

ANALISIS BUTIR SOAL MATA PELAJARAN EKONOMI: VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN, DAN DAYA PEMBEDA

Dian Rachmawati¹, Andika Bayu Pradana²

¹Universitas Negeri Malang, Malang, Indonesia, dian.rachmawati.fe@um.ac.id

²Universitas Negeri Malang, Malang, Indonesia, andika.bayu.2304316@students.um.ac.id

DOI

<https://doi.org/10.26740/jupe.v13n3.p273-284>

Article history

Received

3 July 2025

Revised

20 August 2025

Accepted

7 September 2025

How to cite

Rachmawati, D., & Pradana, 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>

Kata Kunci: validitas, reliabilitas, daya pembeda, tingkat kesukaran, butir soal

Keywords: *validity, reliability, discriminatory power, level of difficulty, test items*

Corresponding author

Dian Rachmawati

dian.rachmawati.fe@um.ac.id

Abstrak

Penelitian ini bertujuan untuk meningkatkan kualitas asesmen pada mata pelajaran ekonomi, khususnya materi kelangkaan pada kelas X SMA. Penelitian ini menggunakan uji validitas, reliabilitas, tingkat kesukaran dan daya pembeda untuk menganalisis 25 butir soal yang telah dikembangkan sebelumnya. Uji coba soal dilakukan pada 30 siswa X SMA. Dari hasil uji validitas dan reliabilitas menggunakan SPSS 26, sebanyak 22 dari 25 butir soal (20 pilihan ganda dan 5 uraian) dinyatakan valid dan reliabel. Sedangkan uji tingkat kesukaran dari 25 soal tersebut digolongkan pada kategori mudah, namun dari perhitungan daya pembeda soal tersebut memiliki kualitas yang baik digunakan untuk penilaian. Dari hasil pengujian tersebut, penelitian ini dapat menjadi rujukan alat asesmen mata pelajaran ekonomi khususnya pada materi kelangkaan kelas X SMA di Indonesia. Namun demikian penelitian lanjutan perlu dilakukan untuk meningkatkan kualitas tingkat kesukaran dari soal ini.

Abstract

This study aims to improve the quality of assessment in economics subjects, particularly the topic of scarcity in grade X of high school. This study uses validity, reliability, difficulty level and discriminating power tests to analyze 25 previously developed questions. The test questions were conducted on 30 students of grade X of high school. From the results of the validity and reliability tests using SPSS 26, as many as 22 of the 25 questions (20 multiple choice and 5 essay) were declared valid and reliable. While the difficulty level test of the 25 questions is classified as easy, but from the calculation of the discriminating power of the questions, they have good quality to be used for assessment. From the test results, this study can be a reference for assessment tools in economics subjects, especially on the topic of scarcity in grade X of high school in Indonesia. However, further research is needed to improve the quality of the difficulty level of these questions.

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).



PENDAHULUAN

Evaluasi pembelajaran merupakan bagian integral dalam proses pendidikan karena dapat memberikan informasi tentang pencapaian kompetensi peserta didik (Apriliana Saputri et al., 2023; Sudijono, 2011). Salah satu bentuk evaluasi yang paling umum digunakan adalah tes tertulis, yang biasanya disusun dalam bentuk butir soal dan digunakan untuk mengukur pemahaman peserta didik terhadap materi pembelajaran (Arikunto, 2010). Dalam konteks mata pelajaran Ekonomi, khususnya di tingkat sekolah menengah atas (SMA), soal yang baik sangat diperlukan untuk mengukur kemampuan berpikir kritis dan analitis siswa terhadap fenomena ekonomi yang kompleks (Mulyasa, 2013). Sayangnya, masih ditemukan banyak soal yang digunakan dalam proses pembelajaran belum melalui analisis kualitas yang memadai, sehingga hasil penilaian dapat kurang tepat dalam menginterpretasikan kemampuan siswa (S. M. Brookhart & Nitko, 2019).

Empat aspek penting yang perlu dianalisis dalam butir soal adalah validitas, reliabilitas, tingkat kesukaran, dan daya pembeda (Arikunto, 2016). Validitas menunjukkan sejauh mana suatu soal mampu mengukur apa yang seharusnya diukur (Sugiyono, 2019), sedangkan reliabilitas berkaitan dengan konsistensi hasil pengukuran yang diperoleh dari soal tersebut (Gay et al., 2012). Tingkat kesukaran digunakan untuk mengetahui proporsi peserta didik yang dapat menjawab soal dengan benar, sedangkan daya pembeda bertujuan untuk melihat seberapa baik soal tersebut membedakan antara peserta didik yang berkemampuan tinggi dan rendah (Haladyna & Rodriguez, 2013). Analisis yang tepat terhadap empat aspek tersebut memungkinkan guru untuk menyusun instrumen penilaian yang adil, objektif, dan akuntabel (S. Brookhart, 2010).

Dalam mata pelajaran Ekonomi, khususnya pada materi *kelangkaan*, kemampuan siswa dalam memahami konsep dasar seperti kebutuhan, sumber daya, dan pilihan alternatif sangat menentukan pemahaman ekonomi yang lebih kompleks di jenjang berikutnya (Lahuri & Dwi Rahayu, 2024; Martha, 2017). Sehingga materi ini menjadi salah satu konsep inti yang mendasari prinsip-prinsip ekonomi. Oleh karena itu, instrumen soal yang digunakan pada topik kelangkaan perlu dianalisis secara mendalam untuk memastikan bahwa soal tersebut mampu mengukur capaian kompetensi dengan tepat.

Analisis butir soal meliputi aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Validitas soal menentukan apakah butir soal benar-benar mengukur pemahaman siswa terhadap konsep kelangkaan dan prinsip ekonomi dasar. Sementara reliabilitas berkaitan dengan konsistensi hasil pengukuran terhadap materi yang sama dalam waktu berbeda (Anshari et al., 2024). Soal yang tidak valid atau tidak reliabel akan menghasilkan data penilaian yang tidak akurat, sehingga dapat mengaburkan gambaran sejati kemampuan siswa. Tingkat kesukaran diperlukan untuk mengetahui sejauh mana siswa dapat menjawab soal kelangkaan dengan benar, apakah soal tersebut terlalu mudah, sedang, atau terlalu sulit. Sementara itu, daya pembeda menunjukkan kemampuan soal dalam membedakan antara siswa dengan tingkat pemahaman tinggi dan rendah pada konsep seperti *opportunity cost* dan alokasi sumber daya terbatas (Hanifah et al., 2014). Studi tersebut menunjukkan bahwa penggunaan bentuk soal pilihan ganda asosiasi mampu meningkatkan daya pembeda dan relevansi kontekstual dibandingkan soal biasa.

