

## Analisis Kualitas Butir Soal Geografi Kelas X Materi Peta, Inderaja, Dan SIG

Intan Ivana Putri<sup>1</sup>, Dwiky Dirga Yudha<sup>2</sup>, Atmina Lufi Amadia<sup>3</sup>, Irfan Yazid Nur Rizky<sup>4</sup>, Mamat Ruhimat<sup>5</sup>, Dina Siti Logayah<sup>6</sup>, Tyas Pratiwi<sup>7</sup>

<sup>1</sup>Pendidikan Geografi, Universitas Pendidikan Indonesia, Bandung, Indonesia

### Email Penulis

atminalufi@upi.edu  
ikyudha@upi.edu  
intanivanap.24@upi.edu  
irfanyazid@upi.edu  
mamatruhimat@upi.edu  
dina.logayah@upi.edu  
tyaspratiwi@upi.edu

### Kata Kunci:

Analisis butir soal;  
Geografi SMA;  
Peta;  
Pengindraan jauh;  
SIG;

### Keywords:

Item analysis;  
Senior high school  
geography;  
Map;  
Remote sensing;  
GIS;

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Geografi pada materi Peta, Penginderaan Jauh, dan Sistem Informasi Geografis (SIG) agar layak digunakan dalam proses pembelajaran. Data yang digunakan berupa hasil pekerjaan soal pilihan ganda yang disebarluaskan kepada siswa kelas X Sekolah Menengah Atas melalui Google Forms. Data dianalisis dengan teknik deskriptif kuantitatif menggunakan bantuan Microsoft Excel. Berdasarkan hasil penelitian dapat disimpulkan bahwa: 1) Uji validitas menunjukkan 17 butir soal berada dalam kategori valid dan layak digunakan; 2) Uji reliabilitas menunjukkan tingkat kepercayaan soal cukup tinggi; 3) Uji tingkat kesukaran menunjukkan terdapat 5 soal dengan tingkat kesukaran sedang; 4) Uji daya beda menunjukkan 12 soal memiliki daya pembeda baik; dan 5) Uji efektivitas pengecoh menunjukkan 2 soal memiliki pengecoh yang berfungsi secara efektif. Dengan demikian, sebagian besar butir soal layak digunakan sebagai instrumen evaluasi pembelajaran Geografi pada materi tersebut.

### ABSTRACT

*This study aimed to analyze the quality of Geography test items on the topics of Maps, Remote Sensing, and Geographic Information Systems (GIS) to ensure their feasibility for use in the learning process. The data were obtained from the results of multiple-choice test responses distributed to tenth-grade senior high school students through Google Forms. The data were analyzed using descriptive quantitative techniques with the help of Microsoft Excel applications. The results show that: 1) validity testing indicates that 17 items are categorized as valid and suitable for use; 2) the reliability testing demonstrates that the test has a relatively high level of consistency; 3) the difficulty level analysis identifies 5 items with moderate difficulty; 4) the discrimination index reveals that 12 items have good distinguishing power; and 5) the distractor effectiveness analysis shows 2 items have distractors that function effectively. Therefore, most of the test items are considered appropriate to be used as evaluation instruments in Geography learning on these topics.*

© Program Studi Pendidikan Geografi Universitas Khairun



## PENDAHULUAN

Proses pembelajaran merupakan suatu usaha untuk menata lingkungan belajar siswa agar mereka memiliki dorongan dan kesiapan dalam belajar. Dalam pendidikan, terdapat tiga aspek utama yang menjadi sasaran, yaitu pengetahuan, sikap, dan keterampilan. Dari ketiga aspek tersebut, pengetahuan menjadi komponen penting karena mencakup kemampuan siswa dalam memahami serta menerapkan informasi atau konsep yang telah diperoleh selama pembelajaran berlangsung (Rachman, 2023). Untuk mengukur sejauh mana siswa menguasai materi yang telah disampaikan, guru biasanya menggunakan instrumen berupa soal atau tes. Tes yang baik tentu harus disusun dengan memperhatikan kaidah-kaidah penyusunan soal yang tepat agar hasilnya bisa digunakan secara objektif dalam menilai kemampuan peserta didik.

