

ANALISIS BUTIR ESAI 10-ITEM PADA KELAS XI SMA MUHAMMADIYAH BATUDAA 2025/2026

Nur'Ain Pratami A. Konijo^{1*}, Tedy Machmud², Emli Rahmi³

^{1,2,3}Magister Pendidikan Matematika, Pascasarjana, Universitas Negeri Gorontalo,
Jl. Jendral Sudirman No.6, Dulalowo, Kota Gorontalo, Gorontalo, 96128, Indonesia
e-mail: ¹ainkonijo091101@gmail.com, ²tedy_m@ung.ac.id, ³emlirahmi@ung.ac.id

*Penulis Korespondensi

Diserahkan: 27-08-2025; Direvisi: 11-09-2025; Diterima: 29-09-2025

Abstrak: Penelitian ini dilakukan untuk mengevaluasi kualitas item matematika pada topik Sistem Persamaan Linear Tiga Variabel dalam *e-book* interaktif berbasis infografis untuk kelas XI SMA Muhammadiyah Batudaa tahun ajaran 2025/2026. Evaluasi dilakukan dengan menganalisis empat aspek utama, yaitu tingkat kesukaran, daya pembeda, validitas, dan reliabilitas. Pendekatan menggunakannya adalah deskriptif kuantitatif, data berasal dari lembar jawaban siswa pada tes esai sebanyak 10 soal. Studi ini melibatkan 27 siswa kelas XIA. Hasil penelitian membuktikan bahwa 50% soal memiliki daya pembeda baik, 10% cukup, 30% kurang, dan 10% kurang sekali. Dilihat dari tingkat kesulitan, 30% soal tergolong mudah, 60% sedang, dan 10% tergolong sukar. Analisis validitas menunjukkan bahwa 50% butir soal dinyatakan valid, sedangkan 50% sisanya tidak valid. Hasil uji reliabilitas yang dilakukan menggunakan rumus Cronbach's Alpha menunjukkan nilai koefisien sebesar 0,803, yang mencakup dalam kategori yang sangat tinggi, menunjukkan bahwa Instruksi memiliki konsistensi yang baik di dalamnya. Berdasarkan temuan ini, disarankan untuk memasukkan pertanyaan yang valid dan berkualitas tinggi ke dalam bank pertanyaan, sementara pertanyaan yang tidak valid atau tidak efektif sebaiknya direvisi atau diganti. Evaluasi dan uji coba instrumen secara berkala sangat disarankan untuk memastikan kualitas penilaian yang adil dan akurat.

Kata Kunci: Butir soal; kesukaran; pembeda; validitas; reliabilitas

Abstract: *In the 2025–2026 academic year, this research is to assess the quality of the math exam problems on the subject of three-variable linear equation systems in the interactive e-book for grade XI at SMA Muhammadiyah Batudaa that is based on infographics. Four primary factors were examined in the evaluation: the degree of difficulty, validity, dependability, and discriminating power. Student response sheets from a ten-question essay test served as the data source for the quantitative descriptive methodology. Twenty-seven XIA students served as the study's subjects. 50% of the questions had strong discriminating power, 10% were sufficient, 30% were poor, and 10% were extremely poor, according to the analysis's findings. Thirty percent of the questions were categorized as easy, sixty percent as moderate, and ten percent as challenging. According to validity analysis, half of the items were deemed valid and the other half were deemed invalid. The instrument has strong internal consistency, as indicated by the reliability test findings using the Cronbach's Alpha formula, which produced a coefficient value of 0.803, falling into the very high range. These results suggest that although incorrect or ineffective questions should be changed or replaced, valid and high-quality questions should be added to the question bank. To guarantee fair and accurate assessment quality, regular evaluation and instrument trials are strongly advised.*

Keywords: *Items; difficulty; differentiators; validity; reliability*

Kutipan: Konijo, N. P., Machmud, T., & Rahmi, E. (2025). Analisis Butir Esai 10-Item Pada Kelas XI SMA Muhammadiyah Batudaa 2025/2026. *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, Vol.11 No.2, (1546-1558). <https://doi.org/10.29100/jp2m.v11i2.9701>



Pendahuluan

Pendidikan menjadi bagian penting dari proses pembangunan sumber daya manusia yang kompetitif dan unggul. Pendidikan merupakan proses yang disusun secara sadar dan terstruktur untuk memfasilitasi pembelajaran yang meliputi pengetahuan, keterampilan, dan sikap, dengan tujuan mendukung pengembangan potensi individu serta mencerdaskan generasi masa depan (Magdalena *et al.*, 2021). Dalam dunia pendidikan, evaluasi berfungsi sebagai instrumen utama untuk menilai efektivitas dan pencapaian hasil pembelajaran (Augustia & Augustia, 2025). Dalam kegiatan pembelajaran, tujuan, proses, dan evaluasi merupakan tiga komponen yang saling terhubung dan tidak dapat dipisahkan. Seluruh aktivitas pembelajaran, termasuk pemahaman terhadap materi, selalu difokuskan pada pencapaian tujuan pembelajaran. Keberhasilan pencapaian tujuan tersebut hanya dapat diketahui melalui evaluasi yang disusun dengan mempertimbangkan materi dan tujuan yang telah dipelajari (Dwipayani, 2013). Evaluasi memungkinkan guru untuk menilai tingkat pencapaian tujuan pembelajaran telah tercapai dan mengidentifikasi kesulitan yang dialami siswa selama proses pembelajaran. Hasil evaluasi tersebut menjadi dasar dalam merancang perbaikan strategi pembelajaran atau memberikan dukungan khusus kepada siswa yang membutuhkannya (Magdalena *et al.*, 2021).

Kata 'evaluasi' secara etimologis terkait dengan bahasa Inggris *evaluation*, yang diturunkan dari kata dasar *value*, yang memiliki arti nilai atau harga. Istilah 'nilai' dalam bahasa Arab disebut *alqiyamah* atau *al-taqdir*, yang keduanya memiliki arti penilaian atau evaluasi. Secara harfiah, istilah evaluasi pendidikan dalam bahasa Arab disebut *al-taqdir at-tarbiyah*, yang merujuk pada proses penilaian dalam konteks pendidikan maupun terhadap berbagai kegiatan yang berkaitan dengan proses pembelajaran. Menurut M. Chabib Thoha, evaluasi adalah proses yang disusun secara sistematis untuk mengetahui kondisi suatu objek melalui instrumen tertentu, lalu membandingkan hasil dengan standar yang telah ditetapkan guna memperoleh kesimpulan. Secara umum, evaluasi merupakan proses yang terstruktur untuk menilai dan menentukan nilai atau mutu suatu aspek tertentu baik itu kebijakan, aktivitas, keputusan, kinerja, proses, individu, objek, dan lain sebagainya berdasarkan kriteria yang telah ditetapkan (Magdalena *et al.*, 2020). Evaluator dapat menilai suatu objek dengan cara langsung membandingkannya terhadap kriteria umum yang berlaku, atau dengan melakukan pengukuran terlebih dahulu terhadap objek tersebut sebelum dibandingkan dengan kriteria tertentu.