Dengan demikian, analisis kualitas soal dalam materi kelangkaan bukan hanya penting untuk meningkatkan mutu penilaian, tetapi juga untuk memastikan bahwa siswa benar-benar memahami konsep inti ekonomi. Selain itu instrumen yang telah dianalisis dengan cermat dapat membantu guru memberikan umpan balik yang lebih tepat, merancang pembelajaran yang lebih responsif, dan memfasilitasi pengambilan keputusan pendidikan yang lebih akurat. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis butir soal pada bukan hanya kegiatan teknis semata, melainkan bagian penting dari upaya meningkatkan kualitas proses evaluasi dalam pembelajaran Ekonomi. Dengan instrumen yang valid dan reliabel, guru dapat memperoleh informasi yang lebih akurat untuk mendukung pengambilan keputusan pendidikan yang tepat.

Untuk itu penelitian ini bertujuan untuk menganalisis butir soal pada materi kelangkaan yang sangat dibutuhkan karena kurangnya referensi instrumen evaluasi yang fokus pada materi ini di tingkat SMA, terutama yang telah melalui pengujian komprehensif validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Selain itu, materi kelangkaan memiliki karakteristik yang dapat menantang penyusunan soal, karena perlu menyeimbangkan antara konsep teoritis dan konteks aplikatif yang relevan dengan kehidupan sehari-hari siswa. Dengan adanya penelitian ini, guru dan pengembang kurikulum dapat memiliki acuan yang kuat dalam menyusun dan memperbaiki soal ekonomi khususnya materi kelangkaan sehingga penilaian menjadi lebih akurat dan bermakna, serta mendukung perencanaan pembelajaran yang lebih responsif.

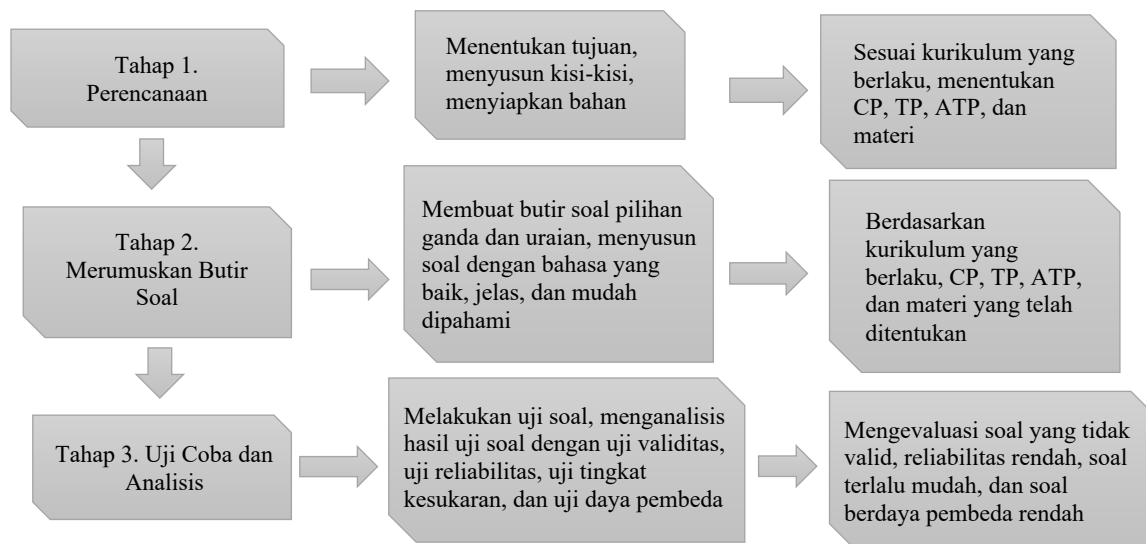
METODE

Penelitian ini menggunakan metode kuantitatif yang bertujuan untuk menganalisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda pada soal-soal ekonomi materi kelangkaan. Penelitian dilakukan di SMAN 1 Garum, salah satu sekolah menengah atas unggulan di Kabupaten Blitar yang sudah menerapkan kurikulum Merdeka dan memiliki mata pelajaran ekonomi sebagai bagian dari peminatan IPS.

Penelitian ini dilakukan di SMAN 1 Garum, SMAN 1 Garum dipilih karena merupakan salah satu sekolah menengah atas yang termasuk dalam lima unggulan sekolah terbaik di Kabupaten Blitar. SMAN 1 Garum juga merupakan sekolah yang sudah menerapkan kurikulum Merdeka dan mempunyai mata Pelajaran ekonomi untuk menjadi bagian dari peminatan IPS. Faktor lain yang menjadikan SMAN 1 Garum dijadikan studi kasus penelitian adalah karena kemudahan akses dan koordinasi antara sekolah dan peneliti. Penelitian yang dilakukan di SMAN 1 Garum, Blitar, dengan jumlah 158 peserta didik. Sampel penelitian diambil dengan teknik purposive sampling, dimana kelas X3 dipilih sebagai sampel utama dengan jumlah 30 peserta didik. Kelas X3 dipilih berdasarkan kriteria representatif yaitu memiliki karakteristik heterogenitas kemampuan akademik yang mewakili variasi keseluruhan populasi kelas X di SMAN 1 Garum. Pemilihan ini bertujuan agar hasil pengujian instrumen dapat menggambarkan kondisi yang realistis dari populasi kelas X.

Instrumen penelitian berupa tes objektif yang terdiri dari 20 butir soal pilihan ganda dan 5 butir soal uraian, yang dikembangkan berdasarkan kisi-kisi soal yang mengacu pada indikator pembelajaran dari kurikulum yang berlaku, khususnya pada materi kelangkaan. Kisi-kisi tersebut memastikan setiap butir soal relevan dengan capaian pembelajaran, tujuan pembelajaran, dan alur tujuan pembelajaran sehingga keterkaitan antar soal dengan indikator pembelajaran dapat dipertanggungjawabkan secara sistematis. Proses pembuatan, perancangan, dan penyusunan 25 butir soal dikerjakan melalui tiga tahapan. Tahap pertama adalah perencanaan, bagian ini dimulai dari menentukan tujuan, membuat kisi-kisi soal, dan menyiapkan bahan pembuatan soal. Pada tahap pertama, perencanaan berdasarkan pada kurikulum yang berlaku, Capaian Pembelajaran (CP), Tujuan Pembelajaran (TP), Alur Tujuan Pembelajaran (ATP), dan materi kelangkaan sebagai permasalahan dasar dalam ekonomi.

Tahap kedua adalah merumuskan butir soal, pada tahap ini dilakukan pembuatan butir soal pilihan ganda sebanyak 20 butir soal dan 5 butir soal uraian. Setelah 25 butir soal sudah disiapkan, maka dilakukan penyusunan butir soal dengan bahasa yang jelas, baik, dan mudah dipahami peserta didik. Perumusan butir soal harus sudah sesuai dengan apa yang telah ditentukan pada Capaian Pembelajaran, Tujuan Pembelajaran, Alur Tujuan Pembelajaran, dan materi yang telah dipilih. Tahap ketiga yaitu uji coba dan analisis butir soal, 25 butir soal yang telah disusun dengan sistematika bahasa yang baik selanjutnya dilakukan uji soal. Hasil dari uji soal kemudian akan dianalisis menggunakan statistika untuk mengetahui tingkat validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal. Setelah dianalisis, maka soal yang tidak valid, memiliki reliabilitas rendah, terlalu mudah, dan memiliki daya pembeda rendah akan dievaluasi untuk selanjutnya digunakan dalam tes penilaian. Hal tersebut terangkum pada diagram alur di bawah ini.



