

ANALISIS BUTIR SOAL GEOGRAFI BERBASIS HOTS PADA MATERI FLORA DAN FAUNA DI INDONESIA DAN DUNIA KELAS XI

Cika Nania Putri Deviani^{1*}, Carissa Windianti², Pini Rianti³, Mamat Ruhimat⁴, Dina Siti Logayah⁵, Igna Juli Triana⁶

^{1,2,3,4,5,6}Universitas Pendidikan Indonesia, Bandung, Indonesia
email Koresponden: cikanania@upi.edu

Diterima:	Direvisi:	Diterbitkan:
-----------	-----------	--------------

Abstrak: Penelitian ini bertujuan untuk menganalisis kualitas 15 butir soal geografi berbasis *Higher Order Thinking Skills* (HOTS) pada materi Flora dan Fauna di Indonesia dan Dunia. Analisis dilakukan berdasarkan lima aspek utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan melibatkan 32 siswa kelas XI SMA Kartika XIX-2 Kota Bandung sebagai responden. Data diperoleh melalui platform *Quizizz* dan dianalisis menggunakan bantuan Microsoft Excel. Hasil penelitian menunjukkan bahwa sebanyak 12 soal (80%) dinyatakan valid, reliabilitas soal mencapai 0,752 yang termasuk kategori tinggi, dan tingkat kesukaran didominasi oleh soal kategori mudah (66,6%). Analisis daya pembeda menunjukkan bahwa 53,3% soal berkategori baik dan 26,6% berkategori sangat baik. Pada aspek efektivitas pengecoh, 23,7% pengecoh dinyatakan tidak efektif karena tidak dipilih sama sekali oleh siswa. Temuan ini menunjukkan bahwa sebagian besar butir soal memiliki kualitas yang baik, meskipun masih terdapat beberapa komponen yang perlu perbaikan untuk meningkatkan efektivitas evaluasi pembelajaran.

Kata Kunci: analisis soal, geografi, HOTS, evaluasi pembelajaran

Abstract: *This study aims to analyze the quality of 15 Higher Order Thinking Skills (HOTS)-based geography questions on Flora and Fauna in Indonesia and the World. The analysis was conducted based on five main aspects, namely validity, reliability, level of difficulty, differentiating power, and effectiveness of exemptions. The research used a descriptive quantitative approach involving 32 students of class XI SMA Kartika XIX-2 Bandung as respondents. Data were obtained through the Quizizz platform and analyzed using the help of Microsoft Excel. The results showed that as many as 12 questions (80%) were declared valid, the reliability of the questions reached 0.752 which was in the high category, and the level of difficulty was dominated by easy category questions (66.6%). Differentiating power analysis showed that 53.3% of questions were categorized as good and 26.6% were categorized as very good. In the aspect of the effectiveness of triggers, 23.7% of triggers were declared ineffective because they were not chosen at all by students. These findings indicate that most of the items have good quality, although there are still some components that need improvement to increase the effectiveness of learning evaluation.*

Keywords: *question analysis, geography, HOTS, learning evaluation*

Pendahuluan

Soal-soal *Higher Order Thinking Skills* (HOTS) merupakan instrumen pengukuran yang digunakan untuk mengukur keterampilan berpikir tingkat tinggi, yaitu keterampilan berpikir yang tidak sekadar mengingat (*remembering*), memahami (*understanding*), atau menerapkan (*applying*). Soal-soal HOTS pada konteks asesmen mengukur kemampuan:

1) transfer satu konsep ke konsep lainnya 2) memproses dan mengintegrasikan informasi, 3) mencari kaitan dari berbagai informasi yang berbeda-beda, 4) menggunakan informasi untuk menyelesaikan masalah, dan 5) menelaah ide dan informasi secara kritis. Dengan demikian soal-soal HOTS menguji kemampuan berpikir menganalisa, mengevaluasi, dan mencipta (Brookhart, 2010). Dalam konteks HOTS, stimulus yang disajikan harus bersifat kontekstual dan menarik. Stimulus dapat bersumber dari isu-isu global seperti masalah teknologi informasi, sains, ekonomi, kesehatan, pendidikan, infrastruktur, dan lain-lain. Stimulus juga dapat bersumber dari permasalahan-permasalahan yang ada di lingkungan sekitar sekolah seperti budaya, adat, kasus-kasus di daerah, atau berbagai keunggulan yang terdapat di daerah tertentu. Stimulus yang baik memuat beberapa informasi/gagasan, yang dibutuhkan untuk mengembangkan kemampuan mencari hubungan antar informasi, transfer informasi, dan terkait langsung dengan pokok pertanyaan (Raiyn & Tilchin, 2015).

