

## **Analisis Kualitas Butir Soal Asesmen Sumatif Geografi pada Materi Peta, Penginderaan Jauh, dan Sistem Informasi Geografis**

**Hadar Rafliansyah Rivaldi<sup>1,\*</sup>, Adip Wahyudi<sup>1</sup>, dan Ratna Arum Nur Indah Sari<sup>2</sup>**

<sup>1</sup> Pendidikan Geografi, Universitas Negeri Malang, Malang, Indonesia

<sup>2</sup> Geografi, SMA Negeri 1 Lawang, Malang, Indonesia

\* Email: [hadar.rafliansyah.2207216@students.um.ac.id](mailto:hadar.rafliansyah.2207216@students.um.ac.id)

### **Abstrak**

*Asesmen sumatif berfungsi untuk mengukur capaian siswa setelah menyelesaikan suatu materi pembelajaran. Diperlukan analisis butir soal untuk memastikan instrumen berkualitas. Penelitian ini bertujuan mengidentifikasi kualitas butir soal pada asesmen sumatif materi peta, penginderaan jauh, dan sistem informasi geografis. Penelitian ini termasuk jenis kuantitatif deskriptif yang dipilih dengan teknik purposive sampling pada 34 siswa kelas X di salah satu SMA Negeri di Malang. Data diperoleh melalui studi dokumentasi instrumen asesmen dan jawaban siswa, teknik analisis menggunakan uji validitas, reliabilitas, indeks kesukaran, daya beda, dan efektivitas distraktor. Hasil menunjukkan 84% butir soal valid dan reliabilitas 0,76 (reliabel). Tingkat kesukaran soal didominasi kategori sedang 88%. Daya beda butir soal optimal sebesar 88% item, sedangkan efektivitas distraktor berfungsi pada 68%. Profil pemahaman siswa menunjukkan bahwa materi peta memiliki rata-rata dan nilai tertinggi dibandingkan penginderaan jauh dan SIG. Secara keseluruhan, asesmen sumatif geografi telah memenuhi standar kualitas instrumen meskipun diperlukan revisi pada butir soal yang bermasalah untuk meningkatkan akurasi penilaian dimasa mendatang. Penelitian ini memberikan kerangka kuantitatif yang dapat diadopsi oleh guru untuk meninjau dan merevisi butir soal, serta mendorong penerapan analisis butir soal secara rutin guna meningkatkan efektivitas penilaian pembelajaran geografi.*

**Kata kunci:** Analisis Butir Soal, Asesmen Sumatif, Pembelajaran Geografi

## ***Analysis of the Quality of Summative Assessment Questions in Geography on Maps, Remote Sensing, and Geographic Information Systems***

### **Abstract**

*Summative assessments serve to measure student achievement after completing a learning subject. Item analysis is required to ensure the quality of the instrument. This study aims to identify the quality of items in summative assessments of maps, remote sensing, and geographic information systems. This study is a descriptive quantitative study using purposive sampling on 34 students in class X in one of the public high schools in Malang. Data were obtained through document analysis of assessment instruments and student responses, with analysis techniques including validity tests, reliability tests, difficulty indices, discrimination indices, and distractor effectiveness. The results showed that 84% of the items were valid, with a reliability of 0.76 (reliable). The difficulty level of the items was dominated by the moderate category at 88%. The discrimination power of the items was optimal at 88%, while the effectiveness of distractors functioned at 68%. The students' understanding profile indicated that the map material had the highest average and score compared to remote sensing and GIS. Overall, the summative assessment of geography has met the quality standards for assessment instruments, although revisions are needed for problematic items to improve the accuracy of future assessments. This study provides a quantitative framework that teachers can adopt to review and revise items, and encourages the routine application of item analysis to enhance the effectiveness of geography learning assessments.*

**Keywords:** Item Analysis, Summative Assessment, Geography Learning

**Histori Naskah**

Diserahkan: 16 Juni 2025

Direvisi: 5 September 2025

Diterima: 8 September 2025

**How to cite:**

Rivaldi, H.R., dkk. (2025). Analisis Kualitas Butir Soal Asesmen Sumatif Geografi pada Materi Peta, Penginderaan Jauh dan Sistem Informasi Geografis. *Jurnal Ilmu Pendidikan dan Pembelajaran*, 4(1), 102-113. DOI: <https://doi.org/10.58706/jipp.v4n1.p102-113>.

**PENDAHULUAN**

Berlandaskan pada Permendikbudristek nomor 21 tahun 2022 yang mengatur standar penilaian pendidikan untuk tingkat sekolah menengah, penilaian merupakan serangkaian proses pengakumulasian serta pengolahan informasi untuk mengidentifikasi sejauh mana ketercapaian hasil belajar siswa yang akan menjadi acuan dalam kegiatan evaluasi pembelajaran (Rahmah & Cahyadi, 2024). Evaluasi pembelajaran merupakan unsur penting untuk mengetahui keberhasilan atau kegagalan dari suatu sistem pembelajaran yang telah diterapkan oleh guru (Sodikin & Gumiandari, 2022). Evaluasi pembelajaran berfungsi untuk mengukur tingkat pemahaman siswa dan sarana untuk menilai efektivitas metode dan kegiatan pembelajaran yang diterapkan oleh guru pada periode tertentu. Dilaksanakan pada akhir unit materi untuk mengetahui capaian tujuan pembelajaran dan pemahaman materi secara keseluruhan (Putri & Rosliyah, 2020). Dengan ini, guru mampu melihat celah lalu mengatasinya dengan intervensi dan modifikasi model pembelajaran.

Sejalan dengan pentingnya evaluasi pembelajaran, asesmen sumatif menjadi salah satu jenis asesmen yang wajib dilakukan guru setelah mempelajari satu rangkaian materi (Mangla et al., 2023). Asesmen sumatif merupakan proses penilaian yang dilaksanakan pada akhir periode atau unit pembelajaran, berfungsi untuk mengukur tingkat capaian siswa terhadap tujuan pembelajaran yang telah ditentukan guru (Darwin et al., 2023). Melalui asesmen sumatif, guru dapat mengevaluasi hasil belajar siswa, memberikan nilai, dan mengambil keputusan terkait kemajuan akademik siswa. Hardianti et al. (2023) menyebutkan bahwa asesmen sumatif berperan dalam mengumpulkan data terkait penguasaan materi siswa setelah kegiatan pembelajaran selesai dilaksanakan, sehingga sanggup menjadi alat penting untuk mendukung keberhasilan pendidikan. Hal tersebut menjadi penting karena dapat memberikan evaluasi program pembelajaran, dan memungkinkan guru untuk menilai efektivitas metode pengajaran berdasarkan pencapaian siswa terhadap tujuan pembelajaran.

Tingkat keberhasilan pembelajaran yang diimplementasikan guru dalam suatu materi pelajaran selama periode tertentu dapat diketahui dari hasil belajar yang didapatkan siswa (Fachruddin et al., 2022). Hasil belajar siswa dapat menginterpretasikan bagaimana metode pembelajaran yang telah diterapkan, jika menghasilkan hasil yang tinggi, maka pembelajaran dapat dikategorikan berhasil (Somayana, 2020). Memberikan keputusan terhadap pembelajaran dari nilai siswa tanpa memperhatikan kualitas butir soal yang digunakan dapat menafsirkan gambaran yang salah tentang kualitas proses belajar-mengajar yang telah berlangsung. Untuk memastikan bahwa hasil belajar tersebut benar-benar mencerminkan kemampuan siswa secara objektif, diperlukan butir soal yang sesuai dengan *classical test theory* yang terdiri dari validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas distaktor.

