

EVALUASI PEMBELAJARAN ANALISIS BUTIR SOAL PILIHAN MAJEMUK MATERI PEMBENTUKAN ALAM SEMESTA

Nazwa Aulia Rahman, Firda Kristiyani, Mamat Ruhimat, Dina Siti Logayah, Ignatius Juli Triana

Universitas Pendidikan Indonesia, Kota Bandung, Indonesia
email Koresponden : nazwaawr29@upi.edu

Diterima: 22-5-2025	Direvisi: 23-10-2025	Diterbitkan: 24-10-2025
---------------------	----------------------	-------------------------

Abstrak: Penelitian ini bertujuan untuk mengkaji kualitas butir soal pilihan majemuk pada materi teori pembentukan alam semesta di kelas X SMA. Latar belakang dari penelitian ini adalah minimnya analisis mendalam terhadap butir-butir soal yang digunakan dalam evaluasi pembelajaran, sehingga berpotensi menurunkan efektivitas alat penilaian dalam mengukur pemahaman siswa secara menyeluruh. Penelitian ini difokuskan pada evaluasi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh dari setiap butir soal. Metode penelitian yang digunakan adalah deskriptif kuantitatif dengan pendekatan analisis butir soal. Data diperoleh dari respons siswa terhadap soal-soal pilihan ganda, kemudian dianalisis menggunakan teknik statistik untuk mengukur aspek-aspek yang diteliti. Hasil menunjukkan bahwa seluruh butir soal memiliki validitas tinggi ($r > 0,30$) dan reliabilitas sangat baik (Cronbach's Alpha = 0,8469), dengan 18 soal (90%) tergolong sedang, 2 soal (10%) mudah, serta 85% pengecoh berfungsi efektif. Kegiatan ini terbukti meningkatkan pemahaman guru terhadap prinsip evaluasi pembelajaran berbasis data, sekaligus memberikan kontribusi nyata dalam peningkatan kompetensi guru dalam menyusun dan menganalisis soal secara ilmiah guna mewujudkan praktik evaluasi yang objektif dan akurat.

Kata Kunci: Evaluasi, Validitas, Reliabilitas

Abstract: This study aims to examine the quality of multiple-choice test items on the topic of the Theory of the Formation of the Universe for 10th-grade students in senior high school. The background of this research lies in the limited in-depth analysis of test items used in learning evaluations, which may reduce the effectiveness of assessment tools in accurately measuring students' understanding. The study focuses on evaluating the validity, reliability, difficulty level, discrimination power, and distractor effectiveness of each test item. A quantitative descriptive method was employed using an item analysis approach. Data were obtained from students' responses to multiple-choice questions and analyzed statistically to assess the examined aspects. The results showed that all test items had high validity ($r > 0.30$) and excellent reliability (Cronbach's Alpha = 0.8469), with 18 items (90%) categorized as moderate, 2 items (10%) as easy, and 85% of distractors functioning effectively. This activity proved to enhance teachers' understanding of data-driven evaluation principles while providing a tangible contribution to improving teachers' competence in scientifically constructing and analyzing test items to achieve an objective and accurate assessment practice.

Keywords: Evaluation, Validation, Reliability

Pendahuluan

Pendidikan merupakan pondasi penting dalam pengembangan sumber daya manusia yang berkualitas. Salah satu komponen krusial dalam proses pendidikan adalah evaluasi pembelajaran, yang berfungsi untuk menilai sejauh mana tujuan pembelajaran telah tercapai. Berdasarkan hasil observasi di beberapa SMA di Kota Bandung, guru masih menghadapi kendala



dalam penyusunan serta analisis butir soal, khususnya pada materi *Teori Pembentukan Alam Semesta Yang* bersifat abstrak dan konseptual. Soal-soal yang digunakan dalam evaluasi seringkali belum dianalisis secara mendalam, baik dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, maupun efektivitas pengecoh. Akibatnya, hasil tes belum sepenuhnya mencerminkan kemampuan kognitif siswa secara akurat (Kumar et al., 2021).

