

Analisis butir soal matematika materi persamaan linier dua variabel berbasis Google Sites di SMP Negeri 3 Telaga

Nurlaila Kaune^{1)*}, Hasan S Panigoro²⁾ Majid³⁾

Department of Mathematics Education, Faculty of Mathematics and Natural Sciences,
Gorontalo State University, Gorontalo, Indonesia

*Correspondence email : nelakaune1@gmail.com

(Received 25-11-2025, Reviewed 15-12-2025, Accepted 23-12-2025)

Abstract

This type of research is descriptive research that uses a quantitative approach. This study aims to determine the quality of mathematics test items for linear equations of two variables based on Google Sites at SMP Negeri 3 Telaga. The research sample consisted of 26 students with a research instrument in the form of a multiple-choice test of 10 questions. Data collection was carried out through online tests available on Google Sites media and analysis using quantitative methods to determine the quality of the test items. The results of the analysis show that the quality of the test items from the level of validity of the questions, which are valid by 90% and invalid by 10%. From the level of reliability of the questions with a reliability value of 0.825 is declared reliable. When viewed in terms of the level of difficulty, most of the questions are in the medium category as much as 55.56%, while 44.44% are classified as easy. Meanwhile, from the aspect of discriminating power, the results of the analysis show that 77.78% of the test items are included in the very good category, 11.11% are in the good category, and another 11.11% are in the sufficient category (needs improvement). The research findings provide a significant contribution to the development of mathematics learning evaluation instruments, with important implications for improving test quality and enhancing the learning outcome assessment process.

Keywords: Item analysis, Multiple choice questions, Learning evaluation, Validity, Reliability

Abstrak

Jenis penelitian ini adalah penelitian deskriptif yang menggunakan pendekatan kuantitatif. Penelitian ini bertujuan pada upaya untuk mengetahui kualitas butir soal matematika untuk materi persamaan linier dua variabel berbasis *Google Sites* di sekolah SMP Negeri 3 Telaga. Sampel penelitian terdiri dari 26 siswa dengan instrumen penelitian berupa tes pilihan ganda sebanyak 10 butir soal. Pengumpulan data dilakukan melalui tes *online* yang terdapat pada media *Google Sites* dan analisis menggunakan metode kuantitatif untuk menentukan kualitas butir soal. Hasil analisis menunjukkan bahwa kualitas butir soal dari tingkat validitas soal, yang valid sebesar 90% dan tidak valid sebesar 10%. Dari tingkat reliabilitas soal dengan nilai reliabilitas sebesar 0.825 dinyatakan reliabel. Jika dilihat dari segi tingkat kesukaran, sebagian besar soal berada pada kategori sedang sebanyak 55,56%, sedangkan 44,44% tergolong mudah. Sedangkan dari aspek daya pembeda, hasil analisis menunjukkan bahwa 77,78% butir soal termasuk dalam kategori sangat baik, 11,11% berkategori baik, dan 11,11% lainnya berada pada kategori cukup (perlu perbaikan). Temuan penelitian memberikan kontribusi signifikan dalam pengembangan instrumen evaluasi pembelajaran matematika, dengan implikasi penting bagi perbaikan kualitas tes dan peningkatan proses penilaian hasil belajar.

Kata kunci: Analisis butir soal, Soal pilihan ganda, Evaluasi pembelajaran, Validitas, Reliabilitas

PENDAHULUAN

Bentuk evaluasi yang paling umum digunakan dalam pendidikan adalah tes pilihan ganda karena dianggap efisien dalam pelaksanaan dan penilaiannya (Hasidik & Utami, 2025). Hasil belajar yang diukur guru melalui tes memberikan informasi mengenai capaian belajar peserta didik, yang mengacu pada standar penilaian sekolah untuk menilai sejauh mana program pembelajaran yang telah diterapkan berjalan efektif (Rahmawati & Rahman, 2025). Namun, kualitas tes sangat ditentukan oleh mutu butir soal yang menyusunnya (Santoso & Badawi, 2022). Soal yang baik adalah soal yang dapat menggambarkan keadaan siswa pada tingkat kemampuan dan keterampilan tertentu. Menurut Qodir et al., (2024).

Kelemahan butir soal tidak terletak pada bentuk atau tipe butir soal, tetapi lebih banyak ditentukan oleh butir soal yang dikonstruksi dengan baik atau tidak baik. Oleh karena itu, analisis butir soal menjadi langkah esensial dalam memastikan bahwa setiap soal yang digunakan mampu mengukur kompetensi secara akurat, adil, dan bermakna. Menurut De Mars (2020), dalam konteks Teori Tes Klasik, analisis butir soal diperlukan untuk menilai karakteristik soal yang akan memengaruhi skor akhir peserta didik dan keputusan pembelajaran yang diambil.

Putriani et al. (2020) menyatakan bahwa evaluasi merupakan sebuah komponen untuk mengetahui hasil kompetensi siswa pada pembelajaran yang telah diberikan. Permasalahan yang muncul dalam praktik pendidikan saat ini adalah masih banyaknya soal yang digunakan dalam evaluasi pembelajaran belum melalui proses analisis yang sistematis. Padahal, salah satu komponen krusial dalam proses pendidikan adalah evaluasi pembelajaran, yang berfungsi untuk menilai sejauh mana tujuan pembelajaran telah tercapai (Rahman dkk., 2025).