Dimensi proses berpikir dalam Taksonomi Bloom sebagaimana yang telah disempurnakan oleh Anderson & Krathwohl, terdiri atas kemampuan: mengingat (*remembering-C1*), memahami (*understanding-C2*), menerapkan (*applying-C3*), menganalisis (*analyzing-C4*), mengevaluasi (*evaluating-C5*), dan mencipta (*creating-C6*). Soal-soal HOTS pada umumnya mengukur kemampuan pada ranah menganalisis (*analyzing-C4*), mengevaluasi (*evaluating-C5*), dan mencipta (*creating-C6*).

Dalam evaluasi kualitas soal, berbagai aspek perlu diperhatikan untuk memastikan bahwa soal yang disusun memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Evaluasi instrumen soal dibutuhkan untuk memperbaiki kualitas dan efektivitas soal. Widari menambahkan bahwa analisis kejelasan instruksi dan ketepatan pemberian soal adalah faktor penting lainnya yang perlu diperhatikan dalam evaluasi (Widari et al., 2021). Oleh karena itu, hasil dari evaluasi soal seharusnya digunakan untuk memperbaiki proses pembelajaran guna mencapai tujuan pendidikan yang diinginkan (Apriyani et al., 2021). Kualitas soal yang baik akan berkontribusi pada penilaian yang adil dan akurat mengenai kemampuan siswa di dalam proses belajar-mengajar. Oleh karena itu, penulis melakukan penelitian terhadap 15 soal geografi yang telah dibuat dan diujikan kepada 32 responden untuk mengetahui tingkat akurasi dan kualitas ke-15 butir soal tersebut dengan rumusan masalah (1) bagaimana teknik pembuatan dan pengujian 15 soal serta (2) Bagaimana hasil analisis kualitas butir 15 soal.

Oleh karena itu, dapat disimpulkan terdapat beberapa hal penting yang harus dikaji terkait analisis butir soal yang telah dibuat dan diberikan kepada siswa/i. Maka dalam penelitian ini akan membahas tentang analisis butir soal sehingga dapat diketahui kualitas dan kelayakan dari butir-butir soal tersebut untuk digunakan sebagai instrumen evaluasi.

Metode

Metode yang digunakan dalam penelitian ini adalah pendekatan kuantitatif bersifat deskriptif yang bertujuan untuk menganalisis kualitas butir soal dan tingkat pemahaman siswa

melalui hasil pengerjaan soal yang disebarakan secara daring menggunakan platform Quizizz. Sebanyak 15 butir soal pilihan ganda dengan materi Flora dan Fauna di Indonesia dan Dunia disebarakan kepada siswa kelas 11 SMA Kartika XIX-2 Kota Bandung, dengan bantuan guru mata pelajaran geografi sebagai perantara.

Responden yang kita dapatkan sebanyak 32 siswa, yang seluruh hasil pengerjaannya langsung terekam secara otomatis di dalam sistem Quizizz, dan data yang dihasilkan dianalisis menggunakan fitur laporan dari Quizizz serta melalui perhitungan menggunakan excel untuk mendukung hasil analisis. Aspek yang dianalisis meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh untuk mengukur tingkat pencapaian siswa terhadap materi Flora dan Fauna di Indonesia dan Dunia. Hasil analisis ini digunakan untuk mengevaluasi efektivitas soal sekaligus mengukur pemahaman siswa terhadap materi pembelajaran.

Hasil dan Pembahasan

Setelah diperoleh data jawaban dari responden, langkah selanjutnya adalah dengan menguji validitas terhadap 15 butir soal geografi dengan materi Flora dan Fauna di Indonesia dan Dunia. Proses ini dilakukan dengan menggunakan aplikasi Excel untuk menghitung r hitung dari setiap soal, selanjutnya nilai r hitung yang diperoleh dibandingkan dengan r tabel tingkat 0,05 dan urutan ke 30 yaitu 0,3494. Sebuah soal dianggap valid jika r hitung lebih besar dari r tabel, sedangkan kalau r hitung lebih kecil soal tersebut akan dianggap tidak valid.