Permasalahan ini menunjukkan adanya kesenjangan (gap analysis) antara praktik evaluasi yang dilakukan guru di lapangan dengan standar pengukuran hasil belajar berbasis data kuantitatif (Nurgiyantoro et al., 2021; Wahyuni & Prasetyo, 2022). Berdasarkan kajian literatur (*state of the art*), penelitian mengenai analisis butir soal banyak dilakukan pada bidang Biologi (Prawiki, 2022; Sitorus et al., 2020), Matematika (Lestari et al., 2023), serta Pendidikan Kesehatan (Rezigalla et al., 2024). Namun, kajian serupa pada bidang Geografi, khususnya pada materi abstrak seperti Teori Pembentukan Alam Semesta, masih sangat terbatas. Hal ini menandakan perlunya tindakan solutif melalui kegiatan pengabdian berbasis riset yang dapat memberikan pemahaman dan pelatihan langsung kepada guru dalam menganalisis kualitas butir soal.

Novelty dari kegiatan ini terletak pada penerapan analisis butir soal pilihan majemuk berbasis data empiris dengan memanfaatkan aplikasi *Quizizz* dan pengolahan statistik sederhana melalui *spreadsheet*. Pendekatan ini memberikan pengalaman langsung kepada guru dalam menganalisis validitas item, reliabilitas (Cronbach's Alpha), indeks kesukaran, daya pembeda, serta efektivitas pengecoh menggunakan data riil siswa. Inovasi ini belum banyak diterapkan pada konteks Geografi SMA, sehingga memberikan nilai kebaruan baik secara praktis maupun akademis.

Tujuan dari pengabdian berbasis riset ini adalah untuk: (1) meningkatkan kemampuan guru SMA dalam melakukan analisis butir soal menggunakan pendekatan kuantitatif, (2) membantu guru mengembangkan bank soal yang valid dan reliabel untuk pembelajaran Geografi, serta (3) memperkuat budaya evaluasi berbasis data di sekolah. Melalui kegiatan ini, diharapkan guru dapat melakukan analisis instrumen evaluasi secara mandiri dan berkelanjutan, sehingga kualitas pembelajaran menjadi lebih objektif, adil, dan terukur.

Metode

Dalam penelitian ini kami menggunakan metode kuantitatif dengan menggunakan tes soal materi geografi kepada 40 siswa pada sekolah menengah atas menggunakan *quizizz* sebanyak 20 soal *multiple choice* atau pilihan majemuk mengenai teori pembentukan alam semesta. Seluruh proses perhitungan data dilakukan sistematis dengan menggunakan bantuan aplikasi *spreadsheet* untuk memastikan akurasi dan efisiensi dalam pengolahan data. Berikut rumus yang kami aplikasikan menggunakan aplikasi *spreadsheet*:

A. Validitas

$$r \text{ (hitung)} = (n * \Sigma XY - \Sigma X * \Sigma Y)$$

Keterangan:

r = koefisien korelasi

n = jumlah responden

Σ^*XY = jumlah hasil perkalian skor item dengan skor total

Σ^*X = jumlah skor item

Σ^*Y = jumlah skor total (seluruh item)

(Sumber: Sugiyono, 2021)

B. Reliabilitas

$$a = (k/(k-1)) * (1 - (\Sigma \sigma^2_i / \sigma^2_t))$$

Keterangan:

a = Cronbach's Alpha

k = item dalam instrumen

σ^2_i = varians skor setiap item

σ^2_t = varians skor total Tingkat Kesukaran

Secara umum, instrumen dianggap reliabel jika nilai Cronbach Alpha lebih besar dari 0,6. Namun, nilai yang lebih tinggi (0,7 atau lebih) menunjukkan reliabilitas yang lebih baik. (Sumber: Sugiyono, 2021)

C. Tingkat Kesukaran

$$P = B / N$$

Keterangan:

P = kesukaran soal

B = Jumlah peserta yang menjawab soal dengan benar

N = Jumlah seluruh peserta tes

Interpretasi Tingkat Kesukaran:

$P \leq 0,30$ Sulit

$0,31 \leq P \leq 0,70$ = Sedang

$P > 0,7$ = Mudah

(Sumber: Fitriyani, 2020)

D. Daya Pembeda

$$D = DA/JA - BB/JB$$

Keterangan:

DA: Jumlah peserta kelompok atas yang menjawab benar

JA: Jumlah peserta dalam kelompok atas

BB: Jumlah peserta kelompok bawah yang menjawab benar

JB: Jumlah peserta dalam kelompok bawah

Interpretasi Daya Pembeda:

0,40 atau lebih = Sangat Baik

0,30-0,39 Baik

0,20-0,29 Cukup, Perlu Perbaikan 0,19-ke bawah Kurang Baik, Dibuang

(Sumber: Fauziah, 2020)

E. Daya Pengecoh

$$D = A/N \times 100\%$$

Keterangan:

D = Tingkat Distraktor (%)

A = Jumlah siswa yang memilih opsi tersebut

N= Jumlah siswa seluruhnya

D Lebih besar dari 5% soal diterima

D sama dengan 0 soal ditolak

(Sumber: Wardoyo W & Supraptono E. 2014)



Bagan 1. Kerangka Pemecahan Masalah

Hasil dan Pembahasan

Berdasarkan pendekatan metodologis yang telah diterapkan secara sistematis dalam

penelitian ini, dilakukan proses analisis yang komprehensif terhadap butir-butir soal yang telah dikembangkan. Analisis tersebut mencakup lima aspek, yaitu uji validitas untuk menilai sejauh mana butir soal mampu mengukur kompetensi yang dimaksudkan, uji reliabilitas untuk mengetahui konsistensi hasil pengukuran, analisis tingkat kesukaran guna menentukan sejauh mana soal dapat dikerjakan oleh peserta didik, analisis daya pengecoh untuk menilai efektivitas pilihan jawaban yang tidak benar dalam menggiring jawaban peserta didik, serta analisis daya pembeda untuk mengukur kemampuan butir soal dalam membedakan antara peserta didik yang memiliki kemampuan tinggi dan rendah. Seluruh hasil analisis dilakukan dengan bantuan *spreadsheet* guna meningkatkan akurasi dan efisiensi pengolahan data. Hasil akhir tersebut kemudian disajikan secara sistematis dalam bentuk tabel, yang bertujuan untuk mempermudah interpretasi dan pembacaan data. Tabel-tabel yang disajikan berikut ini mencerminkan ringkasan dari hasil analisis butir soal bentuk pilihan majemuk yang telah dirancang sebelumnya. Penyajian dilakukan secara langsung terhadap hasil akhir tanpa menguraikan tahapan perhitungan secara terperinci, namun tetap memuat nilai-nilai kunci dari setiap parameter analisis yang telah dilakukan sebagai dasar penilaian kualitas instrumen yang digunakan dalam penelitian ini.

Tabel 1. Uji Validitas

Valid	Tidak Valid
20 Soal	-

(Source : Analysis Result, 2025)

Berdasarkan hasil uji validitas yang ditampilkan pada Tabel 1, diketahui bahwa seluruh item dalam instrumen penelitian dinyatakan valid. Sebanyak 20 soal yang diuji memiliki nilai korelasi item-total yang memenuhi kriteria validitas, yakni nilai korelasi lebih besar dari nilai r-tabel pada taraf signifikansi 5%. Hal ini menunjukkan bahwa setiap butir soal mampu mengukur konstruk yang hendak diteliti secara tepat dan akurat. Tidak terdapat satupun soal yang tergolong tidak valid, sebagaimana ditunjukkan oleh tanda strip (-) pada kolom "Tidak Valid". Hasil ini mencerminkan bahwa instrumen yang digunakan telah memiliki kelayakan secara statistik untuk digunakan dalam pengumpulan data penelitian. Validitas yang tinggi pada seluruh butir soal memberikan jaminan bahwa data yang diperoleh dapat diinterpretasikan secara sahih dan relevan dengan tujuan penelitian.

Tabel 2. Uji Reliabilitas

Hasil dari Rumus Cronbach's Alpha	Keterangan
0.8469695949	> 0,6 (20 Soal Dinyatakan Reliabel)

Source: Analysis Result, 2025)

