



## UJI PSIKOMETRIK INSTRUMEN HOTS UNTUK SISWA KELAS XI PADA MATA PEMBELAJARAN BIOLOGI

Nur Fitri Mitha Sayuti<sup>1)</sup>, Nurkhairo Hidayati<sup>2)</sup>

<sup>1)</sup>Pendidikan Biologi, Fakultas Keguruan, Universitas Islam Riau, Pekanbaru, Indonesia  
Email: [nurfitrimithasayuti@student.uir.ac.id](mailto:nurfitrimithasayuti@student.uir.ac.id)

<sup>2)</sup>Pendidikan Biologi, Fakultas Keguruan, Universitas Islam Riau, Pekanbaru, Indonesia  
Email: [khairbio@edu.uir.ac.id](mailto:khairbio@edu.uir.ac.id)

### Abstract

This study aims to examine the psychometric quality of a Higher Order Thinking Skills (HOTS) instrument in Biology learning for eleventh-grade students, particularly on the topic of the human movement system. The research employed a quantitative approach using psychometric testing methods. The instrument consisted of 21 multiple-choice items developed based on the revised Bloom's Taxonomy at cognitive levels C4 (analyzing), C5 (evaluating), and C6 (creating). The research participants were 40 eleventh-grade students of SMAN 1 Pekanbaru selected through simple random sampling. Data analysis included item validity testing, instrument reliability, difficulty index, and discrimination power. The results indicated that eight items were valid and suitable for use, several items required minor and major revisions, and three items were invalid and eliminated. The reliability coefficient obtained was 0.332, indicating low internal consistency; however, this result is still acceptable for a pilot test in an undergraduate research context. Overall, the HOTS instrument on the movement system topic requires further refinement to achieve better psychometric quality and to function effectively as an assessment tool for measuring students' higher order thinking skills.

**Keywords:** HOTS, psychometric test, assessment instrument, biology education, movement system.

### Abstrak

Penelitian ini bertujuan untuk menguji kualitas psikometrik instrumen Higher Order Thinking Skills (HOTS) pada mata pelajaran Biologi kelas XI, khususnya pada materi sistem gerak. Penelitian menggunakan pendekatan kuantitatif dengan metode uji psikometrik. Instrumen penelitian berupa tes pilihan ganda sebanyak 21 butir soal yang disusun berdasarkan Taksonomi Bloom Revisi pada level kognitif C4 (analisis), C5 (evaluasi), dan C6 (kreasi). Subjek penelitian berjumlah 40 siswa kelas XI SMAN 1 Pekanbaru yang dipilih menggunakan teknik simple random sampling. Analisis data meliputi uji validitas butir, reliabilitas instrumen, tingkat kesukaran, dan daya pembeda. Hasil penelitian menunjukkan bahwa delapan butir soal dinyatakan valid dan layak digunakan, beberapa butir memerlukan revisi ringan maupun revisi besar, serta tiga butir tidak valid dan harus dieliminasi. Nilai reliabilitas instrumen sebesar 0,332 menunjukkan konsistensi internal yang masih rendah, namun masih dapat diterima pada tahap uji coba awal. Secara keseluruhan, instrumen HOTS Biologi pada materi sistem gerak memerlukan penyempurnaan lanjutan agar memenuhi kriteria psikometrik yang lebih baik.

**Kata Kunci:** HOTS, uji psikometrik, instrumen asesmen, Biologi, sistem gerak.



## PENDAHULUAN

Perkembangan abad ke-21 menuntut peserta didik untuk memiliki keterampilan berpikir tingkat tinggi (Higher Order Thinking Skills) yang tercermin dalam kemampuan analisis, evaluasi, dan pemecahan masalah. Tuntutan ini sejalan dengan amanat Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional yang menegaskan bahwa pendidikan berfungsi untuk mengembangkan potensi peserta didik secara optimal melalui penguatan aspek kognitif, afektif dan psikomotorik (Kristanti, 2024). Dalam konteks tersebut, proses pembelajaran tidak hanya berorientasi pada penguasaan pengetahuan faktual tetapi juga pada pengembangan kemampuan berpikir kompleks yang relevan dengan kebutuhan abad ke-21 (Hidayati et al., 2023). Oleh karena itu, sistem pendidikan dituntut untuk menerapkan pendekatan pembelajaran dan penilaian yang mampu mengukur keterampilan berpikir tingkat tinggi secara valid dan reliabel melalui instrumen asesmen yang memiliki kualitas psikometri yang memadai, sehingga hasil belajar yang diperoleh benar-benar mencerminkan capaian kompetensi peserta didik (Herlina et al., 2024).

