

# Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Teori Tes Klasik

Ernie Novriyanti<sup>1\*</sup>, Iva Sarifah<sup>2</sup>, Riyadi<sup>3</sup>

<sup>1\*,2,3</sup>Universitas Negeri Jakarta, Jakarta, Indonesia

---

## Article Info

### Article history:

Received Sep 15, 2024

Accepted Nov 08, 2024

Published Online Des 04, 2024

---

### Keywords:

Analisis butir soal

Teori tes klasik

Indeks kesukaran

Daya beda

Pengecoh

---

## ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi kualitas butir soal Ujian Tengah Semester (UTS) Biologi Umum menggunakan pendekatan Teori Tes Klasik. Tiga parameter utama yang dianalisis adalah indeks kesukaran, daya beda, dan keberfungsian pengecoh. Sampel terdiri dari 75 soal pilihan ganda yang diujikan kepada mahasiswa, dan hasilnya dianalisis untuk melihat efektivitas soal dalam menilai kemampuan mahasiswa. Hasil penelitian menunjukkan bahwa sebagian besar soal berada dalam kategori kesukaran sedang (48%), namun terdapat distribusi soal sulit yang signifikan (40%), dengan sedikit soal yang mudah (12%). Sebagian besar soal memiliki daya beda rendah (43,75%), dan 17% soal memiliki daya beda negatif, yang menunjukkan ketidakefektifan dalam membedakan kemampuan siswa. Selain itu, sebanyak 57% soal memiliki pengecoh yang tidak berfungsi secara optimal. Kesimpulannya, meskipun soal ujian memiliki distribusi kesukaran yang seimbang, aspek daya beda dan pengecoh memerlukan revisi untuk meningkatkan validitas dan reliabilitas soal dalam pengukuran prestasi Mahasiswa. Penelitian ini memberikan kontribusi penting dalam meningkatkan kualitas evaluasi pendidikan melalui perbaikan butir soal yang lebih akurat dan representatif terhadap kemampuan mahasiswa.

*This is an open access under the [CC-BY-SA](#) licence*



---

### Corresponding Author:

Ernie Novriyanti,

Universitas Negeri Jakarta, Jakarta, Indonesia

Jl. R.Mangun Muka Raya No.11, RT.11/RW.14, Rawamangun, Kec. Pulo Gadung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13220

Email: [ernie.novriyanti@mhs.unj.ac.id](mailto:ernie.novriyanti@mhs.unj.ac.id)

---

## *Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Teori Tes Klasik*

### **1. Pendahuluan**

Evaluasi kualitas butir soal merupakan bagian integral dalam sistem penilaian pendidikan. Sebagai alat untuk mengukur pencapaian mahasiswa, soal-soal ujian yang berkualitas sangat penting untuk memberikan gambaran yang akurat mengenai kompetensi yang telah dikuasai mahasiswa (Saha, 2021). Tanpa soal yang memadai, hasil ujian dapat menyesatkan dan memengaruhi keputusan dalam proses pengajaran. Evaluasi terhadap kualitas butir soal perlu dilakukan agar hasil tes dapat dianggap valid dan andal dalam mencerminkan kemampuan mahasiswa yang sesungguhnya (Van Canh et al., 2023).

Teori Tes Klasik (*Classical Test Theory*, CTT) adalah salah satu metode yang umum digunakan untuk mengevaluasi kualitas butir soal. Pendekatan ini memanfaatkan tiga parameter utama, yakni indeks kesukaran, daya beda, dan pengecoh, guna mendapatkan gambaran menyeluruh tentang efektivitas soal dalam mengukur kemampuan mahasiswa (Crocker et al., 2008). Ketiga komponen ini memungkinkan pengajar untuk mengevaluasi apakah soal tersebut dapat memberikan hasil yang akurat dan sesuai dengan tingkat kompetensi mahasiswa (Sainuddin & Ilyas, 2016).

Indeks kesukaran berfungsi untuk menentukan proporsi mahasiswa yang mampu menjawab soal dengan benar, dan nilai indeks ini memberikan informasi tentang seberapa sulit soal tersebut. Idealnya, soal memiliki variasi kesulitan untuk mencerminkan perbedaan kemampuan mahasiswa. Daya beda, di sisi lain, mengukur kemampuan soal dalam membedakan antara mahasiswa yang memiliki kemampuan tinggi dan rendah. Soal dengan daya beda tinggi sangat bermanfaat untuk menggambarkan perbedaan tingkat pemahaman mahasiswa (Crocker et al., 2008).

Selain indeks kesukaran dan daya beda, pengecoh juga memiliki peran penting dalam soal pilihan ganda. Pengecoh yang efektif dapat memengaruhi mahasiswa yang kurang memahami materi untuk memilih jawaban yang salah. Keberfungsian pengecoh dapat memastikan soal tersebut benar-benar mampu menguji pemahaman mahasiswa dengan tepat (Haladyna & Rodriguez, 2013). Jika pengecoh tidak efektif, artinya tidak dipilih oleh mahasiswa, maka soal tersebut menjadi lebih mudah dari yang seharusnya.

Dalam penelitian ini, evaluasi terhadap butir soal Ujian Tengah Semester (UTS) Biologi Umum dilakukan dengan menggunakan teori klasik untuk menilai efektivitas soal dalam mengukur pemahaman konseptual mahasiswa. Analisis ini diharapkan dapat memberikan

rekomendasi untuk meningkatkan kualitas soal-soal ujian di masa mendatang, dengan menitikberatkan pada tiga parameter utama yakni indeks kesukaran, daya beda, dan pengecoh. Rekomendasi yang dihasilkan dapat digunakan untuk memperbaiki dan menyesuaikan soal agar lebih efektif dalam mengukur pencapaian mahasiswa (Hsu et al., 2021).