Penelitian ini menggunakan Teori Tes Klasik (*Classical Test Theory/CTT*) sebagai landasan teoretis utama dalam menganalisis kualitas butir soal asesmen sumatif geografi. Teori Tes Klasik, yang dikembangkan dan dipopulerkan oleh Arikunto (2018), merupakan pendekatan psikometrik yang berfokus pada hubungan antara skor sebenarnya (*true score*) dan skor yang diamati (*observed score*) untuk menjelaskan kemampuan peserta tes. Teori ini dipilih karena kemudahannya serta dapat dilakukan dengan sampel yang relatif kecil. Selain itu, CTT sangat cocok untuk analisis asesmen sumatif karena memberikan informasi praktis tentang kualitas instrumen yang dapat langsung digunakan guru untuk perbaikan pembelajaran (Ajmi et al., 2024). Pemilihan CTT sebagai kerangka teoretis juga didukung oleh karakteristik materi yang membutuhkan pemahaman spasial dan kemampuan berpikir kritis, dimana pendekatan klasik dapat memberikan informasi yang memadai untuk evaluasi kualitas instrumen asesmen pada tingkat sekolah menengah atas.

Kualitas butir soal sangat penting dalam menentukan sejauh mana tes tersebut mampu memberikan gambaran akurat tentang kemampuan siswa (Farida & Musyarofah, 2021). Proses analisis kualitas butir soal menjadi tahap krusial untuk membuktikan bahwa butir soal telah memenuhi ukuran ideal yang diperlukan dalam kegiatan asesmen sumatif. Butir soal yang tidak teruji akan menghasilkan instrumen asesmen yang lemah, sehingga penilaian kompetensi siswa yang diperoleh tidak akurat (Arwansyah et al., 2022). Keberhasilan pembelajaran diukur dari capaian hasil belajar siswa di akhir unit materi melalui penilaian yang objektif oleh guru (Saputra et al., 2022). Dengan butir soal yang bermutu, guru dapat menentukan tingkat

penguasaan materi siswa secara lebih tepat dan menyeluruh sesuai dengan tujuan pembelajaran yang telah ditentukan.

Sebagai salah satu mata pelajaran penting pada jenjang sekolah menengah atas, geografi membutuhkan asesmen sumatif yang berkualitas untuk mengevaluasi capaian pembelajaran pada karakteristik uniknya (Sözcü et al., 2023). Pada materi peta, penginderaan jauh (PJ), dan sistem informasi geografis (SIG), asesmen sumatif harus dirancang untuk meringkas capaian tujuan pembelajaran, yang meliputi: menjelaskan dasar-dasar pemetaan, penginderaan jauh, dan SIG; mengidentifikasi komponen-komponen pemetaan, penginderaan jauh, dan SIG; serta menguraikan pemanfaatan dalam berbagai bidang. Materi-materi ini menuntut siswa memiliki literasi spasial serta kemampuan analisis keruangan yang mendalam, sehingga harus mampu mengintegrasikan pengukuran kompetensi spasial siswa secara objektif (Xiang & Liu, 2017). Dengan asesmen sumatif yang dirancang secara efektif dan memenuhi kelayakan analisis butir soal, guru dapat menjamin penetapan nilai akhir yang mencerminkan capaian belajar siswa secara akurat dan dapat dijadikan dasar evaluasi program pembelajaran secara menyeluruh (Wahyudi et al., 2024).

Meskipun demikian, temuan di lapangan menunjukkan bahwa guru cenderung mengabaikan analisis butir soal karena tidak adanya tuntutan formal dan memerlukan waktu yang cukup banyak. Lusiyanti & Laksono (2021) juga menyampaikan bahwa guru mengaku memiliki waktu yang terbatas dan tidak adanya tuntutan dari instansi sehingga analisis butir soal sering kali tidak dilakukan. Hasil wawancara Savika et al. (2025) juga mengungkapkan bahwa guru tidak melakukan dikarenakan padatnya kegiatan diluar pembelajaran. Teig et al. (2019) menyampaikan bahwa "*time pressure*" dapat mengurangi kemauan guru melakukan analisis butir soal. Kondisi ini mengindikasikan bahwa pelaksanaan asesmen sumatif tidak sesuai dengan standar evaluasi, terbukti ketika dilakukan analisis asesmen sumatif pada materi sebelumnya berjumlah 25 soal dan mendapatkan hasil 76% soal tidak valid serta tidak reliabel, sehingga analisis tidak dilanjutkan. Hasil penelitian (Amalia et al., 2019) menemukan bahwa mayoritas soal yang dianalisis tidak valid, dengan 67,7% memiliki daya pembeda rendah dan 33% soal tergolong sangat sukar. Temuan serupa pada penelitian Jati et al. (2023) dimana dari 30 butir soal yang diuji, hanya 30% termasuk kategori valid, sementara daya pembeda mayoritas berada pada tingkatan lemah hingga baik, tanpa ada yang mencapai kategori kuat. Selain itu, distraktor pada soal-soal tersebut tidak berfungsi optimal akibat pilihan jawaban yang tidak logis atau kurang masuk akal. Penelitian Logayah et al. (2024) menyampaikan bahwa analisis butir soal secara efektif mampu mengidentifikasi soal bermutu tinggi dan rendah, sehingga hal ini seharusnya dilakukan oleh seluruh guru ketika selesai melaksanakan asesmen.

Meskipun berbagai studi telah melakukan analisis butir soal, terdapat beberapa kekurangan, khususnya pada submateri peta, penginderaan jauh, dan SIG. Penelitian sebelumnya sering kali hanya fokus pada indikator analisis butir soal tanpa mengeksplorasi profil pemahaman siswa secara mendalam. Selain itu, penelitian-penelitian terdahulu cenderung mengabaikan efektivitas distraktor yang kurang optimal, sehingga distribusi jawaban tidak sepenuhnya mencerminkan perbedaan kemampuan siswa. Lebih jauh lagi, sebagian besar literatur yang ada belum mengkaji literasi spasial yang diperlukan dalam memahami materi peta, penginderaan jauh dan SIG, sehingga pengukuran kompetensi siswa dalam bidang tersebut menjadi kurang akurat. Oleh karena itu, penelitian ini akan mengisi kekosongan tersebut dengan menambahkan analisis profil pemahaman siswa dan memberikan rekomendasi perbaikan, sehingga diperoleh gambaran yang lebih utuh tentang kualitas asesmen sumatif geografi.

Penelitian ini bertujuan untuk meningkatkan kualitas asesmen sumatif geografi dengan melakukan analisis butir soal pada asesmen sumatif materi peta, penginderaan jauh (PJ), dan sistem informasi geografis (SIG) di kelas X SMAN 1 Lawang. Analisis direalisasikan dengan mengidentifikasi validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas distraktor dari instrumen asesmen sumatif tersebut. Penelitian ini diharapkan mampu berkontribusi dengan memberikan saran perbaikan instrumen asesmen sumatif sekaligus menjadi acuan bagi guru dalam menyusun ataupun memperbaiki asesmen untuk menunjang proses pembelajaran geografi yang lebih adaptif dan bermakna.