Gambar 1. Alur Pembuatan dan Perancangan Soal

Uji instrumen dilakukan dengan empat tahapan utama, yaitu uji validitas, uji reliabilitas, tingkat kesukaran, dan daya pembeda soal. Uji validitas digunakan untuk mengetahui sejauh mana soal-soal pilihan ganda dan uraian benar-benar dapat mengukur kemampuan kognitif siswa dengan objektif dan menyeluruh (Novikasari, 2017). Uji validitas dilakukan dengan analisis statistik untuk menyatakan bentuk soal yang valid dan tidak valid. Berdasarkan analisis statistik, soal yang valid adalah soal yang memiliki korelasi positif dan menunjukkan hasil signifikan dari keseluruhan tes soal. Sedangkan soal yang tidak valid adalah soal yang menunjukkan korelasi negative atau tidak menyatakan korelasi secara kuat.

Tahapan uji yang kedua adalah uji reliabilitas, uji reliabilitas digunakan untuk mengukur sejauh mana soal-soal tersebut mampu memberikan hasil yang konsisten apabila digunakan secara berulang kali. Atau dengan kata lain, uji reliabilitas digunakan untuk mengukur bagaimana kemampuan suatu alat uji berupa soal pilihan ganda dan uraian dapat digunakan dalam waktu yang berbeda, penguji yang berbeda, tetapi dengan kondisi pengukuran yang masih sama (Surapranata, 2005). Uji reliabilitas diukur dengan analisis statistik, soal yang dianggap reliable adalah jenis soal yang memiliki nilai *Cronbach Alpha* lebih dari 0,70.

Tahapan uji yang ketiga adalah tingkat kesukaran soal, tingkat kesukaran soal digunakan untuk mengetahui tingkat proporsi soal tersebut dengan menggolongkan soal yang mudah, sedang, dan sulit atau sukar. Tingkat kesukaran soal perlu diketahui karena digunakan untuk menentukan proporsi soal yang baik untuk mengukur kemampuan kognitif siswa secara menyeluruh terutama siswa yang heterogeny (Hanifah et al., 2014). Dalam pengukuran uji tingkat kesukaran dapat dilakukan dengan cara menganalisis skor rata-rata yang diperoleh oleh siswa, analisis tersebut kemudian digolongkan pada jenis soal mudah, sedang, dan sulit atau sukar. Hasil yang didapatkan dari analisis uji tingkat kesukaran soal dapat digunakan sebagai dasar oleh penguji atau pendidik dalam melakukan revisi soal yang kurang dari standar pembelajaran yang telah ditentukan.

Tahapan uji yang terakhir adalah dengan daya pembeda soal, daya pembeda soal digunakan sebagai alat uji untuk mengetahui sebuah soal dapat membedakan tingkat kemampuan kognitif siswa. Daya pembeda soal dibedakan menjadi soal dengan daya pembeda negatif, rendah, sedang, baik, dan baik sekali, yang ditentukan dengan menghitung indeks diskriminasi soal tersebut. Soal yang baik adalah soal yang mampu memberikan daya pembeda kemampuan kognitif siswa secara optimal, dengan kata lain soal yang memiliki indeks diskriminasi tinggi adalah soal yang mampu membedakan siswa berkemampuan kognitif tinggi dan siswa dengan kemampuan kognitif rendah (Hanifah et al., 2014). Hasil dari uji daya pembeda soal akan dapat digunakan oleh penguji atau pendidik sebagai dasar dalam mengevaluasi bentuk soal yang tidak sesuai kriteria dalam standar pembelajaran.

Data yang diperoleh dari tes kemudian akan dilakukan analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda menggunakan metode statistik deskriptif. Analisis dalam penelitian ini dilakukan dengan menggunakan perangkat lunak uji statistik IBM SPSS versi 26 untuk menguji dan menganalisis empat uji tahapan tersebut.

1. Uji validitas

Pengujian dilakukan dengan menggunakan panduan klasifikasi validitas sebagaimana dijabarkan dalam Tabel 1 berikut ini,

Tabel 1. Klasifikasi Uji Validitas

Nilai Uji	Keterangan
$r \text{ hitung} < r. \text{ tabel}$	Tidak Valid
$r \text{ hitung} > r. \text{ tabel}$	Valid
$Df = (n - 2)$	Nilai $r. \text{ tabel}$

Sumber: Diolah dari Arikunto (2012)

2. Uji reliabilitas

Pengujian reliabilitas digunakan untuk mengukur soal yang reliable atau konsisten dengan menggunakan kaidah pada Tabel 2, berikut ini:

Tabel 2. Klasifikasi Uji Reliabilitas

Nilai Uji	Keterangan
$r < 0,20$	Sangat rendah
$0,20 < r < 0,40$	Rendah
$0,40 < r < 0,70$	Sedang
$0,70 < r < 0,90$	Tinggi
$0,90 < r < 1,00$	Sangat Tinggi

Sumber: Diolah dari Arikunto (2012)

3. Tingkat kesukaran

Tingkat kesukaran digunakan untuk menggolongkan jenis soal yang mudah, sedang, dan sukar atau sulit. Tingkat kesukaran diukur dengan menggunakan panduan klasifikasi pada Tabel 3, berikut ini:

Tabel 3. Klasifikasi Uji Tingkat Kesukaran

Nilai Uji	Keterangan
$P < 0,30$	Soal tersukar
$0,30 < P < 0,70$	Soal sedang
$0,70 < P < 1,00$	Soal termudah

Sumber: Diolah dari Arikunto (2012)

4. Daya pembeda

Pengukuran daya pembeda digunakan untuk mengukur kemampuan soal dalam membedakan kemampuan kognitif siswa. Uji tingkat kesukaran dilakukan dengan mengacu pada klasifikasi nilai uji tingkat kesukaran Tabel 4 berikut ini:

Tabel 4. Klasifikasi Uji Daya Pembeda

Nilai Uji	Keterangan
$D < 0,20$	Jelek
$0,20 < D < 0,40$	Cukup
$0,40 < D < 0,70$	Baik
$0,70 < D < 1,00$	Baik sekali
$D = \text{negatif}$	Jelek

Sumber: Diolah dari Arikunto (2012)