## 2. Tinjauan Pustaka

Teori Tes Klasik merupakan salah satu metode yang paling umum digunakan dalam mengembangkan dan mengevaluasi kualitas butir soal. Pendekatan ini memiliki landasan historis yang kuat dan memberikan cara sederhana namun efektif untuk menilai kualitas soal berdasarkan beberapa parameter kunci, yaitu indeks kesukaran, daya beda, dan pengecoh (Crocker & Algina, 1986; Revelle, 2014). Konsep dasar dari CTT adalah adanya "*true score*" atau skor sebenarnya yang dimiliki oleh setiap mahasiswa. True score ini dipengaruhi oleh berbagai faktor, seperti kondisi tes, suasana hati, dan elemen acak yang dapat muncul selama pengambilan soal, yang secara keseluruhan membentuk "*observed score*" atau skor yang diobservasi (Gronlund & Linn, 1990). CTT memberikan kerangka yang bermanfaat untuk analisis kualitas soal di berbagai konteks pendidikan, termasuk ujian standar dan tes diagnostik.

Indeks kesukaran adalah salah satu parameter utama dalam analisis butir soal berdasarkan teori klasik. Indeks ini mengukur persentase mahasiswa yang menjawab soal dengan benar, dengan nilai berkisar antara 0 hingga 1. Nilai indeks yang tinggi menunjukkan soal yang mudah, sedangkan nilai yang rendah menunjukkan soal yang sulit. Gronlund dan Linn (1990) menyatakan bahwa soal yang baik seharusnya memiliki tingkat kesulitan yang bervariasi untuk mencakup seluruh spektrum kemampuan mahasiswa. Soal dengan indeks kesukaran sedang dianggap ideal karena mampu membedakan tingkat kemampuan mahasiswa dengan lebih baik. Soal yang terlalu mudah atau terlalu sulit, di sisi lain, tidak memberikan informasi yang memadai tentang variasi kemampuan mahasiswa, sehingga kurang efektif dalam mengukur pemahaman mahasiswa terhadap materi. Indeks kesukaran sering kali diinterpretasikan sebagai berikut:

- a. Mudah (nilai indeks di atas 0,70): Soal ini dijawab dengan benar oleh lebih dari 70% mahasiswa.
- b. Sedang (nilai indeks antara 0,30 hingga 0,70): Soal ini dijawab dengan benar oleh 30% hingga 70% mahasiswa.
- c. Sulit (nilai indeks di bawah 0,30): Soal ini dijawab dengan benar oleh kurang dari 30% mahasiswa.

Selain indeks kesukaran, daya beda adalah komponen kunci lain dalam CTT yang

menentukan seberapa baik suatu soal dapat membedakan antara mahasiswa dengan kemampuan tinggi dan rendah. Daya beda yang tinggi menunjukkan bahwa soal tersebut mampu mengidentifikasi mahasiswa yang benar-benar menguasai materi. Sebaliknya, soal dengan daya beda rendah kurang efektif dalam membedakan tingkat pemahaman mahasiswa dan sering kali perlu diperbaiki atau dihapus. Allen dan Yen (1979) menjelaskan bahwa daya beda dihitung dengan membandingkan proporsi mahasiswa dari kelompok berprestasi tinggi dan rendah yang menjawab soal dengan benar. Nilai daya beda berkisar dari -1 hingga +1, di mana nilai positif yang tinggi menunjukkan soal yang efektif dalam membedakan mahasiswa berdasarkan kemampuannya. Daya beda yang baik biasanya diinterpretasikan sebagai berikut:

- a. Daya beda sangat baik (0,40 ke atas): Soal ini sangat efektif dalam membedakan mahasiswa yang menguasai materi dan yang tidak.
- b. Daya beda baik (0,20 hingga 0,39): Soal ini cukup baik dalam membedakan mahasiswa dengan kemampuan yang berbeda.
- c. Daya beda rendah (di bawah 0,20): Soal ini kurang efektif dalam membedakan mahasiswa.

Daya beda negatif: Soal ini cenderung dijawab dengan benar oleh mahasiswa yang kurang memahami materi, yang berarti soal ini harus diperbaiki atau dihapus.

Pengecoh atau distractor dalam soal pilihan ganda memainkan peran penting dalam memastikan kualitas soal. Fungsi utama pengecoh adalah menarik perhatian mahasiswa yang kurang memahami materi, sehingga membantu soal tersebut mengukur pemahaman secara lebih akurat. Haladyna (2004) menegaskan bahwa pengecoh yang efektif akan dipilih oleh sebagian mahasiswa yang tidak memahami materi, sedangkan mahasiswa yang menguasai materi akan cenderung menghindarinya. Jika pengecoh tidak pernah dipilih oleh mahasiswa, pengecoh tersebut dianggap tidak berfungsi, sehingga perlu diperbaiki atau diganti. Keberfungsian pengecoh adalah komponen penting dalam soal pilihan ganda karena dapat memengaruhi tingkat kesulitan soal secara keseluruhan. pengecoh yang baik biasanya dipilih oleh 5-15% mahasiswa, yang menunjukkan bahwa pengecoh tersebut dapat menarik perhatian mahasiswa yang kurang memahami materi tanpa terlalu mudah dikenali sebagai jawaban yang salah.

Meskipun Teori Tes Klasik memiliki banyak keunggulan, teori ini juga memiliki beberapa keterbatasan. Salah satu kelemahan utama CTT adalah ketergantungannya pada sampel tertentu, yang dapat memengaruhi indeks kesukaran dan daya beda suatu butir soal. Hambleton & Jones (1993) menyatakan bahwa hasil analisis butir soal berdasarkan CTT mungkin tidak dapat digeneralisasikan ke populasi lain karena adanya perbedaan dalam karakteristik mahasiswa. Selain itu, teori ini tidak mempertimbangkan bahwa tingkat kesulitan

setiap soal mungkin berbeda bagi masing-masing mahasiswa. Untuk mengatasi keterbatasan ini, Item Response Theory (IRT) telah dikembangkan sebagai pendekatan yang lebih modern Djaali (2008), meskipun IRT membutuhkan sampel yang lebih besar dan perhitungan yang lebih kompleks. Meskipun demikian, CTT tetap relevan untuk analisis butir soal, terutama dalam situasi dengan keterbatasan sumber daya atau ukuran sampel yang kecil (Baker, 2001).

### **3. Metode Penelitian**

#### **Desain Penelitian**

Penelitian ini menggunakan desain penelitian deskriptif kuantitatif, yang bertujuan untuk menggambarkan karakteristik butir soal Ujian Tengah Semester (UTS) Biologi Umum berdasarkan teori tes klasik.