Dari hasil pengujian terhadap 15 soal, terdapat 12 soal yang dianggap valid dan 2 tidak valid serta 1 tidak terdefinisi. Berikut merupakan tabel perhitungan jumlah soal yang valid dan tidak valid.

Tabel 1. Distribusi 15 Butir Soal Berdasarkan Validitas

No.	Validitas	Nomor Butir Soal	Jumlah	Persentase
1	$\geq 0,3494$	2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15	12	80%
2	$< 0,3494$	1, 10	2	13,3%
3	Tak terdefinisi	5	1	6,6%

Terdapat (80%) soal yang valid, hal ini menunjukkan bahwa butir soal tersebut sudah menjalankan fungsinya dengan baik, yakni mengukur kompetensi yang memang ingin diukur. Sementara itu, terdapat (13,3%) dan (6,6%) yang tidak valid dan tidak terdefinisi yang kemungkinan besar disebabkan oleh beberapa faktor. Hal ini menunjukkan bahwa 15 soal geografi yang telah diujikan masih perlu direvisi, karena belum sepenuhnya mengukur yang semestinya. Berikut ini disajikan rincian soal yang valid, tidak valid dan tidak terdefinisi berdasarkan dengan r hitung masing-masing.

Tabel 2. Rincian Validitas Setiap Butir Soal

Butir Soal	r hitung	Kategori
1	0,186317	tidak valid
2	0,507611	valid
3	0,504283	valid
4	0,365866	valid

5		tak terdefinisi
6	0,630998	valid
7	0,572008	valid
8	0,749966	valid
9	0,540897	valid
10	0,252566	tidak valid
11	0,600679	valid
12	0,441039	valid
13	0,444716	valid
14	0,631061	valid
15	0,471827	valid

Pengukuran reliabilitas dilakukan secara manual dengan bantuan program aplikasi Excel dengan menggunakan rumus Cronbach Alpha. Hasil dari perhitungan tersebut, kemudian diinterpretasikan dengan kriteria jika $\geq 0,70$ maka soal tersebut dapat dikatakan reliabel atau memiliki reliabilitas tinggi. Berikut adalah perhitungan dari reliabilitas :

$$r = \left(\frac{15}{15-1} \right) \left(1 - \frac{2,477}{8,318} \right)$$

$$r = 0,752$$

Dari perhitungan tersebut didapatkan reliabilitas sebesar 0,752 yang berarti soal tersebut memiliki tingkat reliabilitas yang tinggi.

Selain mengukur validitas dan reliabilitas, penulis juga mengukur tingkat kesukaran. Tingkat kesukaran adalah untuk mengukur kesulitan pada butir soal yang diajukan. Tingkat kesukaran menyatakan sejauh mana soal tersebut mudah dan sulit bagi siswa. Semakin besar persentase siswa menjawab soal dengan benar maka semakin mudah soal, sebaliknya semakin kecil persentase siswa menjawab soal dengan benar maka semakin sulit soal (Setiyawan & Wijayanti, 2020). Persentase soal mudah, sedang dan sukar dapat dilihat pada tabel berikut ini :

Tabel 3. Distribusi 15 Butir Soal Berdasarkan Tingkat Kesukaran

No.	Tingkat Kesukaran	Nomor Butir Soal	Jumlah	Persentase
1	0,71 – 1,00 (Mudah)	3, 4, 5, 7, 8, 9, 10, 11, 12, 14	10	66,6%
2	0,31 – 0,70 (Sedang)	1, 2, 6, 13, 15	5	33,3%
3	0,00 – 0,30 (Sukar)	-	0	0%

Dari tabel diatas menyatakan bahwa butir soal yang mudah mendapatkan persentase (66,6%), sedang (33,3%) dan (0%) sukar. Hasil dari 15 soal tersebut mayoritas soal berada dalam kategori mudah, ini menunjukkan bahwa sebagian besar peserta tes mampu menjawab soal-soal dengan benar. Meskipun soal yang mudah dapat meningkatkan rasa percaya diri, namun soal seperti ini kurang ideal untuk mengukur tingkat kemampuan siswa secara menyeluruh. Berikut adalah rincian tingkat kesukaran pada setiap butir soal :

