati, 2019). Kemampuan berpikir tingkat tinggi dapat dilatih melalui penggunaan instrument penilaian sebagai alat ukur (Sharif & Cho, 2015).

Instrumen penilaian dapat berupa tes yang digunakan untuk melakukan penilaian atau evaluasi. Proses penilaian tidak hanya berfungsi sebagai alat mengecek hasil belajar, tetapi juga menjadi aspek penting dalam mengukur keberhasilan peserta didik dalam proses belajar (Fidia et al., 2022). Pengukuran kemampuan berpikir tingkat tinggi dapat dilakukan dengan penilaian *Higher Order Thinking Skills* (HOTS). Instrumen HOTS berupa soal dalam tingkat kognitif C-4 (analisis), C-5 (evaluasi), dan C-6 (kreasi) (Salsabila & Trimulyono, 2023). Karakteristik penilaian HOTS berupa soal yang relevan dengan problematika yang terjadi di sekitar peserta didik. Selain itu, dapat pula berupa bentuk soal yang tidak seragam dengan tujuan untuk menguji kemampuan mengolah data (Fanani, 2018).

Menurut hasil survei lembaga assessment internasional (PISA), dalam beberapa dekade terakhir menunjukkan skor peserta didik Indonesia masih menempati urutan terbawah dibanding negara lain. Pada tahun 2015, Indonesia berada di peringkat ke-7 terbawah dari 70 negara (OECD, 2015). Penurunan skor terbukti pada survei yang dilakukan pada 2018 menunjukkan Indonesia berada di peringkat ke-5 terbawah dari 77 negara yang disurvei (OECD, 2018). Sementara itu pada data survei terbaru menurut Kementerian Pendidikan dan Kebudayaan Republik Indonesia, hasil PISA 2022 mengindikasikan penurunan prestasi belajar secara global sebagai dampak dari pandemi COVID-19. Meskipun terjadi penurunan secara global, peringkat Indonesia naik 5-6 posisi dibanding 2018 (Kemendikbud, 2023). Namun demikian, peningkatan skor ini masih tergolong rendah jika dibandingkan negara Asia Tenggara lainnya. Rendahnya tingkat perolehan skor Indonesia diakibatkan karena soal yang digunakan dalam survei berupa tipe soal HOTS sementara peserta didik di Indonesia kurang terbiasa dengan tipe soal tersebut. Kurangnya pembiasaan berpikir tingkat tinggi (HOTS) pada siswa menyebabkan sulit untuk berpikir kritis dan cenderung berpikir pada tingkat rendah (LOTS) (Avina & Winarsih, 2020). Hal ini dapat disebabkan oleh kurangnya instrumen penilaian di sekolah yang berbasis HOTS.

Menurut penelitian yang dilakukan Avina & Winarsih (2020) di SMAN 1 Bangil, sekitar 70% dari total assessment penilaian siswa hanya sebatas mengukur keterampilan berpikir tingkat rendah atau *Lower Order Thinking Skills* (LOTS). Fakta serupa juga dijumpai dalam penelitian Wardani & Ibrahim (2020) di SMAN 1 Menganti. Hasilnya hanya 33,3 % guru biologi kelas XI yang merancang alat penilaian untuk keterampilan berpikir tingkat tinggi (HOTS). Jika tidak diatasi, kondisi ini membuat peserta didik terbiasa untuk berpikir tingkat rendah (Fidia et al., 2022). Akibatnya peserta didik tidak dapat bertahan dengan

ketatnya persaingan global karena kurangnya pembiasaan. Oleh karena itu, penting untuk mengembangkan instrument HOTS untuk meningkatkan kemampuan HOTS pada peserta didik.