#### **Subjek Penelitian**

Subjek penelitian ini adalah butir soal UTS Biologi Umum yang diberikan kepada mahasiswa semester I di Universitas Negeri Padang. Jumlah butir soal yang dianalisis adalah 75 soal pilihan ganda, dengan 5 pilihan jawaban untuk setiap soal. Pemilihan soal UTS Biologi Umum didasarkan pada kebutuhan untuk menilai pemahaman mahasiswa terhadap konsep-konsep penting dalam Biologi, termasuk struktur organisme, fungsi sel, metabolisme, dan biodiversitas.

Pengambilan data dilakukan dengan menggunakan hasil ujian mahasiswa dari UTS tersebut. Populasi mahasiswa yang mengikuti ujian berjumlah 135 mahasiswa semester I, yang kemudian dibagi menjadi kelompok atas dan bawah berdasarkan skor total mereka dalam ujian. Ebel & Frisbie (1991) menyarankan bahwa untuk menghitung daya beda, data dapat dikelompokkan ke dalam dua kategori: kelompok atas (27% mahasiswa dengan skor tertinggi) dan kelompok bawah (27% mahasiswa dengan skor terendah). Kelompok ini dipilih untuk memastikan bahwa daya beda soal dapat terukur secara jelas antara mahasiswa berkemampuan tinggi dan rendah.

#### **Instrumen Penelitian**

Instrumen yang digunakan dalam penelitian ini adalah soal-soal UTS Biologi Umum dalam format pilihan ganda. Setiap soal terdiri dari satu kunci jawaban yang benar dan empat pengecoh. Data yang dianalisis mencakup proporsi mahasiswa yang menjawab setiap opsi dalam soal, termasuk kunci jawaban dan pengecoh. Data ini digunakan untuk menghitung indeks kesukaran, daya beda, dan mengevaluasi keberfungsian pengecoh pada setiap soal.

Tabel 1. Materi dan Sub Materi UAS Biologi Umum

No	Materi	Sub Materi	Nomor Soal	$\Sigma$
1	Struktur dan organisasi tubuh hewan	Echinodermata	1	1
		Arthropoda	2,3	2
		Moluska	4,5	2
		Annelida	6	1
		Nemathelminthes	7	1
		Platyhelminthes	8	1
		Coelenterata	9,10	2
		Porifera	11,12	2
		Ciri-ciri Hewan	13	1
2	Struktur dan organisasi tubuh tumbuhan	Ciri-ciri, contoh, dan manfaat tumbuhan lumut (bryophyta).	14,15,16,17	4
		Ciri-ciri, contoh, dan manfaat tumbuhan biji (spermatophyta).	18,19,20,21,22,23	6
		Tumbuhan Biji (mono/dikotil)	24,25,26,27,28,29	6
3	Biodiversitas	Keanekaragaman ekosistem	30,,31,32,33,34	5
		Endemk: anoa, babirusa, dan maleo, cendrawasih, Komodo, bangkai, rafflesia(parasite)	35	1
		Urutan hirarki	36	1
		Aturan tata nama	37	1
		Perubahan Iklim, Penggunaan Lahan, Deforestasi, urbanisasi, dan konversi lahan untuk pertanian Eksploitasi Berlebihan polusi	38,39	2
4	Metabolisme	Ciri-ciri Makhhluk Hidup	40,41,42,43,44,45	6
		Mekanisme fotosintesis	46,47,48,49	4
		Mekanisme respirasi seluler	50,51,52	3
5	Sel sebagai dasar kehidupan	Sel sebagai dasar kehidupan	53	1
		Metode Pengamatan sel.	54	1
		Sifat fisik dan sifat kimia sel	55	1
		Struktur dan fungsi sel	56,57,58,59,60,61	6
		transport zat pada sel	62,63,64	3
6	Makhhluk hidup dan metode ilmiah	Kedudukan biologi dalam sains	65	1
		Objek biologi	66,67	2
		Cabang-cabang biologi	68,69,70	3
		Langkah metode ilmiah	71,72,73	3
		Teori asal-usul kehidupan (abio, bio dan modern)	74,75	2
<b>Total</b>				<b>75</b>

### Prosedur Pengumpulan Data

Data diperoleh dari hasil ujian yang sudah diselesaikan oleh mahasiswa. Setelah itu, setiap butir soal dianalisis menggunakan CTT. Tahap pertama adalah menghitung indeks

kesukaran, yang diperoleh dengan membagi jumlah mahasiswa yang menjawab benar dengan jumlah total mahasiswa yang menjawab soal tersebut (Nitko, 2001). Kemudian, daya beda dihitung dengan menggunakan formula proporsi mahasiswa dari kelompok atas yang menjawab benar dikurangi dengan proporsi mahasiswa dari kelompok bawah yang menjawab benar (Anne & Urbina, 1997).

Keberfungsian pengecoh dievaluasi dengan melihat persentase mahasiswa yang memilih pengecoh. Pengecoh dianggap berfungsi jika setidaknya 5% dari total mahasiswa memilihnya (Haladyna, 2004). Jika pengecoh tidak dipilih oleh mahasiswa dalam jumlah yang signifikan, maka pengecoh tersebut dianggap tidak berfungsi.

### Teknik Analisis Data

Data yang telah dikumpulkan dianalisis menggunakan statistik deskriptif. Hasil analisis butir soal disajikan dalam bentuk tabel dan grafik untuk menggambarkan distribusi indeks kesukaran, daya beda, dan keberfungsian pengecoh. Hasil ini kemudian diinterpretasikan untuk memberikan rekomendasi perbaikan butir soal yang tidak memenuhi kriteria kualitas berdasarkan teori klasik.