Evaluasi merupakan proses sistematis yang bertujuan untuk mengukur dan menilai suatu objek, kegiatan, atau hasil belajar berdasarkan kriteria tertentu. Dengan menggunakan instrumen yang tepat, evaluasi membantu dalam memperoleh data yang objektif untuk menentukan sejauh mana tujuan atau standar yang telah ditetapkan tercapai. Dalam konteks pendidikan, evaluasi berperan penting untuk mengidentifikasi keberhasilan pembelajaran serta memberikan informasi diagnostik yang berguna bagi perbaikan proses pembelajaran dan pengambilan keputusan. Oleh karena itu, evaluasi menjadi komponen esensial dalam upaya meningkatkan mutu pendidikan secara menyeluruh (Asrul *et al.*, 2022).

Evaluasi dalam kegiatan pembelajaran mencakup tiga domain utama: psikomotorik, afektif, dan kognitif. Ranah kognitif menilai penguasaan pengetahuan dan kemampuan berpikir siswa, yang umumnya diukur melalui berbagai bentuk tes seperti pilihan ganda, esai, atau pertanyaan pengertian. Ranah afektif berhubungan dengan penilaian terhadap sikap, nilai, serta respons emosional siswa terhadap materi pelajaran dan proses belajar. Sementara itu, ranah psikomotorik mengukur kemampuan siswa dalam melakukan keterampilan praktis atau aktivitas fisik tertentu. Penilaian yang mencakup

ketiga ranah ini secara menyeluruh memungkinkan guru memperoleh pemahaman yang utuh terhadap hasil belajar siswa (Augustia & Agustia, 2025).

Menurut Akhmadi M. N. (2021) Selain berfungsi untuk menilai kemajuan belajar peserta didik, evaluasi juga memberikan pijakan yang kuat dalam merumuskan kebijakan dan keputusan pendidikan. Dengan menganalisis hasil evaluasi secara kualitatif dan kuantitatif, guru dapat mengenali aspek-aspek kuat dan lemah dalam pembelajaran, sehingga mampu menyusun langkah-langkah perbaikan yang lebih efektif. Dalam ranah pendidikan, evaluasi dilakukan melalui instrumen yang dapat berupa tes maupun nontes. Menurut Nurkancana dan Sunartana, tes didefinisikan sebagai metode penilaian yang disajikan sebagai tugas-tugas yang perlu dikerjakan oleh siswa baik secara mandiri maupun berkelompok, dengan tujuan mengungkap perilaku atau hasil belajar mereka (Magdalena *et al.*, 2021).

Menurut Kamus Besar Bahasa Indonesia (KBBI) Tes adalah Pelaksanaan tes tertulis, tes lisan, atau wawancara yang digunakan untuk mengukur pengetahuan, keterampilan, bakat, serta kepribadian seseorang; disertai dengan uji coba untuk menilai kelayakan operasional kendaraan bermotor umum. Menurut Grondlund & Linn (1990), tes adalah alat atau proses yang disusun secara sistem untuk mengevaluasi contoh perilaku tertentu. Aswar (2000) mendefinisikan tes sebagai serangkaian pertanyaan yang memerlukan tanggapan, dipilih, atau ditanggapi dari peserta, dengan tujuan menilai aspek-aspek khusus dari perilaku individu tersebut. Mardapi (2008) menyebutkan bahwa tes terdiri dari beberapa pertanyaan yang memerlukan jawaban atau tanggapan, yang berfungsi untuk mengevaluasi kapasitas atau mengungkap bagian tertentu dari peserta. Sedangkan Hamid (2019) menyatakan bahwa tes merupakan instrumen yang memuat kumpulan pertanyaan atau pekerjaan yang harus diselesaikan diselesaikan oleh peserta (Hamid, 2019).

Tes merupakan instrumen atau prosedur yang dirancang secara sistematis berupa kumpulan pertanyaan atau tindakan yang wajib dilakukan peserta untuk mencapai tujuan utama untuk mengukur dan menilai berbagai aspek perilaku, kemampuan, dan karakteristik individu secara akurat. Tes adalah suatu bentuk evaluasi dalam pendidikan yang dilakukan dengan memberikan tugas-tugas tertentu kepada siswa, yang bertujuan untuk mengungkap informasi terkait pencapaian hasil belajar mereka. Tes berfungsi sebagai instrumen evaluatif yang berperan krusial dalam mengevaluasi tingkat pencapaian siswa terhadap tujuan pembelajaran. Tes yang memenuhi kriteria validitas, reliabilitas, daya pembeda, dan tingkat kesukaran dianggap berkualitas. Validitas menyangkut Reliabilitas berkaitan dengan konsistensi hasil tes dan seberapa akurat tes menilai hal yang seharusnya diukur, dan analisis butir soal mengkaji karakteristik kuantitatif setiap soal seperti tingkat kesukaran, daya pembeda (Rahmad & Wijaya, 2020).

Daya pembeda ditentukan oleh sejauh mana suatu soal dapat membedakan siswa berdasarkan tingkat pemahaman mereka, sementara tingkat kesukaran berkaitan dengan aspek teknis yang mempengaruhi mutu soal dalam menilai pemahaman siswa. Tingkat kesukaran (*difficulty index*) adalah ukuran yang menunjukkan seberapa mudah atau sulit suatu soal bagi peserta tes (Ramírez & Brodhead, 2013). Indeks ini penting untuk mengetahui apakah suatu soal dapat dimanfaatkan untuk menilai kemampuan siswa secara efektif. Tingkat kesukaran soal adalah proporsi peserta yang memberikan jawaban tepat pada satu item tes. Makin tinggi nilai indeksnya, makin mudah soal tersebut. Sebaliknya, makin rendah nilainya, makin sulit soal tersebut. Validitas merupakan tingkat ketepatan dan ketelitian suatu instrumen dalam mengukur aspek yang memang menjadi tujuannya. Dengan kata lain, suatu alat dianggap valid jika benar-benar mengukur aspek atau kompetensi yang hendak diukur, bukan aspek lainnya. Reliabilitas merujuk pada derajat keyakinan terhadap hasil sebuah pengukuran. Suatu pengukuran dikatakan reliabel apabila menghasilkan data yang konsisten atau relatif sama ketika dilakukan berulang kali pada kelompok subjek yang sama, selama kondisi atau karakteristik yang diukur tidak mengalami perubahan (Fitra & Maksum, 2021).