Mewujudkan sebuah sumber daya manusia yang berkualitas tidak lepas dari bidang pendidikan. Manusia membutuhkan pendidikan sebagai upaya dalam memberikan pemahaman kepada dirinya sendiri yang dimana belum baik diperbaiki dengan cara diberikan pemahaman (Fathanm et al., 2024). Terdapat hubungan yang erat antara kemampuan peserta didik dan kualitas pendidikan jika kemampuan peserta didik tinggi, maka kualitas pendidikan juga akan baik (Ikawati et al., 2024). Salah satu komponen yang ada didalamnya adalah evaluasi. Evaluasi merupakan sebuah komponen untuk mengetahui hasil kompetensi siswa pada pembelajaran yang telah diberikan (Putriani et al. 2020). Evaluasi dilakukan kepada siswa untuk mengetahui ketercapai siswa dalam pembelajaran secara kognitif, afektif dan psikomotorik (Boiliu & Handayani, 2024). Kriteria yang perlu dipenuhi dalam analisis butir soal pilihan ganda diantaranya aspek

validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh atau distraktor (Augustia & Agustia, 2025).

Salah satu solusi inovatif yang dapat diterapkan dalam pembelajaran matematika adalah penggunaan *Google Sites* sebagai media pembelajaran dan evaluasi. *Google Sites* merupakan media yang ideal untuk menyampaikan materi kepada peserta didik karena mampu memberikan informasi tanpa batasan ruang dan waktu. Karakteristik tersebut memungkinkan peserta didik mengakses materi maupun soal evaluasi secara fleksibel, sehingga berpotensi meningkatkan motivasi belajar dan kemandirian siswa dalam memahami konsep matematika, khususnya pada materi Persamaan Linear Dua Variabel.

Namun demikian, kemudahan penggunaan *Google Sites* sebagai media evaluasi pembelajaran perlu diimbangi dengan kualitas instrumen penilaian yang digunakan. Soal-soal matematika yang disajikan melalui *Google Sites* harus memenuhi kriteria instrumen yang baik agar mampu mengukur kemampuan peserta didik secara objektif dan akurat. Oleh karena itu, analisis butir soal menjadi langkah penting untuk memastikan bahwa soal yang disajikan melalui *Google Sites* memiliki tingkat validitas, reliabilitas, tingkat kesukaran, dan daya pembeda yang sesuai. Tanpa adanya analisis butir soal, penggunaan media digital berpotensi hanya berfungsi sebagai sarana penyajian, tanpa memberikan informasi yang tepat mengenai pencapaian hasil belajar peserta didik.

Beberapa penelitian dahulu telah mengkaji kualitas soal dan kesulitan peserta didik dalam menyelesaikan soal matematika. Penelitian Panoe et al., (2025) menunjukkan bahwa analisis butir soal pada materi transformasi menghasilkan kualitas soal yang tergolong tinggi dari aspek validitas dan reliabilitas. Meskipun demikian, penelitian tersebut masih terbatas pada materi transformasi dan belum mengaitkan analisis butir soal dengan penggunaan platform pembelajaran berbasis digital. Penelitian Maryani & Setiawan (2021) mengungkapkan bahwa peserta didik kelas VIII masih mengalami kesulitan dalam memahami konsep, memodelkan soal cerita, dan menggunakan metode penyelesaian pada materi PLDV. Temuan ini menegaskan bahwa PLDV merupakan materi yang menantang bagi siswa, namun penelitian tersebut belum mengaitkan kesulitan siswa dengan kualitas butir soal yang digunakan dalam evaluasi.

Selanjutnya, penelitian Arsandy et al., (2025) penelitian berbasis tahapan Polya menunjukkan bahwa kesulitan siswa dalam menyelesaikan soal cerita PLDV terjadi secara bertahap, namun belum mengkaji karakteristik butir soal sebagai sumber kesulitan. Sementara itu, kajian berdasarkan Taksonomi SOLO menemukan dominasi kesalahan pada level multistruktural, tetapi masih terbatas pada klasifikasi kesalahan tanpa mengaitkannya dengan kualitas soal evaluasi. Penelitian Lunemo et al., (2025) juga mengidentifikasi Penelitian tersebut mengidentifikasi berbagai jenis kesalahan siswa

dalam menyelesaikan soal PLDV, dengan dominasi kesalahan fakta dan prosedural, namun belum mengkaji kontribusi karakteristik soal terhadap munculnya kesalahan tersebut.

Oleh karena itu, penelitian ini perlu dilakukan untuk menganalisis kualitas butir soal matematika materi Persamaan Linear Dua Variabel yang disajikan melalui *Google Sites* di SMP Negeri 3 Telaga. Penelitian ini diharapkan dapat memberikan gambaran mengenai tingkat validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal, serta menjadi dasar dalam penyempurnaan instrumen evaluasi berbasis digital. Dengan demikian, hasil penelitian ini tidak hanya berkontribusi pada peningkatan kualitas evaluasi pembelajaran matematika, tetapi juga mendukung optimalisasi pemanfaatan teknologi dalam pembelajaran di sekolah menengah pertama

METODE PENELITIAN

Jenis penelitian ini adalah penelitian deskriptif yang menggunakan pendekatan kuantitatif dikarenakan penelitian ini bertujuan untuk mengetahui kelayakan soal ujian semester mata pelajaran matematika pada materi system persamaan linier dua variable berbasis *Google Sites* di SMP Negeri 3 Telaga tahun pelajaran 2025/2026, yang memenuhi kriteria validitas dan reliabilitas. Teknik dalam pengumpulan data menggunakan dokumentasi dan tes tertulis, dilakukan agar mendapatkan data yang dibutuhkan berupa daftar nama siswa, kisi-kisi dan soal, hasil jawaban seluruh siswa kelas VIII SMP Negeri 3 Telaga. Analisis data ini dilakukan dengan mendata nilai yang di dapatkan siswa dalam bentuk tabulasi data, dan selanjutnya peneliti menentukan validitas dan reliabilitas. Analisis ini bertujuan agar tes yang diberikan kepada peserta didik sesuai dengan kemampuan mereka dalam menyerap materi, tingkat kesulitannya, serta keabsahan butir soal yang disajikan (Savika & Zuhriyah, 2024).

