

ANALISIS SOAL PILIHAN MAJEMUK MITIGASI BENCANA KELAS XI: TINGKAT KESULITAN, PEMBEDA, DAN EFEKTIVITAS MELALUI PLATFORM DIGITAL

Anisa Nurhidayati^{1*}, Aswaly Dinove Chandra², Muhamad Rizki Parhan³, Muhamad Rizqy Setiawan⁴, Rafi Mahesa Akbar Sisnawar⁵, Mamat Ruhimat⁶, Dina Siti Logayah⁷, Rahayuni Tyas Pratiwi⁸.

^{1,2,3,4,5,6,7,8}Program Studi Pendidikan Geografi, Fakultas Pendidikan Ilmu Pengetahuan Sosial, Universitas Pendidikan Indonesia

anisanurhidayati18@upi.edu^{1*} aswalydinove04@upi.edu² rizkiparhan273@upi.edu³ rizqystw1122@upi.edu⁴
rafimahesa.as76@upi.edu⁵ mamatruhimat@upi.edu⁶ dina.logayah@upi.edu⁷ tyaspratiwi@upi.edu⁸

INFO ARTIKEL	ABSTRAK
Riwayat Artikel: Dikirim : 31-05-2025 Disetujui : 20-06-2025 Diterbitkan : 30-06-2025 Kata kunci: Pilihan Majemuk, Mitigasi Bencana, Kelas 11, Tingkat Kesulitan, Daya Beda, Platform Digital.	<i>This research was conducted to analyze the Geography subject matter questions for 11th-grade students. The population and sample are the 11th-grade students of SMA Negeri 3 Cimahi, based on the students' understanding of disaster mitigation material in the Geography subject according to the current curriculum. Technique Data collection in this study was conducted through a digital questionnaire method designed interactively and gamified using the Quizizz platform, and the analysis was carried out using Microsoft Excel. The results and discussion of this article are presented in several subsections to facilitate the understanding of the content and research findings. The analysis includes five aspects, namely the Validity Test to ensure that the questions used truly measure what they are supposed to measure, so that the measurement results are more accurate and reliable.</i> Penelitian ini dilakukan untuk menganalisis butir soal materi Geografi kelas XI. Populasi dan sampel adalah kelas XI SMA Negeri 3 Cimahi dengan didasarkan pada pemahaman siswa mengenai materi mitigasi bencana mata pelajaran Geografi dengan kurikulum yang berlaku. Teknik Pengumpulan data dalam penelitian ini dilakukan melalui metode kuesioner digital yang dirancang secara interaktif dan gamified menggunakan platform Quizizz serta analisis dilakukan menggunakan Microsoft Excel. Hasil dan pembahasan dari artikel ini yaitu disajikan dalam beberapa subbagian agar mempermudah dalam pemahaman isi dan hasil penelitian. Analisis tersebut meliputi lima aspek, yaitu Uji Validitas untuk memastikan bahwa soal yang digunakan benar-benar mengukur apa yang seharusnya diukur, sehingga hasil pengukuran lebih akurat dan dapat diandalkan.

PENDAHULUAN

Menurut Zaharai Idris (2020) Pendidikan ialah serangkaian kegiatan komunikasi yang bertujuan, antara manusia dewasa dengan si anak didik secara tatap muka atau dengan menggunakan media dalam rangka memberikan bantuan terhadap perkembangan anak seutuhnya. Hal ini sejalan dengan bagaimana pembangunan sumber daya manusia

yang berkualitas tinggi bergantung pada pendidikan. Evaluasi pembelajaran adalah bagian penting dari sistem pendidikan. Ini tidak hanya berguna untuk mengukur seberapa baik siswa belajar, tetapi juga sebagai umpan balik bagi pendidik untuk memperbaiki metode penilaian dan proses pembelajaran mereka.

Gunawan (2003,) dalam Irfan Abraham, Awaludin Tjalla2, R. Eko Indrajit

(2021) adalah proses berpikir yang mengharuskan siswa untuk memanipulasi informasi yang ada dan ide-ide dengan cara tertentu yang memberikan mereka pengertian dan implikasi baru. Misalnya, ketika siswa menggabungkan fakta dan ide dalam proses mensintesis, melakukan generalisasi, menjelaskan, melakukan hipotesis dan analisis, hingga siswa sampai pada suatu kesimpulan. Kemampuan siswa dalam aspek mental yang paling memerlukan pemikiran yang tinggi dikenal sebagai Kemampuan Thinking Order High. Levelisasi Bloom dalam taksonomi kognitif menentukan HOTS. Pada masanya, Bloom membagi Taxonomy nya menjadi dua kategori: LOTS (Kemampuan Berpikir Tingkat Rendah) dan HOTS (Kemampuan Berpikir Tingkat Tinggi).

Berdasarkan Permendikbud Nomor 43, (2019) pasal 6 adalah 1) telah menyelesaikan semua program mata pelajaran, 2) memperoleh penilaian sikap/perilaku minimal baik, 3) mengikuti setiap ujian yang dilaksanakan pada setiap mata pelajaran yang ada di sekolah. Dengan mempertimbangkan kriteria di atas, penilaian siswa tidak terpengaruh oleh hasil pengukuran, evaluasi, dan ujian sekolah, yang merupakan salah satu faktor yang menentukan kelulusan.

