

# **DEVELOPMENT OF TESTING INSTRUMENTS BASED ON HIGHER ORDER THINKING SKILL (HOTS) ON ECOLOGY CLASS X SENIOR HIGH SCHOOL**

**Restu Sitompul, Darmawati, Yuslim Fauziah**

Email: restu.sitompul6766@student.unri.ac.id, darmawati@lecturer.unri.ac.id,  
yuslim.fauziah@lecturer.unri.ac.id  
Phone Number: +6285206314273

*Study Program of Biology Education  
Department of Mathematics and Natural Sciences  
Faculty of Teacher Training and Education  
Riau University*

**Abstract:** *HOTS-based assessment is one of the demands of the 2013 curriculum. Students are expected to have high reasoning abilities by getting used to solving questions that are also HOTS in nature. However, the results of observations made in schools, HOTS questions are still rarely used by teachers in the process of assessing learning outcomes. This development research aims to produce a quality higher order thinking skill (HOTS) based test instrument. This research method is a research and development (R&D) method with the ADDIE model developed by Dick and Carry which consists of five stages, namely analyze, design, development, implement and evaluated. However, this research was carried out only until the development stage. Data was collected by means of interviews, validation and limited trials. The results of this study were 16 quality HOTS questions based on aspects of validation, reliability, level of difficulty, discriminating power and distracting function, namely 6 practice questions and 10 daily test questions. This is because only the 16 questions match the HOTS test criteria, namely 1) High validity with a total of 25 questions, namely material validity greater than 3.73, construction validity greater than 3.70 and language validity greater than 3.72. 2) High reliability, which is 0.73 for practice questions and 0.84 for daily test questions. 3) Has a moderate to difficult level of difficulty. 4) The discriminatory power of the questions is already good. 5) The distractors are all functioning properly.*

**Key Words:** *Higher Order Thinking Skill, Tes Instrument.*

# PENGEMBANGAN INSTRUMEN TES BERBASIS HIGHER ORDER THINKING SKILL (HOTS) PADA MATERI EKOLOGI KELAS X SEKOLAH MENENGAH ATAS

**Restu Sitompul, Darmawati, Yuslim Fauziah**

Email: restu.sitompul6766@student.unri.ac.id, darmawati@lecturer.unri.ac.id,

yuslim.fauziah@lecturer.unri.ac.id

Nomor Hp: +6285206314273

Program Studi Pendidikan Biologi  
Jurusan Pendidikan Matematika dan Ilmu Pengetahuan Alam  
Fakultas Keguruan dan Ilmu Pendidikan  
Universitas Riau

**Abstrak:** Penilaian berbasis HOTS merupakan salah satu tuntutan kurikulum 2013. Siswa diharapkan memiliki kemampuan penalaran tinggi dengan pembiasaan diri menyelesaikan pertanyaan-pertanyaan yang sifatnya juga HOTS. Namun hasil observasi yang dilakukan di sekolah, soal-soal HOTS masih jarang digunakan oleh guru dalam proses penilaian hasil belajar. Penelitian pengembangan ini bertujuan untuk menghasilkan instrumen tes berbasis higher order thinking skill (HOTS) yang berkualitas. Metode penelitian ini adalah metode penelitian dan pengembangan (R&D) dengan model ADDIE yang dikembangkan oleh Dick dan Carry yaitu terdiri dari lima tahapan yaitu *analyze, design, development, implement and evaluated*. Akan tetapi penelitian ini dilakukan hanya sampai tahap *development* (pengembangan). Pengumpulan data dilakukan dengan wawancara, validasi dan uji coba terbatas. Hasil penelitian ini adalah 16 soal HOTS yang berkualitas berdasarkan aspek validasi, reliabilitas, tingkat kesukaran, daya pembeda dan fungsi pengecoh, yaitu 6 soal latihan dan 10 soal ulangan harian. Hal ini dikarenakan hanya 16 soal tersebut yang sesuai dengan kriteria tes HOTS, yaitu 1) Validitas yang tinggi dengan jumlah 25 soal, yaitu validitas materi lebih besar dari 3,73, validitas konstruksi lebih besar dari 3,70 dan validitas bahasa lebih besar dari 3,72. 2) Reliabilitas yang tinggi yaitu sebesar 0,73 untuk soal latihan dan 0,84 untuk soal ulangan harian. 3) Memiliki tingkat kesukaran yang sedang hingga sukar. 4) Daya pembeda soal yang sudah baik 5) Pengecoh soal seluruhnya sudah berfungsi dengan baik.