Berpikir tingkat tinggi dapat dikenali melalui kemampuan peserta didik menginterpretasikan apa yang diketahuinya untuk menyelesaikan masalah melalui analisis, mengevaluasi, dan mencipta (Desiriah & Setyarsih, 2021). Kemampuan ini juga dibutuhkan dan erat kaitannya dengan fenomena yang terjadi secara nyata. Salah satu materi biologi yang berhubungan dengan fenomena nyata sehari-hari adalah ekologi. Beberapa bahasan yang dipelajari dalam ekologi antara lain aliran energi, ekosistem, interaksi antar komponen ekosistem, dan daur biogeokimia. Cakupan materi yang luas dipandang sebagai salah satu kendala peserta didik dalam mempelajari ekologi (Yazid et al., 2016). Meskipun demikian, saat ini belum banyak peneliti yang mengkaji dan mengembangkan instrumen yang dapat mengukur kemampuan HOTS pada materi tersebut. Penelitian terdahulu umumnya hanya mengembangkan instrument penilaian tanpa memperhatikan kualitas instrument untuk dapat mengukur kemampuan HOTS peserta didik. Seperti penelitian yang dilakukan Hartanto et al., (2014) yang mendeskripsikan kualitas instrument sesuai dengan kurikulum 2013. Dari penelitian tersebut didapatkan instrumen asesmen yang valid dan teruji dengan nilai 0,91. Hal serupa ditemukan dalam penelitian Sensus et al., (2022) yang mengukur kesesuaian instrumen soal dengan kompetensi dasar yang ada pada materi ekologi.

Berdasarkan hal tersebut, maka penelitian ini bertujuan untuk mengembangkan instrumen soal yang valid dan dapat diandalkan dengan mempertimbangkan validitas isi dan validitas konstruksinya. Instrumen yang dikembangkan diharapkan dapat digunakan sebagai contoh referensi oleh lembaga pendidikan dalam pembuatan soal-soal penilaian sehingga tingkatan soal yang dibuat tidak lagi sebatas menilai kemampuan berpikir tingkat dasar.

METODE PENELITIAN

Penelitian ini mengacu pada penelitian pengembangan (R&D) dengan model penelitian 4D yang meliputi tahap *Define, Design, Develop, dan Disseminate* (Thiagarajan, 1974; Salsabila & Trimulyono, 2023). Pelaksanaan penelitian dilakukan pada Maret 2024 – Juni 2024 di SMAN 5 Magelang dengan subyek penelitian sebanyak 36 siswa fase E. Langkah dalam penelitian diuraikan sebagai berikut.

1. *Define*

Tahap *define* mencakup analisis *front-end*, analisis kebutuhan, serta analisis konsep dan tujuan pembelajaran yang kemudian digunakan untuk perumusan instrumen soal. Analisis *front-end* berguna untuk memunculkan dan mendefinisikan permasalahan mendasar yang dihadapi dalam proses pembelajaran (Prajoko et al., 2022). Pada penelitian ini pengembangan instrument HOTS di fokuskan pada materi ekologi.

2. *Design*

Tahap design meliputi penelusuran kajian pustaka, penentuan cakupan materi dalam instrumen penilaian, pemilihan tipe soal, penentuan media, kisi-kisi dan penyusunan *prototipe* instrument penilaian yang sesuai dengan capaian pembelajaran. Pada akhir tahap ini didapatkan desain soal pilihan ganda berjumlah 20 butir yang memenuhi kriteria HOTS. Adapun kriteria HOTS yang digunakan mengacu pada level kognitif Bloom pada Tabel 1.

Tabel 1. Level Kognitif Bloom

Level	Proses Kognitif	Definisi
C4	Menganalisis	Peserta didik dapat memecahkan materi kedalam bagian-bagiannya dan dapat menghubungkan bagian tersebut kedalam struktur atau tujuan keseluruhan
C5	Mengevaluasi	Peserta didik dapat mengevaluasi sebuah masalah dengan mengacu pada kriteria dan standar yang telah ditetapkan.
C6	Mengkreasi	Peserta didik dapat menempatkan unsur-unsur holistik dan menyusun kembali unsur tersebut dalam struktur dan gagasan yang baru.