Implementasi Kurikulum Merdeka menegaskan pentingnya pembelajaran mendalam (*Deep Learning*) yang berorientasi pada pemahaman konseptual, penalaran ilmiah, dan pengembangan keterampilan berpikir tingkat tinggi peserta didik (Rahmayumita & Hidayati, 2023). Dalam kerangka ini, asesmen tidak lagi dipahami sebagai alat pengukur hasil belajar semata, melainkan sebagai bagian integral dari proses pembelajaran yang berfungsi untuk mengevaluasi kualitas penalaran dan kemampuan berpikir kompleks siswa. Banyak pendidik mengalami kesulitan dalam menyusun butir soal HOTS yang kontekstual, selaras dengan karakteristik *Deep Learning*, serta memenuhi standar kualitas psikometrik (Hidayati et al., 2021), seperti validitas dan reliabilitas yang memadai. Kondisi ini menyebabkan pengukuran capaian HOTS siswa belum optimal, sehingga tujuan Kurikulum Merdeka belum sepenuhnya tercapai dan menegaskan pentingnya pengembangan serta pengujian instrumen asesmen HOTS yang kontekstual dan teruji secara psikometrik (Dermawan et al., 2021).

Salah satu materi dalam mata pelajaran Biologi kelas XI yang memiliki potensi tinggi untuk mengembangkan keterampilan berpikir tingkat tinggi adalah materi sistem gerak. Materi ini mencakup kajian struktur dan fungsi tulang, sendi, serta otot, beserta mekanisme kerja koordinatif yang melibatkan berbagai konsep biologis secara terintegrasi. Karakteristik materi sistem gerak bersifat kontekstual dan aplikatif, sehingga memungkinkan peserta didik untuk menganalisis fenomena nyata seperti gangguan sistem gerak, cedera akibat aktivitas olahraga, serta permasalahan postur tubuh yang berkaitan dengan penggunaan perangkat digital. Dengan demikian, diperlukan instrumen asesmen HOTS yang dirancang menuntut kemampuan analisis, evaluasi, dan pemecahan masalah berbasis penalaran ilmiah.

Berdasarkan pengamatan awal di SMAN 1 Pekanbaru, instrumen penilaian HOTS pada mata pelajaran Biologi telah mulai dikembangkan, namun implementasinya belum

optimal. Soal-soal yang digunakan masih didominasi oleh pengukuran kemampuan berpikir tingkat rendah (*Lower Order Thinking Skills*), sehingga belum sepenuhnya mencerminkan tuntutan HOTS. Guru telah menyusun perangkat penilaian yang mencakup aspek kognitif melalui tugas dan kuis, aspek afektif melalui penilaian sikap dan kinerja, serta aspek psikomotorik melalui penilaian keterampilan siswa. Namun demikian, instrumen kognitif yang digunakan belum secara khusus dirancang untuk mengukur keterampilan berpikir tingkat tinggi sehingga pengukuran capaian HOTS siswa belum dapat dilakukan secara optimal.

Hasil wawancara dengan guru Biologi kelas XI di SMAN 1 Pekanbaru menunjukkan bahwa sebagian besar instrumen penilaian disusun dengan mengadaptasi soal dari buku paket dan sumber daring, serta cenderung digunakan secara berulang dari tahun ke tahun akibat keterbatasan waktu dan beban mengajar yang tinggi. Instrumen tersebut umumnya belum melalui proses pengujian kualitas tes, seperti analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Kondisi ini mengindikasikan adanya kesenjangan antara kebutuhan asesmen HOTS yang berkualitas dan ketersediaan instrumen yang teruji secara psikometrik. Oleh karena itu, penelitian ini bertujuan untuk menguji kualitas psikometrik instrumen HOTS pada mata pelajaran Biologi kelas XI, khususnya pada materi sistem gerak, sehingga dihasilkan instrumen penilaian yang valid, reliabel, serta memiliki tingkat kesukaran dan daya pembeda yang memadai dalam mengukur keterampilan berpikir tingkat tinggi peserta didik.