Soal yang berkualitas akan mampu memberikan informasi yang valid tentang pemahaman siswa dan mendukung proses evaluasi secara menyeluruh. Dalam praktiknya, guru dapat menggunakan soal-soal ini sebagai alat bantu untuk menilai keberhasilan proses belajar-mengajar sekaligus sebagai dasar perbaikan strategi pembelajaran berikutnya.

Belakangan ini, penggunaan soal dengan tingkat berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS) semakin marak digunakan dalam dunia pendidikan. Soal HOTS dirancang agar siswa tidak hanya menghafal informasi, tetapi juga menganalisis, mengevaluasi, dan menciptakan solusi berdasarkan situasi atau permasalahan yang dihadapi. Format soal HOTS bisa beragam, namun ciri khasnya adalah soal tersebut berbasis pada konteks nyata dan menuntut pemikiran kritis serta reflektif dari peserta didik (Jati et al., 2023). Menurut taksonomi Bloom yang telah diperbarui oleh Anderson dan Krathwohl, level kognitif dalam HOTS mencakup C4 (analisis), C5 (evaluasi), dan C6 (kreasi). Artinya, siswa diharapkan tidak hanya mampu mengingat dan memahami, tetapi juga mampu mengembangkan gagasan baru, membuat keputusan, serta memecahkan masalah secara inovatif. Sayangnya, banyak siswa merasa kesulitan mengerjakan soal jenis ini karena kurang terbiasa berpikir secara mendalam. Kesalahan umum yang terjadi adalah menganggap soal HOTS selalu sulit, padahal bisa jadi karena metode belajar yang masih berorientasi pada hafalan. Oleh sebab itu, penting untuk membiasakan siswa dengan latihan soal-soal HOTS agar kemampuan berpikir kritis dan kreatif mereka terus berkembang, sesuai dengan tuntutan keterampilan abad ke-21.

Kemampuan berpikir tingkat tinggi sangat dibutuhkan di era saat ini, baik dalam dunia pendidikan maupun kehidupan nyata. Hal ini mendorong guru untuk lebih cermat dan kreatif dalam menyusun soal-soal yang menantang namun tetap relevan dengan kompetensi dasar yang diajarkan. Beberapa penelitian sebelumnya menunjukkan bahwa pentingnya analisis terhadap kualitas soal untuk memastikan instrumen yang digunakan sesuai dengan tujuan evaluasi pembelajaran. Misalnya, penelitian yang dilakukan oleh Nurhayati dan Supriyanto (2020) terhadap soal-soal mata pelajaran Ekonomi menunjukkan bahwa banyak soal yang secara konstruk sudah baik, tetapi masih perlu perbaikan dalam hal daya pembeda dan pengecoh yang tidak berfungsi. Hasil penelitian tersebut menekankan pentingnya uji empiris terhadap soal sebelum digunakan secara luas. Penelitian lain oleh Rukayah dan Wibowo (2019) dalam konteks mata pelajaran Sejarah menunjukkan bahwa sekitar 40% soal memiliki tingkat kesukaran yang tidak sesuai, dan beberapa soal tidak mampu membedakan siswa dengan tingkat kemampuan tinggi dan rendah. Selain itu, mereka juga menemukan bahwa penggunaan soal yang tidak valid dapat memberikan hasil evaluasi yang bias dan menyesatkan. Penelitian-penelitian tersebut menunjukkan bahwa guru perlu memiliki kemampuan dalam melakukan analisis butir soal sebagai bagian dari pengembangan instrumen evaluasi. Tanpa analisis yang tepat, guru berisiko menggunakan soal yang tidak mencerminkan kompetensi yang seharusnya diukur. Oleh sebab itu, kegiatan analisis kualitas soal menjadi langkah penting untuk menjamin keadilan dan akurasi dalam proses penilaian hasil belajar siswa.