3. Develop

Pada tahap *develop* dilakukan uji validitas yang meliputi validitas konstruk dan validitas isi (Prajoko et al., 2022). Uji validitas dilakukan dengan pengujian kepada 36 siswa di SMAN 5 Magelang. Sementara uji validitas isi mencakup uji validitas soal, uji reliabilitas, daya beda, dan indeks kesukaran menggunakan *software* Excel 2016. Untuk menguji validitas soal dapat digunakan persamaan berikut:

$$r_{pbi} = \frac{Mp - Mt}{St} \sqrt{\frac{p}{q}}$$

Hasil perhitungan kemudian disesuaikan dengan kriteria validitas menurut Sari (2018) yang tersaji pada Tabel 2.

Tabel. 2 Kriteria Validitas

Interval	Kategori
$1 \leq V < 2$	Tidak valid
$2 \leq V < 3$	Kurang valid
$3 \leq V < 4$	Cukup valid
$4 \leq V < 5$	Valid
= 5	Sangat valid

Soal yang reliabel dapat diidentifikasi dengan nilai *r* hitung lebih besar daripada *r* tabel (Arikunto, 2013). Hasil perhitungan kemudian disesuaikan dengan kriteria reliabilitas menurut (Asrul et al., 2014) yang tersaji pada Tabel 3.

Tabel. 3 Kriteria Hasil Uji Reliabilitas

Interval	Kategori
$0,00 < \alpha \leq 0,20$	Tidak reliabel
$0,20 < \alpha \leq 0,40$	Kurang reliabel

$0,40 < \alpha \leq 0,60$	Cukup reliabel
$0,60 < \alpha \leq 0,80$	Reliabel
$0,80 < \alpha \leq 1,00$	Sangat reliabel

Analisis daya beda dilakukan dengan cara membandingkan indeks daya beda dengan nilai daya beda dari instrumen soal. Untuk mengitung nilai daya beda dapat menggunakan persamaan berikut;

$$D = \frac{BA}{JA} - \frac{BB}{JA} = PA - PB$$

Hasil perhitungan kemudian dibandingkan dengan indeks kesukaran menurut (Asrul et al., 2014) yang tersaji pada Tabel 4.

Tabel. 4 Indeks Daya Beda Soal

Indeks Daya Beda	Kategori
0,00-0,19	Kurang
0,20-0,29	Cukup
0,30-0,39	Baik
0,40-1,00	Sangat baik

Tingkat kesukaran soal dapat dianalisis menggunakan persamaan menurut Asrul (2014) sebagai berikut;

$$P = \frac{B}{JS}$$

Hasil perhitungan kemudian diinterpretasikan dengan indeks kesukaran butir soal menurut (Asrul et al., 2014) yang tersaji pada Tabel 5.

Tabel 5. Indeks Kesukaran Butir Soal

Interval	Kategori
0,00-30	Sukar
0,31-0,70	Sedang
0,71-1,00	Mudah

Validitas konstruk kemudian diujikan pada 2 orang ahli bahasa dan ahli materi. Validitas konstruk bertujuan untuk menuji format instrumen, isi, konstruksi soal, dan bahasa (Prajoko et al., 2022). Uji validitas teoritis menggunakan skala linkert 1- 4 untuk tiap penilaian kriteria. Nilai rata-rata hasil penilaian kemudian diinterpretasikan dengan kriteria yang diadaptasi dari (Nurdin, 2007; Prajoko, et al 2021) yang disajikan pada Tabel 6.

Tabel . 6 Kriteria Hasil Uji Validitas Konstruk

Skor Rata-Rata	Kriteria
$3,5 \leq M \leq 4,0$	Sangat baik
$2,5 \leq M \leq 3,5$	Baik
$1,5 \leq M \leq 2,5$	Cukup
$M < 1,5$	Buruk

4. Disseminate

Setelah direvisi dan dinyatakan valid oleh validator, instrumen penilaian kemudian diuji coba pada 36 peserta didik fase E di SMAN 5 Magelang secara terbatas. Pelaksanaan ujicoba dilakukan pada tanggal 14 April 2024 dengan durasi pengerjaan selama 90 menit.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil penelitian diperoleh berupa instrument tes tingkat tinggi pada materi ekologi. Rancangan awal produk yang dikembangkan dilakukan dengan menguji kisi-kisi berdasarkan capaian pembelajaran, kompetensi esensial dan indikator pembelajaran yang kemudian dituangkan dalam bentuk soal tes pilihan ganda. Hasil penelitian ini mencakup uji validitas isi dan validitas konstruk. Validitas isi mencakup validitas butir soal, reliabilitas, daya beda, dan tingkat kesukaran. Sementara validitas konstruk meliputi ranah materi, konstruksi dan bahasa.