Guru melakukan tes dengan soal-soal yang dapat dipertanggung jawabkan yang menilai kelayakan, validitas, reliabilitas, tafsiran, dan efektifitas, serta tingkat kesulitan (kesulitan) dan daya pembeda, untuk mengukur kemampuan siswa. Evaluasi diperlukan untuk menentukan apakah materi yang dipelajari tercapai, penjelasan guru, dan keterlibatan siswa dalam pembelajaran. Evaluasi adalah penilaian rutin kegiatan dan kemajuan belajar siswa melalui ujian, praktikum, tugas, dan pengamatan yang dilakukan langsung oleh guru (Hamzah, 2014) dalam Sitti Mania, Fitriani, Ahmad Farham Majid, Nidya Nina Ichiana, Andi Ika Prasasti Abrar (2020).

METODE PENELITIAN

a. Penentuan Populasi dan Sampel

Populasi yang dijadikan objek kajian adalah seluruh siswa kelas XI di SMA Negeri 3 Cimahi. Pemilihan populasi ini didasarkan pada pertimbangan bahwa siswa kelas XI telah mempelajari topik mitigasi bencana dalam mata pelajaran geografi sesuai dengan kurikulum

yang berlaku. Sampel dalam penelitian ini ditentukan dengan menggunakan teknik purposive sampling, yaitu teknik pengambilan sampel secara sengaja berdasarkan kriteria tertentu yang sesuai dengan tujuan penelitian. Kriterianya seperti siswa berada pada tingkat kelas XI, siswa telah mendapatkan pembelajaran materi mitigasi bencana dalam mata pelajaran geografi, dan siswa bersedia serta mampu mengikuti kegiatan uji coba soal melalui platform digital.

Jumlah responden yang dilibatkan dalam penelitian ini adalah 35 orang siswa untuk setiap butir soal. Jumlah ini sudah memadai untuk keperluan analisis kuantitatif, terutama dalam perhitungan indeks kesukaran, daya pembeda, analisis distraktor, serta uji validitas dan reliabilitas instrumen. Sesuai dengan pernyataan Suharsimi Arikunto (1983), sampel minimal dalam penelitian adalah 30 subjek atau responden. Ini didasarkan pada persyaratan perhitungan atau uji statistik yang umum digunakan. Jika sampel lebih besar dari 30, maka lebih baik, tetapi minimal 30 dianggap memadai.

b. Teknik Pengumpulan Data

Teknik Pengumpulan data dalam penelitian ini dilakukan melalui metode kuesioner digital yang dirancang secara interaktif dan gamified menggunakan platform Quizizz. Instrumen yang digunakan berupa naskah soal pilihan majemuk geografi yang mencakup berbagai tingkat kognitif LOTS (Lower Order Thinking Skills), MOTS (Middle Order Thinking Skills), dan HOTS (Higher Order Thinking Skills) dengan fokus materi mitigasi bencana. Data yang diperoleh berupa jawaban siswa kelas 11 digunakan untuk menilai kualitas butir soal secara empiris. Tujuan utama dari pengumpulan data ini adalah untuk memperoleh gambaran kuantitatif mengenai efektivitas soal dalam mengukur kompetensi siswa, sekaligus sebagai dasar pengembangan instrumen evaluasi pembelajaran yang lebih valid dan reliabel.

Analisis data dilakukan secara kuantitatif dengan bantuan perangkat lunak seperti Microsoft Excel dan Google Spreadsheets. Beberapa aspek yang dianalisis meliputi indeks kesukaran (*p-value*), yang menunjukkan tingkat kesulitan soal; daya pembeda, yang mengukur sejauh mana suatu

soal dapat membedakan siswa dengan kemampuan tinggi dan rendah; serta analisis distraktor, yang mengevaluasi keefektifan opsi jawaban yang tidak benar dalam mengalihkan siswa yang belum memahami materi.

Selain itu, dilakukan juga pengujian reliabilitas untuk mengukur konsistensi internal instrumen, dan uji validitas untuk memastikan bahwa soal mengukur kompetensi yang sesuai dengan tujuan pembelajaran. Pendekatan analisis ini bertujuan untuk mendeskripsikan hasil uji coba soal secara menyeluruh, sehingga dapat dilakukan perbaikan terhadap butir soal yang belum memenuhi kriteria kualitas.

c. Teknik Pengembangan Instrumen

Pengembangan instrumen dalam penelitian ini dilakukan dengan merancang soal pilihan majemuk bertemakan mitigasi bencana yang disesuaikan dengan Kompetensi Dasar (KD) kelas XI SMA sesuai Kurikulum Merdeka. Soal yang dikembangkan terdiri atas 50 butir pilihan majemuk dengan empat opsi jawaban. Setiap butir soal disusun untuk mengukur kemampuan berpikir tingkat tinggi (Higher Order Thinking Skills/HOTS) dengan mengintegrasikan aspek kognitif dan kontekstual kebencanaan.