Melalui tes, pendidik dapat memperoleh gambaran mengenai tingkat penguasaan konsep, keterampilan berpikir kritis, dan kemampuan reflektif siswa terhadap materi pembelajaran (Ebrahimi, 2019). Secara umum, tes dapat dibedakan menjadi dua bentuk utama tes objektif seperti pilihan ganda, dan tes subjektif seperti esai. Tes pilihan ganda banyak digunakan karena efisien dalam penilaian massal, mudah dianalisis secara statistik, serta memberikan hasil yang reliabel. Analisis butir (*item analysis*) pada tes pilihan ganda biasanya dilakukan menggunakan teori tes klasik (*Classical Test Theory, CTT*) melalui parameter *item difficulty* dan *discrimination index*, atau menggunakan *Item Response Theory (IRT)* yang lebih kompleks (Gao *et al.*, 2016). Sebaliknya, butir soal esai memiliki karakteristik yang berbeda secara fundamental. Respons esai bersifat terbuka (*constructed response*), memungkinkan variasi jawaban yang luas, baik dari segi strategi penyelesaian, kedalaman argumentasi, maupun representasi matematis yang digunakan siswa. Beberapa studi internasional menegaskan bahwa pendekatan analisis butir yang lazim digunakan pada pilihan ganda tidak sepenuhnya kompatibel ketika diterapkan langsung pada soal esai tanpa penyesuaian metodologis (Ebrahimi, 2019). Tantangan utama terletak pada aspek subjektivitas penskoran, variasi antarpenilai (*inter-rater variability*), serta kesulitan dalam mengontrol konsistensi skor. Penelitian Dugue (2019) menekankan bahwa reliabilitas tes esai sangat dipengaruhi oleh kualitas rubrik penskoran dan konsistensi penerapan rubrik tersebut. Oleh karena itu, dalam konteks soal esai, reliabilitas internal sering kali lebih tepat diukur menggunakan Alpha Cronbach, sementara analisis validitas butir memerlukan interpretasi yang lebih hati-hati karena skor butir sangat bergantung pada konstruksi rubrik dan konteks soal.

Soal matematika materi sistem persamaan linear tiga variabel pada e-book interaktif berbasis infografis kelas X yang di ujikan di SMA Muhammadiyah Batudaa, Karena belum melalui proses uji coba, soal-soal dalam instrumen evaluasi tersebut belum dapat dipastikan tingkat validitas dan reliabilitasnya pada setiap butir. Langkah ini dilakukan untuk memastikan kerahasiaan soal tetap terjaga hingga waktu pelaksanaan tes, guna mencegah terjadinya kebocoran. Karena itu, penting dianalisis terhadap mutu soal, mencakup aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, serta mengevaluasi sejauh mana butir-butir soal tersebut mampu mencerminkan pencapaian siswa terhadap kompetensi dasar yang ditetapkan pada semester tersebut. Secara praktis, daya pembeda butir soal menunjukkan tingkat kemampuan suatu soal dalam secara akurat membedakan antara peserta yang meraih skor tinggi dan peserta yang memperoleh skor rendah. Evaluasi terhadap daya pembeda ini dapat dilakukan melalui berbagai indikator, baik secara eksplisit maupun implisit.

Guru melakukan proses analisis butir soal sebagai langkah untuk mengevaluasi mutu setiap soal yang digunakan dalam penilaian. Hasil dari analisis ini memungkinkan guru untuk memutuskan soal mana yang dapat dipertahankan, perlu direvisi, atau harus dihapus sesuai dengan standar kualitas yang berlaku (Anshari *et al.*, 2024). Tujuan dari analisis butir soal dilakukan guna memperbaiki mutu butir soal sekaligus mendapatkan informasi diagnostik mengenai kemampuan siswa. Soal yang berkualitas adalah yang dapat memberikan data secara tepat, sehingga mampu menentukan siswa yang telah memahami materi dan yang belum.

Berdasarkan penjelasan sebelumnya, sebagian besar penelitian analisis butir soal masih berfokus pada instrumen pilihan ganda, sementara kajian empiris terhadap butir soal esai matematika relatif terbatas, khususnya pada konteks evaluasi pembelajaran di sekolah menengah. Penelitian yang secara khusus mengkaji analisis butir soal esai dengan pendekatan kuantitatif klasik (meliputi tingkat kesukaran, daya pembeda, validitas, dan reliabilitas) masih jarang dilakukan secara komprehensif, terutama pada instrumen yang terintegrasi dalam media pembelajaran digital seperti e-book interaktif berbasis infografis. Padahal, penggunaan media digital dalam pembelajaran menuntut instrumen evaluasi yang tidak hanya menarik secara visual, tetapi juga terbukti berkualitas secara psikometrik. Dalam konteks pembelajaran matematika, khususnya materi Sistem Persamaan Linear Tiga Variabel (SPLTV), sebagian besar evaluasi masih menggunakan soal objektif, sehingga potensi soal esai dalam mengungkap proses berpikir, strategi penyelesaian, dan kedalaman pemahaman siswa belum

dimanfaatkan secara optimal. Akibatnya, informasi diagnostik yang dihasilkan dari penilaian sering kali belum mencerminkan kemampuan berpikir tingkat tinggi siswa secara menyeluruh. Oleh karena itu, dirumuskan masalah utama yang menjadi pusat perhatian penelitian ini, yaitu: bagaimana kualitas butir soal Matematika materi Sistem Persamaan Linear Tiga Variabel pada e-book interaktif berbasis infografis kelas X SMA Muhammadiyah Batudaa berdasarkan analisis kuantitatif (tingkat kesukaran, daya pembeda), Sejauh mana validitas tes dapat dibuktikan secara empiris dan konseptual, Bagaimana konsistensi internal tes berdasarkan reliabilitas. Tujuan utama Studi ini bertujuan untuk menganalisis kualitas instrumen tes berdasarkan karakteristik butir soal, Menilai validitas dan reliabilitas tes secara teoritis dan empiris, dan Memberikan rekomendasi pengembangan instrumen tes yang lebih baik di masa depan.