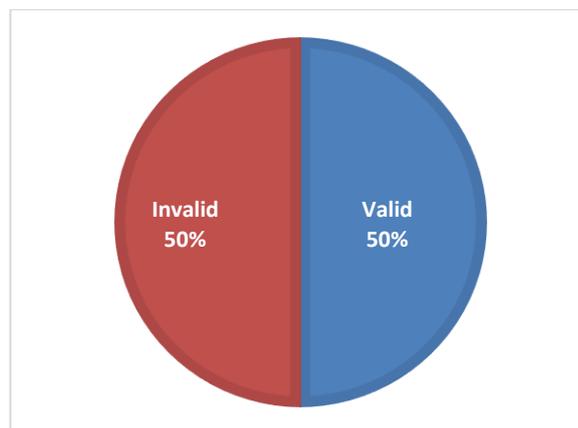
Validitas Isi

Validitas menggambarkan tingkat baiknya butir soal yang sedang dikembangkan. Kevalidan soal dapat berupa hal aktual yang sedang diteliti (Taherdoost, 2016). Analisis validitas butir soal dilakukan dengan *software* Microsoft Excel 2016, adapun hasil perhitungan disajikan pada Tabel 7.

Tabel 7. Uji Validitas

Aspek	Valid	Invalid
Soal Nomor	2,3,4,5,7,8,13,16,18,20	1,6,9,10,11,12,14,15,17,19
Total	10 (50%)	10 (50%)

Dari data pada tabel 7. Uji Validitas maka dapat ditampilkan dalam bentuk diagram lingkaran dapat dilihat pada Gambar 1.



Gambar 1. Diagram Indeks Validitas

Berdasarkan tabel 7 presentase pertanyaan valid sebesar 50%. Soal dianggap valid jika nilai r tabel lebih rendah dibanding r hitung. Berdasarkan uji produk yang telah dilakukan didapatkan 10 soal dengan nilai lebih dari 0,344. Sementara itu, nomor soal 1,6,9,10,11,12,14,15,17, dan 19 dikategorikan sebagai soal yang buruk karena memiliki tingkat validitas yang buruk. Menurut Souza et al., (2017) menyatakan bahwa validitas dapat diartikan sebagai tingkat sejauh mana alat ukur dapat secara akurat

dan tepat mengukur apa yang dimaksudkan untuk diukur. Hal ini serupa dengan penelitian Setyawaty et al., (2018) bahwa suatu instrument layak digunakan apabila dapat menghasilkan hasil yang sama untuk mengevaluasi dari suatu pengukuran Butir soal yang tidak valid tidak dapat dijadikan sebagai instrument tes sehingga harus dilakukan revisi. Proses revisi dapat dilakukan dengan meningkatkan penguasaan teknik peneliti dalam menyusun instrument soal dan penelitian. Sementara itu, butir soal yang dianggap valid dapat dipergunakan untuk pengukuran di masa mendatang. Setelah dinyatakan valid, butir soal dapat dikumpulkan kedalam bank soal (Prajoko et al., 2022). Suatu soal dikategorikan valid apabila butir soal dapat mengukur hal apa yang ingin diukurnya (Asrul et al., 2014)

Reliabilitas

Dalam penelitian ini sampel penelitian sebanyak 36 siswa kelas X di SMA N 5 Magelang yang sudah mempelajari materi ekosistem. Analisis realibilitas soal dilakukan dengan Microsoft Excel 2016 dengan hasil perhitungan disajikan pada Tabel 8.