**Kata kunci:** *Higher order thinkings skill* , Instrumen tes.

## PENDAHULUAN

Proses pembelajaran biologi pada kurikulum 2013 mengharapkan guru dapat memfasilitasi dan memotivasi siswa untuk berpikir kritis, logis dan sistematis serta memiliki kemampuan berpikir tingkat tinggi (HOTS). Pemetaan dari kompetensi dasar mata pelajaran biologi pada kurikulum 2013 menunjukkan bahwa 23 dari 35 materi menargetkan pencapaian tingkat berpikir tinggi (Kemendikbud, 2016). Salah satu materi biologi yang diharapkan dapat meningkatkan kemampuan berpikir tingkat tinggi tersebut adalah materi ekologi. Materi pembelajaran ekologi merupakan materi yang memiliki tingkat kesukaran sedang yang bisa dimanfaatkan untuk meningkatkan kemampuan berpikir tingkat tinggi siswa. Disamping itu, materi ekologi merupakan materi yang berhubungan langsung dengan kehidupan sehari-hari yang akan cukup memudahkan siswa untuk melatih kemampuan berpikir tingkat tinggi (HOTS) mereka.

Taksonomi Bloom mengelompokkan tingkatan berpikir peserta didik menjadi dua, yaitu tingkat tinggi dan tingkat rendah. Kemampuan tingkat rendah terdiri atas pengetahuan, pemahaman, dan aplikasi, sedangkan kemampuan tingkat tinggi meliputi analisis, sintesis, evaluasi dan kreativitas. Dengan demikian, kegiatan peserta didik dalam menghafal termasuk kemampuan tingkat rendah.

Rendahnya kemampuan siswa dalam berpikir, bahkan hanya dapat menghafal, tidak terlepas dari kebiasaan guru dalam melakukan evaluasi atau penilaian yang hanya mengukur tingkat kemampuan yang rendah saja melalui *paper and pencil tes*. Dan kurang tersedianya instrumen tes yang didesain khusus untuk melatih siswa agar menggunakan kemampuan berpikir tingkat tinggi pada aspek pemecahan masalah dalam menjawab setiap permasalahan yang merupakan salah satu penyebab rendahnya kemampuan berpikir tingkat tinggi siswa. Oleh karena itu, upaya yang dapat dilakukan adalah mengembangkan instrumen tes berbasis *higher order thinking skill* yang berkualitas.

Tes dapat didefinisikan sebagai salah satu pertanyaan, atau tugas, ataupun seperangkat tugas yang direncanakan untuk memperoleh informasi tentang elemen dasar atau atribut pendidikan atau psikologi yang setiap butir pertanyaan atau tugas tersebut mempunyai jawaban atau ketentuan yang dianggap benar (Zulkifli, dkk., 2019). Instrumen tes adalah alat ukur yang digunakan dalam rangka pengumpulan data. Dalam pendidikan, instrumen alat ukur yang digunakan untuk mengumpulkan data dapat berupa tes atau non tes. Tes merupakan himpunan pertanyaan yang harus dijawab, harus ditanggapi, atau tugas yang harus dilaksanakan oleh orang yang di tes. Tes digunakan untuk mengukur sejauh mana seorang siswa telah menguasai pelajaran yang telah disampaikan terutama meliputi aspek pengetahuan dan keterampilan.

*Higher order thinking skill* (HOTS) atau kemampuan berpikir tingkat tinggi merupakan kemampuan seseorang dalam mengaitkan antara informasi yang sudah tersimpan didalam ingatannya, kemudian menghubungkan-hubungkan dan atau menata ulang serta mengembangkan informasi tersebut sehingga tercapai suatu tujuan ataupun suatu penyelesaian dari suatu keadaan yang sulit dipecahkan (Rosnawati, 2013). Brookhart (dalam Kemendikbud, 2019) menggunakan tiga istilah dalam mendefinisikan keterampilan berpikir tingkat tinggi (HOTS), yaitu:

- 1) HOTS adalah proses transfer
- 2) HOTS adalah berpikir kritis
- 3) HOTS adalah penyelesaian masalah

Instrumen berbasis HOTS ini jarang diaplikasikan oleh guru disekolah dikarenakan penyusunannya yang sulit dan contoh soal HOTS yang disediakan pada pedoman penyusunan soal HOTS terasa masih belum cukup atau sedikit. Kurangnya atau tidak tersedianya instrumen penilaian yang dirancang untuk meningkatkan kemampuan berpikir tingkat tinggi (HOTS) pada materi ekologi pada sekolah memotivasi peneliti untuk menghasilkan produk instrument tes berbasis *higher order thinking skill* (HOTS) yang berkualitas.