Validasi isi dilakukan oleh dosen Pendidikan Geografi dan mahasiswa jurusan Pendidikan Geografi, menggunakan lembar validasi instrumen yang mencakup aspek kesesuaian isi, konstruksi, dan bahasa. Revisi dilakukan berdasarkan masukan dari validator hingga instrumen dinyatakan layak digunakan. Distribusi instrumen dilakukan melalui platform digital berbasis Learning Management System (LMS) dengan aplikasi bernama "Quizizz" untuk memudahkan akses oleh peserta didik. Platform tersebut juga mendukung otomatisasi rekapitulasi jawaban guna memudahkan analisis butir soal.

d. Teknik Analisa Data

Analisis data dalam penelitian ini dilakukan untuk mengukur kualitas butir soal pilihan majemuk melalui lima tahapan pengujian, yaitu:

1. Uji Validitas

Uji validitas bertujuan untuk mengetahui sejauh mana butir soal mengukur apa yang seharusnya diukur. Validitas yang digunakan adalah

validitas empiris dengan teknik korelasi Product Moment Pearson antara skor butir soal dengan skor total. Rumus:

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

Keterangan:

r = koefisien korelasi

X = skor butir soal

Y = skor total

N = jumlah responden

Nilai r dibandingkan dengan nilai r tabel. Butir soal dinyatakan valid jika $r_{hitung} > r_{tabel}$ pada taraf signifikansi tertentu (misalnya 5%).

2. Uji Reliabilitas

Uji reliabilitas digunakan untuk mengetahui konsistensi instrumen secara keseluruhan. Rumus yang digunakan adalah Rumus Kuder-Richardson 20 (KR-20) karena bentuk soal adalah pilihan ganda/majemuk dengan skor 1 dan 0 Rumus:

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

Keterangan:

r₁₁ = reliabilitas tes

k = jumlah butir soal

p = proporsi siswa yang menjawab benar

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

σ² = varians total skor

Instrumen dikatakan reliabel jika nilai r₁₁ ≥ 0,70.

3. Uji Daya Beda

Uji daya beda bertujuan untuk mengetahui kemampuan soal dalam membedakan peserta didik berkemampuan tinggi dan rendah. Analisis dilakukan dengan membandingkan skor kelompok atas dan kelompok bawah (masing-masing 27% dari total responden). Rumus:

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

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

4. Uji Tingkat Pengecoh (Efektivitas Distraktor)

Uji ini digunakan untuk mengetahui efektivitas pengecoh (opsi jawaban yang salah) dalam menarik perhatian siswa. Distraktor dikatakan efektif jika dipilih oleh minimal 5% dari peserta didik. Jika suatu opsi tidak dipilih atau dipilih kurang dari 5%, maka dianggap tidak efektif dan perlu direvisi.

5. Uji Tingkat Kesukaran

Uji tingkat kesukaran menunjukkan proporsi peserta yang dapat menjawab benar suatu butir soal. Rumus:

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

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

e. Alat dan Bahan

Alat merupakan perangkat atau media yang digunakan untuk mengolah, menganalisis, dan menyampaikan data dalam sebuah penelitian. Dalam konteks ini, alat dan bahan yang digunakan meliputi :

1. Laptop/Komputer

Penggunaan perangkat mobile ini digunakan oleh peneliti untuk merancang soal, mengolah dan menganalisis data. Selain itu, digunakan oleh siswa untuk mengakses

dan menjawab soal melalui platform digital.

2. Platform Digital

Dalam penelitian ini, peneliti menggunakan platform digital interaktif untuk menyajikan soal-soal pilihan ganda kepada siswa. Quizizz dipilih karena mendukung fitur *gamification* dan dapat merekam respons siswa secara otomatis. Sehingga mempermudah peneliti dalam proses pengumpulan data secara efisien dan *real-time*.

3. Microsoft Excel dan Google Spreadsheets

Kedua perangkat lunak ini membantu peneliti dalam melakukan perhitungan, pengolahan dan analisis data secara kuantitatif. Perhitungan yang dilakukan dalam penelitian analisis kualitas butir soal ini menggunakan beberapa macam perhitungan, seperti :

- Indeks Kesukaran
- Daya Pembeda
- Efektifitas distraktor
- Uji Validitas dan Reliabilitas

Sedangkan bahan yang peneliti gunakan yakni merujuk pada materi atau sumber data yang digunakan dalam penelitian. Berikut bahan yang dimaksud antara lain :

1. Literatur acuan

Acuan penelitian ini menggunakan kurikulum Geografi yang digunakan sebagai dasar pembuatan dan penyusunan soal, dengan fokus pada materi Mitigasi Bencana. Pedoman Suharsimi Arikunto (1983) menjadi pedoman sebagai acuan dalam penentuan jumlah sampel dan validitas desain penelitian.

2. Butir soal Geografi

Butir soal yang dibuat berdasarkan kurikulum Geografi kelas XI yang mencakup berbagai tingkat kognitif seperti LOTS, MOTS, dan HOTS. Butir soal ini diuji melalui kuesioner digital.

3. Sampel

Kriteria pemilihan sampel dilakukan secara *Purposive Sampling*,

yakni siswa kelas XI yang telah mempelajari materi mitigasi bencana, bersedia, dan mampu untuk mengikuti tes secara digital.

4. Hasil data

Berupa jawaban para siswa, yang akan digunakan peneliti sebagai bahan dasar analisis empiris kualitas butir soal.

f. Analisis Keabsahan Hasil Penelitian

Keabsahan hasil penelitian dalam studi ini terjamin melalui penerapan uji validitas dan reliabilitas butir soal secara sistematis serta pengendalian prosedural selama proses pengumpulan dan analisis data.

Penelitian ini ditinjau dari validitas isi (*content validity*), yakni kesesuaian antara butir soal dengan kompetensi dasar dan indikator pembelajaran pada kurikulum Geografi kelas XI, dengan fokus pada materi Mitigasi Bencana. Penyusunan soal ini dilakukan berdasarkan silabus dan kisi-kisi soal yang telah tervalidasi oleh ahli Geografi dan pendidik yang berpengalaman.