Tabel 8. Uji Reliabilitas

Cronbach Alfa	Kategori
0,674	Reliabel

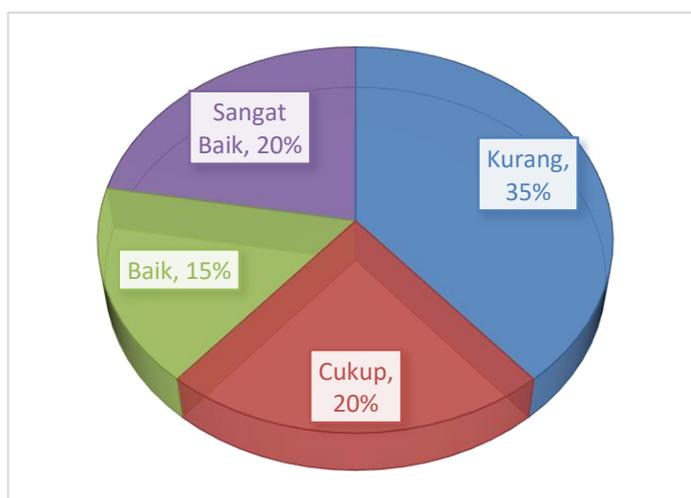
Reliabilitas digunakan untuk menunjukkan tingkat keabsahan suatu butir soal agar dapat dipercaya hasil pengukuran dengan soal tersebut. Hasil pengukuran harus dapat diandalkan dan memiliki konsistensi. Sebagaimana yang dielaskan oleh Bolarinwa (2015) bahwa reliabilitas berhubungan dengan konsistensi skor dari waktu ke waktu. Hasil pengujian pada Tabel 8, menunjukkan instrumen yang diuji termasuk reliabel, artinya hasil pengukuran menggunakan isntumen tersebut dapat dipercaya dan menghasilkan hasil pengukuran yang tepat. Nilai reliabilitas didapatkan dari pengujian rumus Spearman-Brown, didapatkan skor Cronbach alfa dengan nilai 0,674, sementara nilai r tabel sebesar 0,44. Berdasarkan nilai tersebut menunjukkan jika $0,67 > 0,60$, sehingga soal termasuk reliabel. Nilai r hitung lebih tinggi dari r table sehingga insrtumen tes dikatakan reliabel (Arikunto, 2013).

Daya Beda

Daya diskriminatif mengacu pada kemampuan butir-butir untuk membedakan siswa yang memiliki tingkat kemampuan intelektual tertentu dengan siswa yang tingkat kecerdasannya lebih rendah (Mukherjee & Lahiri, 2015). Daya pembeda suatu soal dikatakan baik apabila mampu membedakan siswa berkemampuan tinggi dengan siswa berkemampuan rendah (Prajoko et al., 2022). Hail uji daya beda pada penelitian ini disajikan dalam Tabel 9. Visualisasi hasil uji indeks daya beda juga ditampilkan pada Gambar 2.

Tabel. 9 Uji Daya Beda

Kategori	Nomor Soal	Total (%)
Kurang	6,8,11,12,14,15,17	7 (35%)
Cukup	1,5,7,9,10,19	6 (20%)
Baik	3,13,20	3 (15%)
Sangat baik	2,4,16,18	4 (20%)



Gambar 2. Diagram Indeks Daya Beda

Berdasarkan Tabel 9, diperoleh bahwa 35% butir soal termasuk dalam kategori daya beda yang kurang. Sedangkan soal dengan tingkat “Cukup” sebesar 20 %, persentase soal dengan kategori “Baik” sebesar 15%, persentase soal dengan kategori “Sangat Baik” sebesar 20%. Soal dikategorikan telah memenuhi kriteria apabila sudah termasuk pada kategori cukup, baik dan amat baik (Warju et al., 2020). Soal dengan daya beda yang buruk tidak dapat digunakan untuk mencapai tujuan dalam membedakan tingkat kemampuan siswa. Soal dengan daya beda yang buruk jika diberikan pada siswa akan memberikan hasil yang sama sehingga soal yang kurang memadai sebaiknya segera dihapus dan tidak digunakan.