HASIL DAN PEMBAHASAN

1. Validitas butir soal

Uji validitas merupakan langkah penting dalam analisis butir soal untuk memastikan bahwa setiap pertanyaan dalam tes benar-benar mampu mengukur kompetensi yang ingin dicapai (Siregar et al., 2024). Dalam penelitian ini, uji validitas dilakukan terhadap dua kelompok soal yang disajikan dalam Tabel 5 dan Tabel 6. Tabel 5 memuat 20 butir soal pilihan ganda yang dianalisis menggunakan teknik korelasi antara skor butir soal dengan skor total, dengan batas nilai r tabel sebesar 0,3061 sebagai acuan. Hasilnya menunjukkan bahwa dari 20 butir soal, sebanyak 17 soal memenuhi syarat validitas karena nilai korelasi (r hitung) yang diperoleh lebih tinggi dari r tabel. Soal-soal tersebut mencakup nomor 1 hingga 4, 7 hingga 16, dan 18 hingga 20. Artinya, soal-soal ini sudah mampu merepresentasikan capaian pembelajaran yang ingin diukur. Namun demikian, terdapat tiga soal yang tidak valid, yaitu soal nomor 5 ($r = 0,080$), soal nomor 6 ($r = 0,295$), dan soal nomor 17 ($r = 0,184$), karena nilai korelasinya tidak mencapai batas minimum yang telah ditetapkan. Ketiga soal ini perlu ditinjau kembali, baik dari segi konstruksi kalimat, kesesuaian indikator, maupun tingkat kejelasan jawabannya. Selanjutnya, pada Tabel 6, dilakukan uji validitas terhadap 5 butir soal tambahan. Seluruh soal pada tabel ini menunjukkan nilai r hitung yang lebih besar dari r tabel (r berkisar antara 0,385 hingga 0,702), sehingga seluruhnya dinyatakan valid. Ini menunjukkan bahwa secara keseluruhan, 22 dari 25 butir soal yang diuji telah memenuhi kriteria validitas, suatu indikator yang cukup baik untuk menunjukkan kualitas awal instrumen tes yang dikembangkan.

Validitas butir soal tidak hanya penting dari sisi teknis pengukuran, tetapi juga memiliki peran besar dalam menjamin bahwa proses evaluasi pembelajaran benar-benar menggambarkan penguasaan peserta didik terhadap materi yang diajarkan (Sudaryono et al., 2019). Dalam konteks pembelajaran ekonomi, khususnya pada materi kelangkaan yang menjadi fokus kajian, soal yang valid akan mampu mengevaluasi sejauh mana siswa memahami konsep dasar ekonomi, seperti kebutuhan manusia yang tidak terbatas, keterbatasan sumber daya, serta pilihan dan biaya peluang. Ketika suatu soal dinyatakan valid, ini berarti bahwa siswa yang memiliki pemahaman lebih baik terhadap materi cenderung memperoleh skor lebih tinggi pada soal tersebut, dibandingkan dengan siswa yang pemahamannya masih rendah. Sebaliknya, jika suatu soal tidak valid, maka besar kemungkinan soal tersebut tidak selaras dengan indikator pembelajaran, terlalu membingungkan, atau justru tidak menantang sama sekali. Soal seperti ini tidak hanya gagal memberikan informasi yang akurat tentang penguasaan siswa, tetapi juga berpotensi menciptakan ketidakadilan dalam penilaian. Oleh karena itu, menemukan bahwa sebagian besar soal dalam Tabel 5 dan seluruh soal dalam Tabel 6 dinyatakan valid merupakan indikasi positif bahwa soal-soal tersebut telah disusun dengan mempertimbangkan keterkaitan antara tujuan pembelajaran, materi, dan konstruk kognitif yang tepat. Dalam pengembangan instrumen evaluasi di sekolah, penting bagi guru untuk memastikan bahwa setiap soal yang digunakan benar-benar dapat mengukur kemampuan berpikir dan pemahaman siswa,

bukan hanya aspek menghafal (Suprpto et al., 2020). Dengan validitas yang baik, proses pembelajaran akan lebih bermakna karena guru memiliki alat yang tepat untuk memantau kemajuan siswa dan memberikan umpan balik yang sesuai.

Tabel 5. Uji Validitas Pilihan Ganda

No. Butir Soal	r Tabel	Koefisien Korelasi	Keterangan
1	0,3061	0,873	Valid
2	0,3061	0,554	Valid
3	0,3061	0,670	Valid
4	0,3061	0,444	Valid
5	0,3061	0,080	Tidak Valid
6	0,3061	0,295	Tidak Valid
7	0,3061	0,371	Valid
8	0,3061	0,670	Valid
9	0,3061	0,670	Valid
10	0,3061	0,554	Valid
11	0,3061	0,666	Valid
12	0,3061	0,406	Valid
13	0,3061	0,368	Valid
14	0,3061	0,543	Valid
15	0,3061	0,745	Valid
16	0,3061	0,873	Valid
17	0,3061	0,184	Tidak Valid
18	0,3061	0,482	Valid
19	0,3061	0,462	Valid
20	0,3061	0,745	Valid

Sumber: Output SPSS, diolah oleh peneliti (2025)

Tabel 6. Uji Validitas Uraian

No. Butir Soal	r Tabel	Koefisien Korelasi	Keterangan
1	0,3061	0,385	Valid
2	0,3061	0,651	Valid
3	0,3061	0,564	Valid
4	0,3061	0,564	Valid
5	0,3061	0,702	Valid

Sumber: Output SPSS, diolah oleh peneliti (2025)

Hasil uji validitas dari Tabel 5 dan Tabel 6 memberikan gambaran penting bagi guru dan pengembang instrumen evaluasi, bahwa proses penyusunan soal harus disertai dengan tahapan analisis yang sistematis dan berbasis data. Meskipun sebagian besar soal telah memenuhi syarat validitas, tetap perlu dilakukan revisi terhadap soal-soal yang tidak valid agar kualitas keseluruhan instrumen semakin baik. Revisi tersebut dapat mencakup perbaikan redaksi soal, pemilihan alternatif jawaban yang lebih logis, atau penyesuaian indikator agar lebih sesuai dengan tuntutan kurikulum. Dalam konteks pembelajaran ekonomi, validitas soal memiliki peran strategis karena materi ekonomi sangat dekat dengan kehidupan sehari-hari siswa (Welsandt et al., 2024). Oleh karena itu, butir soal yang valid tidak hanya akan mencerminkan capaian kognitif siswa, tetapi juga dapat memicu refleksi kritis terhadap isu-isu ekonomi di sekitar mereka, seperti keterbatasan barang dan jasa, pilihan-pilihan ekonomi dalam keluarga, serta pengelolaan sumber daya. Dengan demikian, evaluasi tidak sekadar menjadi alat ukur, tetapi juga bagian dari proses belajar itu sendiri (Elliott & Balasubramanyam, 2016). Hasil uji validitas ini dapat menjadi dasar yang kuat bagi pengambilan keputusan dalam perbaikan pembelajaran, baik dalam hal strategi mengajar, penguatan materi, maupun pengembangan soal ke depan yang lebih kontekstual, adil, dan mampu mendorong literasi ekonomi siswa secara menyeluruh (Black & Wiliam, 1998; Wiliam, 2011).