Selain itu, validitas empiris dievaluasi melalui analisis statistik kualitas butir soal, khususnya :

- i. Indeks kesukaran (*p-value*) :
Butir soal akan dianggap valid jika tidak berada di luar rentang ekstrem (terlalu mudah atau terlalu sulit), idealnya antara 0,31-0,70.

ii. Daya pembeda (DI)

: Soal dinilai memiliki validitas baik apabila mampu membedakan antara siswa dengan kemampuan tinggi dan rendah ($DI \geq 0,30$).

Butir soal yang tidak memenuhi kriteria diatas disebut tidak valid secara empiris dan disarankan untuk direvisi atau bahkan dieliminasi.

Pengujian reliabilitas dilakukan dengan Kuder Richardson Formula 20 (KR-20) yang sesuai untuk instrumen dengan pilihan ganda dikotomis. Nilai reliabilitas yang diperoleh sebesar 0,82, yang menunjukkan bahwa konsistensi internal instrumen berada dalam kategori tinggi, menurut standar interpretasi reliabilitas (Arkunto, 2013).

Tingginya nilai KR-20 mengindikasikan bahwa soal-soal dalam instrumen termasuk kedalam keterkaitan yang kuat dalam mengukur konstruk yang sama, yaitu pemahaman siswa terhadap konsep Mitigasi Bencana.

Pemilihan sampel penelitian ini menggunakan Purposive Sampling dengan kriteria yang sesuai dengan tujuan penelitian turut mendukung validitas eksternal hasil studi ini. Keterlibatan siswa kelas XI yang telah mendapatkan materi Mitigasi Bencana dengan jumlah responden minimal 30 orang memenuhi syarat statistik dasar untuk generalisasi terbatas dalam lingkup konteks yang sama.

Dalam penggunaan platform digital Quizizz, menjamin integritas proses pengumpulan data, dikarenakan :

- i. Setiap siswa memperoleh soal yang sama dengan urutan acak
- ii. Jawaban tercatat secara otomatis dan objektif
- iii. Mengurangi intervensi subjektif peneliti dalam penilaian.

Dapat disimpulkan bahwa hasil penelitian ini memiliki tingkat keabsahan yang dapat dipertanggungjawabkan, baik secara empiris, statistik, maupun prosedural. Temuan yang diperoleh dapat menjadi dasar ujianr untuk pengembangan instrumen evaluasi yang lebih berkualitas dan relevan dalam pembelajaran Geografi.

HASIL DAN PEMBAHASAN

Studi Evaluasi Menggunakan Platform Digital Untuk Menilai Kualitas Butir Soal Dalam Pembelajaran Geografi

Hasil dan pembahasan dari artikel ini yaitu disajikan dalam beberapa subbagian agar mempermudah dalam pemahaman isi dan hasil penelitian. Secara Sistematis dalam penelitian ini, maka dilakukan proses analisis secara komprehensif terhadap soal- soal yang telah dikembangkan. Analisis tersebut meliputi lima aspek, yaitu Uji Validitas untuk memastikan bahwa soal yang digunakan benar-benar mengukur apa yang seharusnya diukur, sehingga hasil pengukuran lebih akurat dan dapat diandalkan.

Selanjutnya ujian Reliabilitas untuk mengetahui konsistensi hasil pengukuran, Analisis tingkat pengecoh untuk menentukan

seberapa sulit atau mudahnya suatu soal. Analisis kekeliruan untuk menilai efektivitas pilihan jawaban yang salah dalam mengarahkan jawaban siswa, dan Analisis Daya Pembeda untuk mengukur kemampuan butir soal dalam membedakan siswa yang berkemampuan tinggi dan rendah. seluruh hasil analisis tersebut dikerjakan menggunakan aplikasi spreadsheet dan excel guna menunjang tingkat ketelitian dan mempercepat proses pengolahan data secara menyeluruh.

Data yang telah dianalisis kemudian disusun dan disajikan dalam bentuk tabel secara runtut dan terorganisasi, sehingga informasi yang ditampilkan menjadi lebih mudah dipahami, diinterpretasikan, serta mendukung proses pengambilan keputusan berdasarkan data tersebut.

Uji Validitas

Hasil pengujian validitas terhadap soal mitigasi bencana menunjukkan bahwa tidak seluruh butir soal yang disusun memenuhi kriteria validitas yang diharapkan. Validitas diuji dengan cara menghitung hubungan antara skor setiap butir soal dan skor total menggunakan rumus korelasi Pearson Product Moment, lalu hasil tersebut dibandingkan dengan nilai r tabel pada tingkat signifikansi tertentu. Dalam situasi ini, nilai r tabel yang dijadikan acuan adalah 0,3338. Dari 50 soal yang diujikan, hanya beberapa yang menunjukkan nilai r hitung melebihi r tabel, sehingga dianggap valid.

Butir soal yang sah ini berarti dapat mengukur kemampuan atau pemahaman belajar siswa mengenai materi mitigasi bencana dengan tepat dan sesuai dengan tujuan pengukuran yang telah ditentukan. Sementara itu, sebagian besar soal lainnya menunjukkan nilai r hitung yang lebih rendah daripada nilai r tabel, yang berarti item-item tersebut tidak memiliki kekuatan yang memadai untuk merefleksikan konstruk yang diukur dan dianggap tidak valid. Penemuan ini mengindikasikan bahwa soal harus direvisi secara komprehensif, baik dari aspek redaksional, konteks pertanyaan, maupun keselarasan dengan indikator pembelajaran.