## METODE PENELITIAN

Jenis penelitian yang digunakan adalah kuantitatif deskriptif melalui analisis terhadap butir soal pilihan ganda untuk mengidentifikasi kualitas asesmen sumatif pada materi peta, PJ, dan SIG. Dengan sebaran materi peta terletak pada soal nomor 1 hingga 9, PJ nomor 10 hingga 17, dan SIG nomor 18 hingga 25. Subjek penelitian melibatkan 34 siswa dari kelas X.10 SMAN 1 Lawang yang dipilih melalui teknik *puposive sampling*, berdasarkan pertimbangan: (1) keterwakilan populasi heterogen, (2) kesiapan guru kolaborator, dan (3) ketersediaan dokumen asesmen lengkap. Data dikumpulkan melalui studi dokumentasi terhadap arsip soal

asesmen, kunci jawaban, serta hasil pengerjaan siswa. Instrumen penelitian berupa soal pilihan ganda berjumlah 25 soal. Prosedur penelitian diawali dengan penyusunan butir soal sesuai dengan tujuan pembelajaran, dilakukan validasi, diimplementasikan ke siswa, hasil jawaban dihimpun, dan diakhiri dengan analisis butir soal beserta interpretasinya. Analisis diawali dengan pengurutan nilai peserta dari nilai tertinggi hingga nilai terendah, dilanjutkan dengan uji validitas menggunakan IBM SPSS Statistics 22 melalui Korelasi Pearson. Kriteria validitas ditetapkan dengan membandingkan nilai  $r$  hitung terhadap  $r$  tabel pada tingkat signifikansi 5%, di mana butir soal dinyatakan valid jika  $r$  hitung  $>$   $r$  tabel (0,339). Reliabilitas diukur dengan koefisien Cronbach's Alpha dengan batas minimal 0,70 untuk kategori reliabel. Indeks kesukaran butir soal dihitung menggunakan Persamaan (1):

$$P = \frac{Np}{N} \quad (1)$$

Keterangan:

$Np$  = Jumlah siswa menjawab benar

$N$  = Jumlah total siswa

Klasifikasi tingkat kesukaran butir soal dibagi menjadi tiga kategori, yaitu: sukar (0,00-0,29), sedang (0,30-0,69), dan mudah (0,70-1,00). Selanjutnya, kemampuan daya beda (diskriminasi) butir soal diukur menggunakan Persamaan (2):

$$D = \left( \frac{Ba}{Ja} \right) - \left( \frac{Bb}{Jb} \right) \quad (2)$$

Keterangan:

$Ba$  = Jumlah kelompok atas menjawab benar

$Ja$  = Jumlah siswa kelompok atas

$Bb$  = Jumlah kelompok bawah menjawab benar

$Jb$  = Jumlah kelompok bawah

Klasifikasi daya beda dibagi menjadi empat kategori, yaitu: kurang (0,00-0,19), cukup baik (0,20-0,39), baik (0,40-0,70), dan sangat baik (0,71-1,00). Selanjutnya, pengukuran efektivitas distraktor dilakukan melalui persamaan (3):

$$D = \frac{A}{N} \times 100\% \quad (3)$$

Keterangan:

$A$  = Jumlah siswa pemilih opsi jawaban

$N$  = Jumlah seluruh siswa

Sebuah distraktor dikategorikan berfungsi jika dipilih  $\geq 5\%$  siswa dan lebih dominan pada kelompok bawah. Klasifikasi dibagi menjadi 5 kategori: 4 distraktor berfungsi (sangat baik), 3 distraktor berfungsi (baik), 2 distraktor berfungsi (cukup baik), 1 distraktor berfungsi (kurang baik), 0 distraktor berfungsi (tidak baik). Setelah melakukan analisis, setiap butir soal diinterpretasikan melalui sistem klasifikasi tiga strata: layak pakai (memenuhi 4 kriteria), direvisi (memenuhi 3 kriteria), dan ditolak ( $\leq 2$  kriteria).

## HASIL DAN PEMBAHASAN

Asesmen sumatif geografi pada materi peta, penginderaan jauh, dan SIG di implementasikan ketika kelas X.10 SMAN 1 Lawang selesai mempelajari materi tersebut. Hasil jawaban siswa didapatkan dari kegiatan asesmen, selanjutnya dilakukan analisis untuk mengukur kualitas asesmen melalui analisis butir soal, meliputi: uji validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas distraktor (Arikunto, 2018). Data hasil pengerjaan siswa disusun dalam tabel menggunakan aplikasi *Microsoft Excel*, diurutkan dari siswa dengan total jawaban benar terbanyak di posisi pertama hingga siswa dengan total jawaban benar paling sedikit di posisi terakhir. Selanjutnya data hasil jawaban siswa dianalisis untuk mendapatkan validitas butir soal yang dapat diketahui pada Tabel 1.

**Tabel 1.** Validitas Butir Soal

No.	Kategori	Butir soal	Jumlah	Presentase
1	≥ 0,339 (Valid)	1,2,3,4,5,6,7,8,9,10,11,12,14,15,17,18,19,20,21,22,24	21	84%
2	< 0,339 (Tidak Valid)	13,16,23,25	4	16%

Berdasarkan Tabel 1, sebesar 84% butir soal terbukti valid karena memperlihatkan korelasi positif. Tingkat validitas mencerminkan sejauh mana hasil yang diperoleh konsisten dengan karakteristik tujuan pembelajaran yang diukur, tanpa adanya penyimpangan yang signifikan dari konsep yang mendasarinya (Sugiharni, 2023). Hasil ini memperlihatkan bahwa penerapan asesmen sumatif yang telah dilaksanakan mendapatkan kualitas yang baik dari segi validitas instrumen. Dengan 84% butir soal valid, instrumen ini dapat memberikan pengukuran yang objektif terhadap pencapaian hasil belajar siswa pada akhir periode atau unit pembelajaran. Soal yang valid telah memenuhi syarat sebagai alat asesmen yang dapat diandalkan untuk menentukan tingkat penguasaan kompetensi siswa secara menyeluruh. Temuan ini didukung oleh Brownlie et al. (2024) yang menyampaikan bahwa semakin tinggi validitas semakin baik pula data yang dihasilkan, sehingga hasil asesmen dapat dipercaya. Begitupun sebaliknya, butir soal yang tidak valid harus ditinjau ulang dan diperbaiki guna memastikan keakuratan asesmen sumatif dalam menetapkan hasil belajar siswa. Hasil ini membuktikan bahwa instrumen asesmen sumatif yang digunakan sebagian besar telah memenuhi standar validitas, sehingga dapat menjadi dasar penilaian akademik yang kredibel untuk pemetaan kompetensi siswa secara internal di sekolah.

Setelah proses validitas, tahap berikutnya adalah menguji reliabilitas butir soal dengan menggunakan butir soal valid. Butir soal tidak valid harus dikeluarkan dari analisis untuk memastikan instrumen asesmen memiliki konsistensi internal yang tinggi dan mampu mengukur konstruk penelitian secara andal. Menurut Arikunto (2018) suatu instrumen disebut reliabel jika memiliki nilai reliabilitas memadai dengan nilai Cronbach’s Alpha melebihi 0,70. Dalam penelitian ini, capaian Cronbach’s Alpha yang didapatkan sebesar 0,76. Nilai tersebut menggambarkan bahwa instrumen asesmen memiliki tingkat konsistensi yang baik, sehingga hasil pengukuran dapat diandalkan untuk merepresentasikan kemampuan siswa secara stabil. Konsistensi ini menjamin bahwa jika instrumen diterapkan kembali pada kondisi yang serupa, hasil pengukuran akan cenderung stabil dan tidak mengalami perubahan signifikan (Erfan et al., 2020). Sehingga, guru geografi dapat memanfaatkan instrumen asesmen ini untuk mengukur pemahaman siswa secara berulang dengan tingkat kepercayaan yang tinggi.