Subjek dalam penelitian ini adalah siswa dan siswi kelas VIII SMP Negeri 3 Telaga sebanyak 26 siswa. Sedangkan objek dalam penelitian ini adalah analisis kualitas butir soal pilihan ganda mata pelajaran matematika pada materi Sistem Persamaan Linier Dua Variabel semester ganjil kelas VIII di SMP Negeri 3 Telaga. Teknik analisis data pada butir-butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP Negeri 3 Telaga dengan mencari validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Masing-masing akan dihitung dengan program *Microsoft Excel*. Adapun kriterianya sebagai berikut:

Validitas

Menurut Arikunto (dalam Wahyuni 2025) Untuk menghitung Validitas item pada soal pilihan ganda dapat menggunakan rumus seperti:

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

r_{pbi} = koefisien korelas biserial

M_p = rerata skor dari subjek yang menjawab betul bagi item yang dicari validitasnya

M_t = rerata skor total

SD_t = standar deviasi dari skor total proporsi

$$p = \frac{\text{banyaknya siswa yang menjawab benar}}{\text{jumlah seluruh siswa}}$$

P = proporsi siswa yang menjawab benar

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

Hasil dari perhitungan koefisien korelasi biserial kemudian dibandingkan dengan pada taraf signifikan 5% dan disesuaikan dengan jumlah peserta tes. Apabila $r_{pbi} > r$ tabel maka soal tersebut valid.

Adapun tolak ukur untuk menginterpretasikan derajat validitas instrumen, ditentukan berdasarkan kriteria pada **Tabel 1** (Arif, 2014).

Tabel 1. Kriteria Validitas Butir Soal

Validitas Butir	Kriteria Validitas Butir
$0.80 \leq r_{xy} \leq 1.00$	Sangat Tinggi
$0.60 \leq r_{xy} \leq 0.79$	Tinggi
$0.40 \leq r_{xy} \leq 0.59$	Cukup
$0.20 \leq r_{xy} \leq 0.39$	Rendah
$0.00 \leq r_{xy} \leq 0.19$	Sangat Rendah

Reliabilitas

Menurut (Arikunto, 2013). Untuk mencari reliabilitas pada soal pilihan ganda dapat digunakan rumus sebagai berikut:

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

r_{11} : realibilitas tes keseluruhan

p : proporsi subjek yang menjawab item yang benar

$\sum pq$: jumlah perkalian antara p dan q

q : proporsi subjek yang menjawab item dengan salah

n : jumlah item

S : Standar deviasi atas dasar tes (akar varians)

Kriteria untuk menentukan tinggi rendahnya reliabilitas sebuah perangkat tes, menurut Guilford dalam (Wahyuni, 2025) pada **Tabel 2**.

Tabel 2. Kriteria Realibilitas Butir Soal

Interpretasi	Koefisien Reliabilitas (r)
Sangat Rendah	$0,00 \leq r < 0,20$
Rendah	$0,20 \leq r < 0,40$
Sedang/cukup	$0,40 \leq r < 0,60$
Tinggi	$0,60 \leq r < 0,80$
Sangat Tinggi	$0,80 \leq r < 1,00$

Prosedur penelitian yang dilakukan dalam penelitian ini mencakup beberapa tahapan, yaitu: (1) penyusunan instrumen berdasarkan kisi-kisi soal yang relevan dengan indikator pencapaian kompetensi, (2) Uji coba intrumen pada subjek penelitian (3) pengumpulan data melalui link online yang terdapat pada media *Google Sites* yang diberikan sekolah untuk mempermudah akses dan efisiensi, (4) deskripsi hasil data yang diperoleh, (5) analisis data menggunakan *Microsoft Excel*, serta (6) penarikan kesimpulan berdasarkan hasil analisis (Ramadhan et al., 2023).

Tingkat Kesukaran

Menurut Rachmawati & Pardana (2025) Tingkat kesukaran diperlukan untuk mengetahui sejauh mana siswa dapat menjawab soal kelangkaan dengan benar, apakah soal tersebut terlalu mudah, sedang, atau terlalu sulit. Besarnya indeks kesukaran antara 0,00-1,00. Indeks kesukaran ini menunjukkan taraf kesukaran soal, sehingga soal dengan indeks 0,00 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah. Dalam evaluasi, indeks kesukaran ini diberi simbol P. singkatan dari "proporsa".