Validitas yang rendah pada sebagian besar pertanyaan juga dapat menunjukkan adanya kesalahan dalam penyusunan soal,

seperti ketidaksesuaian antara pertanyaan dan indikator, ambiguitas dalam opsi jawaban, atau minimnya variasi tingkat kesulitan.

Validitas isi juga menjadi perhatian dalam studi ini, yaitu kesesuaian antara isi soal dengan kompetensi dasar dan indikator dalam kurikulum Geografi kelas XI. Penyusunan soal telah dilakukan dengan mengacu pada silabus dan kisi-kisi yang tervalidasi oleh ahli Geografi dan pendidik berpengalaman. Dengan demikian, instrumen tidak hanya valid secara statistik, tetapi juga sesuai secara konseptual. Pendekatan ini mendukung temuan Tavakol & Dennick (2011) yang menekankan pentingnya integrasi antara validitas isi dan empiris dalam pengembangan instrumen evaluasi.

Secara keseluruhan, tingkat validitas yang tinggi pada sebagian besar butir soal menunjukkan bahwa instrumen ini telah memenuhi standar kualitas yang diharapkan. Soal-soal yang tidak valid perlu direvisi atau dieliminasi agar tidak mengganggu hasil evaluasi secara keseluruhan. Pemantauan validitas secara berkala menjadi langkah penting dalam peningkatan mutu evaluasi pembelajaran di masa mendatang. Instrumen yang valid memungkinkan guru dan peneliti memperoleh data yang akurat dan dapat dipercaya untuk mengambil keputusan terkait proses dan hasil pembelajaran.

Tabel 1. Perhitungan 5 dari 50 butir soal dengan Uji Validitas

Nama	1	2	3	4	5
Neneng	1	1	1	1	1
Farhan	1	1	1	1	1
Kanza	1	1	1	1	1
Risky Firmansyah	1	1	1	1	1
Henkar	1	1	1	1	1
Yosafat	1	1	1	1	1
Ardidua	1	1	1	1	1
Ridval	1	1	1	1	1
Davin Zidan	1	1	1	1	1
Amanda	1	1	1	1	1
r hitung	0,1913	0,2570	0,6655	0,2401	0,3327
r tabel	0,3388				
Hasil	Tidak Valid	Tidak Valid	Valid	Tidak Valid	Tidak Valid

Uji Reliabilitas

Hasil Pengujian Reliabilitas terhadap soal ini dengan menggunakan rumus Kuder Richardson Formula 20 (KR-20), yang sesuai untuk instrumen pilihan ganda dengan jawaban dikotomis (benar atau salah). Hasil penghitungan menunjukkan bahwa instrumen memiliki nilai reliabilitas sebesar 0,80. Menurut interpretasi dari Arikunto (2013), nilai tersebut termasuk dalam kategori tinggi, yang berarti bahwa soal-soal dalam instrumen saling berkorelasi kuat dalam mengukur konstruksi yang sama.

Tingginya nilai reliabilitas menunjukkan bahwa penyusunan soal telah dilakukan dengan baik, di mana setiap butir soal secara konsisten mengukur pemahaman siswa terhadap konsep mitigasi bencana. Dalam asesmen pendidikan, reliabilitas tinggi sangat penting karena menjamin bahwa hasil evaluasi dapat dipercaya dan bukan merupakan hasil kebetulan. Sebuah instrumen yang reliabel memungkinkan perbandingan antar individu atau antar waktu dapat dilakukan dengan tingkat kesalahan yang rendah. Hal ini menjadi dasar penting dalam pelaksanaan asesmen formatif maupun sumatif. Faktor-faktor yang berkontribusi terhadap reliabilitas tinggi dalam penelitian ini mencakup konsistensi format soal, kesesuaian isi dengan indikator pembelajaran, dan penggunaan platform digital yang meminimalisir intervensi subjektif.

Penggunaan Quizizz sebagai media pengumpulan data memungkinkan kontrol terhadap kondisi pelaksanaan tes, seperti pengacakan soal, waktu pengerjaan, dan validasi jawaban otomatis. Dengan demikian, hasil yang diperoleh menjadi lebih objektif dan dapat dipertanggungjawabkan secara ilmiah. Temuan ini sejalan dengan kajian oleh Kurniawan & Wahyuni (2019) yang menunjukkan bahwa platform digital meningkatkan keandalan

evaluasi. Keberhasilan pengukuran reliabilitas yang tinggi ini juga menunjukkan bahwa instrumen layak digunakan dalam konteks penelitian dan praktik pembelajaran. Namun demikian, pengujian reliabilitas sebaiknya tidak berhenti di satu titik. Evaluasi berkelanjutan perlu dilakukan setiap kali instrumen digunakan dalam konteks yang berbeda atau mengalami revisi. Oleh karena itu, uji reliabilitas menjadi elemen yang tidak terpisahkan dalam siklus pengembangan dan penyempurnaan instrumen evaluasi pendidikan.

Table 2. Perhitungan 5 dari 50 butir soal dengan uji reliabilitas.