Tingkat kesukaran soal merupakan parameter untuk mengklasifikasikan butir soal termasuk sukar, sedang, atau mudah. Dalam penyusunan soal asesmen, guru harus mempertimbangkan tingkat kesukaran ini agar soal tidak terlalu sulit maupun terlalu mudah (Bano et al., 2022). Soal yang terlalu sukar dapat mempersulit siswa untuk memahami dan mengerjakan soal, sedangkan soal terlalu mudah kurang memotivasi siswa untuk berpikir lebih mendalam. Oleh karena itu, keseimbangan tingkat kesukaran soal sangat penting guna menciptakan asesmen sumatif yang optimal. Hasil uji tingkat kesukaran butir soal dapat diketahui pada Tabel 2.

**Tabel 2.** Tingkat Kesukaran Butir Soal

No.	Kategori	Butir soal	Jumlah	Presentase
1.	0,00 - 0,29 (Sukar)	15,16,24	3	12%
2.	0,30 - 0,69 (Sedang)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,19,20,21,22,23,25	22	88%
3.	0,70 - 1,00 (Mudah)	-	-	-

Berdasarkan Tabel 2, dapat diketahui bahwa soal-soal tersebut telah dirancang dengan baik, searah dengan tujuan utama untuk mengukur atau mengetahui kemampuan siswa secara komprehensif. Menurut Arikunto (2018), distribusi tingkat kesukaran soal yang ideal seharusnya memiliki proporsi seimbang antara soal mudah (30%), sedang (40%), dan sukar (30%). Namun, proporsi kesukaran soal dapat disesuaikan dengan tujuan asesmen yang ingin dicapai. Dalam konteks asesmen sumatif, dominasi soal kategori sedang dapat dibenarkan untuk memperoleh gambaran capaian pembelajaran secara umum. Komposisi soal yang didominasi oleh tingkat kesukaran “sedang” sebesar 88% menunjukkan bahwa instrumen asesmen telah dikembangkan dengan pendekatan yang berorientasi pada kemampuan rata-rata siswa. Temuan ini didukung oleh Hidayati et al. (2022) bahwa soal-soal dengan kategori “sedang” mampu memfasilitasi mayoritas siswa dan memberikan diferensiasi yang cukup untuk mengidentifikasi tingkat pemahaman siswa. Pernyataan tersebut juga didukung

oleh Arifin (2017) bahwa soal dengan kesukaran “sedang” terbukti lebih efisien dalam mengidentifikasi perbedaan kompetensi antara siswa yang memiliki pemahaman mendalam dan siswa yang memerlukan penguatan materi.

Penelitian Basileo et al. (2024) menyampaikan bahwa ketiadaan soal kategori mudah dalam asesmen sumatif perlu menjadi perhatian khusus. Soal kategori mudah memiliki fungsi psikologis untuk membangun kepercayaan diri siswa di awal tes dan memberi kesempatan bagi siswa dengan kemampuan rendah untuk tetap dapat menyelesaikan sebagian soal. Dengan demikian, untuk menyempurnakan instrumen asesmen, perlu dilakukan revisi dengan menambahkan beberapa soal kategori mudah, agar dapat digunakan dan berfungsi secara optimal dalam mengukur hasil belajar siswa pada berbagai tingkat kemampuan. Sedangkan soal kategori sukar perlu ditinjau ulang bukan untuk dieliminasi, melainkan untuk memastikan bahwa tingkat kesulitannya masih dalam jangkauan kemampuan siswa serta relevan dengan materi yang telah diajarkan (Pinaka & Apri, 2024). Soal sukar tetap memiliki daya diskriminasi yang baik, artinya masih mampu mengidentifikasi perbedaan kompetensi antar siswa.

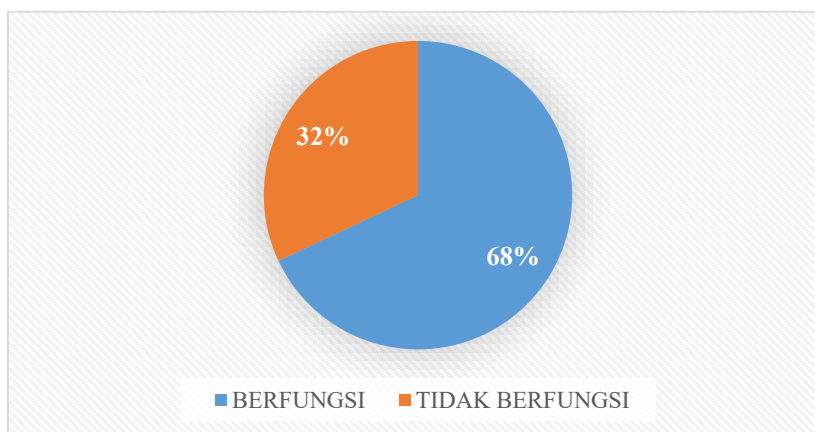
Daya pembeda (diskriminasi) butir soal merupakan indikator untuk mengetahui sejauh mana butir soal mampu membedakan kompetensi akademik antara siswa berkemampuan tinggi dan rendah. Efektivitas soal dalam mengukur pemahaman konsep pembelajaran tercermin dari besarnya nilai daya pembeda yang dihasilkan. Semakin tinggi nilainya, semakin baik kemampuan butir soal tersebut dalam mengidentifikasi perbedaan kompetensi siswa. Soal yang ideal ditandai dengan kecenderungan dijawab benar oleh kelompok atas dan salah oleh kelompok bawah (Bano et al., 2022). Indikator ini menjadikan butir soal sebagai alat ukur yang presisi dalam membedakan tingkat penguasaan materi. Sebaliknya, soal yang dijawab benar oleh kedua kelompok diduga memiliki indeks diskriminatif rendah, karena ketidakmampuannya menyeleksi kemampuan siswa (Sarfraz et al., 2021). Kondisi ini mengindikasikan bahwa soal tersebut gagal menjadi alat asesmen yang efektif, sehingga tidak memenuhi standar evaluasi yang bermutu. Dengan demikian, daya pembeda menjadi tolok ukur penting dalam merancang instrumen asesmen yang mampu menunjukkan perbedaan kemampuan siswa. Hasil analisis daya pembeda butir soal dapat disimak pada Tabel 3.

**Tabel 3.** Daya Pembeda Butir Soal

No.	Kategori	Butir soal	Jumlah	Presentase
1.	0,71 - 1,00 (Sangat Baik)	-	-	-
2.	0,41 - 0,70 (Baik)	2,5,6,7,9,12,14,15,19,23,24	11	44%
3.	0,21 - 0,40 (Sedang)	1,3,4,8,10,11,17,18,20,21,22	11	44%
4.	0,00 - 0,20 (Lemah)	13,16,25	3	12%

Berdasarkan Tabel 3, analisis daya pembeda memperlihatkan variasi yang memadai. Sebanyak 88% butir soal berhasil membedakan tingkat kompetensi siswa dengan optimal, sehingga mampu mengidentifikasi perbedaan pemahaman antar siswa. Butir soal tersebut dapat langsung digunakan kembali pada asesmen mendatang. Keputusan tersebut didukung oleh Nafs et al. (2023) menyampaikan bahwa butir soal dalam kategori sedang hingga baik dapat disimpan untuk asesmen sumatif berikutnya tanpa melakukan perbaikan. Namun, sebesar 12% butir soal termasuk kategori lemah perlu diketahui penyebabnya dan diperbaiki agar mampu membedakan siswa yang menguasai materi secara mendalam dan yang memerlukan penguatan (Hamimi et al., 2020). Daya beda berkorelasi erat dengan validitas soal, perbaikan terhadap butir soal lemah dapat meningkatkan validitas butir soal. Asesmen sumatif ini memberikan bukti kongkret hasil belajar siswa yang dapat menjadi dasar penetapan nilai akhir dan keputusan kenaikan kelas. Dengan memperbaiki soal yang termasuk kategori lemah melalui revisi dan pengujian ulang, asesmen sumatif akan semakin presisi dalam mengukur kompetensi siswa. Hal ini penting sebagai jaminan bahwa instrumen asesmen tetap relevan dan efektif dalam mengukur hasil belajar siswa.