## 4. Hasil dan Pembahasan

### Hasil

#### *Hasil Analisis Indeks Kesukaran*

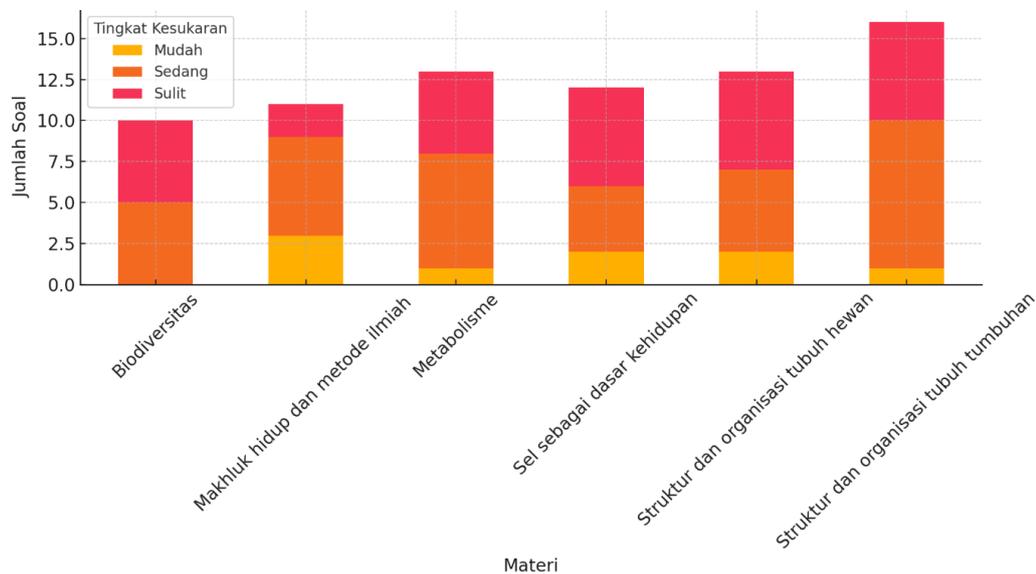
Analisis terhadap tingkat kesukaran soal UTS Biologi Umum pada tabel 2 menunjukkan bahwa soal-soal memiliki distribusi yang tidak merata. Berdasarkan data yang diperoleh, 36 soal (45%) tergolong sedang, 30 soal (37,5%) tergolong sulit, dan hanya 9 soal (11,25%) yang termasuk mudah. Hal ini menunjukkan bahwa mayoritas soal cenderung berada di kategori kesulitan sedang dan sulit.

**Tabel 2.** Distribusi Indeks Kesukaran Butir Soal UTS Biologi Berdasarkan Materi

Indeks Kesukaran	Keseluruhan	Struktur dan organisasi tubuh hewan	Struktur dan organisasi tubuh tumbuhan	Biodiversitas	Metabolisme	Sel sebagai dasar kehidupan	Makhluk hidup dan metode ilmiah
Mudah	9	2	1	0	1	2	3
Sedang	36	5	9	5	7	4	6
Sulit	30	6	6	5	5	6	2

Menurut Nitko (2001), soal yang ideal sebaiknya memiliki proporsi kesukaran yang seimbang antara soal yang mudah, sedang, dan sulit untuk mengukur kemampuan mahasiswa secara komprehensif. Dalam konteks ini, distribusi soal yang terlalu banyak berada pada kategori sulit dapat memberikan tantangan yang berlebihan bagi mahasiswa dan tidak memungkinkan mereka untuk menunjukkan kemampuan maksimal pada berbagai tingkat kemampuan. Anastasi & Urbina (1997) menyarankan bahwa soal-soal dengan tingkat kesulitan yang terlalu tinggi dapat mengakibatkan mahasiswa berkinerja buruk karena soal tersebut tidak memberi kesempatan yang cukup bagi mahasiswa untuk menunjukkan pengetahuan yang mereka kuasai.

Namun, soal yang tergolong dalam kategori sedang masih berada dalam rentang ideal menurut panduan Ebel & Frisbie (1991), yang menunjukkan bahwa soal ini cukup proporsional dalam hal memberikan tantangan yang sesuai untuk mahasiswa. Soal-soal dengan tingkat kesulitan sedang sangat penting karena mereka dapat memberikan informasi yang lebih akurat mengenai variasi kemampuan mahasiswa dalam kelompok yang diuji. Soal-soal ini memungkinkan pengajar untuk menilai apakah mahasiswa benar-benar memahami materi yang telah diajarkan di kelas.



**Gambar 1.** Distribusi Indeks Kesukaran Butir Soal UTS Biologi Berdasarkan Materi

Jika dilihat lebih lanjut berdasarkan materi, materi "Struktur dan Organisasi Tubuh Hewan" memiliki 2 soal mudah, 5 soal sedang, dan 6 soal sulit. Materi "Struktur dan Organisasi Tubuh Tumbuhan" juga menunjukkan variasi yang hampir serupa, dengan 1 soal mudah, 9 soal sedang, dan 6 soal sulit. Materi "Biodiversitas" tidak memiliki soal mudah, tetapi memiliki 5 soal sedang dan 5 soal sulit. Materi "Metabolisme" memiliki distribusi yang cukup seimbang dengan 1 soal mudah, 7 soal sedang, dan 5 soal sulit. Untuk materi "Sel sebagai Dasar

Kehidupan", terdapat 2 soal mudah, 4 soal sedang, dan 6 soal sulit. Sedangkan materi "Makhluk Hidup dan Metode Ilmiah" memiliki 3 soal mudah, 6 soal sedang, dan 2 soal sulit.

Hasil ini menunjukkan bahwa beberapa materi, seperti "Biodiversitas", memiliki lebih banyak soal sulit dan sedang, yang mungkin menandakan bahwa materi ini lebih kompleks untuk dipahami oleh mahasiswa. Sebaliknya, materi seperti "Makhluk Hidup dan Metode Ilmiah" memiliki lebih banyak soal mudah, yang menunjukkan bahwa materi ini lebih sederhana atau dapat diakses oleh mahasiswa dengan lebih baik.

#### **Hasil Analisis Daya Beda:**

Daya beda merupakan parameter yang sangat penting dalam menilai apakah soal tersebut dapat membedakan mahasiswa yang memiliki kemampuan tinggi dengan mahasiswa yang berkemampuan rendah. Berdasarkan hasil analisis pada tabel 3, ditemukan bahwa 35 soal (43,75%) memiliki daya beda rendah, 26 soal (32,5%) memiliki daya beda baik, dan 13 soal (16,25%) memiliki daya beda negatif. Hanya satu soal yang memiliki daya beda sangat baik.