Berdasarkan Hasil uji reliabilitas instrumen penelitian sebagaimana ditunjukkan dalam Tabel 2 menunjukkan bahwa nilai koefisien Cronbach's Alpha sebesar 0,8469695949. Nilai ini berada jauh di atas batas minimum yang umum digunakan dalam penelitian sosial, yaitu 0,60, sehingga dapat disimpulkan bahwa instrumen memiliki tingkat reliabilitas yang tinggi. Reliabilitas merujuk pada konsistensi alat ukur dalam menghasilkan data; semakin tinggi nilai Cronbach's Alpha, maka semakin andal instrumen tersebut dalam mengukur variabel yang dimaksud. Berdasarkan hasil ini, seluruh 20 butir soal yang telah dinyatakan valid sebelumnya juga dinyatakan reliabel, artinya soal-soal tersebut konsisten dalam mengukur aspek yang diteliti meskipun digunakan dalam kondisi atau waktu yang berbeda. Dengan kata lain, instrumen ini tidak hanya sah secara substansi (valid) tetapi juga stabil dan dapat dipercaya dalam penggunaannya. Nilai $> 0,8$ seperti yang diperoleh juga menunjukkan kategori reliabilitas yang sangat baik.

Tabel 3. Tingkat Kesukaran

Sulit	Sedang	Mudah
-	18 Soal	2 Soal

(Source: Analysis Result, 2025)

Berdasarkan hasil Tingkat kesukaran yang ditampilkan pada Tabel 3, menunjukkan distribusi tingkat kesukaran dari 20 butir soal yang dianalisis. Berdasarkan hasil analisis, ditemukan bahwa tidak terdapat soal yang masuk dalam kategori sulit, yang ditunjukkan dengan tanda strip (-) pada kolom "Sulit". Mayoritas soal, yakni sebanyak 18 butir, berada dalam kategori sedang, yang menunjukkan bahwa soal-soal tersebut memiliki tingkat kesukaran yang ideal dan dapat membedakan peserta berdasarkan kemampuan mereka secara proporsional. Sementara itu, terdapat 2 butir soal yang dikategorikan mudah, artinya soal tersebut cenderung dapat dijawab dengan benar oleh sebagian besar peserta. Distribusi tingkat kesukaran yang didominasi oleh soal-soal dalam kategori sedang menunjukkan bahwa instrumen yang digunakan telah dirancang dengan baik dan memiliki kualitas yang memadai untuk mengukur kemampuan secara efektif.

Tabel 4. Daya Pengecoh

Opsi Jawaban	Diterima	Ditolak
A	20 Soal	-
B	19 Soal	1 Soal
C	19 Soal	1 Soal
D	18 Soal	2 Soal
E	17 Soal	3 Soal

(Source : Analysis Result, 2025)

Tabel 4 menyajikan hasil analisis daya pengecoh dari setiap opsi jawaban pada instrumen tes yang terdiri dari 20 soal. Daya pengecoh (distractor effectiveness) mengacu pada kemampuan pilihan jawaban yang salah (distraktor) untuk menarik perhatian peserta yang memiliki pemahaman kurang tepat terhadap materi. Daya pengecoh yang baik adalah yang mampu dipilih oleh peserta dengan kemampuan rendah, sehingga berperan dalam membedakan antara peserta yang memahami materi dengan yang belum. Berdasarkan tabel, terlihat bahwa opsi A memiliki performa terbaik, di mana seluruh 20 soal dengan opsi ini dinyatakan diterima, tanpa ada satupun yang ditolak. Ini menunjukkan bahwa opsi A selalu berfungsi secara efektif sebagai kunci maupun pengecoh yang baik.

Sementara itu, opsi B dan C masing-masing memiliki 1 soal yang ditolak, dengan 19 soal diterima. Artinya, hampir seluruh pengecoh pada pilihan ini berfungsi dengan baik, namun terdapat 1 soal yang perlu diperbaiki karena pengecohnya tidak efektif (kemungkinan tidak menarik peserta sama sekali atau terlalu mencolok sebagai jawaban salah). Opsi D memiliki 2 soal yang ditolak, dan opsi E memiliki daya pengecoh paling lemah dengan 3 soal yang ditolak dari 20. Meskipun sebagian besar soal dengan opsi D dan E tetap diterima, hasil ini mengindikasikan bahwa terdapat beberapa distraktor pada pilihan D dan E yang tidak berfungsi sebagaimana mestinya.

Secara umum, mayoritas pengecoh pada instrumen ini telah memenuhi kriteria efektivitas dan dapat digunakan dalam pengukuran kemampuan peserta. Namun demikian, perlu dilakukan revisi atau perbaikan terhadap sejumlah kecil soal, khususnya pada pengecoh di opsi E, agar efektivitas instrumen semakin optimal. Dengan daya pengecoh yang baik, maka kualitas instrumen sebagai alat evaluasi akan meningkat, karena dapat memberikan informasi yang lebih akurat tentang tingkat penguasaan materi oleh peserta.