2. Reliabilitas

Uji reliabilitas bertujuan untuk mengetahui sejauh mana instrumen penilaian menghasilkan data yang konsisten apabila digunakan dalam kondisi yang serupa (Arbeni et al., 2025). Salah satu teknik yang umum digunakan adalah perhitungan koefisien Cronbach's Alpha, di mana nilai yang semakin mendekati 1 menunjukkan reliabilitas yang semakin tinggi. Berdasarkan data dalam tabel, terdapat dua kelompok soal yang diuji. Kelompok pertama terdiri atas 20 butir soal

yang menunjukkan nilai Cronbach's Alpha sebesar 0,821, yang termasuk dalam kategori reliabilitas tinggi. Artinya, soal-soal dalam kelompok ini konsisten dalam mengukur konstruk yang sama, dan dapat dipercaya untuk digunakan dalam proses evaluasi pembelajaran. Sementara itu, kelompok kedua terdiri dari 5 butir soal dengan nilai Cronbach's Alpha sebesar 0,479, yang berada pada kategori reliabilitas sedang. Hal ini menunjukkan bahwa meskipun soal tersebut masih dapat digunakan, tetapi konsistensi antar butirnya belum optimal sehingga perlu diperhatikan lebih lanjut, khususnya dari aspek kejelasan, kesesuaian indikator, dan keragaman tingkat kognitif.

Tabel 7. Uji Reliabilitas

Cronbach Alpha	N of Items	Keterangan
0,821	20	Reliabilitas Tinggi
0,479	5	Reliabilitas Sedang

Sumber: Output SPSS, diolah oleh peneliti (2025)

Nilai reliabilitas yang tinggi pada 20 butir soal menunjukkan bahwa instrumen tersebut mampu memberikan hasil yang stabil dan dapat diandalkan untuk menilai kemampuan siswa, khususnya dalam konteks materi ekonomi. Konsistensi ini sangat penting agar hasil evaluasi benar-benar mencerminkan kompetensi peserta didik dan tidak dipengaruhi oleh faktor kebetulan. Sebaliknya, nilai reliabilitas sedang pada 5 soal lainnya perlu menjadi perhatian bagi guru dan pengembang instrumen. Meskipun masih dapat digunakan, soal-soal tersebut sebaiknya ditinjau ulang, terutama dalam hal formulasi, relevansi terhadap kompetensi dasar, dan keterkaitan antar soal. Dalam pengembangan instrumen evaluasi, reliabilitas yang tinggi menjadi salah satu syarat agar hasil penilaian dapat dijadikan dasar pengambilan keputusan yang tepat, baik dalam perbaikan strategi pembelajaran, pemetaan tingkat penguasaan materi, maupun penyusunan soal lanjutan yang lebih tajam dan bermakna. Oleh karena itu, guru disarankan untuk tidak hanya berhenti pada uji validitas, tetapi juga melengkapi proses pengembangan instrumen dengan uji reliabilitas guna menjamin kualitas evaluasi secara keseluruhan (Earle & Earle, 2020; Govindasamy et al., 2024; Thompson et al., 2024).

3. Tingkat kesukaran soal

Tingkat kesukaran suatu butir soal menunjukkan sejauh mana siswa mampu menjawab soal tersebut dengan benar, dan dihitung dengan menggunakan rumus $P = \text{Mean} / \text{Skor Maksimum}$. Dalam konteks ini, skor maksimum adalah 4, karena butir soal memiliki rentang nilai dari 0 hingga 4. Berdasarkan data dalam Tabel 8, analisis dilakukan terhadap 20 butir soal yang telah diujikan kepada 30 responden. Hasil penghitungan menunjukkan bahwa 19 dari 20 soal memiliki nilai P lebih besar dari 0,80, yang termasuk ke dalam kategori “mudah”. Hal ini berarti sebagian besar siswa dapat menjawab butir soal dengan benar dalam jumlah yang tinggi. Nilai P tertinggi ditemukan pada soal nomor 3, 8, 9, dan 14 yang masing-masing memiliki skor 0,97, menunjukkan bahwa hampir seluruh siswa menjawab soal ini dengan benar. Nilai P terendah terdapat pada soal nomor 17, yaitu sebesar 0,47 yang masuk dalam kategori “sedang”, karena berada pada rentang $0,30 \leq P \leq 0,70$. Tidak terdapat butir soal yang masuk dalam kategori “sulit” ($P < 0,30$). Dari hasil ini, dapat disimpulkan bahwa secara umum tingkat kesukaran soal berada pada tingkat yang sangat mudah. Ini menunjukkan bahwa soal-soal tersebut cenderung kurang menantang bagi siswa, dan sebagian besar dari mereka memiliki pemahaman dasar yang cukup kuat terhadap materi yang diujikan. Walaupun dari sisi pembelajaran hal ini bisa menjadi indikator positif, namun dari sisi pengukuran, variasi tingkat kesukaran yang terlalu sempit bisa menjadi kelemahan dalam mengukur kemampuan siswa secara komprehensif.

Kecenderungan soal-soal yang tergolong mudah dalam analisis tingkat kesukaran Tabel 8 dan Tabel 9 menunjukkan bahwa mayoritas siswa memiliki kemampuan menjawab soal dengan benar dalam persentase tinggi. Hal ini bisa menunjukkan bahwa materi yang diajarkan telah tersampaikan dengan baik, khususnya dalam konteks pembelajaran ekonomi pada topik kelangkaan. Namun demikian, jika hampir semua soal tergolong mudah, maka ada kemungkinan bahwa instrumen evaluasi ini belum mampu sepenuhnya mengidentifikasi variasi tingkat pemahaman siswa, terutama mereka yang berada pada kelompok menengah dan bawah. Dalam konteks asesmen formatif maupun sumatif, variasi soal berdasarkan tingkat kesukaran menjadi penting agar penilaian tidak hanya sekadar menggambarkan siapa yang bisa menjawab semua soal dengan benar, tetapi juga siapa yang benar-benar memahami konsep di tingkat berpikir yang lebih tinggi (Mustamin, 2024; Ridhwan, 2017). Kehadiran satu soal dengan kategori sedang (soal nomor 17) menjadi poin penting yang menunjukkan adanya sedikit variasi dalam tingkat kesulitan, namun jumlah ini masih terlalu kecil. Oleh karena itu, dalam penyusunan soal ke depan, disarankan untuk menyeimbangkan komposisi soal mudah, sedang, dan sulit dengan proporsi ideal, misalnya 30% mudah, 40% sedang, dan 30% sulit. Dengan begitu, guru dapat memperoleh gambaran yang lebih utuh tentang pencapaian belajar siswa, dan hasil evaluasi bisa digunakan untuk menyusun strategi pembelajaran lanjutan yang lebih tepat sasaran. Tingkat kesukaran yang beragam juga membantu menstimulasi kemampuan berpikir kritis siswa serta menghindari rasa terlalu percaya diri akibat soal yang terlalu ringan.