Beberapa faktor yang menyebabkan butir soal memiliki daya beda yang buruk antara lain, soal mengandung bias, terlalu sulit, dan pengecoh yang tidak logis. Hal ini didukung oleh penelitian Mulyani et al., (2022) butir soal yang memiliki daya beda buruk dapat diakibatkan oleh beberapa faktor, seperti tingkat kesulitan soal sehingga menimbulkan bias. Akibatnya peserta didik yang telah memiliki pemahaman terhadap materi maupun peserta didik yang belum paham dengan materi tidak dapat mengukur menggunakan soal yang berkategori mudah maupun sulit. Dari 35% butir soal yang memiliki daya pembeda kurang rata-rata tergolong kategori soal yang sedang. Penyebab lain butir soal dianggap memiliki daya beda yang buruk dikarenakan karena kunci jawaban soal yang kurang akurat. Faktor ketiga dapat disebabkan tidak berfungsinya pengecoh pada jawaban butir soal.

Pengembangan butir soal HOTS yang dikategorikan kurang seharusnya tidak ditambahkan ke dalam bank soal. Akan tetapi jumlah butir soal dengan daya pembeda yang tergolong kategori cukup, baik, dan sangat baik jumlahnya lebih banyak dibandingkan dengan butir soal yang termasuk kategori kurang, sehingga butir soal tersebut dapat digunakan dalam tes hasil belajar (Oktanin & Sukirno, 2015).

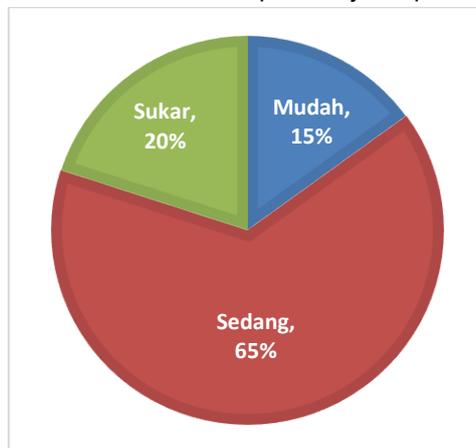
Indeks Kesukaran

Soal dianggap semakin sulit jika indeks tingkat kesukarannya rendah. Sebaliknya, jika indeksnya tinggi maka soal tersebut dianggap lebih mudah (Wantoro et al., 2019). Soal yang baik mempunyai tingkat kesukaran sedang (Arikunto, 2013). Untuk memahami tingkat kesulitan pada tiap butir soal maka diperlukan uji coba soal tersebut. Setiap butir soal memiliki skala kesukaran yang dibagi menjadi kategori mudah, sedang, dan sulit (Pistol et al., 2018). Hasil uji indeks kesukaran disajikan pada Tabel 10.

Tabel 10. Uji Indeks Kesukaran

Kategori	Nomor Soal	Total
Mudah	6,8,19	3 (15%)
Sedang	1,2,3,4,5,7,9,11,13,14,15,16,18	13 (65%)
Sukar	10,12,17,20	4 (20%)

Data pada tabel 10. Uji Indeks Kesukaran dapat disajikan pada diagram lingkaran Gambar 3.



Gambar 3. Diagram Indeks Kesukaran

Berdasarkan hasil uji coba didapatkan rerata soal dengan kriteria “Sedang” 4 soal, dengan interval nilai 0,09 – 0,21. Pada kriteria “Sulit”, 1 soal dengan nilai 0,75. Soal “Mudah”, 15 butir dengan nilai interval 0,30 – 0,69. Hal ini sejalan dengan penelitian Sudjana yang kemukakan oleh Warju et al., (2020) dimana mengarah pada proporsi perbandingan 3-5-2 atau 30 % butir soal kategori mudah, 50% butir soal kategori sedang, dan 20 % butir soal kategori sulit.

Validitas Konstruk

Analisis keabsahan soal dilakukan berdasarkan penilaian validator. Hasil penilaian terjadi pada Tabel 11.