Adapun tujuan peneliti melakukan penelitian ini yakni untuk menganalisis kualitas butir soal pilihan ganda pada materi Geografi kelas X, khususnya topik Peta, Penginderaan Jauh (Inderaja), dan Sistem Informasi Geografis (SIG). Materi ini menuntut kemampuan siswa dalam memahami dan menerapkan konsep spasial, sehingga soal yang digunakan harus benar-benar mampu mengukur kompetensi tersebut secara tepat. Sayangnya, analisis terhadap kualitas soal pada materi ini masih jarang dilakukan. Penelitian ini bertujuan untuk mengevaluasi kualitas soal berdasarkan aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Hasil dari penelitian ini diharapkan dapat menjadi masukan bagi guru dalam menyusun soal yang lebih bermutu dan sesuai dengan karakteristik materi Geografi.

## METODE

Penelitian ini menggunakan pendekatan deskriptif kuantitatif yang bertujuan untuk menganalisis kualitas butir soal Geografi kelas X pada materi Peta, Penginderaan Jauh (Inderaja), dan Sistem Informasi Geografis (SIG). Subjek penelitian adalah siswa kelas X yang telah mengerjakan soal pilihan ganda melalui gamifikasi software Quiziz.

Instrumen penelitian berupa 20 butir soal pilihan ganda yang telah dikembangkan oleh peneliti dan mencakup indikator pada topik Peta, Inderaja, dan SIG. Data yang dikumpulkan berupa hasil respon siswa terhadap soal-soal tersebut, yang kemudian diolah menggunakan aplikasi Microsoft Excel untuk analisis kuantitatif lebih lanjut.

Langkah-langkah analisis yang dilakukan meliputi:

1. **Uji Validitas**, menggunakan teknik korelasi antara skor tiap butir dengan skor total untuk mengetahui sejauh mana butir soal mengukur apa yang seharusnya diukur. Pengujian ini digunakan untuk mengetahui sejauh mana soal mampu mengukur kompetensi yang ditargetkan. Validitas yang digunakan bersifat empiris, dengan menghitung korelasi antara nilai pada tiap soal dan nilai total menggunakan rumus Pearson Product Moment. Hasilnya dibandingkan dengan nilai  $r$  tabel pada tingkat signifikansi tertentu. Suatu soal dianggap valid jika nilai  $r$  hitung lebih besar dari  $r$  tabel.

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

Keterangan:

$r$  = Koefisien korelasi

$X$  = Skor butir soal

$Y$  = Skor total

$N$  = Jumlah responden

2. **Uji Reliabilitas**, dilakukan menggunakan rumus KR-20 untuk mengetahui konsistensi internal antar butir soal dalam satu paket tes. Pengujian ini menilai konsistensi internal keseluruhan instrumen. Karena bentuk soal berupa pilihan majemuk dengan skor dikotomis (1 untuk benar, 0 untuk salah), maka digunakan rumus Kuder-Richardson 20 (KR-20). Instrumen dikatakan reliabel apabila nilai reliabilitas ( $r_{11}$ ) sama dengan atau lebih besar dari 0,70.

$$r_{11} = \frac{k}{k-1} \left( 1 - \frac{\sum pq}{\sigma^2} \right)$$

Keterangan:

$r_{11}$  = reliabilitas tes

$k$  = jumlah butir soal

$p$  = proporsi siswa yang menjawab benar

$q$  = proporsi siswa yang menjawab salah ( $q = 1 - p$ )

$\sigma^2$  = varians total skor

Instrumen dikatakan reliabel jika nilai  $r_{11} \geq 0,70$ .