Analisis efektivitas distraktor merupakan proses analisis seberapa efektif opsi jawaban yang salah mampu menarik perhatian siswa yang belum sepenuhnya memahami materi. Distraktor adalah opsi jawaban yang dirancang tidak benar, dengan tujuan menguji kemampuan siswa terhadap materi yang diujikan. Distraktor yang efektif akan dipilih oleh siswa yang kurang memahami materi, sementara siswa yang memahami materi akan memilih jawaban yang benar (Rokhim et al., 2022). Sebuah distraktor dianggap berfungsi dengan baik jika mampu mengecoh peserta yang kurang memahami materi untuk memilihnya sebagai jawaban. Hasil perhitungan analisis efektivitas distraktor dapat dilihat pada Gambar 1.



Gambar 1. Efektivitas Distraktor Secara Keseluruhan

Berdasarkan Gambar 1, diketahui dari total 125 pilihan jawaban yang tersebar dalam 25 butir soal, terdapat 100 distraktor yang dapat dipilih siswa. Proporsi 68% distraktor berfungsi menunjukkan bahwa instrumen asesmen memiliki kualitas yang cukup baik, namun masih perlu penyempurnaan pada 32% distraktor yang tidak berfungsi. Fungsi distraktor yang relatif tinggi mengindikasikan bahwa mayoritas opsi jawaban salah dalam tes mampu menjalankan fungsinya untuk mengidentifikasi siswa yang memiliki pengetahuan dan yang masih belum. Namun demikian, persentase distraktor tidak berfungsi (32%) melebihi batas toleransi ideal yang direkomendasikan (<10%) (Forthmann et al., 2020).

Adanya 32% distraktor tidak berfungsi menimbulkan konsekuensi terhadap kualitas instrumen asesmen. Distraktor yang tidak berfungsi artinya opsi jawaban tersebut tidak dipilih oleh siswa sehingga harus dikaji ulang. Untuk meningkatkan kualitas distraktor yang tidak berfungsi, beberapa strategi dapat diterapkan, seperti: mengganti distraktor tidak efektif, memastikan homogenitas dan relevansi antar opsi jawaban, membuat distraktor yang secara konseptual lebih mirip dengan jawaban benar, dan merancang distraktor yang informatif namun tetap salah (Trianingih, 2023). Selain analisis efektivitas distraktor secara keseluruhan, juga dilakukan analisis efektivitas distraktor untuk setiap butir soal, dapat disimak pada Tabel 4.

Tabel 4. Efektivitas Distraktor Butir Soal

No.	Kategori	Butir soal	Jumlah	Presentase
1	4 (Sangat Baik)	1,4,15,18,24	5	20%
2	3 (Baik)	5,7,8,9,10,11,13,14,17,23	10	40%
3	2 (Cukup Baik)	2,3,6,12,16,19,21,22	8	32%
4	1 (Kurang Baik)	20,25	2	8%
5	0 (Tidak Baik)	-	-	-

Pada Tabel 4, hasil analisis efektivitas distraktor menunjukkan bahwa 60% opsi pengecoh berfungsi dengan optimal. Hasil ini mengindikasikan bahwa sebagian besar distraktor mampu membedakan peserta berkemampuan tinggi dan rendah sesuai kriteria, sehingga kualitas instrumen keseluruhan dinilai baik hingga sangat baik. Penelitian oleh Putri & Rosliyah (2020) pada butir soal Bunpou Shokyuu Kouhan Semester 2 menegaskan pentingnya kemiripan distraktor dengan kunci jawaban agar efektif mengecoh; dari 57 distraktor yang diuji, 36 dinyatakan efektif dan 21 perlu revisi, atau efektivitas 63% berfungsi 37% tidak berfungsi optimal. Perbedaan persentase ini menggambarkan kondisi siswa dan butir soal yang serupa dengan hanya menghasilkan selisih 3% saja. Sebesar 60% butir soal dapat langsung digunakan kembali pada asesmen mendatang. Sementara 32% soal kategori cukup baik dapat digunakan kembali namun butir-butir soal ini memerlukan revisi untuk meningkatkan efektivitas distraktor pada setiap butir soal. Perbaikan pada kategori ini sangat penting untuk memastikan bahwa setiap butir soal mampu menggambarkan informasi yang akurat tentang pemahaman siswa (Efendi et al., 2024). Evaluasi lanjutan perlu fokus pada revisi 8% butir soal agar mencapai standar minimal yaitu terpilih sebanyak 2 distraktor. Revisi dapat dilakukan dengan mengubah masing masing distraktor dengan menyesuaikan penggunaan kata seperti “sedimentasi” dirubah menjadi “pengendapan”. Kedua kata tersebut memiliki arti yang sama akan tetapi siswa cenderung lebih familiar.

Pemahaman siswa terhadap opsi jawaban mempengaruhi pilihan jawaban yang akan dipilih siswa, sehingga penggunaan istilah tertentu perlu dipertimbangkan kembali.

Analisis terhadap kualitas butir soal asesmen sumatif mengidentifikasi masalah penting yang perlu diperbaiki, meliputi soal tidak valid yang terdiri dari soal nomor 13,16,23, dan 25. Soal terlalu sulit yang terdiri dari soal nomor 15,16, dan 25. Soal yang kurang mampu membedakan kemampuan siswa yang terdiri dari soal nomor 13,16, dan 25. Sebanyak 32 opsi jawaban distraktor yang tidak berfungsi. Serta 2 soal memiliki indeks efektivitas distraktor kurang baik yang terdiri dari soal nomor 20 dan 25. Berdasarkan hasil analisis pemenuhan kriteria, soal nomor 13,15,23, dan 24 perlu ditinjau ulang untuk mengidentifikasi kekurangan yang terjadi dan memperbaiki item soal untuk mendapatkan kualitas maksimal sehingga item soal dapat digunakan kembali. Setiap butir soal ditinjau secara mendalam untuk mengetahui alasan dibalik terciptanya kualitas tersebut sehingga guru dapat memperbaiki soal yang bermasalah. Jika peninjauan mendapatkan hasil yang buruk (sulit diperbaiki), maka butir soal tersebut harus dibuang. Soal nomor 16 dan 25 harus dibuang atau tidak layak digunakan kembali karena tidak memenuhi 3 kriteria minimal. Untuk menganalisis secara mendalam, diperlukan profil pemahaman siswa yang dapat dilihat pada Tabel 5.