## METODE PENELITIAN

Metode penelitian ini adalah metode penelitian dan pengembangan (R&D) dengan model ADDIE yang dikembangkan oleh Dick dan Carry yaitu terdiri dari lima tahapan yaitu *analyze, design, development, implement and evaluated*. Akan tetapi penelitian ini dilakukan hanya sampai tahap *development* (pengembangan). Subjek uji coba penelitian ini adalah kompetensi 3.10 “Menganalisis komponen-komponen ekosistem dan interaksi antar komponen tersebut”. Pengumpulan data dilakukan dengan wawancara, validasi dan uji coba terbatas. Data dianalisis secara deskriptif yaitu melalui analisis validasi dan analisis butir soal pada uji coba terbatas.

### Validasi

Sebelum dilakukan uji coba, soal terlebih dahulu divalidasi oleh ahli evaluasi, ahli materi dan ahli bahasa untuk mendapatkan perbaikan dan mengetahui kelayakan soal sebelum uji coba. Untuk mengetahui hal tersebut maka dilakukannya pencarian persentase kelayakan dari hasil skor validasi dengan menggunakan rumus yaitu:

$$\bar{x} = \frac{\sum x}{n} \quad (\text{Sudjana, 2012})$$

Keterangan:

$\bar{x}$  = rata-rata skor hasil validasi  
 $\sum x$  = jumlah skor pada item ke-i  
 $n$  = skor maksimal item

Kriteria dalam mengambil keputusan dalam validasi instrumen tes berbasis HOTS dapat dilihat pada tabel berikut

Tabel 1. Kriteria Validitas

Interval Rata-rata Skor	Kategori
$3.25 \leq x < 4$	Sangat baik
$2.5 \leq x < 3.25$	baik
$1.75 \leq x < 2.5$	Kurang baik
$1 \leq x < 1.75$	Tidak baik

(Sumber: Sugiyono, 2017)

Setelah mendapatkan hasil dari uji coba peserta didik, maka untuk mencari kelayakan butir soal perlu dilakukannya analisis butir soal. Hal ini sejalan dengan ciri instrumen tes yang baik menurut Arikunto (dalam Widoyoko, 2019) dan Febriana (2019) yaitu tes yang baik adalah tes yang dapat mengukur apa yang harus diukur, reliabilitas, tidak terlalu mudah maupun sukar, dan dapat membedakan antara peserta didik yang memiliki kemampuan tinggi dan peserta didik yang memiliki kemampuan rendah. Oleh karena itu, analisis butir soal dalam penelitian ini meliputi sebagai berikut:

## Reliabilitas

Reliabilitas menunjukkan sejauh mana suatu pengukuran dapat dipercaya. Dapat dipercaya berarti apabila tes tersebut digunakan dalam beberapa kali pelaksanaan pengukuran mendapatkan hasil pengukuran yang relatif sama (Muzakki, 2019). Reliabilitas soal pilihan ganda dianalisis menggunakan rumus sebagai berikut:

$$r_1 = \frac{n}{(n-1)} \left[ 1 - \frac{\sum pq}{s^2} \right]$$

(Arikunto, 2013)

Keterangan:

$r_1$  = reliabilitas tes secara keseluruhan soal pilihan ganda

$p$  = proporsi subjek menjawab item dengan benar

$q$  = proporsi subyek menjawab item dengan salah ( $q = 1 - p$ )

$\sum pq$  = jumlah hasil perkalian antara  $p$  dan  $q$   $n$  = banyaknya item

$S$  = standar deviasi dari tes (standar deviasi adalah akar varians)

Kategori tingkat reliabilitas butir soal menurut Arikunto (2013) dapat dilihat pada tabel berikut.

**Tabel 2. Kriteria Reliabilitas**

Interval rata-rata skor	Kategori Reliabilitas
$0,50 < r_{11} \leq 1,00$	Sangat tinggi
$0,40 \leq r_{11} \leq 0,50$	Tinggi
$0,30 \leq r_{11} \leq 0,40$	Sedang
$0,20 \leq r_{11} \leq 0,30$	Rendah
$r_{11} \leq 0,20$	Sangat rendah

(Sumber: Arikunto, 2013)

## Daya Pembeda

Menurut Febriana (2019) daya pembeda soal adalah kemampuan sesuatu soal untuk membedakan antara peserta didik yang pandai (berkemampuan tinggi) dengan peserta didik yang kurang pintar (berkemampuan rendah).

$$DP = \frac{\sum A}{nA} - \frac{\sum B}{nB}$$

(Febriana, 2019)

Keterangan:

DP = Indeks daya beda

$\Sigma A$  = Jumlah peserta tes yang menjawab benar pada kelompok atas

$\Sigma B$  = Jumlah peserta tes yang menjawab benar pada kelompok bawah

nA = Jumlah peserta tes pada kelompok atas

nB = Jumlah peserta tes pada kelompok bawah

Kategori tingkat daya pembeda butir soal menurut Febriana (2019) dapat dilihat pada tabel berikut.