Metode

Dalam penelitian ini diterapkan pendekatan deskriptif kuantitatif yang berfokus pada evaluasi data hasil tes matematika materi sistem persamaan linear tiga variabel. Data yang dianalisis berupa lembar jawaban peserta didik pada tes essay berjumlah 10 soal. Hal ini bertujuan menarik kesimpulan dari fenomena yang tampak selama pelaksanaannya, dengan melakukan observasi, telaah, serta pemaparan terhadap objek yang dikaji, menggunakan data kuantitatif sebagai dasar analisis (Sapitri *et al.*, 2022). Metode ini dimanfaatkan untuk mengevaluasi mutu butir-butir soal Matematika pada materi sistem persamaan linear tiga variabel pada e-book interaktif berbasis infografis kelas X. Subjek dalam penelitian ini terdiri atas 27 siswa yang dipilih menggunakan teknik sampling jenuh (total sampling) karena seluruh kelas XI-A di SMA Muhammadiyah digunakan sebagai subjek penelitian. Sementara itu, objek penelitian ini meliputi analisis terhadap validitas isi, daya pembeda, tingkat kesulitan, serta reliabilitas dari setiap butir soal.

Langkah-langkah prosedural dalam penulisan dan analisis ini mencakup:

Uji Daya Pembeda

Daya pembeda adalah indikator yang merefleksikan kapasitas suatu butir soal dalam mengklasifikasikan peserta didik berdasarkan perbedaan tingkat penguasaan materi (Akhmadi, M. N., 2021). Suatu soal dinyatakan memiliki daya pembeda yang baik apabila mampu memberikan skor tinggi kepada siswa dengan tingkat kemampuan tinggi dan skor rendah siswa yang memiliki kemampuan rendah. Daya pembeda yang tinggi menandakan soal berkualitas karena dapat membedakan tingkat pemahaman siswa. Sebaliknya, indeks daya pembeda negatif (kurang dari 0) menunjukkan ketidaksesuaian, di mana siswa yang memiliki kemampuan rendah justru lebih sering menjawab benar dibanding siswa dengan tingkat kemampuan tinggi (Mawardi *et al.*, 2023). Berikut ini adalah tabel klasifikasi kategori daya pembeda.

Tabel 1. Kriteria Daya Pembeda

Kriteria Daya Pembeda	
< 0.00	Sangat Jelek
0.01 – 0.20	Jelek
0.21 – 0.40	Cukup
0.41 – 0.70	Baik
>0.71	Sangat Baik

Namun, penelitian ini menggunakan metode langsung, yaitu dengan membandingkan skor kelompok atas dan bawah. Nilai indeks daya beda suatu butir soal dilambangkan dengan huruf DP, dan perhitungannya dilakukan menggunakan rumus sebagai berikut.

$$DP = \frac{\bar{X}_A - \bar{X}_B}{SMI} \tag{1}$$

Keterangan:

DP = indeks daya beda

\bar{X}_A = rata-rata skor kelompok atas

\bar{X}_B = rata-rata skor kelompok bawah

SMI = skor maksimum (Arikunto, 2018)

Dalam analisis daya pembeda, peserta tes dibagi menjadi kelompok atas dan kelompok bawah berdasarkan peringkat skor total tes (Alfansuri *et al.*, 2025). Prosedur pembagian dilakukan dengan menggunakan pendekatan persentil ekstrem, yang lazim digunakan dalam analisis butir berbasis CTT. Sebanyak 27% siswa dengan skor tertinggi ditetapkan sebagai kelompok atas dan 27% siswa dengan skor terendah ditetapkan sebagai kelompok bawah. Dengan jumlah responden 27 orang, maka masing-masing kelompok terdiri atas 7 siswa, sementara siswa pada kelompok tengah tidak dilibatkan dalam perhitungan daya pembeda. Penggunaan batas 27% didasarkan pada pertimbangan statistik bahwa proporsi tersebut mampu memaksimalkan perbedaan rerata skor antara kelompok atas dan bawah, sehingga estimasi daya pembeda menjadi lebih sensitif dan stabil. Selisih rerata skor antara kedua kelompok kemudian digunakan untuk menghitung indeks daya pembeda setiap butir soal esai. Pendekatan ini umum digunakan dalam penelitian evaluasi pendidikan dan dinilai sesuai untuk instrumen berskala kecil seperti tes kelas (Alfansuri *et al.*, 2025).

Uji Tingkat Kesukaran

Ukuran tingkat kesulitan soal ditetapkan berdasarkan banyaknya individu yang menjawab benar dibandingkan dengan total seluruh peserta. Soal yang baik memiliki tingkat kesulitan seimbang, yaitu tidak tergolong mudah sampai semua peserta bisa menjawab dengan benar, namun juga hanya sedikit sulit peserta yang mampu menjawabnya (Bano *et al.*, 2022). Butir soal yang dikategorikan sulit sebaiknya direvisi atau diperbaiki. Demikian pula, soal yang terlalu mudah perlu dievaluasi kembali. Jika seluruh siswa dapat menjawab soal secara tepat, ini menunjukkan bahwa item tersebut tidak cukup menantang dan tidak efektif dalam mengukur pemahaman siswa. Oleh karena itu, guru dianjurkan untuk mengganti atau memperbaiki soal tersebut agar lebih menstimulasi dan dapat menilai penguasaan materi siswa dengan lebih tepat. Adapun soal yang berada pada tingkat kesulitan yang berkategori sedang dapat dipertahankan di dalam bank soal untuk digunakan pada tes mendatang. Berikut ini adalah tabel klasifikasi kategori tingkat kesukaran.

Tabel 2. Interpretasi Tingkat Kesukaran

Interpretasi Tingkat Kesukaran	
< 0.00	Terlalu Sukar
0.01 – 0.30	Sukar
0.31 – 0.70	Sedang
0.71 - 1	Mudah
= 1	Terlalu Mudah

dan perhitungannya dilakukan menggunakan rumus sebagai berikut

$$TK = \frac{\bar{x}}{SMI} \tag{2}$$

Keterangan:

TK = indeks tingkat kesukaran

\bar{x} = nilai rata-rata tiap butir soal

SMI = skor maksimum ideal (Arikunto, 2018).

Uji Validitas

Validitas empiris sering digunakan untuk memverifikasi validitas item soal berdasarkan data respon siswa melalui analisis statistik, biasanya dengan menghitung nilai korelasi antara skor suatu butir dan skor total. Rumus *Korelasi Pearson Product Moment (r)*:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}} \quad (3)$$

Keterangan:

X = skor butir soal

Y = skor total

N = jumlah peserta

Validitas menunjukkan seberapa besar suatu instrumen dapat secara mengukur secara benar aspek yang memang menjadi tujuannya, seperti dalam menilai hasil belajar siswa. Sebuah alat ukur dinyatakan valid apabila mampu mencerminkan tujuan pengukuran yang telah ditentukan dengan benar. Menentukan validitas berdasarkan tabel nilai kritis (*r-tabel*). Butir soal dengan *r-hitung* > *r-tabel* dinyatakan valid.