## METODE PENELITIAN

Metode Penelitian ini menggunakan pendekatan kuantitatif untuk menguji psikometrik instrumen *High Order Thinking Skills* (HOTS) yang dirancang khusus untuk siswa kelas XI pada mata pelajaran Biologi. Dalam penelitian kuantitatif, data diperoleh melalui pengukuran sistematis terhadap variabel tertentu dimana variabel tersebut merupakan fenomena atau konstruk yang menjadi fokus utama pengamatan penelitian, seperti kemampuan berpikir kritis siswa, analitis, dan kreasi dalam konteks biologi (Prananto, 2025). Pendekatan ini memungkinkan pengumpulan data numerik seperti instrumen tes.

Dalam penelitian ini data dikumpulkan dengan cara menguji kualitas psikometrik instrumen HOTS untuk mata pelajaran biologi yang mencakup aspek-aspek seperti validitas (kemampuan instrumen untuk mengukur kemampuan berpikir siswa dengan akurat), reliabilitas (konsistensi hasil pengukuran), tingkat daya beda (kemampuan instrumen untuk membedakan antara siswa dengan kemampuan tinggi dan rendah dalam berpikir kritis), tingkat kesulitan (seberapa menantang item instrumen bagi siswa) serta analisis pengecoh (untuk mendeteksi item yang menyesatkan/tidak relevan dengan konteks biologi). Pengujian psikometrik ini sangat penting bagi evaluasi guru dalam mengajar biologi (Umami et al., 2021).

Populasi dalam penelitian ini melibatkan siswa kelas XI SMAN 1 Pekanbaru, yaitu kelas XI.3 dan XI.8. Sampel penelitian terdiri atas 40 siswa, masing-masing 20 siswa



dari kelas XI.3 dan 20 siswa dari kelas XI.8. Pemilihan sampel dilakukan menggunakan teknik *simple random sampling*. Instrumen penelitian yang digunakan berupa tes pilihan ganda sebanyak 21 butir soal yang disusun berdasarkan Taksonomi Bloom revisi pada ranah kognitif C4 (analisis), C5 (evaluasi), dan C6 (kreasi).

Teknik analisis data dalam penelitian ini dilakukan untuk menguji psikemtrik instrumen HOTS pada siswa dengan menggunakan program *EXCEL* untuk pengolahan data awal kemudian di analisis statistik lebih lanjut di *SPSS*.

#### 1. Validitas butir,

Validitas dihitung menggunakan korelasi Product Moment Pearson antara skor setiap butir soal dengan skor total, butir dianggap valid jika nilai  $r$ -hitung lebih besar dari  $r$ -tabel pada tingkat signifikansi  $\alpha = 0.05$  dengan memastikan konsistensi instrument dapat mengukur kemampuan HOTS.

Kriteria: butir valid jika  $r\text{-hitung} > r\text{-tabel}$  pada  $\alpha = 0.05$ .

- Uji reliabilitas instrumen untuk memastikan konsistensi instrumen dalam mengukur kemampuan berpikir kritis siswa di bidang Biologi dengan dihitung menggunakan Cronbach's Alpha dengan interpretasi berdasarkan tabel 1 :

**Tabel 1** interpretasi Cronbach's Alpha

0.90	Sangat Tinggi
0.70-0.90	Tinggi
0.60-0.70	Sedang
<0.60	Rendah

- Indeks kesukaran untuk menilai seberapa menantang instrumen HOTS bagi siswa kelas XI dalam konteks pembelajaran Biologi dengan rumus :

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

Klasifikasi:

- 0.00–0.30 = sulit
- 0.31–0.70 = sedang
- 0.71–1.00 = mudah

- Daya pembeda, dihitung menggunakan indeks diskriminasi antara kelompok atas dan bawah yang membantu menentukan kemampuan instrumen untuk membedakan siswa dengan kemampuan tinggi dan rendah dalam berpikir analitis dengan kategori :

- $\geq 0.40$  = sangat baik
- 0.30–0.39 = baik
- 0.20–0.29 = cukup
- $< 0.20$  = kurang

- Korelasi Item-Total, Menunjukkan kontribusi tiap butir terhadap keseluruhan instrumen memastikan instrumen secara keseluruhan efektif dan dapat diandalkan untuk evaluasi kemampuan HOTS siswa.
  - Item dipertahankan jika  $r\text{-item-total} > 0.30$ .

## HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis kualitas psikometrik instrumen HOTS yang digunakan pada siswa

kelas XI SMAN 1 Pekanbaru. Instrumen berjumlah 21 butir soal pilihan ganda, yang terdiri dari indikator C4 (12 soal), C5 (7 soal), dan C6 (2 soal). Instrumen diuji menggunakan empat analisis utama, yaitu validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda. Berikut tabel 2 deskripsi instrument C4,C5 dan C6

**Tabel 2** Deskripsi Instrumen (C4–C6)

Level Kognitif	Jumlah Butir	Nomor Soal
C4 (Analisis)	12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
C5 (Evaluasi)	7	13, 14, 15, 16, 17, 18, 19
C6 (Kreasi)	2	20, 21

#### 1. Uji Validitas Empiris

Berdasarkan hasil uji validitas empiris menggunakan korelasi item-total (Corrected Item Total), diperoleh bahwa butir soal dinyatakan valid apabila memiliki nilai  $r \geq 0,30$  atau nilai signifikansi  $< 0,05$ . Hasil analisis menunjukkan bahwa dari 21 butir soal, terdapat 8 butir yang dinyatakan valid dan dapat digunakan tanpa revisi, yaitu nomor 5, 7, 14, 15, 16, 17, 18, dan 21. Selain itu, terdapat beberapa butir yang berada pada kategori revisi ringan, yaitu nomor 2, 9, 12, dan 21, karena nilai  $r$  yang mendekati batas minimal validitas. Sementara itu, butir nomor 1, 4, 6, 8, 10, 11, dan 20 termasuk kategori revisi besar karena memiliki nilai validitas dan daya pembeda yang rendah. Adapun butir nomor 3, 13, dan 19 dinyatakan tidak valid sama sekali (item mati) karena memiliki nilai  $r = 0$  sehingga tidak dapat digunakan dalam instrumen. Dengan demikian, instrumen ini menunjukkan variasi kualitas butir, di mana beberapa item seperti nomor 16, 17, dan 18 memiliki validitas terbaik dan daya pembeda tinggi, sehingga layak dipertahankan berdasarkan table 3 :

**Tabel 3.** Hasil Uji Validitas Empiris

Kategori	Nomor Butir
Dipakai (Valid)	5, 7, 14, 15, 16, 17, 18, 21
Revisi Ringan	2, 9, 12, 21
Revisi Besar	1, 4, 6, 8, 10, 11, 20
Tidak Valid (Dibuang)	3, 13, 19

#### 2. Uji Realibilitas

Hasil uji reliabilitas dengan menggunakan Cronbach's Alpha menghasilkan nilai sebesar **0,332**. Nilai ini termasuk kategori rendah, sehingga menunjukkan bahwa instrumen memiliki konsistensi internal yang masih perlu ditingkatkan. Meski demikian, pada penelitian eksploratif atau pengembangan awal instrumen, nilai ini masih dapat diterima dengan syarat dilakukan revisi terhadap item yang tidak valid berdasarkan tabel 4 :

**Tabel 4.** Hasil Uji Realibilitas

Jenis Uji	Hasil
Cronbach's Alpha	0,332
Kategori	Reliabilitas rendah (perlu perbaikan item)