Tabel 8. Uji Tingkat Kesukaran Pilihan Ganda

No. Butir Soal	N		Mean	Tingkat Kesukaran (P = Mean/4)	Keterangan
	Valid	Missing			
1	30	0	3.73	0.93	Mudah
2	30	0	3.73	0.93	Mudah
3	30	0	3.87	0.97	Mudah
4	30	0	3.47	0.87	Mudah
5	30	0	3.07	0.77	Mudah
6	30	0	3.07	0.77	Mudah
7	30	0	3.73	0.93	Mudah
8	30	0	3.87	0.97	Mudah
9	30	0	3.87	0.97	Mudah
10	30	0	3.73	0.93	Mudah
11	30	0	3.2	0.80	Mudah
12	30	0	3.6	0.90	Mudah
13	30	0	3.6	0.90	Mudah
14	30	0	3.87	0.97	Mudah
15	30	0	3.47	0.87	Mudah
16	30	0	3.73	0.93	Mudah
17	30	0	1.87	0.47	Sedang
18	30	0	3.6	0.90	Mudah
19	30	0	3.73	0.93	Mudah
20	30	0	3.47	0.87	Mudah

Sumber: Output SPSS, diolah oleh peneliti (2025)

Tabel 9. Uji Tingkat Kesukaran Uraian

No. Butir Soal	N		Mean	Tingkat Kesukaran (P = Mean/4)	Keterangan
	Valid	Missing			
1	30	0	3.8	0.95	Mudah
2	30	0	3.53	0.882	Mudah
3	30	0	3.5	0.88	Mudah
4	30	0	3.5	0.88	Mudah
5	30	0	3	0.75	Mudah

Sumber: Output SPSS, diolah oleh peneliti (2025)

4. Daya pembeda soal

Uji daya pembeda bertujuan untuk mengetahui kemampuan suatu butir soal dalam membedakan antara siswa dengan tingkat penguasaan tinggi dan rendah terhadap materi. Dalam Tabel 10, nilai *Corrected Item-Total Correlation* (CITC) digunakan sebagai indikator daya pembeda. Semakin tinggi nilai CITC suatu soal, semakin baik daya pembeda butir tersebut. Hasil analisis menunjukkan bahwa butir nomor 1 dan 16 memiliki nilai 0,850, yang tergolong dalam kategori "sangat baik", menunjukkan bahwa kedua soal ini sangat efektif dalam membedakan siswa berdasarkan tingkat penguasaannya. Butir soal lain seperti nomor 3, 8, 9, 15, dan 20 juga termasuk kategori "baik" karena memiliki nilai CITC di atas 0,60. Namun, beberapa butir menunjukkan kualitas sedang atau kurang, seperti butir 4, 7, 12, 13, 18, dan 19 dengan nilai CITC antara 0,275–0,397 (kategori cukup), serta butir 5, 6, dan 17 yang memiliki nilai di bawah 0,20 bahkan negatif yang termasuk dalam kategori jelek, dan secara teoretis perlu direvisi atau dibuang. Sejalan dengan pendapat Nunnally & Bernstein (1994), butir soal dengan nilai diskriminasi di bawah 0,20 sebaiknya tidak digunakan dalam instrumen akhir karena tidak mampu berkontribusi dalam membedakan performa siswa secara bermakna (Triola & Holloway, 2011)

Temuan dari Tabel 10 memberikan gambaran penting bagi guru atau pengembang instrumen evaluasi dalam meningkatkan kualitas asesmen pembelajaran. Butir-butir dengan daya pembeda tinggi dapat digunakan sebagai model dalam menyusun soal evaluasi yang menuntut kemampuan berpikir kritis, bukan sekadar menghafal. Sementara itu, butir dengan daya pembeda rendah atau negatif perlu ditinjau kembali dari aspek redaksi soal, konteks yang digunakan, atau tingkat kesulitan yang tidak seimbang. Menurut (Crocker & Algina, 2008) daya pembeda yang tinggi merupakan syarat esensial dalam instrumen yang bertujuan untuk mengklasifikasikan kemampuan siswa secara akurat, baik dalam asesmen formatif maupun sumatif. Oleh karena itu, pengujian daya pembeda merupakan bagian penting dalam validasi internal

instrumen, karena hasilnya dapat menjadi dasar pengambilan keputusan untuk memperbaiki soal, menyesuaikan strategi pembelajaran, serta memperkuat akuntabilitas evaluasi. Dalam konteks pembelajaran ekonomi, soal dengan daya pembeda baik sangat diperlukan untuk menilai seberapa jauh siswa mampu menerapkan konsep ekonomi dalam situasi nyata, bukan hanya memahami definisinya. Maka, proses analisis daya pembeda bukan hanya teknis, tetapi juga strategis dalam menciptakan pembelajaran yang berkualitas dan berorientasi pada penguatan literasi ekonomi siswa.

Tabel 10. Uji Daya Pembeda Pilihan Ganda

No. Butir Soal	<i>Corrected Item-Total Correlation</i>	Keterangan
1	0.850	Baik sekali
2	0.490	Baik
3	0.634	Baik
4	0.344	Cukup
5	-0.065	Jelek
6	0.156	Jelek
7	0.294	Cukup
8	0.634	Baik
9	0.634	Baik
10	0.490	Baik
11	0.578	Baik
12	0.315	Cukup
13	0.275	Cukup
14	0.498	Baik
15	0.687	Baik
16	0.850	Baik sekali
17	0.014	Jelek
18	0.397	Cukup
19	0.391	Cukup
20	0.687	Baik

Sumber: Output SPSS, diolah oleh peneliti (2025)

Tabel 11 merupakan analisis daya pembeda butir soal uraian dengan menggunakan indikator *Corrected Item-Total Correlation* (CITC). Dari lima butir soal, hanya butir nomor 5 yang tergolong memiliki daya pembeda “baik” (0,510), sedangkan butir 2, 3, dan 4 tergolong “cukup” karena nilai CITC-nya berkisar antara 0,213–0,332. Sementara itu, butir nomor 1 menunjukkan performa paling rendah dengan nilai 0,083 dan tergolong “jelek”. Hasil ini konsisten dengan temuan sebelumnya pada Tabel 10, yang menunjukkan masih terdapat soal-soal dengan kualitas diskriminatif rendah atau bahkan negatif. Daya pembeda yang rendah menunjukkan bahwa soal tersebut tidak efektif dalam membedakan siswa yang benar-benar memahami materi dan yang belum, sehingga mengganggu validitas hasil evaluasi. Oleh karena itu, butir-butir semacam ini sebaiknya direvisi atau diganti. Menurut (Triola & Holloway, 2011), CITC di bawah 0,20 sebaiknya dieliminasi karena tidak berkontribusi terhadap keandalan instrumen. Sementara itu (Black & Wiliam, 2009) menekankan bahwa penilaian berkualitas tinggi harus mampu memberikan informasi bermakna bagi guru dalam mengidentifikasi capaian dan kekurangan siswa, serta mendorong evaluasi pembelajaran yang lebih reflektif dan akurat.