Uji Reliabilitas

Tujuan uji reliabilitas adalah untuk menilai seberapa jauh suatu instrumen tes menghasilkan output yang sama ketika diberikan pada kondisi dan subjek yang sama secara berulang. Instrumen yang memiliki reliabilitas baik akan memberikan hasil yang stabil, mencerminkan keandalan alat ukur dalam menilai objek yang dimaksud. *Alpha Cronbach* digunakan jika butir soal dalam bentuk essay.

Tabel 3. Interpretasi Koefisien Reliabilitas

Interpretasi Koefisien Reliabilitas	
< 0.20	Sangat Rendah
0.21 – 0.40	Rendah
0.41 – 0.60	Sedang
0.61 – 0.80	Tinggi
>0.81	Sangat Tinggi

Rumus *Alpha Cronbach*:

$$r_{11} = \frac{k}{k-1} \left[1 - \frac{\sum S_i^2}{S^2} \right] \quad (4)$$

Keterangan

r_{11} = koefisien reliabilitas (disebut juga reliabilitas internal)

k = jumlah butir soal

$\sum S_i^2$ = varians tiap butir soal

S^2 = varians total skor (Arikunto, 2018).

Penskoran soal esai dalam penelitian ini dilakukan menggunakan rubrik analitik, yang dirancang untuk mengurangi subjektivitas penilaian dan meningkatkan konsistensi skor. Setiap butir soal dilengkapi dengan rubrik yang memuat indikator kemampuan pemecahan masalah menurut Polya yaitu memahami masalah, menyusun rencana, melaksanakan rencana, meninjau kembali (Hidayah, S., *et al.*, 2023). Indikator memahami masalah diberi rentang skor 0 – 3, indikator menyusun rencana diberi rentang 0 – 2, indikator melaksanakan rencana diberi rentang skor 0 – 3, dan indikator meninjau kembali diberi rentang skor 0 – 2.

Untuk menjamin konsistensi dan keandalan skor, proses penskoran dilakukan dengan prosedur Penilai diberikan pelatihan awal, berupa penjelasan rubrik, contoh jawaban siswa, serta simulasi penskoran, Penilai melakukan penskoran bersama (*calibration scoring*) terhadap beberapa lembar

jawaban sebagai sampel, kemudian mendiskusikan perbedaan skor hingga diperoleh kesepakatan interpretasi rubrik, Setelah konsistensi awal tercapai, penskoran dilakukan secara sistematis terhadap seluruh lembar jawaban siswa dengan tetap berpedoman pada rubrik yang telah disepakati. Pendekatan ini bertujuan untuk meminimalkan bias subjektif dan variasi antarpemilai, yang merupakan salah satu tantangan utama dalam penilaian tes esai. Konsistensi skor yang dihasilkan selanjutnya diuji secara kuantitatif melalui koefisien reliabilitas *Alpha Cronbach*, sehingga keandalan instrumen dapat dibuktikan secara empiris.

Hasil dan Pembahasan

Dalam studi ini, SMA Muhammadiyah Batudaa Kelas XI-A yang berjumlah 27 siswa berfungsi sebagai lokasi dan sekaligus subjek yang dianalisis. Yang menjadi objek pada penelitian ini ialah aspek validitas isi, daya beda, tingkat kesukaran, dan reliabilitas dari butir-butir soal Matematika pada materi sistem persamaan linear tiga variabel pada e-book interaktif berbasis infografis tahun pelajaran 2025/2026 sejumlah 10 butir soal essay yang ditelaah.

Uji Daya Pembeda

Guna menilai kemampuan setiap butir soal dalam membedakan tingkat kemampuan peserta didik, dilakukan analisis terhadap daya pembeda setiap soal. Analisis ini penting guna menjamin bahwa soal-soal yang digunakan dalam penilaian dapat memberikan gambaran yang akurat mengenai kemampuan siswa. Berdasarkan hasil pengolahan data, diperoleh kategori-kategori daya pembeda yang menunjukkan kualitas setiap butir soal sebagai berikut:

Tabel 4. Analisis Daya Pembeda

Kriteria	Frekuensi	Persentase
Sangat Jelek	1	10%
Jelek	3	30%
Cukup	1	10%
Baik	5	50%
Sangat Baik	0	0%

Dari Tabel 4 berdasarkan hasil analisis daya pembeda terhadap butir-butir soal yang digunakan dalam penelitian ini, diperoleh gambaran mengenai kemampuan setiap soal dalam membedakan peserta didik berdasarkan tingkat kemampuannya. Berdasarkan kriteria yang telah ditetapkan, terdapat beberapa kategori daya pembeda yang menggambarkan kualitas butir soal secara keseluruhan. Sebanyak 10% item soal masuk dalam kategori dalam kategori daya pembeda berkategori sangat jelek, yakni pada soal nomor 9 dengan persentase 30% butir soal termasuk dalam kategori daya pembeda jelek yaitu soal nomor 1, 4, dan 6, yang menunjukkan bahwa soal-soal tersebut kurang efektif dalam membedakan antara peserta didik berkemampuan tinggi dan berkemampuan rendah. Butir dengan kategori ini direkomendasikan untuk direvisi, bukan langsung dihapus, terutama karena bentuk soal esai memungkinkan perbaikan pada redaksi, stimulus, atau tingkat tuntutan kognitif. Selanjutnya, 10% butir soal masuk dalam kategori daya pembeda cukup yaitu soal nomor 3, yang berarti soal tersebut memiliki kemampuan sedang dalam membedakan tingkat kemampuan siswa, meskipun belum optimal. Sementara itu, sebanyak 50% butir soal dikategorikan berkategori daya pembeda baik, yaitu pada soal nomor 2, 5, 7, 8, dan 10, yang menunjukkan menunjukkan bahwa butir-butir soal tersebut mampu secara efektif membedakan peserta didik berdasarkan kemampuan mereka. Dengan demikian, mayoritas butir soal yang dianalisis menunjukkan kualitas yang baik dalam hal daya pembeda, namun masih terdapat sebagian kecil soal yang perlu diperbaiki untuk meningkatkan efektivitas penilaian.

Temuan ini sejaland dengan penelitian Magdalena *et al.* (2021) dan Solichin (2017) yang melaporkan bahwa pada evaluasi kelas, proporsi butir dengan daya pembeda baik umumnya berada pada rentang 40–60%, khususnya ketika instrumen belum melalui uji coba awal. Kesamaan proporsi ini menunjukkan bahwa kualitas daya pembeda dalam penelitian ini masih berada dalam pola umum yang dilaporkan oleh studi sejenis.