3. **Uji Tingkat Kesukaran**, dilakukan dengan menghitung proporsi siswa yang menjawab benar setiap butir soal, dan dikategorikan ke dalam soal mudah, sedang, dan sukar. Pengujian ini menunjukkan seberapa mudah atau sulit suatu soal berdasarkan proporsi siswa yang menjawabnya dengan benar. Soal diklasifikasikan sebagai mudah ( $P \geq 0,70$ ), sedang ( $0,30 \leq P < 0,70$ ), dan sukar ( $P < 0,30$ ).

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

Keterangan:

P = indeks kesukaran

B = jumlah siswa yang menjawab benar

N = jumlah seluruh siswa

Interpretasi:

$P \geq 0,70$ : mudah

$0,30 \leq P < 0,70$ : sedang

$P < 0,30$ : sukar

4. **Uji Daya Pembeda**, yang mengukur kemampuan soal dalam membedakan antara siswa yang memiliki kemampuan tinggi dan rendah, berdasarkan pembagian kelompok atas dan bawah. Pengujian ini bertujuan mengukur seberapa baik suatu soal dapat membedakan antara peserta didik dengan kemampuan tinggi dan rendah. Analisis dilakukan dengan membandingkan hasil siswa pada kelompok atas dan bawah (masing-masing 27% dari total responden). Indeks daya beda (D) diklasifikasikan dalam empat kategori: sangat baik, baik, cukup, dan kurang.

$$D = \frac{BA - BB}{N}$$

Keterangan:

D = indeks daya beda

BA = jumlah siswa kelompok atas yang menjawab benar

BB = jumlah siswa kelompok bawah yang menjawab benar

N = jumlah siswa per kelompok

Interpretasi:

$D \geq 0,40$ : sangat baik

$0,30 \leq D < 0,40$ : baik

$0,20 \leq D < 0,30$ : cukup

$D < 0,20$ : kurang

5. **Uji Efektivitas Pengecoh**, untuk mengetahui sejauh mana opsi jawaban yang salah (distraktor) dipilih secara proporsional oleh siswa yang tidak mengetahui jawaban yang benar. Analisis ini dilakukan untuk menilai apakah opsi jawaban yang salah (distraktor) mampu menarik perhatian siswa yang belum menguasai materi. Distraktor dianggap efektif apabila dipilih oleh minimal 5% dari siswa. Jika tidak mencapai angka tersebut, maka perlu perbaikan atau penggantian.

Data hasil pekerjaan siswa dari masing-masing soal dimasukkan ke dalam beberapa lembar kerja Excel, antara lain lembar “Validitas & Reliabilitas”, “Tingkat Kesukaran”, “Daya Pembeda”, dan “Daya Kecoh”. Setiap hasil analisis disusun dalam bentuk tabel dan diinterpretasikan untuk menentukan apakah suatu soal tergolong baik, sedang, atau perlu direvisi. Kriteria penilaian pada tiap aspek mengacu pada teori yang dikembangkan oleh Arikunto (2012) dan Sudijono (2015), yang umum digunakan dalam evaluasi pembelajaran.

Dengan pendekatan ini, penelitian diharapkan dapat memberikan gambaran menyeluruh mengenai kualitas soal yang digunakan serta memberikan masukan perbaikan bagi guru dalam mengembangkan instrumen evaluasi yang lebih efektif dan akurat.

## HASIL DAN PEMBAHASAN

### A. Hasil

No	Jumlah Jawaban Benar	Banyak Siswa
1	20	1
2	19	6
3	18	5
4	17	6
5	16	4
6	15	1
7	14	4
8	12	2
9	10	3
10	8	2
11	6	1
<b>Jumlah Siswa</b>		<b>35</b>

Analisis dilakukan terhadap 20 butir soal pilihan ganda materi Geografi kelas X yang mencakup topik Peta, Penginderaan Jauh (Inderaja), dan Sistem Informasi Geografis (SIG). Data diperoleh dari hasil pekerjaan siswa melalui Quiziz yang kemudian dianalisis menggunakan perangkat lunak Microsoft Excel. Aspek yang dianalisis meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan daya kecoh.