Tabel 5. Daya Pembeda

Sangat Baik	Baik	Cukup	Kurang Baik
9	4	6	1

(Source: Analysis Result, 2024)

Tabel 5 menunjukkan hasil analisis terhadap daya pembeda masing-masing butir soal dalam instrumen yang digunakan. Daya pembeda merupakan indikator penting yang menunjukkan kemampuan suatu soal dalam membedakan antara peserta dengan kemampuan tinggi dan peserta dengan kemampuan rendah. Butir soal yang memiliki daya pembeda yang baik akan dijawab dengan benar oleh peserta yang menguasai materi dan dijawab salah oleh peserta yang tidak menguasai materi. Berdasarkan data dalam tabel, diketahui bahwa sebanyak 9 soal memiliki daya pembeda dalam kategori sangat baik, yang berarti soal-soal tersebut sangat efektif dalam memilah kemampuan peserta. Selain itu, terdapat 4 soal yang masuk dalam kategori baik, dan 6 soal dalam kategori cukup, yang secara umum masih dapat diterima dan digunakan meskipun efektivitasnya dalam membedakan peserta tidak sekuat dua kategori

sebelumnya.

Namun, terdapat 1 soal yang tergolong kurang baik dalam hal daya pembeda. Soal dalam kategori ini cenderung tidak mampu secara efektif membedakan antara peserta dengan tingkat pemahaman yang berbeda. Hal ini bisa disebabkan oleh berbagai faktor, seperti redaksi soal yang ambigu, materi yang belum sepenuhnya dikuasai oleh peserta, atau distribusi jawaban yang merata di seluruh tingkat kemampuan. Oleh karena itu, soal ini sebaiknya ditinjau ulang, baik untuk diperbaiki maupun untuk dipertimbangkan untuk dihapus.

Hasil analisis terhadap instrumen evaluasi pembelajaran Geografi tingkat SMA menunjukkan bahwa seluruh 20 butir soal yang diuji memiliki validitas yang memadai. Validitas merupakan aspek krusial dalam pengembangan instrumen evaluasi, karena memastikan bahwa setiap butir soal benar-benar mengukur kompetensi yang dimaksud. Validitas yang baik mencerminkan kesesuaian antara isi soal dengan indikator pembelajaran yang telah ditetapkan (Rahayu, 2021).

Selanjutnya, reliabilitas instrumen diuji menggunakan koefisien Cronbach Alpha dan menghasilkan nilai sebesar 0,8469. Nilai ini menunjukkan bahwa instrumen memiliki konsistensi internal yang tinggi, yang berarti bahwa hasil tes dapat dipercaya untuk mengukur pencapaian belajar siswa secara konsisten (Kartika, Wahyuni, & Artina, 2025).

Analisis tingkat kesukaran menunjukkan bahwa dari 20 butir soal, 18 berada dalam kategori sedang, 2 dalam kategori mudah, dan tidak ada soal yang tergolong sulit. Komposisi ini dianggap ideal, karena soal dengan tingkat kesukaran sedang mampu membedakan siswa dengan kemampuan berbeda secara efektif, sementara soal mudah dapat meningkatkan kepercayaan diri siswa (Rahayu, 2021).

Daya pengecoh, atau efektivitas distraktor, dianalisis untuk memastikan bahwa opsi jawaban yang salah dapat menarik siswa yang kurang memahami materi. Hasil menunjukkan bahwa opsi A diterima pada seluruh 20 soal, opsi B dan C masing-masing ditolak pada 1 soal, opsi D ditolak pada 2 soal, dan opsi E ditolak pada 3 soal. Efektivitas distraktor yang baik penting untuk meningkatkan daya pembeda soal (Rahayu, 2021).

Daya pembeda soal dianalisis untuk menilai kemampuan soal dalam membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Hasil menunjukkan bahwa 9 soal memiliki daya pembeda sangat baik, 4 soal baik, 6 soal cukup, dan 1 soal kurang baik. Soal dengan daya pembeda tinggi sangat penting untuk mengidentifikasi tingkat pemahaman siswa secara akurat (Rahayu, 2021).