Uji Tingkat Kesukaran

Guna menilai derajat kemampuan butir soal dalam mengukur kemampuan peserta didik secara optimal, dilakukan analisis tingkat kesukaran terhadap soal-soal pada soal matematika materi sistem persamaan linier tiga variabel pada e-book interaktif berbasis infograsi di kelas XI-A. Analisis ini bertujuan untuk mengidentifikasi proporsi soal yang tergolong mudah, sedang, maupun sukar, sehingga dapat dievaluasi keseimbangan dan mutu butir soal yang dipakai dalam mengukur capaian belajar peserta didik. Hasil analisis disajikan sebagai berikut.

Tabel 5. Analisis Tingkat Kesukaran

Kriteria	Frekuensi	Persentase
Terlalu Sukar	0	0%
Sukar	1	10%
Sedang	6	60%
Mudah	3	30%
Terlalu Mudah	0	0%

Berdasarkan Tabel 5 di atas, hasil analisis terhadap tingkat kesukaran butir soal dalam tes matematika materi sistem persamaan linear tiga variabel pada e-book interaktif berbasis infografis untuk kelas X yang dilakukan terhadap kelas XI-A, diperoleh informasi mengenai sebaran tingkat kesulitan soal yang digunakan. Hasil analisis menekankan bahwa sebesar 30% soal tersebut termasuk dalam tingkat kesulitan mudah diantaranya soal nomor 1, 2, dan 4. Soal-soal dalam kategori ini cenderung dapat dijawab mampu dijawab dengan benar oleh mayoritas peserta didik, sehingga kurang optimal dalam mengukur kemampuan secara menyeluruh. Selanjutnya, sebanyak 60% butir soal dikategorikan sebagai soal sedang diantaranya soal nomor 3, 5, 7, 8, 9, dan 10. Ini menunjukkan bahwa hampir separuh dari seluruh soal memiliki tingkat kesukaran yang seimbang, yang berarti memiliki efektivitas yang cukup dalam mengukur kemampuan siswa secara keseluruhan karena soal berada pada tingkat kesulitan sedang. Sementara itu, sebanyak 10% butir soal dikategorikan sebagai soal sukar, yakni soal nomor 6. Soal-soal ini mungkin hanya bisa dijawab oleh peserta didik dengan kemampuan tinggi, sehingga meskipun penting untuk membedakan siswa dengan penguasaan materi yang lebih baik, proporsi soal sukar yang tinggi dapat menyulitkan banyak siswa. Oleh karena itu, proporsi soal sedang yang mendominasi merupakan indikator positif, namun perlu diimbangi dengan evaluasi terhadap soal mudah dan sukar agar distribusi tingkat kesukaran lebih proporsional dan adil.

Hasil ini konsisten dengan temuan Saputri, H. A., *et al.* (2023) dan Bano *et al.* (2022) yang melaporkan bahwa instrumen evaluasi kelas yang baik umumnya didominasi oleh butir berkategori sedang, karena butir pada kategori ini paling efektif dalam mendukung daya pembeda. Secara kuantitatif, butir dengan tingkat kesukaran sedang dalam penelitian ini juga cenderung memiliki nilai daya pembeda lebih tinggi dibandingkan butir yang terlalu mudah atau terlalu sukar.

Uji Validitas

Untuk menjamin bahwa setiap butir soal dalam ujian benar-benar mampu mengukur kompetensi yang menjadi tujuan pembelajaran, dilakukan analisis validitas secara sistematis dengan memanfaatkan perangkat lunak Microsoft Excel. Analisis ini menjadi langkah penting dalam proses evaluasi instrumen, karena validitas berperan sebagai indikator utama dari keabsahan hasil pengukuran. Validitas mengacu

pada seberapa jauh suatu soal mampu mencerminkan atau merepresentasikan kemampuan yang seharusnya diukur, sehingga hasil tes benar-benar mencerminkan pencapaian peserta didik yang sesungguhnya. Sebuah instrumen dinyatakan valid apabila data yang dihasilkan mampu memberikan gambaran yang akurat dan dapat dipercaya mengenai performa peserta tes.

Melalui perhitungan hubungan antara skor tiap butir soal dengan skor total dapat diketahui sejauh mana keterkaitan dan kontribusi setiap soal terhadap keseluruhan hasil tes. Dengan demikian, analisis validitas ini tidak sekadar bermanfaat untuk mengetahui kualitas soal, tetapi juga sebagai dasar untuk penyempurnaan dan perbaikan instrumen evaluasi di masa mendatang. Hasil lengkap dari analisis validitas disajikan pada tabel berikut, yang menunjukkan performa masing-masing soal berdasarkan tingkat keterkaitannya dengan keseluruhan instrumen tes.

Tabel 6. Analisis Validitas Butir

Kategori	Frekuensi	Persentase
Valid	5	50%
Gugur	5	50%

Berdasarkan hasil analisis yang dilakukan menggunakan perangkat lunak Microsoft Excel, dari total 10 butir soal yang dianalisis, diperoleh temuan bahwa sebanyak 5 soal dinyatakan valid. Soal-soal tersebut adalah nomor 2, 5, 7, 8, dan 10. Sementara itu, ada 5 soal yang tidak memenuhi kriteria validitas, yaitu soal nomor 1, 3, 4, 6, dan 9. Evaluasi validitas dilakukan melalui teknik hubungan antara skor setiap butir soal dan skor total, yang menggambarkan seberapa jauh suatu soal dapat mengukur aspek yang relevan dengan keseluruhan kompetensi yang diujikan.

Soal-soal yang telah terbukti valid memiliki potensi untuk dimasukkan ke dalam bank soal dan dapat dipakai kembali dalam kegiatan evaluasi pada periode selanjutnya. Sebaliknya, soal-soal yang tidak valid sebaiknya tidak digunakan kembali tanpa revisi, karena berisiko menghasilkan pengukuran yang tidak akurat terhadap kemampuan siswa. Oleh karena itu, soal-soal tidak valid perlu ditinjau ulang secara menyeluruh, baik dari segi materi, konstruksi kalimat, maupun tingkat kesesuaian dengan indikator pembelajaran. Analisis ini penting sebagai dasar untuk penyempurnaan kualitas instrumen penilaian secara berkelanjutan.

Proporsi validitas sebesar 50% ini sejalan dengan hasil penelitian Sari, A. P., (2024) dan Farida, *et al.* (2021) yang melaporkan bahwa pada instrumen yang baru pertama kali digunakan, jumlah butir valid sering kali berada pada kisaran 40–60%. Hal ini menguatkan bahwa hasil penelitian ini tidak bersifat anomali, melainkan mencerminkan kondisi empiris yang wajar pada tahap pengembangan awal instrumen.