**Tabel 3.** Distribusi Indeks Daya Beda Butir Soal UTS Biologi Berdasarkan Materi

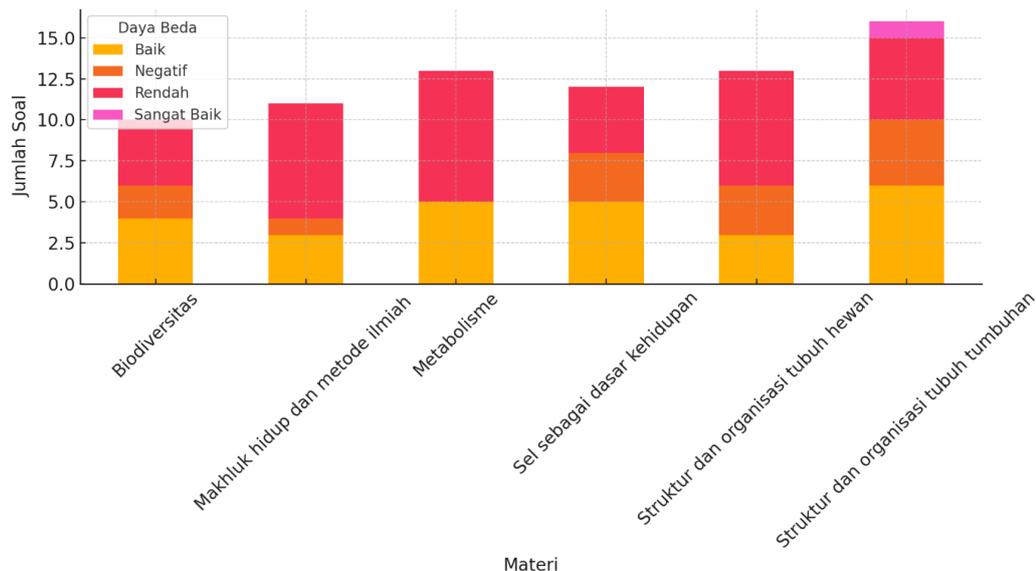
Daya Beda	Keseluruhan	Struktur dan organisasi tubuh hewan	Struktur dan organisasi tubuh tumbuhan	Biodiversitas	Metabolisme	Sel sebagai dasar kehidupan	Makhluk hidup dan metode ilmiah
<b>Negatif</b>	13	3	4	2	0	3	1
<b>Rendah</b>	35	7	5	4	8	4	7
<b>Baik</b>	26	3	6	4	5	5	3
<b>Sangat Baik</b>	1	0	1	0	0	0	0

Gronlund & Linn (1990) menjelaskan bahwa soal dengan daya beda rendah menunjukkan bahwa soal tersebut kurang efektif dalam membedakan antara mahasiswa dengan kemampuan yang berbeda. Soal dengan daya beda negatif bahkan lebih bermasalah, karena menunjukkan bahwa mahasiswa dengan kemampuan rendah justru lebih cenderung menjawab benar dibandingkan mahasiswa dengan kemampuan tinggi. Ini sering kali disebabkan oleh adanya kesalahan pada kunci jawaban atau pengecoh yang terlalu mudah dikenali oleh mahasiswa.

Daya beda yang rendah dan negatif menunjukkan bahwa soal tersebut memerlukan perbaikan. Anastasi dan Urbina (1997) menyarankan bahwa soal dengan daya beda negatif harus segera direvisi atau dihilangkan dari ujian, karena soal ini merusak validitas tes secara keseluruhan. Jika soal tidak mampu membedakan antara mahasiswa yang memahami materi

dan yang tidak, maka hasil tes tersebut tidak dapat dijadikan dasar yang kuat untuk menilai pencapaian mahasiswa.

Nitko (2001) menjelaskan bahwa idealnya, soal dengan daya beda baik harus mencapai proporsi yang lebih tinggi dalam suatu tes. Hal ini penting agar tes dapat berfungsi secara optimal dalam menilai dan mengelompokkan mahasiswa berdasarkan tingkat pencapaian mereka. Oleh karena itu, soal dengan daya beda rendah harus direvisi agar kualitas pengukuran yang dihasilkan lebih andal.



**Gambar 2.** Distribusi Indeks Daya Beda Butir Soal UTS Biologi Berdasarkan Materi

Berdasarkan materi, soal-soal dengan daya beda negatif banyak ditemukan pada materi "Struktur dan Organisasi Tubuh Hewan" (3 soal), "Struktur dan Organisasi Tubuh Tumbuhan" (4 soal), dan "Sel sebagai Dasar Kehidupan" (3 soal). Hal ini menandakan bahwa ada beberapa soal dalam materi tersebut yang perlu direvisi agar dapat berfungsi dengan lebih baik dalam menilai kemampuan mahasiswa. Materi "Metabolisme" tidak memiliki soal dengan daya beda negatif, yang menunjukkan bahwa soal pada materi ini lebih baik dalam membedakan mahasiswa dengan kemampuan yang berbeda.

Soal dengan daya beda rendah mendominasi hampir semua materi, dengan distribusi terbesar terdapat pada materi "Metabolisme" dengan 8 soal, dan "Makhluk Hidup dan Metode Ilmiah" dengan 7 soal. Gronlund & Linn (1990) menyebutkan bahwa soal dengan daya beda rendah perlu diperbaiki, karena soal ini tidak efektif dalam mengidentifikasi perbedaan kemampuan mahasiswa. Sementara itu, soal dengan daya beda baik ditemukan cukup merata di berbagai materi, dengan jumlah tertinggi pada "Struktur dan Organisasi Tubuh Tumbuhan" (6 soal) dan "Metabolisme" (5 soal).