Tabel 3. Kriteria Daya Pembeda

No	Range Daya Pembeda	Kategori	Keterangan
1	0,40 – 1,00	Sangat Baik	Tidak direvisi
2	0,30 – 0,39	Baik	Tidak direvisi
3	0,20 – 0,29	Cukup	Revisi
4	0,00 – 0,19	Jelek	Revisi

(Sumber: Febriana, 2019)

### Tingkat Kesukaran

Tingkat kesukaran soal adalah peluang menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Adapun rumus menghitung tingkat kesukaran bentuk pilihan ganda menurut Kusaeri (2014).

$$TK = \frac{\text{Peserta didik yang menjawab benar}}{\text{Jumlah peserta didik yang mengikuti tes}}$$

Kategori tingkat reliabilitas butir soal menurut Arikunto (2013) dapat dilihat pada tabel berikut.

Tabel 4 Kriteria Tingkat Kesukaran

No	Range Tingkat Kesukaran	Kategori
1	0,71 – 1,00	Mudah
2	0,31 – 0,71	Sedang
3	0,00 – 0,30	Sukar

(Sumber: Kusaeri, 2014)

## Fungsi Distraktor (Pengecoh)

Menurut Suharsimi Arikunto (2013), suatu distraktor dapat dikatakan berfungsi dengan baik jika paling sedikit dipilih oleh 5% dari pengikut test ( $5\% \times \text{Jumlah Peserta} = \dots$ ). Apabila persentase kecil dari 5% maka pengecoh harus diperbaiki.

## HASIL DAN PEMBAHASAN

### Hasil Penelitian

#### Validasi

Berdasarkan hasil validasi yang telah dianalisis diperoleh rata-rata skor total validasi instrumen tes berbasis HOTS sebagai berikut.

**Tabel 5** Rekapitulasi Hasil Validasi Butir Soal Latihan

No	Aspek Penilaian	Skor Rata-rata			Rata-rata Skor Total	Kategori
		V.1	V.2	V.3		
1	Materi	3,38	3,87	4,00	3,75	Sangat Baik
2	Konstruksi	3,34	3,97	4,00	3,77	Sangat Baik
3	Bahasa	3,30	3,88	4,00	3,73	Sangat Baik
		<b>Rata-rata</b>			<b>3,75</b>	Sangat Baik

#### Keterangan:

V.1: Validator 1, V.2: Validator 2, V.3: Validator 3

Hasil rekapitulasi pada tabel 5 menunjukkan bahwa butir soal dinyatakan sangat baik dengan skor rata-rata adalah 3,75. Hasil skor rata-rata dari seluruh validator pada masing-masing aspek menunjukkan  $\bar{x} > 3,25$ .

**Tabel 6** Rekapitulasi Hasil Validasi Butir Soal Ulangan Harian

No	Aspek Penilaian	Skor Rata-rata			Rata-rata Skor Total	Kategori
		V.1	V.2	V.3		
1	Materi	3,30	3,93	4,00	3,74	Sangat Baik
2	Konstruksi	3,19	3,93	4,00	3,71	Sangat Baik
3	Bahasa	3,30	4,00	4,00	3,77	Sangat Baik
		<b>Rata-rata</b>			<b>3,74</b>	Sangat Baik

#### Keterangan:

V.1: Validator 1, V.2: Validator 2, V.3: Validator 3

Hasil rekapitulasi pada tabel 6 menunjukkan bahwa butir soal dinyatakan sangat valid dengan masing-masing skor rata-rata 3,74. Hasil skor rata-rata dari seluruh

validator pada masing-masing aspek menunjukkan  $\bar{x} > 3,25$ . Meskipun produk awal instrumen tes memiliki tingkat validitas yang sangat baik, produk awal ini masih perlu direvisi. Revisi dilakukan sesuai saran yang diberikan para ahli (validator).

## Uji coba terbatas I

### Reliabilitas

Berdasarkan analisis data terhadap soal latihan dan soal ulangan harian yang telah diujicobakan diperoleh hasil reliabilitas seperti pada tabel 7 berikut.

Tabel 7. Reliabilitas Soal HOTS pada Uji Coba I

Jenis Soal	Nilai Reliabilitas	Kategori
Soal Latihan	0,92	Sangat Tinggi
Soal Ulangan Harian	0,91	Sangat Tinggi

Berdasarkan analisis data tersebut reliabilitas tes yang diperoleh adalah 0,92 pada soal latihan dan 0,91 pada soal ulangan harian dengan interpretasi reliabilitas tinggi. Sehingga berdasarkan analisis reliabilitas tes tersebut, maka tidak perlu dilakukan revisi instrumen tes.