Uji Reliabilitas

Untuk mengetahui sejauh mana instruksi tes konsisten dan keandalan untuk mengukur kemampuan peserta didik, dilakukan analisis reliabilitas terhadap seluruh butir soal yang dinyatakan valid yaitu berjumlah 5 soal. Uji reliabilitas merupakan langkah penting dalam evaluasi instrumen, karena bertujuan memastikan bahwa tes menghasilkan data yang dapat diandalkan dan stabil bila diujikan kembali dalam situasi yang sama. Analisis ini dilakukan menggunakan formula *alpha cronbach*, yang sesuai untuk instrumen tes berbentuk essay. Hasil uji reliabilitas tersebut disajikan dan dianalisis pada bagian berikut.

Tabel 7. Analisis Uji Reliabilitas

Jumlah Butir Soal	5
Jumlah varian butir	24,63
Varians total skor	88,69
<i>Alpha cronbach</i>	0.803

Berdasarkan perhitungan koefisien reliabilitas diperoleh dengan menggunakan rumus *alpha cronbach* sebesar 0.803. Nilai ini menunjukkan bahwa instrumen tes tergolong dalam kategori sangat tinggi dan cukup andal digunakan untuk mengukur kemampuan siswa. Dengan kata lain, 5 butir soal yang digunakan dalam ujian ini memiliki konsistensi internal yang baik, di mana antarbutir soal memiliki kesesuaian dan mengukur aspek yang sama secara relatif stabil. Nilai reliabilitas sebesar 0.803 juga mengindikasikan bahwa apabila tes ini diujikan kembali dalam kondisi serupa kepada kelompok siswa yang sama, maka hasilnya akan sebanding konsisten. Hal ini tentu menjadi indikator penting bahwa tes telah memenuhi salah satu kriteria kualitas alat ukur yang baik, yaitu reliabel. Tes yang reliabel dapat memberikan pemahaman yang tepat mengenai pencapaian belajar siswa dan mendukung pengambilan keputusan dalam proses evaluasi pembelajaran. Namun demikian, meskipun nilai reliabilitas berada dalam kategori sangat tinggi, pengembang soal tetap disarankan untuk meninjau kembali butir-butir soal secara individual, khususnya terhadap soal yang memiliki tingkat kesukaran atau daya beda yang kurang optimal. Perbaikan ini bertujuan untuk lebih meningkatkan kualitas instrumen di masa mendatang dan menjaga agar seluruh aspek penilaian berjalan dengan objektif dan bermakna.

Jika dibandingkan dengan penelitian sejenis, nilai reliabilitas ini relatif lebih tinggi dibandingkan temuan Ariyanti, E. & Bhakti (2020) dan Anshari *et al.* (2024) yang melaporkan koefisien reliabilitas pada rentang 0,65–0,78. Hal ini mengindikasikan bahwa meskipun jumlah butir relatif sedikit, penggunaan rubrik esai yang terstruktur dan konsisten mampu meningkatkan keandalan instrumen.

Kesimpulan

Berdasarkan hasil analisis terhadap 10 butir soal essay pada tes mata pelajaran Matematika materi sistem persamaan linear tiga variabel pada e-book interaktif berbasis infografis kelas X, dapat disimpulkan bahwa kualitas instrumen tes secara umum berada pada kategori baik. Dari aspek daya pembeda, mayoritas soal (50%) tergolong baik, menunjukkan bahwa sebagian besar soal mampu membedakan antara peserta didik berkemampuan tinggi dan rendah. Tingkat kesukaran soal menunjukkan proporsi yang cukup berimbang, dengan 30% tergolong mudah, 60% sedang, dan 10% sukar. Namun demikian, proporsi soal sukar yang cukup tinggi perlu mendapat perhatian agar tidak menghambat pengukuran capaian belajar secara menyeluruh. Dari segi validitas, sebanyak 5 soal (50%) dinyatakan valid berdasarkan analisis korelasi terhadap skor total, sedangkan sisanya (50%) tidak valid. Sementara itu, hasil uji reliabilitas menunjukkan nilai koefisien sebesar 0.803, yang menandakan bahwa instrumen memiliki konsistensi internal yang sangat tinggi dan layak digunakan sebagai alat ukur.

Sebagai tindak lanjut, butir-butir soal yang valid dan memiliki daya pembeda yang baik dapat dimasukkan ke dalam bank soal dan digunakan kembali pada penilaian berikutnya. Soal-soal yang tidak valid serta memiliki daya pembeda rendah atau negatif sebaiknya direvisi secara menyeluruh atau dihapus karena tidak efektif dalam mengukur kompetensi siswa. Selain itu, butir soal dengan tingkat kesukaran yang terlalu tinggi atau terlalu rendah perlu ditinjau kembali agar distribusi kesukaran soal menjadi lebih proporsional. Revisi yang dilakukan hendaknya memperhatikan kesesuaian soal dengan indikator pembelajaran, kejelasan redaksi, serta ketepatan konten terhadap kompetensi dasar yang diukur. Guru juga disarankan untuk melakukan uji coba terhadap instrumen sebelum digunakan secara resmi dan melaksanakan evaluasi secara berkala sebagai upaya peningkatan kualitas soal. Dengan demikian, instrumen yang digunakan benar-benar mampu memberikan gambaran yang akurat dan adil terhadap capaian belajar peserta didik.

Penelitian berikutnya disarankan untuk mengeksplorasi metode penilaian alternatif dalam mengevaluasi soal esai, seperti penggunaan rubrik berbasis level kognitif, penilaian berbantuan teknologi (*computer-assisted scoring*), atau penilaian ganda (*multi-rater scoring*) untuk menguji konsistensi antarpenilai. Pendekatan ini diharapkan dapat memperkuat objektivitas dan reliabilitas

penskoran, sekaligus memberikan gambaran yang lebih komprehensif mengenai kemampuan pemecahan masalah dan penalaran matematis peserta didik. Dengan dilaksanakannya penelitian lanjutan sebagaimana direkomendasikan, diharapkan pengembangan instrumen evaluasi berbasis e-book interaktif tidak hanya menghasilkan soal yang menarik secara visual, tetapi juga instrumen penilaian yang semakin valid, reliabel, dan berdaya guna dalam mengukur capaian belajar peserta didik secara adil dan akurat.