Tabel 4. Rincian Tingkat Kesukaran Setiap Butir Soal

Butir Soal	Tingkat Kesukaran	Kategori
1	0,59	Sedang
2	0,56	Sedang
3	0,78	Mudah
4	0,91	Mudah
5	1,00	Mudah
6	0,63	Sedang
7	0,75	Mudah
8	0,75	Mudah
9	0,88	Mudah
10	0,91	Mudah
11	0,84	Mudah
12	0,88	Mudah
13	0,66	Sedang
14	0,81	Mudah
15	0,63	Sedang

Analisis kualitas soal berikutnya dilakukan dengan menganalisis daya pembeda. Pengukuran daya pembeda ini dilakukan secara manual menggunakan program Excel, yaitu dengan menghitung selisih antara rata-rata jumlah jawaban benar dari kelompok siswa berkemampuan tinggi dan kelompok siswa berkemampuan rendah, lalu hasilnya dibagi dengan skor maksimal tiap butir soal.

Tabel 5. Distribusi 15 Butir Soal Berdasarkan Daya Pembeda

No.	Daya Pembeda	Nomor Butir Soal	Jumlah	Persentase
1	$\leq 0,19$ (Kurang Baik)	5	1	6,6%
2	0,20 – 0,29 (Cukup)	9, 12	2	13,3%
3	0,30 – 0,39 (Baik)	1, 3, 6, 7, 11, 13, 14, 15	8	53,3%
4	$\geq 0,40$ (Sangat Baik)	2, 4, 8, 10,	4	26,6%

Tabel 6. Rincian Daya Pembeda Setiap Butir Soal

Butir Soal	Rata-Rata Atas	Rata-Rata Bawah	Daya Pembeda	Kategori
1	0,75	0,44	0,31	Baik
2	0,88	0,25	0,63	Sangat Baik
3	0,94	0,63	0,31	Baik
4	0,94	0,88	0,06	Sangat Baik
5	1,00	1,00	0,00	Kurang Baik
6	0,81	0,44	0,38	Baik
7	0,94	0,56	0,38	Baik
8	1,00	0,50	0,50	Sangat Baik
9	1,00	0,75	0,25	Cukup
10	0,94	0,88	0,06	Sangat Baik
11	1,00	0,69	0,31	Baik
12	1,00	0,75	0,25	Cukup
13	0,81	0,50	0,31	Baik
14	1,00	0,63	0,38	Baik
15	0,81	0,44	0,38	Baik

Berdasarkan Tabel 5 dan Tabel 6, analisis daya pembeda terhadap 15 butir soal menunjukkan bahwa sebagian besar soal memiliki kemampuan yang baik dalam membedakan siswa berkemampuan tinggi dan rendah. Dari Tabel 5, diketahui bahwa sebanyak 8 soal (53,3%) berada pada kategori baik dengan indeks daya pembeda antara 0,30–0,39, dan 4 soal (26,6%) termasuk dalam kategori sangat baik dengan indeks $\geq 0,40$. Sementara itu, 2 soal (13,3%) tergolong cukup dan hanya 1 soal (6,6%), yaitu butir soal nomor 5, yang masuk dalam kategori kurang baik dan perlu direvisi atau diganti karena daya pembedanya rendah.

Tabel 7. Distribusi 15 Butir Soal Berdasarkan Efektivitas Pengecoh

No.	0% (Ditolak)	0,1– 4,9% (Direvisi)	$\geq 5\%$ (Diterima)	Jumlah Ditolak	Persentase Ditolak
1	-	-	A, B, C, D, E	0	0%
2	-	E	A, B, C, D	0	0%
3	-	B, E	A, C, D	0	0%
4	B, D, E	-	A, C	3	4%
5	A, B, C, E	-	D	4	5,3%
6	-	-	A, B, C, D, E	0	0%
7	A, C	E	B, D	2	2,6%
8	-	E	A, B, C, D	0	0%
9	D, E	-	A, B, C	2	2,6%
10	D	E	A, B, C	1	1,3%
11	A, C	D	B, E	2	2,6%
12	D	B	A, C, E	1	1,3%
13	-	A, C	B, D, E	0	0%
14	A, C, D	E	B	3	4%
15	-	B, D	A, C, E	0	0%
Jumlah				18	23,7%