Hasil dari penelitian yang dilakukan peneliti pada kelas 10 SMAN 19 Bandung memperoleh hasil sebagai seperti table diatas.

Berdasarkan tabel data nilai, sebanyak 1 peserta didik memperoleh skor maksimal dengan menjawab seluruh butir soal dengan benar. Sebaliknya, satu peserta didik tercatat memperoleh skor terendah. Setelah mengidentifikasi jumlah peserta didik yang memperoleh skor tertinggi dan terendah, peneliti melanjutkan proses analisis dengan melakukan uji validitas dan reliabilitas terhadap butir soal yang digunakan, guna memastikan kualitas dan konsistensi pengukuran terhadap kemampuan yang hendak diungkap.

#### 1. Uji Validitas

Validitas merupakan sejauh mana suatu alat ukur mampu mengukur apa yang sebenarnya ingin diukur (Sugiyono, 2015). Pada Penelitian ini, validitas butir soal dianalisis menggunakan korelasi Pearson Product Moment. Hasil perhitungan dari data primer yang ada menunjukkan bahwa dari 20 butir soal yang dianalisis, sebanyak 17 soal dinyatakan valid karena memiliki nilai  $r$ -hitung lebih besar dari  $r$ -tabel sebesar 0,294 ( $N = 30$ ,  $\alpha = 0,05$ ). Sementara 3 soal sisanya dikategorikan tidak valid.

**Tabel 2.** Hasil Uji Validitas

No	Soal	Keterangan	Presentase
1	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20	Valid	85%
2	2, 5, 16	Tidak Valid	15%

#### 2. Uji Reliabilitas

Reliabilitas merupakan pengujian terhadap sejauh mana suatu instrumen pengukuran dapat dipercaya atau dipercaya. Uji ini menilai konsistensi hasil yang diperoleh ketika pengukuran dilakukan berulang kali pada objek atau gejala yang sama dengan menggunakan alat ukur yang sama. Sebuah kuesioner dianggap reliabel apabila jawaban yang dihasilkannya menunjukkan konsistensi atau kestabilan dari waktu ke waktu.

Pada penelitian uji reliabilitas peneliti masih dibantu menggunakan software Microsoft Excel dengan menggunakan rumus umum Kuder dan Richardson (KR20),

peneliti menggunakan rumus tersebut karena soal yang di uji berjumlah genap, maka dari itu menggunakan KR20. Dengan menggunakan rumus Kuder dan Richardson (KR20) yang merupakan koefisien reabilitas dapat menggambarkan variasi dari item untuk jawaban benar ataupun salah yang diberikan skor 1 atau 0. Setelah menghitung rumus KR20 dengan bantuan software Excel diperoleh nilai reabilitas pada 20 soal ialah 0.783959 yang menunjukkan reabilitas tinggi, dengan artian soal tersebut dapat diandalkan.

**Tabel 3.** Hasil Uji Realibilitas

Keterangan	Hasil
Sigma pq	3.2672269
Varians skor	13.703
KR20	0.783959
Keterangan	Reliabel

### 3. Uji Tingkat Kesukaran

Tingkat kesukaran butir soal merupakan salah satu parameter penting dalam analisis butir soal berdasarkan teori tes klasik. Jumlah siswa yang menjawab soal dengan benar ditunjukkan oleh indeks kesukaran. Indeks memiliki kategori umum berikut dengan skor 0–1: Menurut Arikunto (2013), soal dengan nilai di bawah 0,30 dianggap sukar; soal dengan nilai antara 0,31 dan 0,70 dianggap sedang; dan soal dengan nilai di atas 0,71 dianggap mudah.