### Daya Pembeda

Berikut ini adalah nilai daya pembeda instrumen tes HOTS pada materi ekologi yang dikembangkan

Tabel 8. Nilai Daya Pembeda Soal Latihan Uji Coba I

Kategori	Range Daya Pembeda	Nomor Soal	Jumlah Soal
Sangat Baik	0,40 – 1,00	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10
Baik	0,30 – 0,39	-	0
Cukup	0,20 – 0,29	-	0
Jelek	0,00 – 0,19	-	0

Berdasarkan tabel 8 diketahui bahwa seluruh soal latihan yang diujicobakan kepada 10 responden mahasiswa menunjukkan hasil sangat baik sesuai kriteria nilai daya pembeda pada bab sebelumnya yaitu dengan persentase 100% memiliki daya pembeda yang baik.

Tabel 9. Nilai Daya Pembeda Soal Latihan Uji Coba I

Kategori	Range Daya Pembeda	Nomor Soal	Jumlah Soal
Sangat Baik	0,40 – 1,00	1,2,3,4,5,6, 7, 8, 9,,12,13, 14, 15	13
Baik	0,30 – 0,39	-	0
Cukup	0,20 – 0,29	-	0
Jelek	0,00 – 0,19	10, 11	2

Tabel 9 menunjukkan bahwa dari 15 butir soal yang diujicobakan terdapat 2 butir soal yang memiliki daya pembeda dengan kategori jelek sedangkan untuk 13 soal lainnya memiliki kategori daya pembeda yang sangat baik. Hal ini berarti 2 soal dengan kategori jelek perlu dilakukan revisi.

#### Tingkat Kesukaran

Berikut hasil analisis tingkat kesukaran pada tes HOTS yang dikembangkan.

Tabel 10. Tingkat Kesukaran Soal HOTS pada Uji Coba I

	Kategori Tingkat Kesukaran	Range Tingkat Kesukaran	Jumlah Soal	Persentase
<b>Soal Latihan</b>	Mudah	0,71 – 1,00	3	30%
	Sedang	0,31-0,70	4	40%
	Sukar	0,00 – 0,30	3	30%
<b>Soal Ulangan Harian</b>	Mudah	0,71 – 1,00	5	30,3%
	Sedang	0,31-0,70	5	30,3%
	Sukar	0,00 – 0,30	5	30,3%

Berdasarkan tabel 10 diketahui tingkat kesukaran soal latihan dan soal ulangan harian memiliki tingkat kesukaran 1:1:1. Sedangkan untuk kriteria soal HOTS minimal memiliki tingkat kesukaran sedang. Hal ini menunjukkan dibutuhkan perbaikan pada butir soal yang memiliki tingkat kesukaran mudah.

#### Fungsi pengecoh

Hasil analisis fungsi pengecoh soal HOTS yang dikembangkan pada materi ekologi dapat dilihat pada tabel berikut.

Tabel 11. Hasil Analisis Fungsi Pengecoh Soal HOTS pada Uji Coba 1

	Distraktor	Jumlah Soal	Persentase
Soal Latihan	D > 5%	6	60%
	D < 5%	4	40%
Soal Ulangan Harian	D > 5%	7	47%
	D < 5%	8	53%

Tabel 11 menunjukkan bahwa soal yang memiliki pengecoh yang baik ( $D > 5\%$ ) pada soal latihan adalah 60% (6 butir soal) dan pada soal ulangan harian sebanyak 47%

(7 butir soal). Sedangkan untuk butir soal dengan pengecoh yang tidak baik ( $D < 5\%$ ) pada soal latihan adalah 40% (4 butir soal) dan pada soal ulangan harian sebanyak 53% (8 butir soal).

## Uji coba terbatas II

### Reliabilitas

Berdasarkan analisis data terhadap soal latihan dan soal ulangan harian yang telah diujicobakan diperoleh hasil reliabilitas seperti pada tabel 12 berikut.

**Tabel 12** Reliabilitas Soal HOTS pada Uji Coba II

Jenis Soal	Nilai Reliabilitas	Kategori
Soal Latihan	0,73	Sangat Tinggi
Soal Ulangan Harian	0,84	Sangat Tinggi

Berdasarkan analisis data tersebut reliabilitas tes yang diperoleh adalah 0,73 pada soal latihan dan 0,84 pada soal ulangan harian dengan interpretasi reliabilitas tinggi. Sehingga berdasarkan analisis reliabilitas tes tersebut, maka tidak perlu dilakukan revisi instrumen tes.

### Daya Pembeda

Hasil analisis daya pembeda butir soal pada uji coba II dapat dilihat pada tabel 13 berikut.