Angla indeks kesukaran item dapat diperoleh dengan rumus yang dikemukakan oleh Du Bois, yaitu :

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

Keterangan :

P : Proporsi indeks kesukaran item

Np : Banyaknya siswa yang menjawab benar

N : Jumlah siswa yang mengikuti tes hasil belajar

Daya Pembeda

Secara operasional dalam teori pengukuran klasik (*Classical Test Theory*, CTT), daya pembeda sering dihitung dengan membandingkan persentase peserta di kelompok atas (*upper*) dan kelompok bawah (*lower*) yang menjawab benar soal tersebut (metode “upper-lower 27 %”) atau melalui korelasi skor butir dengan skor total. Nilai daya pembeda berkisar antara $-1,00$ hingga $+1,00$; nilai positif mendekati $+1,00$ menunjukkan daya pembeda sangat baik, sedangkan nilai negatif menunjukkan butir soal bermasalah (lebih banyak peserta lemah menjawab benar daripada peserta kuat). Kriteria daya pembeda disajikan pada **Tabel 3**.

Tabel 3. Kriteria Daya Pembeda

Kriteria Daya Pembeda	
0,00 – 0,19	kurang baik
0,20 – 0,39	cukup baik
0,40 – 0,70	baik
0,71 – 1,00	sangat baik

HASIL DAN PEMBAHASAN

Validitas Soal

Analisis validitas pada penelitian ini menggunakan validitas isi, dimana perhitungan dilakukan dengan bantuan microsoft excel. Selanjutnya validitas butir soal ini dikorelasikan dengan nilai r_{tabel} . Jika $r_{hitung} > r_{tabel}$, maka butir soal valid, sebaliknya jika $r_{hitung} < r_{tabel}$ maka butir soal tidak valid. Dimana hasil dari analisis validitas butir soal pilihan ganda didapatkan butir-butir soal yang termasuk dalam kategori valid dan tidak valid dengan jumlah siswa (N) = 26 dan taraf signifikansi 5% yang tersaji pada **Tabel 4**.

Tabel 4. Distribusi Frekuensi dan Persentasi Validitas Soal UAS

Kelompok	Frekuensi	Nomor Soal	Presentasi (%)
Valid	9	1, 2, 4, 5, 6, 7, 8, 9, 10	90%
Tidak Valid	1	3	10%
Jumlah	10 soal	10 soal	100%

Berdasarkan hasil analisis validitas butir soal Ujian Semester Tahun Pelajaran 2024/2025, dari sepuluh butir soal yang dianalisis terdapat sembilan butir soal (90%) yang dinyatakan valid dan satu butir soal (10%) yang dinyatakan tidak valid. Hasil ini menunjukkan bahwa secara umum instrumen tes telah mampu mengukur hasil belajar peserta didik sesuai dengan tujuan pembelajaran. Butir soal yang valid dapat dipahami dengan baik oleh peserta didik dan sesuai dengan tingkat kemampuan mereka. Soal-soal tersebut mampu membedakan peserta didik yang menguasai materi dengan yang belum menguasai, sehingga skor setiap butir soal memiliki keterkaitan yang jelas dengan skor total tes.

Sebaliknya, butir soal nomor 3 dinyatakan tidak valid karena tidak mampu menggambarkan kemampuan peserta didik secara tepat. Sebagian besar peserta didik mengalami kesulitan memahami dan menyelesaikan soal tersebut, sehingga baik peserta didik berkemampuan tinggi maupun rendah menunjukkan pola jawaban yang relatif sama. Akibatnya, butir soal tersebut tidak mampu membedakan tingkat kemampuan peserta didik. Sebagai tindak lanjut, peneliti melakukan perbaikan dengan menyesuaikan tuntutan soal agar sesuai dengan kemampuan peserta didik serta memperjelas konteks soal. Apabila setelah direvisi butir soal tersebut tetap tidak valid, maka soal tersebut dikeluarkan dan digantikan dengan butir soal yang lebih sesuai.

Menurut Endah et al. (2023) menyatakan faktor-faktor yang mempengaruhi validitas adalah faktor internal dan eksternal, yang dimaksud internal pada tes seperti alur atau petunjuk tes yang disajikan memberikan makna yang tidak jelas, sedangkan faktor eksternal adalah waktu pengerjaan tesnya yang kurang cukup sehingga membuat siswa menjadi tegang dan gugup dikarenakan merasa tergesagesa pada saat mengerjakan soal dan bisa menimbulkan kecurangan. Hal ini menunjukkan bahwa validitas dan reliabilitas instrumen tidak hanya ditentukan oleh kualitas soal itu sendiri, tetapi juga oleh pendekatan pembelajaran dan pelaksanaan evaluasi yang mempertimbangkan kondisi internal dan eksternal siswa (Sujiwo et al., 2025).

Tabel 5. Analisis Kriteria Hasil Uji Validitas Tes

Butir Soal	r_{hitung}	r_{tabel}	Kategori
1	0.673	0.379	Tinggi
2	0.748	0.379	Tinggi
3	0.320	0.379	Rendah
4	0.619	0.379	Tinggi
5	0.434	0.379	Cukup
6	0.755	0.379	Tinggi
7	0.674	0.379	Tinggi
8	0.477	0.379	Cukup
9	0.389	0.379	Cukup
10	0.755	0.379	Tinggi