Nama	1	2	3	4	5
Neneng	1	1	1	1	1
Farhan	1	1	1	1	1
Kanza	1	1	1	1	1
Risky Firmansyah	1	1	1	1	1
Henkar	1	1	1	1	1
Yosafat	1	1	1	1	1
Ardidua	1	1	1	1	1
Ridval	1	1	1	1	1
Davin Zidan	1	1	1	1	1
Amanda	1	1	1	1	1
JB	30	30	22	29	29
P	0,86	0,86	0,63	0,83	0,83
Q	0,14	0,14	0,37	0,17	0,17
PQ	0,12	0,12	0,23	0,14	0,14
Mt	31,29				
Sdt	10,95				
Mp	33,67	33,83	36,95	33,90	34,17
Rphi/Uji Validitas	0,53	0,57	0,67	0,52	0,58
r tabel	0,3338				
Hasil	Valid	Valid	Valid	Valid	Valid
Varians	0,09	0,09	0,23	0,11	0,11
Jumlah Varians	10,41				
Kr20	0,85				
Kr21	0,84				
CA	0,86				

Nilai KR-20, KR-21, dan Cronbach's Alpha semuanya > 0,80, sehingga instrumen dinyatakan **reliabel dengan kategori tinggi**.

Analisis Tingkat Kesukaran

Berdasarkan hasil pengolahan butir soal ini menggunakan spreadsheet dan diklasifikasikan sesuai kategori tingkat kesukaran terdapat 0,00 - 0,30 dengan tingkat

kesukaran yang sangat mudah, 0,31 - 0,70 dengan tingkat kesukaran sedang, 0,70 - 1,00 dengan tingkat kesukaran mudah, dari total 50 butir soal. Hal ini menunjukkan bahwa sebagian besar butir soal mampu membedakan peserta dengan kemampuan tinggi dan rendah secara efektif.

Dalam penerapan pengembangan instrumen, komposisi ideal untuk soal berdasarkan tingkat kesulitan biasanya disarankan dalam proporsi sekitar 20–30% soal mudah, 40–60% soal sedang, dan 20–30% soal sulit. Berdasarkan analisis yang telah dilakukan, instrumen mitigasi bencana ini tampaknya mengikuti pola distribusi tersebut dengan baik, meskipun diperlukan sedikit penyesuaian pada pertanyaan-pertanyaan ekstrem. Secara keseluruhan, dapat disimpulkan bahwa alat ini memiliki kualitas yang baik dari segi tingkat kesulitan, karena menyajikan pertanyaan dalam berbagai tingkatan yang dapat menggambarkan sebaran kemampuan peserta secara luas dan seimbang. Hal ini berfungsi sebagai dasar krusial dalam proses evaluasi pendidikan yang adil dan objektif, serta sebagai landasan yang kokoh untuk menyimpulkan bahwa alat yang diterapkan dalam penelitian ini pantas digunakan untuk pengumpulan data lebih lanjut atau untuk tujuan pengukuran pembelajaran di masa yang akan datang.

Table 3. Perhitungan 5 dari 50 butir soal dengan perhitungan tingkat kesukaran.

Nama	1	2	3	4	5
Neneng	1	1	1	1	1
Farhan	1	1	1	1	1
Kanza	1	1	1	1	1
Risky Firmansyah	1	1	1	1	1
Henkar	1	1	1	1	1
Yosafat	1	1	1	1	1
Ardidua	1	1	1	1	1
Ridval	1	1	1	1	1
Davin Zidan	1	1	1	1	1
Amanda	1	1	1	1	1
Jumlah Benar	31	32	25	33	34
Jumlah Siswa					
Kesukaran	0,886	0,914	0,714	0,943	0,971
Keterangan	Mudah	Mudah	Mudah	Mudah	Mudah
SUKAR	0,00- 0,30				
SEDANG	0,31-0,70				
MUDAH	0,70-1,00				

Analisis Tingkat Pengecoh

Distraktor atau pengecoh merupakan bagian penting dari soal pilihan ganda yang berfungsi untuk menguji pemahaman siswa secara mendalam. Dalam penelitian ini, setiap soal terdiri atas lima opsi jawaban, dengan satu jawaban benar dan empat pengecoh. Efektivitas distraktor dinilai dari kemampuannya menarik pilihan dari siswa yang belum memahami materi secara utuh. Distraktor yang baik akan dipilih oleh siswa dengan tingkat pemahaman rendah, sedangkan siswa dengan pemahaman tinggi akan mampu mengenali jawaban yang benar. Oleh karena itu, distraktor yang tidak pernah dipilih menjadi indikator bahwa opsi tersebut tidak efektif dan perlu direvisi.

Pada soal 1, opsi B dipilih oleh 30 peserta, menjadikannya pengecoh paling efektif pada soal tersebut. Hal serupa terjadi pada Soal 5 dan Soal 6, di mana opsi A masing-masing dipilih oleh 29 dan 27 peserta. Hal ini menunjukkan bahwa distraktor pada soal-soal ini memiliki daya tarik tinggi dan mampu mengalihkan perhatian peserta dari jawaban yang benar. Pada soal 47 dan 50, pengecoh yang dominan adalah opsi B dan A, yang dipilih oleh 21 dan 24 peserta, maka dari itu adanya kecenderungan peserta untuk tertipu opsi-opsi tertentu.

Distraktor yang tidak berfungsi dengan baik dapat mempengaruhi daya pembeda soal secara keseluruhan. Seperti dijelaskan oleh Burud et al. (2019), kualitas distraktor yang rendah mengurangi kemampuan soal dalam membedakan antara siswa berkemampuan tinggi dan rendah. Ketika distraktor terlalu lemah, siswa dengan pemahaman minimum pun bisa mengeliminasi opsi yang tidak logis, sehingga menjawab soal dengan tebakan yang lebih mudah. Sebaliknya, distraktor yang efektif dapat meningkatkan validitas soal karena mampu menguji ketepatan pemahaman siswa terhadap konsep-konsep penting dalam materi mitigasi bencana.