Tabel 5. Profil Pemahaman Siswa

Indikator	Submateri			Keseluruhan (1-25)
	Peta (1-9)	PJ (10-17)	SIG (18-25)	
Mean	20,35	14,71	17,65	52,71
Median	24	16	20	56
Modus	28	16	20	44
Max	32	28	32	84
Min	4	0	4	20
Standar deviasi	8,49	6,73	6,69	17,79
Varians	72,11	45,30	44,72	316,58

Berdasarkan Tabel 5 tentang sebaran nomor soal, butir soal yang bermasalah terletak pada materi penginderaan jauh dan sistem informasi geografis, sedangkan materi peta seluruhnya telah memenuhi kriteria kualitas minimal. Hal ini sejalan dengan data pada Tabel 5 yang menunjukkan bahwa di antara ketiga submateri tersebut, materi peta memiliki nilai statistik tertinggi dibandingkan dengan materi penginderaan jauh dan sistem informasi geografis. Hasil ini mengindikasikan bahwa siswa lebih menguasai materi peta. Soal nomor 8 memperoleh tingkat ketepatan jawaban tertinggi dengan 24 siswa menjawab benar. Soal tersebut membahas skala peta dengan pertanyaan: "Pada peta berskala 1:30.000.000, jarak kota A ke B pada peta adalah 4 cm. Berapa jarak sebenarnya?". Tingginya penguasaan siswa karena konsep ini mengandalkan kemampuan berhitung yang telah dipelajari dalam mata pelajaran matematika. Dengan demikian, siswa dapat dengan mudah mengaplikasikan pengetahuan matematika yang dimiliki ke dalam perhitungan skala peta. Sebaliknya, soal nomor 16 memiliki tingkat ketepatan jawaban terendah dengan hanya 7 siswa yang menjawab benar. Soal ini menguji pemahaman siswa tentang interpretasi citra dengan pertanyaan: "Setiap objek memantulkan tenaga ke sensor. Jika objek di citra terlihat gelap dan buram, hal ini terjadi karena?". Interpretasi citra merupakan materi yang membahas teknik membaca dan menganalisis citra, di mana siswa dituntut memahami faktor-faktor yang memengaruhi karakteristik visual objek dalam citra. Rendahnya tingkat penguasaan siswa terhadap materi ini disebabkan oleh kompleksitas konsep interpretasi citra yang masih relatif baru bagi siswa (Garcia et al., 2021). Oleh karena itu, diperlukan inovasi dalam model, media, dan pendekatan pembelajaran yang akan diterapkan oleh guru.

Penelitian yang dilakukan oleh Saputro et al. (2020) mendapatkan temuan yang sama dimana nilai peta lebih tinggi dibandingkan nilai penginderaan jauh dan SIG. Permasalahan tersebut dapat terjadi karena materi peta memiliki karakteristik yang lebih nyata dan mudah dipahami secara visual dibandingkan dengan penginderaan jauh dan SIG. Penelitian Xiang & Liu (2017) menunjukkan bahwa konsep-konsep geografis dapat dikategorikan dalam dimensi abstrak-konkret, dimana konsep yang lebih konkret cenderung lebih mudah dipahami oleh siswa. Materi peta melibatkan objek yang dapat langsung diobservasi dan diinterpretasi secara visual, sementara penginderaan jauh dan SIG memerlukan pemahaman konseptual yang lebih abstrak tentang teknologi dan analisis spasial yang ada didalamnya. Pembelajaran penginderaan jauh dan SIG memerlukan

keterampilan berpikir spasial yang lebih kompleks dibandingkan dengan memahami peta. Penelitian Mark & Philip (2017) menunjukkan bahwa kemampuan visualisasi spasial yang diperlukan untuk memahami data penginderaan jauh lebih tinggi daripada interpretasi peta. Temuan ini menciptakan tantangan tambahan dalam pembuatan soal yang mampu mengukur pemahaman konsep secara akurat. Guru perlu melakukan analisis butir soal secara berkala sebagai bagian dari pengembangan asesmen sumatif yang berkualitas dan berkelanjutan untuk meningkatkan kualitas asesmen sumatif di masa mendatang.

Penelitian ini dibatasi pada analisis butir soal asesmen sumatif materi peta, penginderaan jauh, dan sistem informasi geografis di kelas X.10 SMAN 1 Lawang dengan jumlah responden hanya 34 siswa yang dipilih melalui purposive sampling. Instrumen yang dianalisis mencakup 25 soal objektif yang hanya mewakili tiga submateri tersebut, sehingga temuan tidak dapat digeneralisasi ke materi geografi lain atau jenjang pendidikan berbeda. Proses validitas dan reliabilitas hanya menggunakan aplikasi IBM SPSS Statistics 22 dan Cronbach's Alpha, sehingga tidak mempertimbangkan metode triangulasi alat ukur lain. Selain itu, efektivitas distraktor dianalisis berdasarkan taksonomi pilihan ganda standar ( $\geq 5\%$  pemilih), tanpa menguji pilah kualitas butir soal secara kualitatif melalui wawancara guru atau siswa.

Temuan penelitian ini memberikan kontribusi penting dalam meningkatkan mutu asesmen sumatif geografi di tingkat menengah atas dengan menyediakan kerangka kerja analisis kuantitatif yang dapat diadopsi oleh guru untuk meninjau dan merevisi butir soal sebelum dan setelah digunakan. Dengan mengetahui tingkat kesulitan, daya beda, dan efektivitas distraktor, guru dapat merancang soal yang lebih seimbang dan mendukung pembelajaran diferensiasi, sehingga hasil penilaian merefleksikan kemampuan siswa secara lebih akurat. Selain itu, rekomendasi revisi butir soal pada submateri penginderaan jauh dan SIG mendorong perbaikan kurikulum dan pelatihan guru dalam memanfaatkan teknologi spasial, yang dapat meningkatkan literasi spasial siswa. Implementasi rutin analisis butir soal juga berpotensi membangun budaya profesionalisme guru melalui refleksi instrumen penilaian, mendukung keputusan berbasis data dalam evaluasi pembelajaran, serta memperkuat akuntabilitas dan transparansi penilaian di lingkungan sekolah.

## KESIMPULAN

Berdasarkan hasil analisis serta pembahasan yang telah dijabarkan, dapat disimpulkan secara keseluruhan asesmen sumatif geografi telah menunjukkan kualitas yang baik, meskipun diperlukan beberapa revisi untuk memaksimalkan hasil asesmen dimasa mendatang. Temuan menunjukkan bahwa butir soal yang bermasalah merupakan soal submateri penginderaan jauh dan SIG, yang artinya bahwa pada submateri tersebut terdapat permasalahan dan diperlukan revisi untuk memperbaiki kualitas asesmen di masa mendatang. Jika revisi tidak dilakukan, maka hasil pengukuran tidak akan optimal, pembelajaran tidak bermakna, dan keputusan akademik tidak sesuai. Apabila hal tersebut terjadi, maka fungsi asesmen sumatif sebagai sarana evaluasi tidak dapat dicapai. Revisi dapat dilakukan dengan memperbaiki butir soal maupun memperbaiki kegiatan pembelajaran yang dilakukan dikelas melalui *self evaluation* yang dilakukan oleh guru. Sebagai tindak lanjut, guru dapat secara praktis merevisi butir soal bermasalah seperti menyederhanakan soal nomor 15, 16, dan 25 agar sesuai kosakata siswa, merubah istilah seperti "sedimentasi" menjadi "pengendapan," serta memperkuat distraktor pada soal nomor 13, 16, dan 23 dengan pilihan jawaban yang mencerminkan miskonsepsi umum siswa untuk meningkatkan daya pembeda. Selain itu, penerapan rutin analisis butir soal setelah asesmen akan membantu guru membangun bank soal berkualitas. Dengan langkah-langkah ini, instrumen asesmen geografi akan semakin valid, reliabel, dan mampu mencerminkan capaian belajar siswa secara akurat. Potensi penelitian lebih lanjut dapat melakukan inovasi dalam kegiatan pembelajaran khususnya pada materi PJ dan SIG untuk meningkatkan kemampuan berpikir kritis dan literasi spasial siswa guna mencapai tujuan pembelajaran. Serta dipadukan dengan pemanfaatan teknologi digital ketika melaksanakan asesmen untuk meningkatkan efisiensi dan akurasi analisis butir soal.