Tabel 11. Uji Validitas Konstruk

Aspek	Nilai Rata-Rata	Kriteria
Bahan	3,33	Baik
Konstruksi	3,5	Sangat baik
Bahasa	4	Sangat baik
Total	3,61	Sangat baik

Berdasarkan tabel diatas menunjukan butir soal yang dikembangkan termasuk kedalam katagori yang sangat baik dengan skor perolehan 3,61 sehingga memenuhi validitas konstruk yang baik dan dapat diuji. Soal dinyatakan valid apabila dapat mengukur kopetensi yang ingin dicapai (Prajoko et al., 2022). Validitas konstruk pada penelitia ini meliputi ranah materi, konstruksi dan kebahasaan. Perolehan nilai 3,61 berdasarkan hasil butir soal yang dikembangkan dan diselaraskan dengan indikator yang

dikembangkan melalui integrasi HOTS pada materi ekologi untuk kelas X.. Instrumen juga menggunakan penomoran butir soal yang jelas dan ukuran huruf yang mengikuti standar. Selain itu, tata letak dan penulisan sudah disesuaikan dengan pedoman penulisan soal pilihan ganda. Dalam segi kebahasaan, soal yang dikembangkan sudah menerapkan kaidah kebahasaan yang baik dan benar.

SIMPULAN

Berdasarkan pada 20 butir soal HOTS yang dikembangkan, beberapa di antaranya memerlukan perbaikan atau tidak layak digunakan. Dari 20 soal pilihan ganda yang disusun, 50% di antaranya termasuk dalam kategori valid, sementara 50% lainnya tergolong tidak valid. Selain itu, uji reliabilitas instrumen tes menghasilkan nilai koefisien 0,674, yang menunjukkan bahwa instrumen tersebut dapat diandalkan. Untuk daya beda, 35% butir soal berada dalam kategori kurang, 20% dalam kategori cukup, 15% dalam kategori baik, dan 20% dalam kategori sangat baik. Dalam hal tingkat kesulitan, 15% butir soal termasuk kategori mudah, 65% dalam kategori sedang, dan 20% dalam kategori sulit. Validitas konstruksi memperoleh skor 3,61, yang termasuk dalam kategori sangat baik.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada SMAN 5 Magelang dan semua pihak yang telah berkontribusi dalam proses penelitian ini.

RUJUKAN

- Afrita, M., & Rahmawati, D. (2019). Validitas Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi Materi Sistem Respirasi di Kelas XI SMA. *Jurnal Mangifera Edu*, 4(2), 129–142. <https://doi.org/10.31943/mangiferaeduv4i2.83>
- Arikunto, S. (2013). *Prosedur Penelitian, Suatu Pendekatan Praktik*. Rineka Cipta.
- Asrul, Ananda, R., & Rosnita. (2014). *Evaluasi Pembelajaran*. Citapustaka Media.
- Avina, Y., & Winarsih. (2020). Pengembangan Instrumen Penilaian sebagai Contoh Paket Soal Higher Order Thinking Skills (HOTS) Materi Pencemaran Lingkungan Kelas X SMA. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 9(1), 217–223. <https://doi.org/https://doi.org/10.26740/bioedu.v9n2.p217-223>
- Bolarinwa, O. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 22(4), 195. <https://doi.org/10.4103/1117-1936.173959>
- Desiriah, E., & Setyarsih, W. (2021). Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (HOTS) Fisika di SMA. *ORBITA*, 7(1).
- Fanani, M. (2018). Strategi Pengembangan Soal HOTS dalam Kurikulum 2013. *Journal of Islamic Religious Education*, 2(1), 57–65.
- Fidia, F., Rinie Pratiwi Puspitawati, R., & Yakub, P. (2022). Pengembangan Instrumen Soal Higher Order Thinking Skills (HOTS) Materi Jaringan dan Organ pada Tumbuhan Kelas XI SMA. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 11(3). <https://doi.org/https://doi.org/10.26740/bioedu.v11n1.p138-147>

- Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan Instrumen Penilaian Berbasis HOTS Menggunakan Aplikasi "QUIZIZZ." *Efisiensi: Kajian Ilmu Administrasi*, 18(1), 105–124. <https://doi.org/10.21831/efisiensi.v17i1.34895>
- Hartanto, R., Fitrihidajati, H., & Prastiwi, M. (2014). Kualitas Instrumen Asesment Berbasis Kurikulum 2013 Untuk Menilai Materi Ekologi SMA. *BioEdu*, 3(3).
- Kemendikbud. (2023, December 5). *Peringkat Indonesia pada PISA 2022 Naik 5-6 Posisi Dibanding 2018*.
- Mukherjee, P., & Lahiri, S. K. (2015). Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal. *IOSR Journal of Dental and Medical Sciences (IOSR-JDMS) e-ISSN*, 14(12), 47–52. <https://doi.org/10.9790/0853-141264752>
- Mulyani, S., Krismonita, M., & Yamtinah, S. (2022). Analisis Butir Soal dan Kecukupan HOTS Soal Ujian Akhir Semester Mata Pelajaran Kimia SMK Kelas X. *PAEDAGOGIA*, 25(2), 162. <https://doi.org/10.20961/paedagogia.v25i2.60913>
- OECD. (2015). *PISA 2015 Assesment and Analytical and Financial Library*.
- OECD. (2018). *PISA 2018 Insthings and Interprwetations*.
- Oktanin, W. S., & Sukirno, S. (2015). Analisis Butir Soal Ujian Akhir Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 13(1). <https://doi.org/10.21831/jpai.v13i1.5183>
- Pistol, I., Trandabăț, D., & Răschip, M. (2018). Medi-Test: Generating Tests from Medical Reference Texts. *Data*, 3(4), 70. <https://doi.org/10.3390/data3040070>
- Prajoko, S., Erawati, A., Amali, F., Zunaena, M., & Arganingtias, I. (2022). The Development of High Order Thinking Skills Test based on Google Forms on Cell Biology Materials. *JOBE*, 5(1), 64–79.
- Salsabila, P. &, Trimulyono, G. (2023). Pengembangan Instrumen Penilaian HOTS Materi Virus untuk Mengukur Kemampuan Berpikir Kritis Siswa Kelas X SMA. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 12(2), 287–297. <https://doi.org/https://doi.org/10.26740/bioedu.v12n2.p287-297>
- Sensus, M., Arifin, K., & Munir, A. (2022). Validitas Soal pada Asesmen Kompetensi Minimum Materi Ekologi SMA Kelas X. *Wahana-Bio*, 14(1), 1–10.
- Setyawaty, R., Sulistyorini, T. B., . M., & Rahmawati, L. E. (2018). Validity Test and Reliability of Indonesian Language Multiple Choice in Final Term Examination. *KnE Social Sciences*, 3(9), 43. <https://doi.org/10.18502/kss.v3i9.2609>
- Sharif, A., & Cho, S. (2015). 21st-Century Instructional Designers: Bridging the Perceptual Gaps between Identity, Practice, Impact and Professional Development. *RUSC. Universities and Knowledge Society Journal*, 3, 72–85.
- Sibirian, J., Corebima, A. D., Ibrohim, & Saptasari, M. (2019). The Correlation Between Critical and Creative Thinking Skills on Cognitive Learning Results. *Eurasian Journal of Educational Research*, 2019(81), 99–114. <https://doi.org/10.14689/ejer.2019.81.6>
- Souza, A. C. de, Alexandre, N. M. C., Guirardello, E. de B., Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação

- da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3205040>
- Wantoro, J., Utama, S., Zuhriah, S., & Hafida, S. H. N. (2019). Pengembangan Instrumen Penilaian Pendidikan Profesi Guru Sekolah Dasar Berbasis HOTS. *Profesi Pendidikan Dasar*, 1(1), 11–20. <https://doi.org/10.23917/ppd.v1i1.8453>
- Wardani, A., & Ibrahim, M. (2020). Karakteristik Higher Order Thinking Skills (HOTS) Materi Dampak Penyalahgunaan Psikotropika Untuk SMA. *Jurnal BioEdu*, 9(1), 60–67.
- Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95. <https://doi.org/10.23887/jptk-undiksha.v17i1.22914>
- Yazid, K., Susantini, E., & Fitrihidajati, H. (2016). Validitas Buku Saku Materi Ekologi Untuk Siswa Kelas X SMA. *BIOEDU*, 5(3).