Keterangan	Masih dapat diterima untuk penelitian eksploratif
------------	---

### 3. Uji Tingkat Kesukaran

Analisis tingkat kesukaran menunjukkan bahwa sebagian besar item berada pada kategori mudah hingga sedang. Sebanyak 10 butir soal berada pada kategori mudah ( $p > 0,70$ ), yaitu nomor 4, 5, 6, 10, 14, 15, 16, 18, 19, dan 21. Tujuh butir termasuk kategori sedang ( $0,31-0,70$ ), yaitu nomor 1, 7, 8, 9, 14, 17, dan 20. Lima butir tergolong sulit ( $p < 0,30$ ), yaitu nomor 2, 3, 11, 12, dan 13. Hasil ini menunjukkan bahwa sebagian besar soal terlalu mudah bagi siswa sehingga kurang optimal untuk mengukur HOTS. Berikut hasil uji Tingkat kesukaran tabel 5 :

**Tabel 5.** Hasil Uji Tingkat Kesukaran

Kategori Tingkat Kesukaran	Nomor Butir
Mudah ( $p > 0,70$ )	4, 5, 6, 10, 14, 15, 16, 18, 19, 21
Sedang ( $0,31-0,70$ )	1, 7, 8, 9, 14, 17, 20
Sulit ( $p < 0,30$ )	2, 3, 11, 12, 13

### 4. Uji Daya Pembeda

Hasil uji daya pembeda memperlihatkan bahwa enam butir soal memiliki kemampuan pembeda sangat baik ( $D \geq 0,40$ ), yaitu nomor 5, 7, 14, 16, 17, dan 18. Dua butir berada pada kategori baik ( $0,30-0,39$ ), yaitu nomor 15 dan 21. Tiga butir masuk kategori cukup ( $0,20-0,29$ ), yaitu nomor 2, 9, dan 12. Sementara itu, tujuh butir memiliki daya pembeda rendah ( $< 0,20$ ), yaitu nomor 1, 4, 6, 8, 10, 11, dan 20. Tiga butir lainnya (3, 13, 19) memiliki daya pembeda 0, yang berarti tidak mampu membedakan kelompok siswa berkemampuan tinggi dan rendah. Hal ini menunjukkan bahwa sejumlah butir perlu diperbaiki terutama pada aspek pengecoh dan kejelasan stimulus berdasarkan tabel 6 :

**Tabel 6.** Hasil Uji Daya Pembeda

Kategori Daya Pembeda	Nomor Butir
Sangat Baik ( $D \geq 0,40$ )	5, 7, 14, 16, 17, 18
Baik ( $0,30-0,39$ )	15, 21
Cukup ( $0,20-0,29$ )	2, 9, 12
Rendah ( $< 0,20$ )	1, 4, 6, 8, 10, 11, 20
Tidak Membeda ( $D = 0$ )	3, 13, 19

Berikut rekapitulasi hasil uji psikometerik pada tabel 7 berikut :

**Tabel 7.** Tabel rekapitulasi uji psikometrik

No Butir	p	D	r	Keputusan	Alasan
1	0,65	0,170	0,170	Revisi Besar	Validitas & daya pembeda rendah

2	0,23	0,268	0,268	Revisi Ringan	Mendekati valid, perlu perbaikan
3	0,00	0,000	0,000	Tidak Dipakai	Item mati, semua siswa salah
4	0,93	0,049	0,049	Revisi Besar	Terlalu mudah & tidak valid
5	0,88	0,470	0,325	Dipakai	Valid & daya pembeda baik
6	0,93	0,192	0,192	Revisi Besar	Tidak valid & pembeda rendah
7	0,70	0,441	0,441	Dipakai	Valid & membedakan dengan baik
8	0,68	0,074	0,074	Revisi Besar	Daya pembeda sangat rendah
9	0,70	0,277	0,277	Revisi Ringan	Mendekati valid
10	0,98	0,135	0,135	Revisi Besar	Terlalu mudah & tidak valid
11	0,30	0,052	0,052	Revisi Besar	Tidak mampu membedakan siswa
12	0,13	0,251	0,251	Revisi Ringan	Validitas cukup, perlu revisi redaksi
13	0,00	0,000	0,000	Tidak Dipakai	Item mati (tidak bekerja)
14	0,73	0,433	0,433	Dipakai	Valid & daya pembeda baik
15	0,95	0,366	0,366	Dipakai	Cukup valid
16	0,78	0,543	0,543	Dipakai	Sangat valid & pembeda kuat
17	0,75	0,572	0,572	Dipakai	Item terbaik & stabil
18	0,78	0,513	0,513	Dipakai	Valid & daya pembeda baik
19	1,00	0,000	0,000	Tidak Dipakai	Semua benar, tidak membedakan