## KONTRIBUSI PENULIS

**Hadar Rafliansyah Rivaldi:** Conceptualization, Methodology, Writing - Original Draft, Software, Formal Analysis, Investigation, Data Curation, Visualization, and Project Administration; **Adip Wahyudi:** Methodology, Formal Analysis, Resources, Writing - Review & Editing, Validation, and Supervision; and **Ratna Arum Nur Indah Sari:** Conceptualization, Writing - Review & Editing, Validation, Resources, and Supervision. Semua penulis telah membaca dan menyetujui versi akhir dari naskah ini.

**PERNYATAAN BEBAS KONFLIK KEPENTINGAN**

Para penulis menyatakan bahwa tidak ada konflik kepentingan finansial maupun hubungan pribadi yang dapat mempengaruhi hasil yang dilaporkan dalam naskah ini.

**PERNYATAAN ETIKA PENELITIAN DAN PUBLIKASI**

Para penulis menyatakan bahwa penelitian dan penulisan naskah ini telah mematuhi standar etika penelitian dan publikasi, sesuai dengan prinsip ilmiah, serta bebas dari plagiasi.

**PERNYATAAN PEMANFAATAN TEKNOLOGI ASISTIF**

Para penulis menyatakan bahwa Kecerdasan Buatan Generatif (*Generative Artificial Intelligence*) dan teknologi asistif lainnya tidak digunakan secara berlebihan dalam proses penelitian dan penulisan naskah ini. Secara khusus, ChatGPT dan Deepseek digunakan untuk brainstorming serta mengembangkan ide awal dalam penyusunan outline; Perplexity digunakan untuk mengembangkan gaya bahasa berdasarkan ide pokok yang dimiliki penulis dan menemukan artikel/buku yang relevan dengan penelitian; Scite digunakan untuk menemukan artikel pendukung. Para penulis telah meninjau dan menyunting semua konten yang dihasilkan AI guna memastikan ketepatan, kelengkapan, serta kepatuhan terhadap standar etika dan ilmiah, dan bertanggung jawab penuh atas naskah versi akhir.

**DAFTAR PUSTAKA**

- Ajmi, M.A., Mustakim, S.S., Roslan, S., & Almehrizi, R. (2024). Psychometric characteristics of the numerical ability test for Gulf students. *International Journal of Evaluation and Research in Education (IJERE)*, **13**(4), 2552-2561. DOI: <https://doi.org/10.11591/ijere.v13i4.28917>.
- Amalia, N., Kamaruddin, T., & Amri, A. (2019). Analisis butir soal geografi kelas XI semester ganjil tahun ajaran 2017/2018 di SMA Negeri Kota Lhokseumawe. *Jurnal Ilmiah Mahasiswa Pendidikan Geografi FKIP Unsyiah*, **4**(1), 105–112. Retrieved from: <https://jim.usk.ac.id/geografi/article/view/13243>.
- Arifin, Z. (2017). *Evaluasi pembelajaran (edisi ke-1)*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Arwansyah, Y.B., Setiawan, K.E.P., & Yuda, R.K. (2022). Analisis butir soal penilaian akhir tahun (PAT) bahasa indonesia kelas XI SMA Negeri 1 Polanharjo Klaten. *Lingua Rima: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, **11**(2), 25-33. DOI: <https://doi.org/10.31000/lgrm.v11i2.6620>.
- Bano, V.O., Marambaawang, D.N., & Njoeroemana, Y. (2022). Analisis kriteria butir soal ujian sekolah mata pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, dan Budaya*, **8**(1), 145–152. Retrieved from: <https://jurnal.ideaspublishing.co.id/index.php/ideas/article/view/660>.
- Basileo, L.D., Otto, B., Lyons, M., Vannini, N., & Toth, M.D. (2024). The role of self-efficacy, motivation, and perceived support of students' basic psychological needs in academic achievement. *Frontiers in Education*, **9**, 1385442. DOI: <https://doi.org/10.3389/feduc.2024.1385442>.
- Brownlie, N., Burke, K., & van der Laan, L. (2024). Quality indicators of effective teacher-created summative assessment. *Quality Assurance in Education*, **32**(1), 30–45. DOI: <https://doi.org/10.1108/QAE-04-2023-0062>.
- Darwin, D., Boeriswati, E., & Murtadho, F. (2023). Asesmen pembelajaran bahasa dalam kurikulum merdeka belajar pada siswa SMA. *Lingua Rima: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, **12**(2), 25-36. DOI: <https://doi.org/10.31000/lgrm.v12i2.8639>.
- Efendi, M., Zulhimmah, Z., Lubis, N., & Harahap, H.A. (2024). Penerapan asesmen formatif dan sumatif dalam kurikulum merdeka di Madrasah Aliyah Swasta Darul Hadits Huta Baringin. *Cognoscere: Jurnal Komunikasi dan Media Pendidikan*, **2**(2), 64–72. DOI: <https://doi.org/10.61292/cognoscere.169>.
- Erfan, M., Mauluda, M.A., Hidayati, V.R., Astria, F.P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model Rasch. *Indonesian Journal of Educational Research and Review*, **3**(1), 11-19. DOI: <https://doi.org/10.23887/ijerr.v3i1.24080>.
- Fachruddin, F., Amiruddin, A., Lidani, A., Putra, E., Nasution, S., & Yuliana, Y. (2022). Pengaruh manajemen kesiswaan terhadap prestasi belajar peserta didik. *Edukatif: Jurnal Ilmu Pendidikan*, **4**(1), 1443–1450. DOI: <https://doi.org/10.31004/edukatif.v4i1.2166>.
- Farida & Musyarofah, A. (2021). Validitas dan reliabilitas dalam analisis butir soal. *Al-Mu'arrif: Jurnal Pendidikan Bahasa Arab*, **1**(1), 34–44. Retrieved from: <https://jurnal.lp2msasbabel.ac.id/AL-MUARRIB/article/view/2100>.