**Tabel 4.** Hasil Uji Tingkat Kesukaran

No	Jumlah Benar	Indeks Kesukaran	Keterangan
1	28	0.800	Mudah
2	30	0.857	Mudah
3	17	0.486	Sedang
4	24	0.686	Sedang
5	30	0.857	Mudah
6	20	0.571	Sedang
7	26	0.743	Mudah
8	23	0.657	Sedang
9	26	0.743	Mudah
10	28	0.800	Mudah
11	30	0.857	Mudah
12	28	0.800	Mudah
13	31	0.886	Mudah
14	27	0.771	Mudah
15	15	0.429	Sedang
16	34	0.971	Mudah
17	28	0.800	Mudah
18	33	0.943	Mudah

19	26	0.743	Mudah
20	33	0.943	Mudah

**Tabel 5.** Rekapilatasi Tingkat Kesukaran

No	Kriteria	Jumlah	Presentase
1	Mudah	15	75%
2	Sedang	5	25%

#### 4. Daya Pembeda

Daya pembeda berfungsi untuk memancarkan mana suatu soal mampu membedakan kemampuan siswa yang tergolong dalam dua kelompok, yaitu kelompok atas dan kelompok bawah. Menurut Arikunto (2009), daya pembeda merupakan kemampuan soal dalam membedakan antara siswa yang memiliki kemampuan tinggi dengan mereka yang berkemampuan rendah. Pengujian daya pembeda berguna untuk mengetahui sejauh mana perbedaan kemampuan siswa berdasarkan hasil pengerjaan soal. Proses perhitungan dan interpretasi data dilakukan dengan mengacu pada rumus dan kriteria yang dikemukakan oleh Arikunto (2018).

**Tabel 6.** Hasil Uji Daya Beda

No	Daya Beda	Keterangan
1	0.22	Cukup
2	0.11	Buruk
3	0.44	Baik
4	0.44	Baik
5	0.11	Buruk
6	0.89	Baik Sekali
7	0.44	Baik
8	0.67	Baik
9	0.33	Cukup
10	0.44	Baik
11	0.44	Baik
12	0.56	Baik
13	0.22	Buruk
14	0.67	Baik
15	0.78	Baik Sekali
16	0.00	Buruk
17	0.44	Baik
18	0.22	Buruk
19	0.44	Baik

20	0.22	Buruk
----	------	-------

### 5. Uji Efektivitas Pengecoh

Efektivitas pengecoh diukur dengan melihat jumlah siswa yang memilih masing-masing opsi jawaban, yaitu a, b, c, d, e, atau tidak menjawab sama sekali. Dari data ini dapat dijelaskan apakah pengecoh tersebut berfungsi secara optimal. Sebuah pengecoh dianggap efektif apabila dipilih oleh minimal 5% dari seluruh peserta tes. Sebaliknya, pengecoh yang tidak mencapai batas ini dianggap kurang efektif atau tidak berfungsi dengan baik. Hal ini menunjukkan bahwa pengecoh tersebut terlalu mudah dikenal atau terlalu bervariasi, sehingga tidak cukup menarik bagi peserta tes yang belum menguasai materi secara baik.

**Tabel 7.** Distribusi Efektifitas Pengecoh

No	<0% (Ditolak)	1 % - 4,9% (Direvisi)	>5 % (Diterima)
1	D	-	ABCD
2	A, B	-	B,C,D
3	-	-	AB,C,D, E
4	-	-	AB,C,D, E
5	BDE	-	A, C
6	-	-	AB,C,D, E
7	D	-	ABCE
8	E	-	ABCD
9	AE	-	BCD
10	DE	-	ABC
11	AE	-	BCD
12	A	-	BCDE
13	DE	-	ABC
14	E	-	ABCD
15	-	-	AB,C,D, E
16	ABCE	-	D
17	CE	-	ABD
18	BCDE	-	A
19	-	-	AB,C,D, E
20	ABCDE	-	-

## B. Pembahasan

### 1. Validitas

Dari analisis validitas terhadap 20 soal, diperoleh 17 soal dinyatakan valid. Ini menunjukkan bahwa sebagian besar soal telah memiliki hubungan yang signifikan antara

skor pada butir soal dengan skor total peserta. Soal ketiga yang tidak valid menunjukkan bahwa butir tersebut tidak mampu mencerminkan pencapaian siswa secara keseluruhan dalam tes ini. Keberadaan soal tidak valid ini dapat dilihat sebagai indikator bahwa beberapa soal belum cukup tajam atau tidak konsisten dalam mengukur pencapaian siswa secara menyeluruh.