Daftar Pustaka

- Alfansuri, D. U., Hasibuan, S. W., Azizah, N., & Mustikawati, L. (2025). *Kualitas Soal: Tingkat Kesukaran Daya Pembeda Pengecoh Soal*. 1(4), 92–101. <https://jurnal.permata-edukasi.com/TLRJ/article/view/81/68>
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964–975. <https://doi.org/10.31004/edukatif.v6i1.5931>
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. PT. Bumi Aksara. https://books.google.co.id/books?id=j5EmEAAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Ariyanti, E. & Bhakti, Y. B. (2020). Perbandingan Bentuk Tes Pilihan Ganda dan Teknik Penskoran Terhadap Reliabilitas Tes Mata Pelajaran Kimia. *Titian Ilmu: Jurnal Ilmiah Multi Sciences*, 12(2), 66–76. <https://doi.org/10.30599/jti.v12i2.627>
- Asrul, Sarigih, A. H., & Mukhtar. (2022). Evaluasi Pembelajaran. In *Perdana Publishing*. [http://repo.iain-tulungagung.ac.id/5510/5/BAB 2.pdf](http://repo.iain-tulungagung.ac.id/5510/5/BAB%202.pdf)
- Augustia, A. D., & Agustia, C. N. (2025). Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan. 5(1), 250–265. <https://doi.org/10.57235/arrumman.v1i2.4398>
- 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. <https://doi.org/10.32884/ideas.v8i1.660>
- Dugue, L. (2019). 2'-Deoxyribonucleoside 5'-triphosphates bearing 4-phenyl and 4-pyrimidinyl imidazoles as DNA polymerase substrates. *Organic & Biomolecular Chemistry*, 17(2). <https://doi.org/10.1039/c8ob02464b>
- Dwipayani, A. A. (2013). Analisis Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Bahasa Indonesia Kelas X.D SMA N 1 Terhadap Pencapaian Kompetensi. *Jurnal Pendidikan Bahasa dan Sastra UNDIKSHA*, 1(5), 1-18. <https://ejournal.undiksha.ac.id/index.php/JJPBS/article/view/578/463>
- Ebrahimi, A. (2019). *User Behavior Evaluation in Robot-Assisted Retinal Surgery*. 174–179. <https://doi.org/10.1109/ROMAN.2018.8525638>.User
- Farida, Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrif: Jurnal Pendidikan Bahasa Arab*, 1(1), 34–44. <https://doi.org/10.32923/al-muarrif.v1i1.2100>
- Fitra, J., & Maksum, H. (2021). Efektivitas Media Pembelajaran Interaktif dengan Aplikasi Powtoon pada Mata Pelajaran Bimbingan TIK. *Jurnal Pedagogi dan Pembelajaran*, 4(1), 1. <https://doi.org/10.23887/jp2.v4i1.31524>
- Gao, Z., Putnam, A. A., Bowers, H. A., Guenther, U., Ye, X., Kindsfather, A., Hilliker, A. K., & States, U. (2016). *Coupling between the DEAD-box RNA helicases Ded1p and eIF4A*. 1–22. <https://doi.org/10.7554/eLife.16408>
- Hamid, A. (2019). *Penyusunan Tes Tertulis (Paper and Pencil Test)*. Uwais Inspriasi Indonesia. https://books.google.co.id/books?hl=en&lr=&id=V2azDwAAQBAJ&oi=fnd&pg=PA1&dq=pengetian+tes&ots=MySPkRaRcO&sig=fHJv0KjZ2fpgbaX-acQiOA6Gydw&redir_esc=y#v=onepage&q&f=false

- Hidayah, S., Purwoko, R. Y., Ngazizah, N. (2023). *Analisis Kemampuan Pemecahan Masalah Berdasarkan Teori Polya*. 1(1), 155–161. <https://jurnal.permataedukasi.com/TLRJ/article/view/81/68>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nopus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas Iii Sdn Karet 1 Sepatan. *BINTANG : Jurnal Pendidikan dan Sains*, 3(2), 198–214. <https://ejournal.stitpn.ac.id/index.php/bintang>
- Magdalena, I., Ridwanita, A., & Aulia, B. (2020). Evaluasi belajar peserta didik. *Jurnal Pendidikan dan Dakwah*, 2(1), 117–127. <https://ejournal.stitpn.ac.id/index.php/pandawa>
- Mawardi, M. sholeh, Fuady, A., & Sunismi, S. (2023). Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang. *Wahana*, 75(1), 31–41. <https://doi.org/10.36456/wahana.v75i1.6820>
- Mochammad Noor Akhmadi. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, 6(1), 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>
- Rahmad, E., & Wijaya, A. (2020). Keefektifan pembelajaran matematika realistik ditinjau dari kemampuan pemodelan matematika dan prestasi belajar. *Pythagoras: Jurnal Pendidikan Matematika*, 15(1), 100–110. <https://doi.org/10.21831/pg.v15i1.34593>
- Ramírez, R., & Brodhead, D. (2013). *Utilization Focused Evaluation A primer for evaluators*. July, 132.
- Sapitri, A., Kurniati, T., & Yuliawati, A. (2022). Analisis Kualitas Soal UAS Biologi SMA Kelas X dan XI MIA. *Bioeduca : Journal of Biology Education*, 4(1), 45–56. <https://doi.org/10.21580/bioeduca.v4i1.8433>
- Saputri, H. A., Zuhijrah, Larasati, N. J., S. (2023). ANALISIS INSTRUMEN ASSESMEN : VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09(5), 2986–2995. <https://doi.org/10.36989/didaktik.v9i5.2268>
- Sari, A. P. I., Y. (2024). Validitas dan Reliabilitas Soal Ujian Sumatif Tengah Semester (STS) Kelas X Fase-E SMA Negeri 8 Palembang Tahun Pelajaran 2023 / 2024. *Alfabet: Jurnal Bahasa, Sastra, dan Pembelajarannya*, 7(2), 337–347. <https://doi.org/10.33503/alfabet.v7i2.134>
- Solichin, M. (2017). ANALISIS DAYA BEDA SOAL, TARAF KESUKARAN, VALIDITAS BUTIR TES, INTERPRETASI HASIL TES DAN VALIDITAS RAMALAN DALAM EVALUASI PENDIDIKAN. *Dirasat: Jurnal Manajemen & Pendidikan Islam*, 2(2), 192–213. [https://download.garuda.kemdikbud.go.id/article.php?article=563543&val=9596&title=Analisis Daya Beda Soal Taraf Kesukaran Validitas Butir Tes Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan](https://download.garuda.kemdikbud.go.id/article.php?article=563543&val=9596&title=Analisis%20Daya%20Beda%20Soal%20Taraf%20Kesukaran%20Validitas%20Butir%20Tes%20Interpretasi%20Hasil%20Tes%20dan%20Validitas%20Ramalan%20dalam%20Evaluasi%20Pendidikan)