Berdasarkan **Tabel 5** dapat diketahui bahwa dari 10 butir soal yang dianalisis, terdapat 6 butir soal yang memiliki tingkat validitas kategori tinggi (butir 1, 2, 4, 6, 7, dan 10), yaitu dengan nilai r_{hitung} lebih besar dari r_{tabel} (0,379) dan berada di atas rentang 0,60. Butir-butir ini dapat dikatakan sangat baik dalam mengukur apa yang seharusnya diukur sesuai indikator yang ditetapkan. Selanjutnya, terdapat 3 butir soal yang berada pada kategori cukup (butir 5, 8, dan 9), yang berarti soal-soal tersebut cukup layak digunakan, meskipun masih memungkinkan untuk ditingkatkan kualitasnya, khususnya dari segi keselarasan dengan kompetensi yang diukur atau kejelasan perumusan soal. Sementara itu, terdapat 1 butir soal yang tergolong dalam kategori rendah (butir 3) yang artinya validitas soal-soal ini masih lemah dan tidak cukup baik dalam mencerminkan kemampuan yang ingin diukur. Soal-soal dengan validitas rendah tersebut perlu direvisi atau bahkan diganti, karena kemungkinan besar tidak sesuai dengan indikator pembelajaran, kurang jelas, atau tidak sesuai dengan tingkat kemampuan peserta didik.

Secara keseluruhan, sebagian besar butir soal sudah menunjukkan validitas yang memadai, namun masih ada beberapa soal yang perlu diperbaiki agar keseluruhan instrumen tes memiliki kualitas yang lebih optimal dalam mengukur kompetensi peserta didik. Tingginya validitas soal dapat berdampak positif pada proses pembelajaran. Guru akan mendapatkan gambaran akurat tentang kemampuan siswa, umpan balik menjadi kurang efektif, dan strategi pengajaran sulit disesuaikan dengan kebutuhan siswa. Menurut Khasanah et al., (2023) menyatakan bahwa butir soal yang memiliki nilai validitas tinggi secara jelas menunjukkan kemampuan dalam mengukur keterampilan siswa dengan tepat dan akurat. Di sisi lain, butir soal yang memiliki validitas sangat rendah memerlukan tindakan lanjutan, karena soal tersebut dianggap tidak valid. Soal-soal dengan validitas rendah dan sangat rendah perlu direvisi.

Secara teoritis, validitas yang rendah dapat dipengaruhi oleh beberapa faktor penting yang saling berkaitan. Pertama, adanya ketidaksesuaian antara perumusan soal dengan indikator kompetensi menyebabkan instrumen evaluasi tidak mampu mengukur aspek yang dituju secara tepat. Kedua, materi yang terlalu kompleks dan tidak sebanding dengan kemampuan kognitif siswa dapat menyulitkan mereka dalam memahami dan menjawab soal secara benar. Ketiga, konsep akuntansi yang bersifat abstrak dan tidak disajikan secara konkret dalam soal juga turut menyebabkan rendahnya tingkat validitas.

Reliabilitas Tes

Pengujian ini dilakukan untuk mengetahui apakah instrumen reliabel atau tidak. Penelitian ini, untuk uji reliabilitas dengan bantuan *Microsoft Excel*. **Tabel 6** merupakan hasil perhitungan reliabilitas.

Tabel 6. Hasil Analisis Reliabilitas

Jumlah Soal	10 Soal
Varians Total (St^2)	6.890
Reliabilitas Tes (r_{11})	0.825
Interpretasi	Sangat Tinggi

Hasil perhitungan reliabilitas menunjukkan bahwa instrumen yang digunakan memiliki tingkat keandalan yang sangat tinggi dalam mengukur kemampuan peserta didik. Koefisien reliabilitas sebesar 0,825 mengindikasikan bahwa instrumen mampu menghasilkan skor yang konsisten dan stabil ketika digunakan pada kondisi pengukuran yang relatif sama. Hal ini menunjukkan bahwa butir-butir soal dalam instrumen memiliki tingkat keseragaman internal yang baik, sehingga layak digunakan sebagai alat ukur dalam proses evaluasi pembelajaran.

Temuan ini sejalan dengan teori pengukuran yang menyatakan bahwa reliabilitas merupakan indikator utama konsistensi hasil pengukuran suatu instrumen. Menurut Azwar (2021), koefisien reliabilitas yang berada di atas 0,80 menunjukkan bahwa instrumen memiliki tingkat keandalan yang sangat baik dan dapat dipercaya untuk mengukur kemampuan yang sama secara berulang. Selain itu, Sugiyono (2022) menegaskan bahwa instrumen dengan reliabilitas tinggi mencerminkan stabilitas hasil pengukuran dan minimnya kesalahan pengukuran (*measurement error*).

Secara khusus, klasifikasi koefisien reliabilitas yang diperoleh juga sejalan dengan kriteria yang dikemukakan oleh Sundayana (2014), yang menyatakan bahwa nilai reliabilitas pada rentang 0,80 hingga kurang dari 1,00 termasuk dalam kategori sangat tinggi. Dengan demikian, hasil penelitian ini memperkuat pandangan bahwa instrumen evaluasi yang memiliki reliabilitas tinggi dapat memberikan informasi yang akurat dan konsisten mengenai kemampuan peserta didik, sekaligus mendukung kualitas proses penilaian dalam pembelajaran.

Tingkat Kesukaran

Soal dengan tingkat kesukaran yang terlalu tinggi atau terlalu rendah dianggap kurang layak digunakan. Soal yang sangat sulit dapat menurunkan motivasi peserta didik dalam mengerjakan, sedangkan soal yang terlalu mudah tidak mampu menstimulasi kemampuan berpikir kritis dan pemecahan masalah mereka (Nurhalimah et al., 2022). Berdasarkan hasil analisis butir soal pada elemen Job Profile/Peluang Pekerjaan & Usaha di Bidang Akuntansi yang disajikan pada **Tabel 7**.