Tabel 11. Uji Daya Pembeda Uraian

No. Butir Soal	<i>Corrected Item-Total Correlation</i>	Keterangan
1	0.083	Jelek

2	0.332	Cukup
3	0.213	Cukup
4	0.213	Cukup
5	0.510	Baik

Sumber: Output SPSS, diolah oleh peneliti (2025)

Berdasarkan analisis daya pembeda soal yang ditunjukkan dalam Tabel 10 dan 11, sebagian besar butir soal berada pada kategori cukup hingga baik, dengan beberapa menunjukkan performa sangat baik. Tabel 10 memperlihatkan bahwa dari 20 butir soal, mayoritas memiliki daya pembeda tinggi, seperti butir nomor 1, 3, 8, 15, dan 16, yang tergolong baik hingga sangat baik. Meskipun demikian, terdapat beberapa butir dengan daya pembeda rendah hingga negatif (butir 5, 6, dan 17), yang dapat menurunkan kualitas instrumen dan berisiko memberikan informasi yang tidak akurat mengenai kemampuan siswa. Tabel 11, yang memuat lima butir soal tambahan, menunjukkan bahwa hanya satu butir tergolong baik (butir 5), sementara lainnya tergolong cukup dan jelek. Temuan ini menegaskan pentingnya proses analisis daya pembeda dalam evaluasi soal untuk menjamin bahwa setiap item mampu membedakan peserta didik berdasarkan tingkat penguasaan konsep. Tavakol dan Dennick (2011) menekankan bahwa nilai *corrected item-total correlation* di bawah 0,20 menunjukkan butir yang lemah dan sebaiknya tidak digunakan dalam asesmen yang valid. Selain itu, Black dan Wiliam (2009) menyatakan bahwa instrumen yang baik harus dapat mengungkap informasi mendalam tentang capaian belajar siswa dan mendukung pengambilan keputusan pedagogis. Oleh karena itu, hasil dari Tabel 10 dan 11 menjadi dasar kuat bagi guru atau pengembang instrumen untuk merevisi soal yang tidak efektif dan mempertahankan soal yang memiliki kualitas diskriminatif tinggi demi tercapainya evaluasi pembelajaran yang lebih adil, bermakna, dan akurat.

Hasil analisis butir soal menunjukkan bahwa soal dengan daya pembeda negatif atau sangat rendah cenderung tidak efektif dalam membedakan antara peserta didik yang berkemampuan tinggi dan rendah. Menurut Arikunto (2013), daya pembeda merupakan ukuran kemampuan soal untuk secara valid membedakan siswa berdasarkan kompetensinya. Dengan demikian, guru perlu merevisi soal-soal tersebut dengan mempertimbangkan aspek konstruksi, konteks, dan tingkat kognitif. Alternatif perbaikan, seperti menyusun ulang soal agar lebih jelas, menambahkan konteks situasional yang relevan, serta menaikkan level kognitif ke ranah aplikasi atau analisis (sesuai Taksonomi Bloom revisi), dapat mengoptimalkan daya pembeda dan validitas instrumen evaluasi (Damayanti et al., 2021; Oktavianto Putra et al., 2020).

Instrumen evaluasi yang valid dan reliabel tidak hanya dinilai dari kualitas tiap butir soal, tetapi juga dari distribusi proporsional soal mudah, sedang, dan sukar. Sudjana (dalam Giani et al., 2012) menyarankan proporsi ideal 30 % mudah, 40 % sedang, dan 30 % sulit agar tes dapat menggambarkan rentang kemampuan peserta didik secara komprehensif (Oktavianto Putra et al., 2020). Hal ini diperkuat oleh temuan bahwa soal yang dominan mudah atau sukar tidak dapat mengukur level penguasaan materi secara akurat dan dapat menurunkan kualitas evaluasi pendidikan (Damayanti et al., 2021). Oleh karena itu, penelaahan tingkat kesukaran butir soal idealnya dimasukkan pada tahap perencanaan kisi-kisi agar tuntas saat pengujian.

Analisis butir soal tidak sekadar menjadi alat evaluasi, tetapi juga berfungsi sebagai umpan balik (feedback) bagi rancangan pembelajaran. Utami dan Nurgiyantoro (2016) menyatakan bahwa analisis jawaban butir soal berdasarkan indikator kompetensi dasar dapat digunakan sebagai bahan refleksi pedagogis untuk mengidentifikasi materi yang telah dikuasai dan yang memerlukan penguatan melalui pembelajaran remedial atau pengayaan (Damayanti et al., 2021). Dengan demikian, guru dapat merancang strategi pembelajaran yang lebih berbasis data misalnya kelompok remedial khusus untuk kompetensi yang rendah, atau kegiatan pengayaan bagi siswa yang sudah unggul sehingga proses pembelajaran menjadi adaptif dan responsif terhadap kebutuhan peserta didik.

Pengembangan kompetensi guru dalam menyusun butir soal sangat penting, terutama untuk menciptakan variasi ranah kognitif yang lebih luas, tidak hanya terbatas pada kemampuan mengingat (C1), tetapi juga mendorong keterampilan berpikir tingkat tinggi seperti mencipta atau mengevaluasi (C5–C6) sesuai Taksonomi Bloom revisi (Damayanti et al., 2021). Literasi guru terhadap teknik analisis item—validitas, reliabilitas, daya pembeda, dan tingkat kesukaran—juga krusial untuk menghasilkan instrumen evaluasi yang baik dan bermakna (Arifin, 2017; Trianingsih, 2023). Oleh karena itu, pelatihan atau workshop berbasis bukti (evidence-based) tentang pengembangan soal, integrasi teknologi pembelajaran, serta refleksi terhadap hasil analisis butir soal sangat layak diadakan secara berkala di tingkat sekolah atau wilayah.

SIMPULAN

Penelitian ini dilakukan untuk mengembangkan soal dan menganalisis setiap butir soal pilihan ganda dan uraian. Berdasarkan pada soal yang dikembangkan pada materi kelangkaan dengan rincian 20 soal pilihan ganda, dan 5 soal uraian, didapatkan bahwa soal yang tidak valid sejumlah 3 butir soal. Sedangkan keseluruhan soal telah memenuhi kaidah reliabel untuk digunakan. Dalam pengujian tingkat kesukaran hanya 4% soal dengan kategori sedang, yang artinya 96% soal dalam kategori mudah. Sedangkan dari uji daya pembeda soal, didapatkan bahwa hanya 16% dari soal tersebut atau 4 butir soal yang dikategorikan jelek atau kurang relevan digunakan untuk membedakan kemampuan siswa. Sehingga sisanya 84% atau 21 soal telah mampu untuk dijadikan alat untuk membedakan kemampuan siswa.

Dengan begitu penelitian ini memberikan kontribusi terhadap referensi soal yang dapat digunakan oleh guru ekonomi khususnya SMA dalam memberikan asesmen terhadap siswanya khususnya materi kelangkaan. Sehingga soal yang dijadikan alat untuk melakukan pengukuran dan penilaian kognitif untuk para siswa merupakan soal yang telah melalui kaidah analisis butir soal yang komprehensif. Namun bagaimanapun, penelitian ini memerlukan kajian lanjutan untuk merubah dan merivisi beberapa butir soal dalam kategori uji pembeda, dan tingkat kesukaran. Selain itu kajian lanjutan sangat perlu dilakukan untuk mengembangkan beragam materi ekonomi yang lain.