20	0.7 5	0.02 2	0.02 2	Revisi Besar	Tidak valid & tidak membedakan
21	0.8 8	0.35 6	0.35 6	Dipakai	Cukup valid

Hasil penelitian menunjukkan bahwa kualitas psikometrik instrumen HOTS berada pada kategori **cukup**, dengan sebagian butir soal memenuhi standar validitas dan daya pembeda. Dari 21 butir soal yang diuji, delapan butir dinyatakan valid dan layak digunakan, yaitu nomor 5, 7, 14, 15, 16, 17, 18, dan 21. Butir-butir tersebut didominasi oleh indikator kognitif C4 (analisis), yang mengindikasikan bahwa kemampuan berpikir analitis siswa relatif lebih mudah diukur melalui format soal pilihan ganda dibandingkan kemampuan evaluatif (C5) dan kreatif (C6). Temuan ini sejalan dengan karakteristik asesmen HOTS yang dilaporkan dalam penelitian Rahman, Rusnayati dan Muslim (2021) di mana level analisis cenderung lebih stabil dan konsisten diukur secara objektif dibandingkan level berpikir tingkat tinggi yang lebih kompleks.

Nilai reliabilitas instrumen yang diperoleh sebesar  $\alpha = 0,332$  menunjukkan bahwa konsistensi internal instrumen masih tergolong rendah. Namun demikian, nilai reliabilitas yang rendah pada penelitian ini masih dapat diterima karena instrumen berada pada tahap uji coba awal (*pilot test*) dan digunakan khusus untuk kepentingan penelitian skripsi. Rendahnya reliabilitas dapat dipengaruhi oleh beberapa faktor, antara lain jumlah item yang relatif terbatas (21 butir), sebaran kualitas butir yang tidak merata, serta heterogenitas indikator HOTS yang diukur (C4, C5, dan C6) sehingga konstruk yang diukur belum sepenuhnya homogen. Selain itu, keberadaan butir yang sangat mudah dan sangat sulit turut berkontribusi terhadap rendahnya konsistensi internal instrumen. Nilai reliabilitas yang rendah dapat disebabkan oleh jumlah item yang sedikit (21 butir), sebaran kualitas butir yang tidak merata, adanya butir yang sangat mudah dan sulit, indikator HOTS yang heterogen (C4, C5, C6) sehingga tidak sepenuhnya homogen. Prinsip ini didukung oleh penelitian psychometric bahwa reliabilitas meningkat ketika inter-item correlation tinggi dan tingkat kesukaran item lebih beragam, serta asumsi homogenitas konstruk terpenuhi (Rahman et al., 2021).

Ditinjau dari tingkat kesukaran, hasil analisis menunjukkan bahwa sebagian besar butir soal berada pada kategori mudah hingga sedang. Dominasi soal dengan indeks kesukaran tinggi ( $p > 0,70$ ), seperti pada butir 4, 10, dan 19, mengindikasikan bahwa soal-soal tersebut kurang menantang bagi siswa dan berpotensi tidak optimal dalam mengukur keterampilan berpikir tingkat tinggi. Soal dengan tingkat kesukaran yang terlalu mudah atau terlalu sulit cenderung memiliki daya pembeda yang rendah sehingga kurang efektif dalam membedakan kemampuan siswa berkemampuan tinggi dan rendah. Oleh karena itu, keseimbangan tingkat kesukaran menjadi faktor penting dalam pengembangan instrumen asesmen agar instrumen mampu mengukur kemampuan berpikir siswa secara

optimal (Wahyuni, Hidayati, 2023). Temuan ini menegaskan bahwa keseimbangan tingkat kesukaran merupakan aspek penting dalam pengembangan instrumen HOTS. Beberapa butir soal memerlukan revisi ringan, seperti perbaikan redaksi, pengecoh, atau penyesuaian konteks. Sementara butir yang memerlukan revisi besar memiliki kelemahan pada struktur logika pertanyaan, ketidakjelasan stimulus, atau opsi jawaban yang tidak berfungsi dengan baik.