Secara keseluruhan, instrumen evaluasi pembelajaran Geografi ini menunjukkan kualitas yang baik dalam hal validitas, reliabilitas, tingkat kesukaran, daya pengecoh, dan daya pembeda. Instrumen ini dapat digunakan sebagai alat evaluasi yang efektif untuk mengukur pencapaian belajar siswa dan memberikan umpan balik yang berguna bagi perbaikan proses pembelajaran.

Grafik 1. Tingkat Kesukaran



(Source : Analysis Result, 2025)

Grafik ini menunjukkan jumlah soal untuk setiap kategori tingkat kesukaran. terdapat 18 soal dengan tingkat kesukaran "Sedang" dan 2 soal dengan tingkat kesukaran "Mudah". Ini memberikan gambaran umum tentang distribusi kesulitan soal dalam tes.

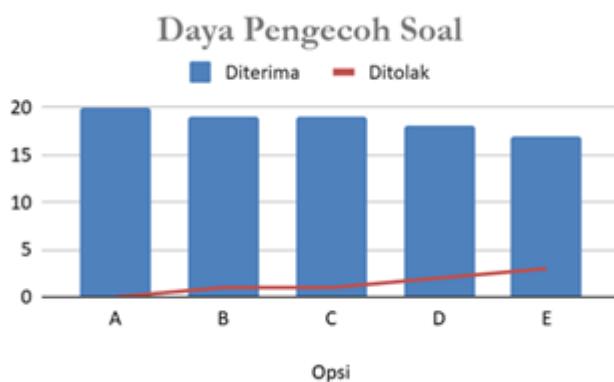
Grafik 2. Daya Pembeda



(Source : Analysis Result, 2025)

Grafik ini menunjukkan jumlah soal untuk setiap kategori daya pembeda. terdapat 9 soal dengan daya pembeda "Sangat baik", 6 soal dengan daya pembeda "Baik", dan seterusnya. Ini memberikan gambaran tentang seberapa efektif soal-soal tersebut dalam membedakan kemampuan siswa.

Grafik 3. Daya Pengecoh



(Source : Analysis Result, 2025)

Grafik ini menunjukkan seberapa banyak siswa yang memilih setiap pilihan jawaban, termasuk pengecoh. Pilihan jawaban yang baik sebagai pengecoh seharusnya dipilih oleh sebagian siswa, tetapi tidak sebanyak jawaban yang benar.

Kesimpulan

Hasil kegiatan menunjukkan bahwa guru mampu menginterpretasikan nilai validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh secara benar menggunakan data empiris siswa. Nilai Cronbach's Alpha sebesar 0,8469 menunjukkan konsistensi tinggi, sedangkan 90% soal tergolong sedang dan 85% pengecoh berfungsi efektif. Temuan ini membuktikan bahwa kegiatan pelatihan dan pendampingan berbasis riset efektif dalam meningkatkan pemahaman guru terhadap prinsip evaluasi pembelajaran berbasis data, sesuai dengan tujuan program pengabdian yang telah dirumuskan. Kegiatan serupa disarankan untuk diperluas ke bidang studi lain, terutama yang memerlukan kemampuan analisis instrumen evaluasi seperti Fisika, Kimia, dan Biologi. Selain itu, perlu dikembangkan modul pelatihan lanjutan dengan integrasi perangkat lunak statistik sederhana seperti Google Sheets atau SPSS agar guru dapat melakukan analisis butir soal secara mandiri dan berkelanjutan. Hasil kegiatan ini juga diharapkan menjadi dasar bagi penyusunan kebijakan pelatihan nasional dalam bidang evaluasi pembelajaran berbasis riset, guna memperkuat kualitas asesmen pendidikan di sekolah menengah.

Referensi

- Fauziah, A., Arikunto, S., & Daryanto. (2020). Analisis kualitas butir soal hasil belajar peserta didik. *Jurnal Pendidikan dan Pembelajaran*, 5(2), 45-60. <https://ojs.unkriswina.ac.id/index.php/Edusavana/article/download/671/394/3363>
- Fitriyani, N., & Hidayat, W. (2020). Analysis of item validity, reliability, difficulty level, and discrimination index of mathematics test. *Journal of Physics: 1657(1), Conference Series*, 012022. <https://doi.org/10.1088/1742-6596/1657/1/012022>
- Gliem, J. A., & Gliem, R. R. (2003, October). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *MidwestResearch-to-Practice Conference in Adult, Continuing, and Community Education*.