**Tabel 13** Nilai Daya Pembeda Soal Latihan Uji Coba II

Kategori	Range Daya Pembeda	Jumlah Soal	Persentase
Sangat Baik	0,40 – 1,00	8	80%
Baik	0,30 – 0,39	0	0%
Cukup	0,20 – 0,29	1	10%
Jelek	0,00 – 0,19	1	10%

Berdasarkan tabel 13 diketahui bahwa daya pembeda soal latihan berkisar antara 0,2 hingga 1,00 dengan kategori cukup berjumlah 10% dan persentase sangat baik berjumlah 80%. Sedangkan soal dengan kategori jelek memiliki persentase 10%. Daya pembeda pada soal ulangan harian dapat dilihat pada tabel 14 berikut.

**Tabel 14.** Nilai Daya Pembeda Soal Ulangan Harian Uji Coba II

Kategori	Range Daya Pembeda	Jumlah Soal	Persentase
Sangat Baik	0,40 – 1,00	9	60%
Baik	0,30 – 0,39	3	20%
Cukup	0,20 – 0,29	2	13%
Jelek	0,00 – 0,19	1	7%

Berdasarkan tabel 14 diketahui bahwa daya pembeda soal ulangan harian berkisar antara 0,2 hingga 1,00 dengan kategori cukup berjumlah 13%, kategori baik 20% dan persentase sangat baik berjumlah 60%. Sedangkan soal dengan kategori jelek memiliki persentase 7%.

#### Tingkat Kesukaran

Proporsi tingkat kesukaran pada uji coba II dapat dilihat pada tabel 15 berikut.

Tabel 15. Proporsi dan Persentase Tingkat Kesukaran Butir Soal HOTS Uji coba II

	Kategori Kesukaran	Tingkat Kesukaran	Range Kesukaran	Tingkat Kesukaran	Jumlah Soal	Persentase
<b>Soal Latihan</b>	Mudah		0,71 – 1,00		2	20%
	Sedang		0,31-0,70		5	50%
	Sukar		0,00 – 0,30		3	30%
<b>Soal Ulangan Harian</b>	Mudah		0,71 – 1,00		5	30,3%
	Sedang		0,31-0,70		5	30,3%
	Sukar		0,00 – 0,30		5	30,3%

Tabel 15 menunjukkan bahwa proporsi dan persentase tingkat kesukaran instrumen tes berbasis HOTS yang dikembangkan seimbang, yakni mudah : sedang : sukar = 1:1:1. Hal ini tidak sesuai dengan kriteria kualitas soal HOTS yang memiliki kategori daya pembeda minimal sedang.

#### Fungsi Pengecoh (Distraktor)

Hasil analisis fungsi pengecoh pada soal yang dikembangkan dapat dilihat pada tabel berikut.

Tabel 16. Hasil Analisis Fungsi Pengecoh Soal HOTS pada Uji Coba 1

	Distraktor	Jumlah Soal	Persentase
<b>Soal Latihan</b>	D > 5%	10	100%
	D < 5%	0	0%
<b>Soal Ulangan Harian</b>	D > 5%	15	100%
	D < 5%	0	0%

Dari tabel 16 dapat diketahui bahwa pengecoh soal sudah bekerja dengan baik pada soal latihan maupun soal ulangan harian yang dikembangkan dengan persentase fungsi pengecoh pada ke dua bentuk soal 100% berfungsi.

## PEMBAHASAN

Hasil analisis pada tabel 5 dan tabel 6 menunjukkan bahwa butir soal yang telah divalidasi memiliki skor rata-rata 3,75 pada soal latihan dan 3,74 pada soal ulangan harian. Hasil yang diperoleh tersebut dinyatakan memiliki validitas yang sangat baik

dikarenakan nilai yang diperoleh memiliki nilai rata-rata  $3,25 \leq \bar{x} < 4$  berdasarkan kriteria validitas pada tabel 3.3. Hal ini menunjukkan bahwa keseluruhan butir soal yang dikembangkan sudah sesuai aspek-aspek penulisan butir soal HOTS pada lembar validasi dan dinyatakan masuk dalam kriteria sangat baik dari aspek materi, konstruksi dan bahasa.

Meskipun produk awal instrumen tes yang dikembangkan telah memenuhi kategori validasi yang baik, produk awal ini masih perlu direvisi. Revisi dilakukan sesuai saran yang diberikan para ahli (validator) terutama pada bagian tata tulis yang belum sepenuhnya sesuai dengan pedoman umum ejaan bahasa Indonesia (PUEBI). Hasil validasi keseluruhan telah menunjukkan bahwa butir soal latihan dan soal ulangan yang telah dikembangkan sudah berkualitas dan layak digunakan untuk uji coba terbatas setelah dilakukan perbaikan.