Berdasarkan tabel tersebut, sebagian efektivitas pengecoh sudah berfungsi dengan baik atau diterima. Hal ini terlihat dari banyaknya pengecoh yang masuk dalam kategori lebih dari 5% dan dipilih oleh siswa yaitu sebesar 58,6%, yang berarti pengecoh tersebut berfungsi. Sebaliknya, terdapat 18 pengecoh yang tidak dipilih sama sekali oleh siswa (kategori 0%), sehingga dianggap tidak efektif dan perlu diperbaiki. Persentasenya mencapai 23,7% dari keseluruhan pengecoh yang dianalisis. Selain itu, ada juga beberapa pengecoh yang hanya dipilih oleh kurang dari 5% siswa, sehingga masuk kategori perlu direvisi yaitu sebesar 17,3%.

Kesimpulan

Berdasarkan hasil analisis terhadap 15 butir soal geografi, dapat disimpulkan bahwa mayoritas soal telah memenuhi kriteria kualitas yang baik. Sebanyak 80% soal tergolong valid, reliabilitas instrumen tinggi dengan nilai 0,752, dan tingkat kesukaran didominasi oleh soal mudah. Dari segi daya pembeda, sebagian besar soal mampu membedakan siswa dengan kemampuan tinggi dan rendah, dengan 53,3% soal berkategori baik dan 26,6% sangat baik. Meskipun demikian, terdapat 23,7% pengecoh yang tidak dipilih sama sekali, menandakan perlunya perbaikan pada opsi jawaban. Oleh karena itu, beberapa soal dan pengecoh perlu direvisi agar instrumen evaluasi dapat lebih optimal dalam mengukur pemahaman siswa terhadap materi pembelajaran.

Referensi

- Apriyani, R., Simatupang, G. M., & Pasaribu, F. T. (2021). Analisis kemampuan penalaran matematis siswa dalam menyelesaikan soal cerita matematika di mts negeri 5 merangin. *Absis: Mathematics Education Journal*, 3(2), 48-56. <https://doi.org/10.32585/absis.v3i2.1383>
- Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.
- Jati, P., Ruhimat, M., & Logayah, D. S. (2023). ANALISIS KUALITAS BUTIR SOAL GEOGRAFI. *Jurnal Pendidikan Geosfer*, 8(2), 287-298.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis butir soal ujian akhir sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274-284. <https://journal3.uin-alauddin.ac.id/index.php/alasma/article/view/16569>
- Raiyn, J., & Tilchin, O. (2015). Stimulating higher-order thinking through personalized learning. *Journal of Education and Training Studies*, 3(4), 93-100.
- Ruhimat, M., Logayah, D.S., & Suhendro. (2024). Isu Isu Kontemporer Asesmen Pembelajaran Geografi. Bandung: Referensi Cendekia.
- Saputri, H. A. S., & Larasati, N. J. (2023). ANALISIS INSTRUMEN ASSESMENT: VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986-2995.
- Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis kualitas instrumen untuk mengukur kemampuan pemecahan masalah siswa selama pembelajaran daring di masa pandemi. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 130-139.
- Sulistiyo, S., & Mawardi, I. (2024). Portable Smart Biogas Digester Using Pressure Sensor and Safety Valve Based on Internet of Things. *Journal of Electrical Engineering and Computer (JEECOM)*, 6(1), 243-251.
- Widana, I. W. (2020). Pengaruh pemahaman konsep asesmen HOTS terhadap kemampuan guru matematika SMA/SMK menyusun soal HOTS. *Emasains: Jurnal Edukasi Matematika dan Sains*, 9(1), 1-10. <https://ojs.mahadewa.ac.id/index.php/emasains/article/view/618/632>
- Widari, N., Astawan, I., & Sumantri, M. (2021). Bahan ajar interaktif bermuatan pendidikan karakter pada materi sistem pernapasan pada manusia dan hewan. *Mimbar Ilmu*, 26(3), 364. <https://doi.org/10.23887/mi.v26i3.37088>