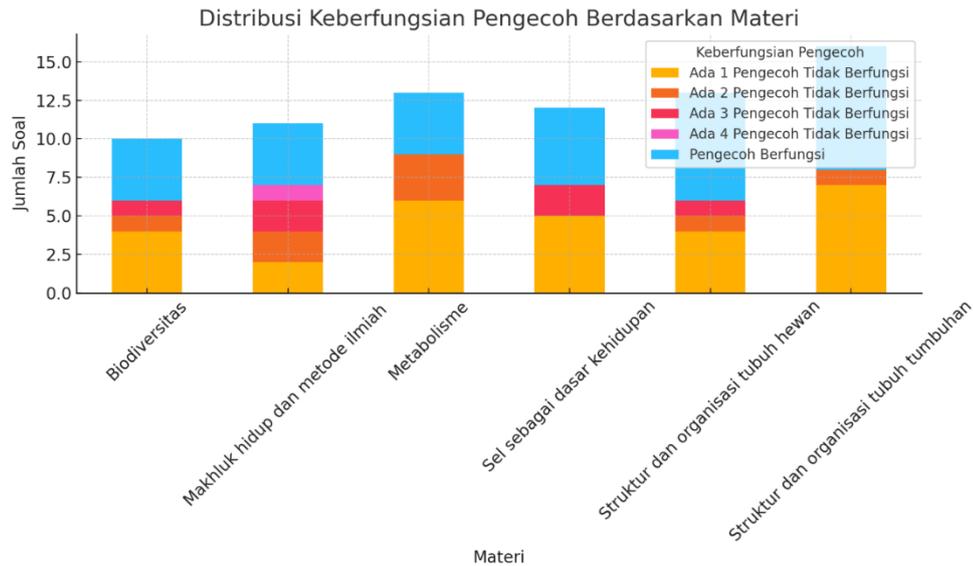
**Hasil Analisis Keberfungsian Pengecoh:**

Keberfungsian pengecoh adalah aspek yang tidak kalah penting dalam soal pilihan ganda. Pengecoh yang baik harus mampu menarik perhatian mahasiswa yang tidak memahami materi, sementara mahasiswa yang memahami materi akan menghindarinya. Berdasarkan hasil analisis, sebanyak 32 soal (40%) memiliki pengecoh yang berfungsi dengan baik, sedangkan 28 soal (35%) memiliki 1 pengecoh yang tidak berfungsi, 8 soal (10%) memiliki 2 pengecoh yang tidak berfungsi, dan 6 soal (7,5%) memiliki 3 pengecoh yang tidak berfungsi. Terdapat 1 soal dengan 4 pengecoh yang tidak berfungsi, yang berarti tidak ada pengecoh yang berhasil menarik perhatian mahasiswa.

**Tabel 4.** Distribusi Keberfungsian Pengecoh Butir Soal UTS Biologi Berdasarkan Materi

Fungsi Pengecoh	Materi						
	Keseluruhan	Struktur dan organisasi tubuh hewan	Struktur dan organisasi tubuh tumbuhan	Biodiversitas	Metabolisme	Sel sebagai dasar kehidupan	Makhluk hidup dan metode ilmiah
<b>Pengecoh Berfungsi</b>	32	7	8	4	4	5	4
<b>Ada 1 Pengecoh Tidak Berfungsi</b>	28	4	7	4	6	5	2
<b>Ada 2 Pengecoh Tidak Berfungsi</b>	8	1	1	1	3	0	2
<b>Ada 3 Pengecoh Tidak Berfungsi</b>	6	1	0	1	0	2	2
<b>Ada 4 Pengecoh Tidak Berfungsi</b>	1	0	0	0	0	0	1

Menurut Haladyna (2004), pengecoh yang tidak berfungsi mengindikasikan bahwa pilihan jawaban alternatif tersebut terlalu mudah dikenali sebagai jawaban yang salah atau tidak cukup menarik perhatian mahasiswa. Ini menurunkan kualitas soal secara keseluruhan, karena soal tersebut menjadi lebih mudah dari yang diharapkan. Pengecoh yang baik seharusnya dipilih oleh 5-15% mahasiswa yang tidak memahami materi, sehingga soal tetap menantang dan dapat menguji pemahaman mahasiswa secara lebih efektif.



Jika dilihat per materi, soal-soal dengan pengecoh berfungsi paling banyak ditemukan pada materi "Struktur dan Organisasi Tubuh Hewan" (7 soal) dan "Struktur dan Organisasi Tubuh Tumbuhan" (8 soal). Namun, materi "Biodiversitas", "Metabolisme", dan "Makhluk Hidup dan Metode Ilmiah" menunjukkan adanya pengecoh yang tidak berfungsi pada beberapa soal. Haladyna (2004) menyatakan bahwa pengecoh yang tidak berfungsi menandakan bahwa pilihan jawaban tersebut terlalu mudah dikenali sebagai salah, sehingga mahasiswa dapat dengan mudah mengabaikannya. Pengecoh seperti ini harus direvisi atau diganti agar soal tetap menantang dan efektif dalam menilai pemahaman mahasiswa.

Materi "Metabolisme" memiliki 4 soal dengan pengecoh berfungsi, namun terdapat juga 3 soal dengan 2 pengecoh yang tidak berfungsi, yang menunjukkan bahwa soal dalam materi ini membutuhkan perbaikan dalam opsi jawaban pengecoh. Sebaliknya, materi "Struktur dan Organisasi Tubuh Tumbuhan" memiliki distribusi yang lebih baik, dengan 8 soal yang pengecohnya berfungsi dengan baik dan hanya sedikit yang memerlukan perbaikan.

### Pembahasan

Hasil penelitian ini menunjukkan bahwa kualitas butir soal UTS Biologi Umum masih memiliki beberapa kelemahan signifikan, terutama dalam aspek daya beda dan keberfungsian pengecoh. Meski sebagian besar soal berada pada tingkat kesukaran yang diharapkan, proporsi soal dengan daya beda rendah dan pengecoh yang tidak berfungsi menunjukkan bahwa perlu dilakukan revisi pada beberapa butir soal untuk meningkatkan efektivitas pengukuran. Untuk memahami temuan ini dengan lebih baik, penting untuk membandingkannya dengan penelitian serupa dalam konteks kualitas butir soal.

### *Indeks Kesukaran: Seimbang tetapi Tidak Ideal*

Dalam penelitian ini, soal dengan tingkat kesulitan sedang mendominasi (48%),

sementara soal mudah (12%) dan sulit (40%) menunjukkan variasi yang cukup. Komposisi ini tidak sepenuhnya seimbang dengan rekomendasi umum, di mana soal ideal sebaiknya memiliki proporsi soal mudah, sedang, dan sulit yang lebih merata. Ebel & Frisbie (1991) mengusulkan bahwa tes yang baik seharusnya memiliki distribusi yang seimbang antara tingkat kesulitan agar dapat mengukur mahasiswa di berbagai level kemampuan. Namun, proporsi soal sulit yang cukup tinggi dapat menjadi tantangan bagi mahasiswa berkemampuan rendah.