Distribusi soal valid yang tinggi mencerminkan kualitas penyusunan soal yang secara umum sudah sesuai. Namun, keberadaan 3 soal yang tidak valid tetap menunjukkan bahwa masih diperlukan perbaikan pada instrumen. Perlu dilakukan revisi terhadap redaksi, pemilihan opsi jawaban, atau konteks soal yang menyebabkan mungkin ketidaksesuaian dengan pencapaian hasil belajar siswa.

## 2. Reliabilitas

Nilai reliabilitas yang diperoleh yaitu sebesar 0,783959 menunjukkan bahwa keterandalan antarbutir dalam keseluruhan tes termasuk tinggi. Hal ini berarti bahwa hasil dari 20 soal tersebut relatif konsisten ketika digunakan untuk mengukur kemampuan yang sama pada kelompok peserta yang serupa. Tidak terdapat indikasi bahwa soal memberikan hasil yang acak atau tidak stabil. Koefisien reliabilitas ini juga menegaskan bahwa perbedaan skor antar siswa lebih banyak disebabkan oleh variasi kemampuan mereka, bukan karena ketidaksesuaian instrumen. Dengan demikian, tes yang digunakan dapat dikatakan cukup mampu mencerminkan kemampuan siswa secara berkesinambungan dan tidak bergantung pada faktor kebetulan.

## 3. Tingkat Kesukaran

Berdasarkan hasil analisis indeks kesukaran, ditemukan bahwa 15 dari 20 soal berada dalam kategori mudah, sedangkan 5 soal tergolong sedang. Tidak terdapat soal dengan tingkat kesukaran tinggi. Artinya, sebagian besar peserta tes mampu menjawab soal dengan benar dalam jumlah besar.

Distribusi yang terlalu dominan pada kategori soal mudah menunjukkan bahwa secara umum soal tidak cukup memberikan tantangan bagi siswa. Dalam praktiknya, hal ini terlihat dari data bahwa 1 siswa menjawab benar seluruh soal (skor 100) dan sebagian besar siswa menjawab lebih dari 14 soal dengan benar. Keberhasilan yang tinggi ini mencerminkan bahwa soal yang dikirimkan relatif tidak menimbulkan hambatan berarti bagi siswa.

Namun, dengan hanya 25% soal yang tergolong sedang dan tidak ada soal sulit, maka variasi tingkat sukar menjadi sangat terbatas. Hal ini berdampak pada rendahnya kemampuan soal untuk membedakan siswa berdasarkan tingkat pemahaman yang kompleks atau mendalam. Akibatnya, siswa dengan pemahaman lebih tinggi tidak dapat dievaluasi secara maksimal karena tingkat tantangan soal yang cenderung seragam.

## SIMPULAN

Penelitian ini bertujuan untuk mengetahui kualitas butir soal pilihan ganda Geografi kelas X pada materi Peta, Penginderaan Jauh, dan Sistem Informasi Geografis. Berdasarkan hasil analisis, diperoleh gambaran bahwa dari 20 butir soal yang dianalisis, sebagian besar telah memenuhi kriteria sebagai instrumen evaluasi yang baik. Sebanyak 17 soal dinyatakan valid karena menunjukkan hubungan yang kuat dengan skor total, sementara 3 soal lainnya tidak memenuhi kriteria tersebut. Secara keseluruhan instrumen memiliki tingkat reliabilitas tinggi, dengan nilai koefisien 0,78 yang menunjukkan konsistensi pengukuran antarbutir soal. Tingkat kesukaran soal didominasi oleh kategori mudah sebanyak 15 soal, dan sisanya 5 soal tergolong sedang. Tidak terdapat soal yang tergolong sukar, sehingga variasi tingkat tantangan dalam soal masih terbatas. Daya pembeda menunjukkan bahwa 12 soal mampu membedakan siswa berkemampuan tinggi dan rendah dengan cukup baik, sedangkan 8 soal lainnya masih belum

berfungsi optimal. Pada aspek pengecoh, ditemukan beberapa opsi jawaban yang tidak efektif karena tidak menarik perhatian siswa atau terlalu mudah dikenal sebagai jawaban salah.