Hasil uji daya pembeda menunjukkan bahwa hanya sebagian butir soal yang memiliki kemampuan diskriminatif baik hingga sangat baik, seperti butir 5, 7, 14, 16, 17, dan 18. Butir-butir tersebut memiliki nilai daya pembeda  $\geq 0,40$  dan korelasi item-total yang memadai, sehingga mampu membedakan siswa berkemampuan tinggi dan rendah secara signifikan. Sebaliknya, beberapa butir memiliki daya pembeda sangat rendah bahkan nol, seperti butir 3, 13, dan 19, yang dikategorikan sebagai item mati karena tidak berfungsi dalam mengukur perbedaan kemampuan siswa. Kondisi ini menunjukkan bahwa perbaikan pada aspek stimulus, kejelasan redaksi, serta efektivitas pengecoh sangat diperlukan. Soal yang terlalu mudah atau terlalu sulit cenderung memiliki daya pembeda yang rendah sehingga kurang efektif dalam membedakan kemampuan siswa berkemampuan tinggi dan rendah. Oleh karena itu, keseimbangan tingkat kesukaran merupakan aspek penting dalam pengembangan instrumen HOTS agar instrumen mampu mengukur kemampuan berpikir tingkat tinggi secara optimal (Hasibuan et al., 2024).

Berdasarkan rekapitulasi uji psikometrik, butir yang memerlukan revisi ringan umumnya memiliki validitas mendekati batas minimal dan daya pembeda cukup, sehingga perbaikan dapat difokuskan pada penyempurnaan redaksi, konteks soal, dan pengecoh. Sementara itu, butir yang memerlukan revisi besar menunjukkan kelemahan mendasar pada struktur logika pertanyaan, kesesuaian indikator HOTS, serta fungsi opsi jawaban (Suhady et al., 2020). Oleh karena itu, pengembangan lanjutan instrumen HOTS pada materi sistem gerak perlu dilakukan secara sistematis agar menghasilkan alat ukur yang lebih valid, reliabel, dan representatif dalam mengukur keterampilan berpikir tingkat tinggi siswa.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa pengembangan instrumen HOTS Biologi memerlukan proses uji psikometrik yang berkelanjutan. Instrumen yang telah direvisi berdasarkan temuan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda diharapkan dapat digunakan sebagai alat ukur yang lebih akurat dalam menilai capaian HOTS siswa, sekaligus mendukung implementasi Kurikulum Merdeka yang menekankan pembelajaran bermakna dan penguatan keterampilan berpikir tingkat tinggi.

## KESIMPULAN

Hasil penelitian menunjukkan bahwa kualitas psikometrik instrumen Higher Order Thinking Skills (HOTS) Biologi kelas XI pada materi sistem gerak masih bervariasi. Dari 21 butir soal yang dianalisis, delapan butir dinyatakan valid dan layak digunakan, sementara beberapa butir lainnya memerlukan revisi ringan maupun revisi besar, serta tiga butir tidak valid dan harus dieliminasi. Nilai



reliabilitas instrumen sebesar 0,332 menunjukkan konsistensi internal yang masih rendah, namun masih dapat diterima pada tahap uji coba awal (*pilot test*) dan untuk keperluan penelitian skripsi.

Secara praktis, hasil ini mengimplikasikan bahwa guru perlu melakukan analisis psikometrik secara sederhana sebelum menggunakan instrumen asesmen HOTS, khususnya dengan memperhatikan validitas, tingkat kesukaran, dan daya pembeda butir soal. Instrumen yang telah direvisi diharapkan dapat menjadi acuan awal bagi guru Biologi dalam menyusun asesmen HOTS yang lebih berkualitas dan mendukung pembelajaran bermakna sesuai Kurikulum Merdeka.

#### UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga artikel ini dapat diselesaikan dengan baik. Penulis juga menyampaikan apresiasi dan terima kasih kepada guru Biologi dan peserta didik serta pihak SMA Negeri 1 Pekanbaru yang telah memberikan izin, bantuan, dan kerja sama dalam proses pengumpulan data dan pelaksanaan wawancara, sehingga penelitian ini dapat terlaksana dan diselesaikan dengan lancar. Semoga segala bantuan, bimbingan dan dukungan yang telah diberikan mendapatkan balasan pahala yang berlipat ganda dari Allah SWT.

#### DAFTAR PUSTAKA

- Dermawan, D. D., Wardani, S., Kurniawati, Y., Pranoto, S., & Kunci, K. (2021). Pengembangan Instrumen Asesmen HOTS Pada Pembelajaran IPA Kelas V Sekolah Dasar. *JIKAP PGSD: Jurnal Ilmiah Ilmu Kependidikan*, 5(3).
- Hasibuan, Y. V., Hasairin, A., & Restuati, M. (2024). Pengembangan Instrumen Penilaian Hasil Belajar Bioteknologi Kelas XII SMA Berbasis HOTS dan Kemampuan Berpikir Kritis. In *Didaktika: Jurnal Kependidikan* (Vol. 13, Issue 4). <https://jurnaldidaktika.org>
- Herlina, Arismunandar, & Tolla, I. (2024). Education Character in the Era of Globalization: Facing the Challenges of the Modern World. *International Journal of Engineering, Science and Information Technology*, 4(4), 230–236. <https://doi.org/10.52088/ijesty.v4i4.649>
- Hidayati, N., Zubaidah, S., & Amnah, S. (2021). The PBL vs. Digital Mind Maps Integrated PBL: Choosing Between the two with a view to Enhance Learners' Critical Thinking. *Participatory Educational Research*, 9(3), 330–343. <https://doi.org/10.17275/per.22.69.9.3>
- Hidayati, N., Zubaidah, S., & Amnah, S. (2023). Effective learning model bases problem based learning and digital mind maps to improve student's collaboration skills. *International Journal of Evaluation and Research in Education*, 12(3), 1307–1314. <https://doi.org/10.11591/ijere.v12i3.22654>
- Kristanti, K. (2024). Rancang bangun dan uji psikometrik instrumen sikap nasionalisme siswa sekolah dasar.

*Teacher in Educational Research*, 6(1), 22–29. <https://doi.org/10.33292/ter.v6i1.500>

- Prananto, K. (2025). Uji Psikometrik Skala University Student Engagement Inventory (USEI) Versi Bahasa Indonesia: Validitas dan Reliabilitas pada Mahasiswa di Indonesia. *Humanitas*, 9(1), 41–56.
- Rahman, A., Rusnayati, H., & Muslim, M. (2021). Analysis of the Characteristics of Higher Order Thinking Skills (HOTS) Test on Momentum and Impulse for Senior High School Student Using Item Response Theory. *Jurnal Penelitian Fisika Dan Aplikasinya (JPFA)*, 11(2), 127–137. <https://doi.org/10.26740/jpfa.v11n2.p127-137>
- Rahmayumita, R., & Hidayati, N. (2023). Kurikulum Merdeka: Tantangan dan Implementasinya pada Pembelajaran Biologi. In *Biology And Education Journal* (Vol. 3, Issue 1).
- Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan Soal untuk Mengukur Higher Order Thinking Skill (HOTS) Siswa. *Jurnal Gantang*, 5(2), 143–150. <https://doi.org/10.31629/jg.v5i2.2518>
- Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan instrumen tes untuk mengukur higher order thinking skills (HOTS) berorientasi programme for international student assessment (PISA) pada peserta didik. *JP3M (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, 7(1), 57–68. <https://doi.org/10.37058/jp3m.v7i1.2069>
- Wahyuni, F., & Hidayati, N. (2023). Analisis Kemampuan HOTS (Hight Order Thingking Skills) Siswa Kelas XI pada Pembelajaran Biologi di SMAN 2 Pekanbaru Tahun Ajaran 2021/2022. *Universitas Islam Riau*.