Penelitian serupa yang dilakukan oleh Sainuddin & Ilyas (2016) dalam konteks pendidikan menengah menunjukkan temuan yang hampir serupa, di mana soal-soal ujian didominasi oleh soal dengan tingkat kesulitan sedang dan sulit. Ini mengindikasikan bahwa pengujian sering kali lebih berfokus pada aspek penguasaan mahasiswa terhadap materi yang lebih sulit, yang mungkin tidak sepenuhnya representatif dalam mengukur pemahaman mahasiswa di berbagai tingkatan kemampuan. Temuan ini menggarisbawahi perlunya menambahkan lebih banyak soal mudah untuk memberikan distribusi kesulitan yang lebih seimbang.

#### ***Daya Beda: Proporsi yang Perlu Ditingkatkan***

Hasil penelitian menunjukkan bahwa 43,75% dari total soal memiliki daya beda rendah, dan 17% memiliki daya beda negatif. Soal-soal dengan daya beda rendah tidak mampu membedakan mahasiswa berkemampuan tinggi dan rendah, dan soal dengan daya beda negatif menunjukkan adanya kesalahan struktural, seperti pengecoh yang membingungkan atau jawaban benar yang tidak jelas. Gronlund & Linn (1990) menekankan bahwa daya beda yang buruk pada soal tidak hanya merusak reliabilitas soal, tetapi juga mencerminkan kesalahan dalam mengukur perbedaan kemampuan mahasiswa.

Penelitian oleh Purnamasari et al. (2019) yang menganalisis soal ujian di tingkat perguruan tinggi menemukan bahwa sekitar 40% soal memiliki daya beda rendah, sementara penelitian ini menemukan persentase yang lebih tinggi (43,75%). Temuan ini menunjukkan bahwa masalah soal dengan daya beda rendah cukup umum dalam pengembangan soal ujian di berbagai jenjang pendidikan. Kegagalan dalam mengidentifikasi perbedaan kemampuan mahasiswa menunjukkan bahwa revisi soal sangat diperlukan. Dalam konteks yang lebih luas, hal ini juga mengindikasikan bahwa banyak soal ujian yang dibuat belum sepenuhnya memenuhi standar pengujian yang efektif.

Dari perspektif praktis, daya beda negatif pada soal seharusnya menjadi perhatian utama dalam revisi soal. Anastasi & Urbina (1997) menyatakan bahwa soal dengan daya beda negatif sebaiknya dihilangkan atau direvisi secara substansial, karena soal tersebut secara fundamental salah dalam membedakan mahasiswa dengan kemampuan yang berbeda. Penelitian ini

menunjukkan bahwa materi "Struktur dan Organisasi Tubuh Tumbuhan" dan "Sel sebagai Dasar Kehidupan" memiliki proporsi yang cukup tinggi dari soal dengan daya beda negatif, yang mengisyaratkan bahwa revisi pada soal-soal ini sangat diperlukan.

### ***Keberfungsian Pengecoh: Masalah yang Konsisten***

Keberfungsian pengecoh juga menjadi perhatian dalam penelitian ini. Sebanyak 37% soal memiliki 1 pengecoh yang tidak berfungsi, sementara 8% soal memiliki 3 pengecoh yang tidak berfungsi. Hal ini menunjukkan bahwa banyak soal pilihan ganda tidak memiliki pengecoh yang efektif, yang berarti opsi jawaban yang salah terlalu mudah dikenali oleh mahasiswa, sehingga mereka dapat dengan mudah mengabaikannya. Haladyna (2004) menekankan pentingnya pengecoh yang berfungsi dengan baik dalam meningkatkan efektivitas soal pilihan ganda. Pengecoh yang baik harus menarik perhatian mahasiswa yang tidak memahami materi dengan cukup baik, sementara mahasiswa yang memahami materi dapat menghindari pengecoh tersebut.

Dalam penelitian serupa oleh Sainuddin & Ilyas (2016), ditemukan bahwa lebih dari 30% soal ujian di tingkat menengah memiliki pengecoh yang tidak berfungsi, yang secara signifikan menurunkan tingkat kesulitan soal. Penelitian ini menemukan angka yang lebih tinggi, dengan total 57% soal memiliki pengecoh yang tidak berfungsi sebagian (dengan 1 atau lebih pengecoh tidak dipilih mahasiswa). Hal ini menegaskan bahwa masalah pengecoh yang tidak berfungsi adalah masalah yang umum dalam pengembangan soal pilihan ganda di berbagai tingkat pendidikan.

Revelle (2014) menyebutkan bahwa pengecoh yang tidak berfungsi dapat membuat soal menjadi terlalu mudah dan mengurangi validitas tes. Dalam penelitian ini, pengecoh yang tidak berfungsi ditemukan terutama pada soal di materi "Biodiversitas", "Metabolisme", dan "Makhluk Hidup dan Metode Ilmiah". Revisi pengecoh ini menjadi prioritas untuk memastikan bahwa soal tersebut lebih menantang dan efektif dalam mengevaluasi kemampuan mahasiswa.

Penelitian ini memiliki kesamaan dengan penelitian lain, baik dari segi temuan maupun rekomendasi. Sebagai contoh, penelitian oleh Purnamasari et al. (2019) juga menemukan masalah serupa dengan daya beda rendah dan pengecoh yang tidak berfungsi, terutama dalam konteks soal ujian berbasis pilihan ganda. Baik penelitian ini maupun penelitian sebelumnya menyoroti perlunya meningkatkan kualitas soal dengan memfokuskan pada revisi soal yang memiliki kelemahan signifikan, terutama pada pengecoh yang tidak berfungsi dan soal dengan daya beda rendah atau negatif.