Tabel 7. Hasil Analisis Tingkat Kesukaran

Kriteria Tingkat Kesukaran	No. Butir Soal	Jumlah	Persentase
Sedang	1,2,6,7, dan 8	5	55.56%
Mudah	3, 4, 5, dan 9	4	44.44%

Hasil analisis menunjukkan bahwa 55,56% butir soal berada pada kategori sedang, sedangkan 44,44% tergolong mudah. Komposisi ini menggambarkan bahwa tingkat kesukaran soal sudah proporsional dan mampu mengukur kemampuan peserta didik secara optimal. Temuan ini sejalan dengan teori evaluasi pendidikan yang menyatakan bahwa tes yang baik seharusnya didominasi oleh butir soal berkategori sedang, disertai dengan sebagian soal mudah dan sukar. Menurut Arikunto dalam (Wahyuni, 2025), butir soal dengan tingkat kesukaran sedang merupakan soal yang paling baik karena mampu membedakan peserta didik yang berkemampuan tinggi dan rendah secara lebih efektif.

Soal yang terlalu mudah cenderung tidak memberikan informasi yang cukup tentang kemampuan peserta didik, sedangkan soal yang terlalu sukar dapat menurunkan motivasi dan tidak mencerminkan kemampuan sebenarnya. Soal dengan tingkat kesukaran sedang dapat menilai pemahaman konsep secara mendalam, sementara soal mudah membantu peserta didik beradaptasi dengan konteks tes. Tidak ditemukannya soal dengan kategori sukar menunjukkan bahwa seluruh soal masih sesuai dengan kemampuan peserta. Dengan demikian, instrumen tes ini dapat dinyatakan layak digunakan sebagai alat evaluasi pembelajaran.

Daya Pembeda

Dalam soal berbentuk pilihan ganda, setiap alternatif jawaban harus memiliki kemampuan membedakan yang baik agar dapat mengidentifikasi perbedaan tingkat pemahaman dan kemampuan peserta didik dalam memilih jawaban yang benar (Merlin Purwati et al., 2021). Daya pembeda pada suatu butir soal berfungsi untuk menilai kualitas soal secara empiris serta mengetahui sejauh mana soal tersebut mampu membedakan peserta didik yang telah Daya pembeda soal diperoleh dari hasil perhitungan 10 butir soal dalam kategori sangat baik terdapat 7 butir soal dengan presentase 77.78%, soal dalam kategori baik terdapat 1 butir soal dengan presentase 11.11%, soal dengan kategori cukup dan perlu diperbaiki terdapat 1 butir soal dengan presentase 11.11%. Berikut disajikan dalam bentuk **Tabel 8**.

Tabel 8. Hasil Analisis Daya Pembeda

Kriteria Tingkat Kesukaran	No. Butir Soal	Jumlah	Persentase
Sangat Baik	1, 2, 3, 5, 6, 8, dan 9	7	77.78%
Baik	7	1	11.11%
Cukup (Perlu Perbaikan)	4	1	11.11%

Hasil analisis daya pembeda menunjukkan bahwa tujuh butir soal, yaitu nomor 1, 2, 3, 5, 6, 8, dan 9, berada pada kategori sangat baik dengan persentase sebesar 77,78%. Satu butir soal, yaitu nomor 7, termasuk dalam kategori baik dengan persentase 11,11%, sedangkan satu butir soal lainnya, yaitu nomor 4, berada pada kategori cukup atau perlu perbaikan dengan persentase yang sama, yaitu 11,11%. Temuan ini menunjukkan bahwa secara umum butir soal mampu membedakan peserta didik yang berkemampuan tinggi dan rendah secara efektif.

Secara teoretis, daya pembeda merupakan salah satu indikator utama kualitas butir soal karena menunjukkan kemampuan suatu soal dalam mengidentifikasi perbedaan tingkat penguasaan materi antar peserta didik. Azwar (2021) menegaskan bahwa butir soal dengan daya pembeda tinggi menunjukkan fungsi evaluatif yang optimal, karena peserta didik berkemampuan tinggi cenderung menjawab benar, sementara peserta didik berkemampuan rendah menjawab salah. Dengan demikian, dominasi butir soal berkategori sangat baik dalam penelitian ini mengindikasikan bahwa instrumen tes telah memenuhi kriteria kualitas pengukuran yang baik.

Lebih lanjut, Retnawati (2022) menyatakan bahwa daya pembeda yang baik mencerminkan keselarasan antara indikator soal, tingkat kesukaran, dan kemampuan peserta didik, sehingga soal tidak bersifat ambigu atau menebak. Hasil penelitian ini yang hanya menemukan satu butir soal dengan kategori cukup menunjukkan bahwa sebagian kecil soal masih memerlukan revisi, terutama dalam hal kejelasan stimulus atau kesesuaian tingkat kognitif yang diukur.

Temuan ini juga sejalan dengan pandangan Brookhart (2023) yang menyatakan bahwa butir soal dengan daya pembeda rendah umumnya disebabkan oleh ketidaktepatan konstruksi soal, alternatif jawaban yang kurang berfungsi, atau materi yang belum sepenuhnya dikuasai peserta didik. Oleh karena itu, keberadaan satu butir soal yang berada pada kategori cukup dalam penelitian ini merupakan hal yang wajar dan dapat diperbaiki melalui revisi redaksi atau penyesuaian indikator.