Tingkat reliabilitas dalam penyusunan soal dapat menunjukkan sejauh mana soal dapat dipercaya dan menguji kekonsistenan pertanyaan tes. Untuk uji coba terbatas I dengan jumlah soal 10 soal latihan dan 15 soal ulangan harian dengan menggunakan program *SPSS 23* pada tabel 7, diperoleh nilai reliabilitas soal masing masing 0,92 dan 0,91 dengan kategori “sangat tinggi”. Sehingga tidak perlu dilakukan revisi soal menurut hasil reliabilitas yang diperoleh. Sedangkan pada uji coba terbatas II dengan jumlah soal yang sama pada tabel 12 diperoleh hasil reliabilitas tes 0,73 untuk soal latihan dan 0,84 untuk soal ulangan harian dengan kategori “sangat tinggi”. Berdasarkan koefisien reliabilitas berarti soal HOTS yang dikembangkan bersifat reliabel dengan kategori tinggi. Artinya tes yang telah dikembangkan dapat memberikan hasil yang sama bila diberikan pada kelompok yang sama meskipun dilakukan oleh orang yang berbeda sehingga konsistensi instrumen tes HOTS yang dikembangkan dianggap tinggi dan dapat dipercaya. Penilaian HOTS dalam pembelajaran efektif untuk mengukur keterampilan berpikir jika memiliki tingkat konsistensi yang tinggi (Arikunto dalam Putro Widoyoko, 2019).

Daya pembeda soal merupakan kemampuan suatu soal untuk membedakan peserta didik dengan kemampuan tinggi dan peserta didik dengan kemampuan rendah. Menurut Kusaeri (2014) daya pembeda pada soal dikatakan sangat baik jika memiliki rentang nilai 0,40-1,00, daya pembeda pada soal dikatakan baik jika memiliki rentang nilai 0,30-0,39, daya pembeda pada soal dikatakan cukup jika memiliki rentang nilai 0,20-0,29, dan daya pembeda pada soal dikatakan jelek jika memiliki rentang nilai 0,00-0,19. Maka dari itu berdasarkan pada tabel 8 dan tabel 9 dapat disimpulkan pada uji coba terbatas I seluruh soal latihan memiliki daya pembeda yang sangat baik sedangkan pada soal ulangan harian terdapat 2 butir soal dengan kriteria “jelek” yaitu soal nomor 10 dan 11. Soal nomor 10 dan nomor 11 memiliki nilai  $\bar{x} < 0,20$ .

Tingkat kesukaran pada soal menunjukkan peluang menjawab benar suatu soal pada tingkatan kemampuan tertentu. Menurut Aviory (2021), soal HOTS memiliki karakteristik dengan tingkat kesukaran minimal “sedang” dengan daya pembeda yang “baik”. Pada uji coba terbatas I diperoleh hasil perbandingan tingkat kesukaran soal mudah : soal sedang : soal sukar pada tabel 10 adalah 3:4:3 untuk soal latihan dan 5:5:5 pada soal ulangan harian. Soal latihan memiliki 3 butir soal yang dikategorikan mudah yaitu soal nomor 8, 9 dan 10. Soal ulangan harian memiliki 5 butir soal yang dikategorikan mudah yaitu soal nomor 2, 7, 10, 11 dan 14. Sedangkan pada uji coba terbatas II, berdasarkan tabel 15 diperoleh hasil perbandingan tingkat kesukaran soal mudah : soal sedang : soal sukar untuk soal latihan 2:5:3 dan untuk soal ulangan harian

5:5:5. Soal latihan memiliki 2 butir soal yang dikategorikan mudah yaitu soal nomor 7 dan 8. Soal ulangan harian memiliki 5 butir soal yang dikategorikan mudah yaitu soal nomor 2, 7, 8, 10, 11. Berdasarkan hasil uji coba terbatas yang dilakukan tingkat kesukaran yang diperoleh tidak sesuai dengan kriteria instrumen tes berbasis HOTS sehingga dilakukan penelaahan lebih dalam terhadap soal-soal tersebut.

Fungsi pengecoh digunakan untuk mengecoh peserta tes saat melakukan tes sehingga dapat membedakan siswa yang benar-benar memahami materi pembelajaran (siswa pandai) dan siswa yang belum memahami materi dengan baik. Hasil uji coba terbatas I pada tabel 11 menunjukkan hasil bahwa fungsi distraktor pada soal belum seluruhnya berfungsi dengan baik. Terdapat 4 soal pada soal latihan yang memiliki pengecoh yang belum berfungsi dengan baik yang memiliki nilai  $D < 5\%$ . Sejalan dengan pendapat Basuki & Hariyanto (2016), pengecoh dikatakan berfungsi jika opsi pilihan jawaban dipilih oleh 5% dari jumlah peserta tes.