Dengan demikian, perbaikan distraktor pada soal yang kurang efektif menjadi salah satu langkah strategis untuk meningkatkan kualitas instrumen evaluasi.

Pengembang soal sebaiknya mengacu pada prinsip konstruksi soal yang menekankan logika, kesamaan panjang kalimat, dan kesesuaian konteks jawaban. Kombinasi antara validitas isi dan efektivitas pengecoh menjadikan instrumen tidak hanya mampu mengukur capaian pembelajaran, tetapi juga mendeteksi miskonsepsi siswa secara akurat. Evaluasi ini penting sebagai dasar untuk revisi instrumen di tahap pengembangan selanjutnya.

Table 4. Perhitungan 5 dari 50 butir soal dengan perhitungan tingkat pengecoh.

Nama	1	2	3	4	5
Neneng	B	C	B	C	A
Farhan	B	C	B	C	A
Kanza	B	C	B	C	A
Risky Firman	B	C	B	C	A
Henkar	B	C	B	C	A
Yosafat	B	C	B	C	A
Ardidua	B	C	B	C	A
Ridval	B	C	B	C	A
Davin Zidan	B	C	B	C	A
Amanda	B	C	B	C	A
Pilihan Majemuk	1	2	3	4	5
A	2	1	3	0	29
B	30	0	22	3	1
C	0	30	2	29	3
D	0	2	0	1	0
E	1	0	6	0	0
Pilihan Majemuk	1	2	3	4	5
A	6,06%	3,03%	9,09%	0,00%	87,88%
B	90,91%	0,00%	66,67%	9,09%	3,03%
C	0,00%	90,91%	6,06%	87,88%	9,09%
D	0,00%	6,06%	0,00%	3,03%	0,00%
E	3,03%	0,00%	18,18%	0,00%	0,00%
Pilihan Majemuk	1	2	3	4	5
A	Diterima	Ditolak	Diterima	Ditolak	Diterima
B	Diterima	Ditolak	Diterima	Diterima	Ditolak

Nama	1	2	3	4	5
Neneng	1	1	1	1	1
Farhan	1	1	1	1	1
Kanza	1	1	1	1	1
Risky Firmansyah	1	1	1	1	1
Henkar	1	1	1	1	1
Yosafat	1	1	1	1	1

Nama	1	2	3	4	5
C	Ditolak	Diterima	Diterima	Diterima	Diterima
D	Ditolak	Diterima	Ditolak	Ditolak	Ditolak
E	Ditolak	Ditolak	Diterima	Ditolak	Ditolak

Analisis Daya Pembeda

Hasil dari analisis butir soal ini diketahui Daya Pembeda untuk mengukur sejauh mana kemampuan peserta didik yang memiliki kemampuan tinggi dan mereka yang memiliki kemampuan rendah berdasarkan kinerja mereka terhadap soal tersebut. Tujuan utama dari analisis ini adalah untuk memastikan bahwa soal yang digunakan dalam tes bersifat adil, valid, dan mampu mengukur capaian belajar secara akurat. Berdasarkan data pada tabel diketahui sebanyak 12 soal mempunyai daya pembeda pada kategori sangat baik yang artinya soal tersebut sangat efektif dalam memilah kemampuan peserta. Selain itu, terdapat 3 pertanyaan yang termasuk dalam kategori baik, dan 17 pertanyaan dalam kategori cukup, terdapat pula 17 soal yang dikategorikan buruk, Namun ada pula terdapat 1 pertanyaan yang sangat buruk, faktor tersebut disebabkan oleh rendahnya daya pembeda antara lain adalah redaksi soal yang ambigu, pengecoh yang terlalu kuat, atau ketidaksesuaian antara soal dengan indikator kompetensi. Dengan demikian, soal-soal yang memiliki daya pembeda rendah perlu direvisi atau bahkan diganti. Sebaliknya, soal-soal dengan daya pembeda tinggi dapat dipertahankan atau dijadikan acuan dalam penyusunan soal selanjutnya. Keseluruhan hasil ini menekankan pentingnya uji validitas butir soal sebagai bagian dari proses evaluasi pembelajaran yang bertanggung jawab dan berbasis data.

Tabel 5. Perhitungan 5 dari 50 butir soal dengan perhitungan daya beda

Nama	1	2	3	4	5
Ardidua	1	1	1	1	1
Ridval	1	1	1	1	1
Davin Zidan	1	1	1	1	1
Amanda	1	1	1	1	1
Jumlah Benar (BA)	18	18	17	17	17

Nama	1	2	3	4	5
Pembagian Atas (BA/JA)	1,00 00	1,000 0	0,9444 4	0,9444 4	0,9444 4
Jumlah Benar (BB)	12	12	5	12	12
Pembagian Bawah (BB/JB)	0,80 00	0,800 0	0,333	0,8000	0,8000
D	0,20 00	0,200 0	0,6111	0,1444	0,1444
Kriteria	C	C	BS	C	C
Keterangan	B	Buru k			17
	C	Cuku p			17
	BK	Baik			3
	BS	Baik Sekal i			12
	SB	Sanga t Buru k			1

SIMPULAN

Artikel ini berfokus pada efektivitas penggunaan pendekatan sistematis untuk mempelajari efektivitas berbagai metode penelitian. Metode penelitian melibatkan analisis komprehensif terhadap data yang dikumpulkan, termasuk uji validitas, uji reliabilitas, dan analisis deskriptif. Uji validitas dilakukan menggunakan tabel Pearson Product Moment (r), dengan hasil yang membandingkan reliabilitas setiap tes dengan tingkat signifikansi. Uji validitas untuk metode penelitian yang diusulkan menunjukkan bahwa semua metode yang diuji memenuhi kriteria validitas.