Dengan demikian, hasil penelitian ini memperkuat teori bahwa instrumen evaluasi yang baik ditandai oleh dominasi butir soal dengan daya pembeda tinggi, serta menunjukkan bahwa proses penyusunan soal yang dilakukan telah sesuai dengan prinsip-prinsip pengembangan tes yang valid dan reliabel.

SIMPULAN

Berdasarkan hasil analisis kualitas butir soal matematika semester ganjil kelas VIII SMP Negeri 3 Telaga, dari sepuluh butir soal yang dianalisis diperoleh sembilan butir soal (90%) berkategori valid dan satu butir soal (10%) tidak valid, dengan reliabilitas tes yang sangat tinggi sebesar 0,837. Dari aspek validitas, sebagian besar butir soal berada pada kategori tinggi hingga cukup, sedangkan dari tingkat kesukaran menunjukkan komposisi yang proporsional, yaitu kategori sedang (55,56%) dan mudah (44,44%). Sementara itu, hasil analisis daya pembeda menunjukkan dominasi kategori sangat baik (77,78%), yang mengindikasikan bahwa mayoritas butir soal mampu membedakan peserta didik berdasarkan tingkat penguasaan materi secara efektif.

Secara ilmiah, temuan ini menegaskan bahwa keterpaduan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda merupakan indikator utama kualitas instrumen evaluasi pembelajaran matematika. Instrumen tes yang dihasilkan tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai sarana diagnostik kemampuan peserta didik. Penelitian ini berimplikasi pada pentingnya analisis butir soal secara sistematis dalam pengembangan instrumen evaluasi, serta membuka peluang penelitian lanjutan melalui penerapan teori tes modern, pengujian pada konteks dan jenjang yang berbeda, dan analisis fungsi pengecoh secara lebih mendalam.

DAFTAR PUSTAKA

- Akhmadi, M. N. (2021). Analisis butir soal evaluasi Tema 1 kelas IV SDN Plumbungan menggunakan program Anates. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, 6(1), 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>
- Arsandy, Q. S., Siregar, A. R., Sinuhaji, R. D., Tabita, Z., Sinaga, V. R., & Siregar, B. H. (2025). Analisis kesulitan siswa dalam pemecahan soal cerita sistem persamaan linear dua variabel berdasarkan tahapan polya. *Papanda Journal of Mathematics and Science Research*, 4(2), 337–350. <https://doi.org/10.56916/pjmsr.v4i2.2432>
- Augustia, A. D., & Agustia, C. N. (2025). Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software Anates pada mata pelajaran Perpajakan. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250–265. <https://doi.org/10.37481/jmeb.v5i1.1165>.
- Azwar, S. (2021). *Reliabilitas dan validitas* (Edisi revisi). Pustaka Pelajar.
- Balau, M., Pesik, A., & Damai, I. W. (2021). Analisis kualitas butir soal buatan guru mata pelajaran matematika kelas VIII SMP Negeri Satap Matabulu Kabupaten Bolaang Mongondow Timur. *Marisekola: Jurnal Matematika Riset Edukasi dan Kolaborasi*, 2(1), 13–18. <https://doi.org/10.53682/marisekola.v2i1.1112>
- Boiliu, F. M., & Handayani, S. (2024). Implementasi manajemen pendidikan agama Kristen untuk meningkatkan hasil belajar siswa. *JMPK : Jurnal Manajemen Pendidikan Kristen*, 4(1), 01-15. <https://doi.org/10.51667/jmpk.v4i1.1723>

- Brookhart, S. M. (2020). *How to make decisions with different kinds of student assessment data*. ASCD.
- Brookhart, S. M. (2023). *Classroom assessment: Principles and practice for effective instruction* (9th ed.). Pearson Education.
- De Mars, C. (2020). *Classical test theory: Applications and insights*. Routledge.
- Dianova, F. R., & Anwar, N. (2024). Analisis butir uji validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal sumatif bahasa Arab SD Islam. *Jurnal Bahasa Daerah Indonesia*, 1(3), 1–13. <https://doi.org/10.47134/jbdi.v1i3.2863>
- Hidayati, K. & Nisa, Z.E.K., (2023). Analisis butir soal penilaian akhir semester matematika. *Aksioma: Jurnal Program Studi Pendidikan Matematika*, 12(3), 3516–3529. <http://dx.doi.org/10.24127/ajpm.v12i3.7575>
- Fathan, A., Yuliany, N., & Hasrianti, A. (2024). Perbandingan media evaluasi hasil belajar matematika siswa dengan menggunakan aplikasi Quizizz dan Kahoot pada siswa kelas XI IPA SMAN 1 Mamuju. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 8(2), 1853–1864. <https://doi.org/10.31004/cendekia.v8i2.2875>
- Hasidik, H., & Utami, N. P. (2025). Evaluasi pembelajaran pada mata pelajaran matematika: Antara harapan dan kenyataan di kelas. *Algoritma: Jurnal Matematika, Ilmu pengetahuan Alam, Kebumihan dan Angkasa*, 3(4), 127–132. <https://doi.org/10.62383/algoritma.v3i4.629>
- Ikawati, D. S., Agustin, E. W., Hakim, L., Pratiwi, V., & Surabaya, U. N. (2024). Analisis butir soal pilihan ganda pada siswa SMK menggunakan Anates. *Edukatif: Jurnal Ilmu Pendidikan*, 6(6), 6773–6781.
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2022). *Panduan pembelajaran dan asesmen*. Kemdikbudristek.
- Khasanah, I., Fuady, A., & Universitas Islam Malang. (2023). Analisis soal ulangan harian matematika bentuk pilihan ganda. *Mathema Journal*, 5(2).
- Lunemo, M., Kaunang, D., Regar, V. E., & Manado, U. N. (2025). Analisis kesalahan peserta didik dalam menyelesaikan soal-soal sistem persamaan linear. *Pasundan Journal of Mathematics Education : Jurnal Pendidikan Matematika*, 10(1), 18–24. Retrieved from <https://journal.unpas.ac.id/index.php/pjme/article/view/2444>
- Maryani, A., & Setiawan, W. (2021). Analisis kesulitan peserta didik kelas VIII dalam menyelesaikan soal sistem persamaan linear dua variabel (SPLDV) di MTs Atsauri Sindangkerta. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 5(3), 2619–2627 <https://doi.org/10.31004/cendekia.v5i3.915>
- Nurhalimah, S., Hidayati, Y., Rosidi, I., Wiwin, D., & Hadi, P. (2022). Hubungan antara validitas item dengan daya pembeda dan tingkat kesukaran soal pilihan ganda PAS. *Jurnal Natural Science Educational Research*, 4(3). <https://doi.org/10.21107/nser.v4i3.8682>
- Panoe, W., Machmud, T., & Majid. (2025). Analisis butir soal matematika materi transformasi berbasis Geogebra di SMPN 3 Suwawa. *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, 11(2), 1559–1570. <https://doi.org/10.31004/cendekia.v8i2.2875>