Namun, penelitian ini juga menunjukkan perbedaan yang menarik, terutama dalam proporsi soal sulit dan pengecoh yang tidak berfungsi. Jika dibandingkan dengan penelitian

oleh Sainuddin & Ilyas (2016), proporsi soal sulit dalam penelitian ini lebih tinggi, yang menunjukkan bahwa soal UTS Biologi Umum mungkin lebih menantang daripada soal ujian di mata pelajaran lain atau tingkat pendidikan yang lebih rendah. Ini menandakan bahwa distribusi kesulitan soal dalam ujian ini masih dapat dioptimalkan lebih lanjut agar lebih proporsional dan dapat mencakup berbagai tingkatan kemampuan mahasiswa.

### ***Implikasi untuk Pengembangan Soal***

Berdasarkan temuan ini, pengembangan soal di masa mendatang harus lebih memperhatikan aspek daya beda dan keberfungsian pengecoh. Nitko (2001) menyarankan bahwa revisi soal dengan daya beda rendah harus difokuskan pada perbaikan pengecoh dan struktur soal itu sendiri. Peningkatan kualitas pengecoh dapat memastikan bahwa soal menjadi lebih efektif dalam membedakan mahasiswa yang memahami materi dan yang tidak (Romm et al., 2019).

Selain itu, Anastasi & Urbina (1997) menekankan pentingnya melakukan uji coba dan analisis butir soal secara berkala untuk memastikan kualitas soal tetap terjaga. Hasil penelitian ini mendukung perlunya evaluasi berkelanjutan terhadap soal ujian di berbagai mata pelajaran untuk memastikan bahwa instrumen evaluasi yang digunakan oleh sekolah dan lembaga pendidikan memenuhi standar yang diharapkan.

## **5. Kesimpulan dan Saran**

Hasil analisis butir soal UTS Biologi Umum menggunakan teori tes klasik menunjukkan bahwa distribusi tingkat kesukaran soal sebagian besar berada pada kategori sedang (48%), dengan 40% soal tergolong sulit dan 12% soal mudah. Distribusi ini mengindikasikan dominasi soal-soal yang lebih menantang, namun menunjukkan kurangnya keseimbangan. Agar ujian lebih aksesibel dan sesuai dengan variasi kemampuan mahasiswa, perlu ditambah soal-soal yang lebih mudah untuk mencapai distribusi yang seimbang antara soal mudah, sedang, dan sulit.

Dari sisi daya beda, mayoritas soal memiliki daya beda rendah (43,75%) dan 17% bahkan memiliki daya beda negatif, menunjukkan ketidakmampuan soal dalam membedakan antara mahasiswa berkemampuan tinggi dan rendah. Hanya 35% soal yang memiliki daya beda baik atau sangat baik, mengindikasikan kebutuhan perbaikan signifikan untuk meningkatkan ketepatan soal dalam mengevaluasi perbedaan kemampuan. Pada aspek pengecoh, hanya 43% pengecoh yang berfungsi dengan baik, sementara sisanya menunjukkan beberapa pengecoh tidak efektif dalam menarik perhatian mahasiswa yang kurang memahami materi, sehingga peninjauan dan perbaikan pengecoh diperlukan untuk meningkatkan kualitas evaluasi dalam

soal pilihan ganda.

Ke depan, penggunaan Item Response Theory (IRT) disarankan untuk memperoleh analisis yang lebih komprehensif terkait karakteristik soal dan kemampuan mahasiswa. IRT menawarkan pendekatan lebih mendalam yang dapat meningkatkan keakuratan evaluasi. Selain itu, penelitian selanjutnya bisa memperluas sampel ke berbagai tingkatan pendidikan untuk meningkatkan keterwakilan data, serta melakukan analisis kualitatif terhadap pengecoh, guna memahami dan memperbaiki alasan ketidakefektifan pengecoh dalam meningkatkan kualitas soal.

## 6. Konflik Kepentingan

Penulis menyatakan tidak ada konflik kepentingan.

## DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. : Wadsworth.
- Anne, A., & Urbina, S. (1997). *Tes Psikologi* (Benyamin Molan, Ed.). Prenhalindo.
- Baker, F. B. (2001). The Basics of Item Response Theory. In *ERIC Clearinghouse on Assessment and Evaluation*. <https://doi.org/10.1111/j.1365-2362.2010.02362.x>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
- Crocker, L., Algina, J., Staudt, M., Mercurio, S., Hintz, K., & Walker, R. A. (2008). Introduction to Classical and Modern Test Theory. In *Cengage Learning*.
- Djaali, H., & Muljono, P. (2008). *Pengukuran dalam Bidang Pendidikan* (Y. B. Sudarmanto, Ed.). Grasindo.
- Ebel, R. L., & Frisbie, D. A. (1991). Essentials of Educational Measurement. In *Educational Researcher* (Fifth Edit). Rajkamal Electirc Press. <https://doi.org/10.2307/1175572>
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Evaluation in Teaching*. McMillan Company.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Routledge.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47.
- Hsu, J. L., Lo, S. M., & Sato, B. K. (2021). Defining understanding: Perspectives from biology instructors & biology education researchers. *The American Biology Teacher*, 83(6), 372–376.
- Nitko, A. J. (2001). *Educational Assessment of Students. Third Edition*.
- Purnamasari, U. D., & Kartowagiran, B. (2019). Application rasch model using R program in analyze the characteristics of chemical items. *Jurnal Inovasi Pendidikan IPA*, 5(2), 147–156.
- Revelle, W. (2014). Classical test theory and the measurement of reliability. *Psychometric Theory*, 205–239.
- Romm, A. T., Kika, J. C., & Schoer, V. (2019). A test taker's gamble: The effect of average grade to date on guessing behaviour in a multiple choice test with a negative marking rule.

- South African Journal of Economic and Management Sciences*, 22(1), 1–12.
- Saha, S. K. (2021). Towards development of a system for automatic assessment of the quality of a question paper. *Smart Learning Environments*, 8(1), 4.
- Sainuddin, S., & Ilyas, M. (2016). Karakteristik Butir Tes Matematika pada Tes Buatan MGMP Matematika Kota Palopo Berdasarkan Teori Klasik. *Pedagogy: Jurnal Pendidikan Matematika*, 1(1).
- Van Canh, N., Van Tac, P., & Tuan, N. Q. (2023). Analysis and evaluation of question items: A solution to enhance the quality of multiple-choice test. *Tạp Chí Khoa Học Đại Học Đồng Tháp*, 12(3), 18–28.