- Forthmann, B., Förster, N., Schütze, B., Hebbecke, K., Flessner, J., Peters, M. T., & Souvignier, E. (2020). How much g is in the distractor? Re-thinking item-analysis of multiple-choice items. *Journal of Intelligence*, *8*(1), 11. DOI: <https://doi.org/10.3390/jintelligence8010011>.
- Garcia, C., Olivas, L., Olivella, R., & Peliova, A. (2021). Edusat: remote sensing as an educational resource. The use of data from the Copernicus program as an innovative teaching method for students, teachers, and researchers. *AGILE: GIScience Series*, *2*, 1–6. DOI: <https://doi.org/10.5194/agile-giss-2-26-2021>.
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Analisis butir soal ujian matematika kelas VII semester ganjil tahun pelajaran 2017/2018. *MATHEMA: Jurnal Pendidikan Matematika*, *2*(1), 57-66. DOI: <https://doi.org/10.33365/jm.v2i1.459>.
- Hardianti, H., Liliawati, W., & Tayubi, Y.R. (2023). Karakteristik tes kemampuan berpikir kritis siswa SMA pada materi momentum dan impuls: Perbandingan classical theory test (CTT) dan model Rasch. *WaPFI (Wahana Pendidikan Fisika)*, *8*(1), 21–28. DOI: <https://doi.org/10.17509/wapfi.v8i1.30958>.
- Hidayati, A., Firdaussiah, L., Utami, I.R., & Ayu, N.P. (2022). Analisis kesulitan mahasiswa dalam menyelesaikan soal matematika ekonomi pada materi fungsi permintaan dan penawaran ditinjau dari kemampuan koneksi matematis. *Jurnal Jendela Pendidikan*, *2*(3), 451–458. DOI: <https://doi.org/10.57008/jjp.v2i03.261>.
- Jati, P., Alawiati, A., Jibril H.A., Mudin, A., & Prisma S.C. (2023). Analisis kualitas butir soal geografi. *Jurnal Pendidikan Geosfer*, *8*(2), 287–298. DOI: <https://doi.org/10.24815/jpg.v8i2.32500>.
- Logayah, D.S., Ruhimat, M., Arrasyid, R., & Islamy, M.R.F. (2024). Item analysis of national geography olympiad Multiple-Choice Questions (MCQs) in Indonesia. *Cogent Social Sciences*, *10*(1), 2354971. DOI: <https://doi.org/10.1080/23311886.2024.2354971>.
- Lusiyanti, Amilda, & Laksono, P.J. (2021). Analisis butir soal pilihan ganda pada soal ulangan semester ganjil mata pelajaran kimia kelas X tahun ajaran 2018/2019. *Jurnal Al'Ilmi: Jurnal Pendidikan MIPA*, *10*(2), 11-18. Retrieved from: <https://jurnal.radenfatah.ac.id/index.php/alilmi/article/view/14422>.
- Mangla, M., Mehta, V., Pattnaik, C.R., & Mohanty, S.N. (2023). Students' feedback- An effective tool towards enhancing the Teaching Learning Process. *EAI Endorsed Transactions on Scalable Information Systems*, *10*(5), 3347. DOI: <https://doi.org/10.4108/eetsis.3347>.
- Mark, J. & Philip, M. (2017). *Debates in geography education*. Oxfordshire: Routledge. DOI: <https://doi.org/10.4324/9781315562452>.
- Nafs, H., Sridana, N., Hikmah, N., & Soeprianto, H. (2023). Analisis kualitas butir soal ulangan akhir semester genap mata pelajaran matematika kelas VII SMPN 6 Mataram tahun ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, *8*(4), 2324–2331. DOI: <https://doi.org/10.29303/jipp.v8i4.1785>.
- Pinaka, T. & Apri, D.S.K. (2024). Analisis butir soal model klasik pada Penilaian Harian Bersama (PHB) Mata Pelajaran bahasa indonesia kelas VII MTS Negeri 5 Magelang. *Indonesian Journal of Education and Learning*, *7*(1), 31–45. DOI: <https://doi.org/10.31002/ijel.v7i1.835>.
- Putri, R.A. & Rosliyah, Y. (2020). Analisis distraktor butir soal bunpou shokyuu kouhan semester 2 tahun 2018/2019 prodi pendidikan bahasa jepang UNNES. *KIRYOKU*, *4*(1), 18–25. DOI: <https://doi.org/10.14710/kiryoku.v4i1.18-25>.
- Rahmah, R. & Cahyadi, A. (2024). Analisis implementasi permendikbud no. 21 tahun 2022 dalam standar penilaian pendidikan di Indonesia. *Al-Madrasah Jurnal Pendidikan Madrasah Ibtidaiyah*, *8*(2), 831-843. DOI: <https://doi.org/10.35931/am.v8i2.3460>.
- Rokhim, D.A., Tyas, F.K., Rahayu, S., & Habiddin, H. (2022). Perspektif siswa dan guru dalam pelaksanaan AKM (Asesmen Kompetensi Minimum) Pada mata pelajaran kimia. *JAMP: Jurnal Administrasi dan Manajemen Pendidikan*, *5*(1), 46–52. DOI: <https://doi.org/10.17977/um027v5i12022p46>.
- Saputra, H.D., Purwanto, W., Setiawan, D., Fernandez, D., & Putra, R. (2022). Hasil belajar mahasiswa: Analisis butir soal tes. *Edukasi: Jurnal Pendidikan*, *20*(1), 15–27. DOI: <https://doi.org/10.31571/edukasi.v20i1.3432>.
- Saputro, R., Liesnoor, D., Setyowati, & Hardati, P. (2020). The students spatial critical thinking skill by using map and remote sensing imagery on geography lesson. *Proceedings of the International Conference on Science and Education and Technology (ISET 2019)*, 250-254. DOI: <https://doi.org/10.2991/assehr.k.200620.049>.
- Sarfraz, F., Sarfraz, F., Jawad, I., Zia-Ul-Miraj, M., Ahmad, R.Z., & Saleem, J. (2021). OSCE: An effective tool of assessment for medical students. *Pakistan Journal of Medical and Health Sciences*, *15*(8), 2235–2239. DOI: <https://doi.org/10.53350/pjmhs211582235>.

- Savika, H.I., Zuhriyah, I.A., & Susilawati, S. (2025). Peran guru dalam analisis butir soal di sekolah dasar. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, **8**(3), 3313–3319. DOI: <https://doi.org/10.54371/jiip.v8i3.7534>.
- Sodikin, S., & Gumiandari, S. (2022). Analisis SWOT mutu evaluasi pembelajaran. *JDMP (Jurnal Dinamika Manajemen Pendidikan)*, **6**(1), 59-69. DOI: <https://doi.org/10.26740/jdmp.v6n1.p59-69>.
- Somayana, W. (2020). Peningkatan hasil belajar siswa melalui metode PAKEM. *Jurnal Pendidikan Indonesia*, **1**(3), 350–361. DOI: <https://doi.org/10.36418/japendi.v1i3.33>.
- Sözçü, U., Türker, A., & Dünder, E. (2023). High school students' evaluations on the reflections of geographical information on daily life. *Journal of Teacher Education and Lifelong Learning*, **5**(1), 275–294. DOI: <https://doi.org/10.51535/tell.1264253>.
- Sugiharni, G.A.D. (2023). Content validity of flipped learning-based statistical learning evaluation instruments at tourism colleges. *Journal of Education Research and Evaluation*, **7**(2), 335–344. DOI: <https://doi.org/10.23887/jere.v7i2.61240>.
- Teig, N., Scherer, R., & Nilsen, T. (2019). I know i can, but do i have the time? The role of teachers' self-efficacy and perceived time constraints in implementing cognitive-activation strategies in science. *Frontiers in Psychology*, **10**, 1697. DOI: <https://doi.org/10.3389/fpsyg.2019.01697>.
- Trianingsih, L. (2023). Analisis kualitas tes dan butir soal integral pada evaluasi formatif matematika teknik. *Indonesian Journal of Civil Engineering Education*, **9**(2), 44-57. DOI: <https://doi.org/10.20961/ijcee.v9i2.84711>.
- Wahyudi, M.W.P., Indryani, I., & Amanda, R.S. (2024). Analisis penggunaan asesmen sumatif sebagai alat penilaian perkembangan dan pembelajaran pada anak usia dini di TK Kirana Kota Jambi. *JECIE (Journal of Early Childhood and Inclusive Education)*, **8**(1), 64–72. DOI: <https://doi.org/10.31537/jecie.v8i1.1536>.
- Xiang, X. & Liu, Y. (2017). Understanding 'change' through spatial thinking using Google Earth in secondary geography. *Journal of Computer Assisted Learning*, **33**(1), 65–78. DOI: <https://doi.org/10.1111/jcal.12166>.