- Purwati, L. M., Arianty, R., Syakilah, D. M., Ridlo, S., & Susilaningsih, E. (2021). Analisis soal tes pilihan ganda berbasis higher order thinking skill menggunakan aplikasi Anates Windows versi 4.0.9 for Windows. *Jurnal Pendidikan UNIGA*, 15 (2), 460-473. <https://doi.org/10.52434/jp.v15i2.1287>
- Putriani, D., Turahmah, F., Sunarti, N. R., Ismarliana, P., & Walid, A. (2020). Analisis soal UAS Biologi 2018/2019 kelas X dan XI SMAN 11 Kota Bengkulu. *Journal of Biology Learning*, 2(1), 1–7. <https://doi.org/10.32585/jbl.v2i1.559>
- Qodir, A., Huda, N., & Hermina, D. (2024). Analisis butir tes: tingkat kesukaran, daya pembeda dan efektivitas pengecoh. *Al-Furqan: Jurnal Agama, Sosial, Dan Budaya*, 3(3), 1450-1467.
- Qomariyah, R. S., Putri, R. G. A. T. K., Putri, D. R., Putri, D. S., & Triya, M. R. (2022). Analisis tingkat kesukaran dan daya pembeda pada butir soal pilihan ganda mata pelajaran Bahasa Indonesia kelas V semester 1 SDN Kedungdalem 2. *Jurnal Pendidikan*, 1(2), 74–80. <https://doi.org/10.47233/jpst.v1i2.278>
- Rachmawati, D., & Pardana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE)*, 13(3), 273-284. <https://doi.org/10.26740/jupe.v13n3.p273-284>
- Rahman, N. A., Kristiyani, F., Ruhimat, M., Logayah, D. S., & Juli, I. (2025). Evaluasi pembelajaran analisis butir soal pilihan Majemuk materi pembentukan alam semesta. *Reskilling*, 1(2), 65_75-65_75.
- Rahmawati, N. S., & Rahman, M. F. (2025). Analisis butir soal objektif penilaian sumatif akhir semester gasal mata pelajaran ekonomi. *JlIP-Jurnal Ilmiah Ilmu Pendidikan*, 8(8), 8863-8873. <https://doi.org/10.54371/jiip.v8i8.8837>
- Rajagukguk, M. J. T., & Naibaho, D. (2023). Mampu memilih soal berdasarkan tingkat kesukaran. *Jurnal Pendidikan Sosial dan Humaniora*, 2(4). *Jurnal Pendidikan Sosial Dan Humaniora*, 2(4), 12736-12747. Retrieved from <https://publisherqu.com/index.php/pediaqu/article/view/701>
- Retnawati, H. (2022). *Teori dan aplikasi analisis butir*. Parama Publishing.
- Santoso, S., & Badawi, J. A. (2022). Analisis butir soal pilihan ganda tema pertumbuhan dan perkembangan makhluk hidup kelas III Madrasah Ibtidaiyah. *Jurnal Basicedu*, 6(2), 1678-1686. <https://doi.org/10.31004/basicedu.v6i2.2206>
- Savika, H. I., & Zuhriyah, I. A. (2024). Peran analisis butir soal terhadap kualitas soal, kompetensi guru, dan prestasi belajar peserta didik di sekolah dasar. *PANDU: Jurnal Pendidikan Anak dan Pendidikan Umum*, 2(2), 43-51. <https://doi.org/10.59966/pandu.v2i2.856>
- Sujiwo, I. K. A., Rasna, I. W., & Paramarta, I. K. (2025). Validitas dan reliabilitas instrumen penilaian dalam mata pelajaran bahasa Bali. *Widya Laksana*, 14(1), 16-23. <https://doi.org/10.23887/jwl.v14i1.73094>