- Kartika, D., Wahyuni, I., & Artina, A. (2025). Analisis kesulitan mahasiswa pendidikan geografi dalam memahami materi ajar SMA berbasis kurikulum merdeka di kampus UNADA Banda Aceh. *Journal of Smart Education and Learning*, 2(1). 1-9. [<https://journal.nurscienceinstitute.id/index.php/jsel/article/download/1648/558>]
- Kumar, D., Singh, M., & Sharma, N. (2021). *Item analysis of multiple-choice questions: A quality assurance measure*. Advances in Medical Education and Practice, 12, 1049–1056. <https://doi.org/10.2147/AMEP.S287293>
- Lestari, A. S., Fitrianna, A. Y., & Zanthy, L. S. (2023). Analisis butir soal tes materi sistem persamaan linear dua variabel pada siswa kelas VIII. JPMI Jurnal Pembelajaran Matematika Inovatif, 6(1), 367-376. <https://journal.ikipsiliwangi.ac.id/index.php/jpmi/article/view/12389/5756>
- Lions, S., Fedeli, P., & Papi, S. (2021). *The position of distractors in multiple-choice test items*. Frontiers in Education, 6, 731763. <https://doi.org/10.3389/feduc.2021.731763>
- Munandar, A., Maryani, E., Rohmat, D., & Ruhimat, M. (2019). Assessment on Geography Field study at Universities in Indonesia 19(1), 3-9. vol
- Nurgiyantoro, B., Gunawan, W., & Marzuki, M. (2021). Pengembangan soal pilihan ganda dan analisis kualitas butir soal. Jurnal Kependidikan: Penelitian Inovasi Pembelajaran, 55(1), 10-22. <https://doi.org/10.21831/jk.v55i1.34567>
- Prawiki, R. A. (2022). Analisis Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Biologi Kelas X SMA Negeri 1 Teluk Sebong. Biodidaktika: Jurnal Biologi dan Pembelajarannya, 17(2), 120-130. <https://doi.org/10.22437/bio.v17i2.1.6493>
- Rahayu, S. (2021). Analisis butir soal elemen prinsip-prinsip dan konsep geografi. Jurnal Pendidikan dan Kajian Pembelajaran*, 4(1), 45-56. [<https://journalpedia.com/1/index.php/jpkp/article/download/4057/4234>]
- Rezigalla, A. A., Osman, A. H., & Hamid, A. H. (2024). *Distractor efficiency and its effect on the difficulty and discrimination indices of multiple-choice questions*. Education Sciences, 14(2), 214–224. <https://doi.org/10.3390/educsci14020214>
- Ruhimat, M. (2018). Kompetensi Pembuatan Instrumen Pengukuran Hasil Belajar Oleh Guru IPS SMP di Kota Bandung. Jurnal Pendidikan. vol 21(3), 176-187.
- Sajjad, M., Khan, M. A., & Ahmad, S. (2020). *Nonfunctional distractor analysis: An indicator for quality multiple-choice questions*. Advances in Medical Education and Practice, 11, 541–547. <https://doi.org/10.2147/AMEP.S254153>
- Sitorus, M. T., Tumanduk, T. M., & Rotty, V. E. (2020). Analisis Kualitas Butir Soal Ulangan Harian Biologi Kelas X SMA Negeri 1 Remboken. Jurnal Sains Matematika dan Edukasi (JSME), 8(2), 113-120. <https://ejournal.unima.ac.id/index.php/jsme/article/view/294>
- Sugiyono. (2021). Metode penelitian pendidikan (Pendekatan kuantitatif, kualitatif, dan R&D). Bandung: Alfabeta..
- Wahyuni, S., & Prasetyo, Z. K. (2022). Item analysis of multiple choice questions: Difficulty, discrimination, and distractor analysis. International Journal of Instruction, 15(3), 1-16. <https://doi.org/10.29333/iji.2022.1531a>
- Wardoyo, W., & Supraptono, E. (2014). Evaluasi Pembelajaran. Yogyakarta: UNY Press.