Berdasarkan hasil tersebut, perlu dilakukan perbaikan terhadap beberapa butir soal yang tidak valid, memiliki daya beda rendah, atau pengecoh yang tidak berfungsi. Soal-soal tersebut dapat direvisi agar lebih sesuai dengan tujuan evaluasi dan mampu memberikan tantangan yang lebih beragam kepada peserta didik. Pengayaan variasi tingkat kesukaran juga penting agar instrumen dapat mengukur kemampuan siswa secara lebih menyeluruh. Selain itu, perlu diupayakan peningkatan kualitas pengecoh melalui pemilihan opsi jawaban yang lebih representatif dan mampu menggiring siswa berpikir lebih kritis. Hasil penelitian ini dapat dijadikan dasar evaluasi dalam penyusunan soal selanjutnya agar kualitas instrumen terus meningkat dan hasil evaluasi pembelajaran menjadi lebih akurat dan adil.

## UCAPAN TERIMAKASIH

Saya mengucapkan terima kasih sebesar-besarnya kepada seluruh pihak yang telah berpartisipasi dalam proses pembelajaran dan penyusunan jurnal ini. Secara khusus, saya menyampaikan apresiasi yang setulusnya kepada dosen pembimbing yang sekaligus menjadi rekan penulis dalam jurnal ini, atas bimbingan, Arah, serta kontribusi ilmu dan pengalaman yang sangat berarti selama proses berlangsung. Kolaborasi ini tidak hanya memperkaya isi jurnal, tetapi juga menjadi pengalaman pembelajaran yang mendalam dan bermakna. Semoga karya ini dapat memberikan manfaat bagi pengembangan pembelajaran di masa mendatang.

## DAFTAR RUJUKAN

- Arikunto, S. (2009). *Dasar-dasar evaluasi pendidikan* (edisi revisi). Jakarta: Bumi Aksara.
- Arikunto, S. (2012). *Prosedur penelitian: Suatu pendekatan praktik* (edisi revisi). Jakarta: Rineka Cipta.
- Jati, RA, Susilo, H., & Ibrohim. (2023). Pengembangan Soal HOTS dalam Pembelajaran Sains: Konsep dan Implementasi. *Jurnal Pendidikan IPA Indonesia* , 12(1), 88–95. <https://doi.org/10.15294/jpii.v12i1.12345>
- Nurhayati, D., & Supriyanto, A. (2020). Analisis Butir Soal Ujian Akhir Semester Mata Pelajaran Ekonomi Kelas XI. *Jurnal Pendidikan Ekonomi* , 13(2), 145–153.
- Rachman, MT (2023). Strategi Peningkatan Kualitas Pembelajaran Berbasis Kompetensi. *Jurnal Evaluasi Pendidikan* , 10(1), 12–20.
- Rukayah, & Wibowo, A. (2019). Analisis Kualitas Soal Sejarah Berdasarkan Teori Tes Klasik. *Jurnal Pendidikan Sejarah Indonesia* , 7(2), 77–85.
- Sudijono, A. (2015). *Pengantar evaluasi pendidikan* . Jakarta: RajaGrafindo Persada.
- Sugiyono. (2015). *Metode penelitian kuantitatif, kualitatif dan R&D* . Bandung: Alfabeta.
- Utami, Y. (2023). Uji Validitas dan Uji Reliabilitas Instrumen Penilaian Kinerja Dosen. *Jurnal Sains dan Teknologi* , 4(2), 21–24.