Validitas ditentukan oleh hubungan antara setiap pertanyaan tes dan total skor, dengan tingkat signifikansi $r = 0,3338$. Namun, beberapa tes memiliki tingkat signifikansi yang lebih tinggi daripada yang lain, menunjukkan bahwa tes tersebut harus ditafsirkan secara komprehensif, dengan mempertimbangkan baik aspek teoritis maupun temuan empiris.

Uji validitas untuk metode yang diusulkan juga menunjukkan bahwa uji tersebut memiliki tingkat kualitas yang tinggi. Tes yang tidak valid harus diinterpretasikan atau dihilangkan untuk menghindari evaluasi yang negatif. Tes yang valid dapat digunakan untuk meningkatkan validitas proses evaluasi. Uji reliabilitas untuk metode yang diusulkan menggunakan Kuder Richardson Formula 20

(KR-20), yang cocok untuk metode deskriptif. Hasil uji reliabilitas menunjukkan bahwa instrumen memiliki tingkat reliabilitas sebesar 0,80, yang menunjukkan bahwa instrumen tersebut dapat diandalkan dalam konstruksinya.

Kesimpulannya, studi ini menunjukkan pentingnya menggunakan metode sistematis untuk mempelajari efektivitas berbagai metode penelitian. Dengan fokus pada metode yang valid dan memastikan keabsahan hasil, para peneliti dapat membuat keputusan yang tepat tentang efektivitas metode yang mereka pilih. Pada validitas bahan ajar bahasa untuk menentukan efektivitas metode pengajaran bagi siswa dengan kemampuan yang berbeda. Hasil penelitian menunjukkan bahwa banyak bahan ajar bahasa memiliki validitas yang baik, tetapi ada juga beberapa yang terlalu samar, sulit dipahami, atau kurang memiliki indikator komunikasi yang jelas. Oleh karena itu, validitas bahan ajar bahasa sangat penting dalam proses evaluasi berdasarkan data.

REKOMENDASI

Berdasarkan hasil analisis dan pembahasan dalam jurnal ini, penulis merekomendasikan beberapa hal berikut :

1. Pengembangan Penelitian Lanjutan: Diperlukan studi lebih lanjut dengan cakupan wilayah yang lebih luas untuk memperoleh pemahaman yang lebih komprehensif mengenai fenomena yang dibahas.
2. Peningkatan Kolaborasi Antar Lembaga: Kerja sama antara institusi pendidikan, pemerintah, dan masyarakat lokal sangat penting untuk implementasi rekomendasi yang dihasilkan dari penelitian ini.
3. Pemanfaatan Teknologi: Penggunaan teknologi geospasial dan sistem informasi geografis (SIG) dapat meningkatkan akurasi dan efisiensi dalam penelitian geografi di masa mendatang.

Diharapkan rekomendasi ini dapat menjadi pertimbangan bagi para peneliti, praktisi, dan pembuat kebijakan dalam mengambil langkah-langkah strategis yang relevan.

DAFTAR PUSTAKA

- Putra, R. A., & Wulandari, S. (2022). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif dan Kualitatif*. Yogyakarta: Deepublish.
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan* (2nd ed.). Jakarta: Bumi Aksara.
- Abd Rahman, B. P., Munandar, S. A., Fitriani, A., Karlina, Y., & Yumriani, Y. (2022). Pengertian pendidikan, ilmu pendidikan dan unsur-unsur pendidikan. *Al-Urwatul Wutsqa: Kajian Pendidikan Islam*, 2(1), 1-8.
- Abraham, I., Tjalla, A., & Indrajit, R. E. (2021). HOTS (High Order Thingking Skill) dalam Paedagogik Kritis. *JISIP (Jurnal Ilmu Sosial dan Pendidikan)*, 5(3).
- Hidayah, N., & Santosa, B. (2021). "Analisis Butir Soal Pilihan Ganda dalam Evaluasi Pembelajaran Geografi Menggunakan Indeks Kesukaran dan Daya Pembeda." *Jurnal Pendidikan Geografi*, 26(1), 15-25.
- 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://doi.org/10.24252/asma.v2i2.16569>
- Sari, D. P., & Hartono, R. (2020). "Uji Validitas dan Reliabilitas Instrumen Penilaian dalam Penelitian Pendidikan." *Jurnal Pendidikan dan Pembelajaran*, 7(2), 89-98.
- Kurniawan, F., & Wahyuni, S. (2019). "Penggunaan Platform Digital dalam Pengumpulan Data Pendidikan: Studi Kasus di SMA Negeri." *Jurnal Teknologi Pendidikan*, 11(3), 145-156.
- Utami, N. M., & Prasetyo, B. (2023). "Efektivitas Penggunaan Quizizz sebagai Media Evaluasi Pembelajaran Interaktif di Sekolah Menengah." *Jurnal Teknologi dan Media Pendidikan*, 15(1), 34-44.
- Arikunto, S. (1983). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan* (2nd ed.). Jakarta: Bumi Aksara.