Setelah dilakukan revisi pada pengecoh soal latihan dan ulangan harian pada uji coba terbatas I, selanjutnya soal kembali diujicobakan. Pada uji coba terbatas II seperti yang ditunjukkan pada tabel 16 diperoleh hasil bahwa keseluruhan soal memiliki nilai  $D > 5\%$ . Hal ini menunjukkan bahwa fungsi pengecoh soal telah berfungsi dengan baik. Dari 20 subjek uji coba telah memilih opsi yang tersedia secara merata sehingga dapat dibedakan siswa yang memiliki kemampuan tinggi dan siswa yang memiliki kemampuan rendah. Dengan demikian instrumen tes berbasis HOTS yang dikembangkan telah memenuhi kriteria kualitas fungsi pengecoh yang bekerja dengan efektif.

Setelah dilakukan analisis validasi dan analisis butir soal penelitian ini maka diperoleh instrumen tes berbasis *higher order thinking skill* (HOTS) pada materi ekologi kelas X yang berkualitas. Tes HOTS yang dikembangkan pada penelitian ini berjumlah 25 soal, 10 soal diantaranya adalah soal latihan dan 15 soal ulangan harian. Namun dari 25 soal tersebut hanya 16 soal yang dapat dikategorikan soal HOTS yang berkualitas yaitu 6 soal latihan (*lampiran 19*) dan 10 soal ulangan harian (*lampiran 20*). Hal ini dikarenakan hanya 16 soal tersebut yang sesuai dengan kriteria tes HOTS, yaitu 1) Validitas yang tinggi dengan jumlah 25 soal, yaitu validitas materi lebih besar dari 3,73, validitas konstruksi lebih besar dari 3,70 dan validitas bahasa lebih besar dari 3,72. 2) Reliabilitas yang tinggi yaitu sebesar 0,73 untuk soal latihan dan 0,84 untuk soal ulangan harian. 3) Memiliki tingkat kesukaran yang sedang hingga sukar. 4) Daya pembeda soal yang sudah baik 5) Pengecoh soal seluruhnya sudah berfungsi dengan baik.

## **SIMPULAN DAN REKOMENDASI**

### **Simpulan**

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan sebagai berikut: 1) Kualitas instrumen tes berbasis *higher order thinking skill* (HOTS) yang dikembangkan pada materi ekologi kelas X SMA sangat baik berdasarkan aspek materi, konstruksi maupun bahasa. 2) Butir soal yang dikembangkan telah memenuhi kriteria kualitas tes berbasis *higher order thinking skill* (HOTS) yang baik dari aspek reliabilitas, daya pembeda, tingkat kesukaran dan fungsi pengecoh atau distraktor.

## Rekomendasi

Rekomendasi yang dirumuskan setelah melakukan penelitian ini adalah 1) Untuk penelitian selanjutnya apabila hendak menganalisis hasil tes pilihan ganda sebaiknya menggunakan sampel uji coba yang lebih besar minimal 15 orang. 2) Penelitian selanjutnya diharapkan dapat mengembangkan instrumen penilaian berbasis HOTS dalam bentuk soal yang berbeda dan materi yang lebih abstrak. 3) Penelitian selanjutnya diharapkan dapat melakukan implementasi terhadap instrumen tes berbasis HOTS untuk mengukur kemampuan berpikir tingkat tinggi siswa.

## DAFTAR PUSTAKA

- Arikunto, Suharsimi. 2013. *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Aviory, K., & Susetyawati, M. E. 2021. Kualitas Soal HOTS (Higher Order Thinking Skill) pada Siswa SMP Kelas VII. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(2), 639-647.
- Basuki dan Hariyanto. 2016. *Asesmen Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Febriana, R. 2019. *Evaluasi Pembelajaran*. Jakarta: Bumi Aksara.
- Kemendikbud. 2016. *Permendikbud No 20 tahun 2016 Tentang Standar Kompetensi Lulusan Pendidikan Dasar Dan Menengah*. Jakarta: Kemendikbud.
- Kusaeri. 2014. *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013*. Yogyakarta: Ar-Ruzz Media.
- Rosnawati, R. 2013. Enam Tahapan Aktivitas dalam Pembelajaran Matematika untuk Mendayagunakan Berpikir Tingkat Tinggi Siswa. Universitas Negeri Yogyakarta. *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*. 16 Mei 2019. Fakultas FMIPA UNY. Yogyakarta.
- Sudjana, Nana. 2012. *Penelitian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Sugiyono. 2017. *Metode Penelitian dan Pengembangan*. Bandung: Alfabeta.
- Widoyoko, E. P. 2019. *Evaluasi Program Pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Zulkifli Matondang. 2019. *Evaluasi Hasil Belajar*. Jakarta: Yayasan Kita Menulis.