DAFTAR PUSTAKA

- 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>
- Apriliana Saputri, H., Joti Larasati, N., Guru Madrasah Ibtidaiyah, P., & Ilmu Tarbiyah Dan Keguruan Universitas Islam Negeri Sunan Kalijaga, F. (2023). ANALISIS INSTRUMEN ASSESMENT: VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986–2995. <https://doi.org/10.36989/DIDAKTIK.V9I5.2268>
- Arbeni, W., Windiani, A., Sariyani Br Sihotang, D., Anggraini, N., Wulandari, S., Nugroho, A., & Syekh Abdul Halim Hasan Binjai, I. (2025). Test Reliability Analysis in Educational Evaluation: A Quantitative Approach to Consistency and Validity. *Holistic Science*, 5(1), 59–64. <https://doi.org/10.56495/HS.V5I1.838>
- Arifin, Z. (2017). *Kriteria Instrumen dalam suatu Penelitian*. 2(1), 28.
- Arikunto, S. (2010). *Prosedur Penelitian: Suatu pendekatan praktik*.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *International Journal of Phytoremediation*, 21(1), 7–74. <https://doi.org/10.1080/0969595980050102;WGROUPE:STRING:PUBLICATON>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/S11092-008-9068-5/METRICS>
- Brookhart, S. (2010). *How to assess higher-order thinking skills in your classroom*. <https://www.sciepub.com/reference/361905>
- Brookhart, S. M., & Nitko, A. J. (2019). *Educational Assessment of Students*.
- Crocker, L. M., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory (M.B.M. Staudt & M. Stranz, Ed.)*. 8–11.
- Damayanti, W. W., Halidjah, S., Program, R. P., Pendidikan, S., Sekolah, G., Fkip, D., & Pontianak, U. (2021). *ANALISIS TINGKAT KESUKARAN BUTIR SOAL PILIHAN GANDAPADA PENILAIAN TENGAH SEMESTER KELAS IV*.
- Earle, S., & Earle, S. (2020). Balancing the demands of validity and reliability in practice: Case study of a changing system of primary science summative assessment. *London Review of Education*, 18(2), 221–235. <https://doi.org/10.14324/LRE.18.2.06>
- Elliott, C., & Balasubramanyam, V. (2016). Assessing students: Real-world analyses underpinned by economic theory. *Cogent Economics and Finance*, 4(1). <https://doi.org/10.1080/23322039.2016.1151171>
- Gay, L., Mills, G. E., & Airasian, P. (2012). *EDUCATIONAL RESEARCH*.
- Govindasamy, P., Cumming, T. M., & Abdullah, N. (2024). Validity and reliability of a needs analysis questionnaire for the development of a creativity module. *Journal of Research in Special Educational Needs*, 24(3), 637–652. <https://doi.org/10.1111/1471-3802.12659>
- Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. *Developing and Validating Test Items*, 1–446. <https://doi.org/10.4324/9780203850381>
- Hanifah, N., Studi, P., & Konseling, B. (2014a). PERBANDINGAN TINGKAT KESUKARAN, DAYA PEMBEDA BUTIR SOAL DAN RELIABILITAS TES BENTUK PILIHAN GANDA BIASA DAN PILIHAN GANDA ASOSIASI MATA PELAJARAN EKONOMI. *SOSIO E-KONS*, 6(1), 41–55.
- Hanifah, N., Studi, P., & Konseling, B. (2014b). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *SOSIO E-KONS*, 6(1), 41–55.

- Lahuri, S. Bin, & Dwi Rahayu, H. M. (2024). Concept of Scarcity in Conventional Economics; A Critical Analysis With Islamic Economic Perspective-Konsep Scarcity Dalam Ekonomi Konvensional; Sebuah Analisis Kritis Dengan Perspektif Ekonomi Islam. *Jurnal Ilmiah Ekonomi Islam*, 10(1), 472. <https://doi.org/10.29040/jiei.v10i1.11950>
- Martha, J. (2017). ISU KELANGKAAN AIR DAN ANCAMANNYA TERHADAP KEAMANAN GLOBAL. In *Jurnal Ilmu Politik dan Komunikasi: Vol. VII* (Issue 2). <http://water.usgs.gov/edu/earthwherewater.html>
- Mustamin, R. (2024). The role of formative and summative assessment in improving learning quality and student learning outcomes. *Journal of Strategy and Transformation in Educational Management*, 1(1), 17–21. <https://jostem.professoronline.com/index.php/journal/article/view/4>
- Novikasari, I. (2017). Uji Validitas Instrumen. *Seminar Nasional Riset Inovatif 2017*, 1(1), 530–535.
- Oktavianto Putra, S., Qudsiyah, K., Purnomo Susanto, H., Matematika, P., & PGRI Pacitan, S. (2020). *PRESENTASE KESEIMBANGAN TINGKAT KESULITAN SOAL-SOAL UJI KOMPETENSI MATERI FUNGSI PADA BUKU TEKS MATEMATIKA SISWA KELAS X SMA BERDASARKAN TAKSONOMI BLOOM*. <http://repository.stkippacitan.ac.id>
- Ridhwan, M. (2017). Understanding Formative and Summative Assessment for EFL Teachers: Theoretical Reflections on Assessment for Learning. *J-SHMIC: Journal of English for Academic*, 4(1), 40–50. [https://doi.org/10.25299/JSHMIC.2017.VOL4\(1\).505](https://doi.org/10.25299/JSHMIC.2017.VOL4(1).505)
- Siregar, T. M., Agustina, N., Siahaan, F. B., Sinaga, L. L., Prabawa, I., & Sagala, R. Z. (2024). *Analisis Validitas dan Reliabilitas Butir Soal Pilihan Ganda Dalam Mata Pelajaran Matematika Kelas 12 Semester Genap di SMAN 1 Adiankoting*. 1(2).
- Sudaryono, Rahardja, U., Aini, Q., Isma Graha, Y., & Lutfiani, N. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, 1364(1), 012050. <https://doi.org/10.1088/1742-6596/1364/1/012050>
- Sudijono. (2011). *Pengantar Evaluasi Pendidikan*. <https://www.rajagrafindo.co.id/produk/pengantar-evaluasi-pendidikan/>
- Suprpto, E., Saryanto, S., Sumiharsono, R., & Ramadhan, S. (2020). The analysis of instrument quality to measure the students' higher order thinking skill in physics learning. *Journal of Turkish Science Education*, 17(4), 520–527. <https://doi.org/10.36681/tused.2020.42>
- Thompson, W. J., Clark, A. K., & Jake, W. (2024). *Improving Instructional Decision-Making Using Diagnostic Classification Models*. <https://doi.org/10.1111/emip.12619>
- Trianingsih, L. (2023). *Analisis Kualitas Tes dan Butir Soal Integral pada Evaluasi Formatif Matematika Teknik*. 9(2), 44–57.
- Triola, M. M., & Holloway, W. J. (2011). Enhanced virtual microscopy for collaborative education. *BMC Medical Education*, 11(1), 1–7. <https://doi.org/10.1186/1472-6920-11-4/TABLES/2>
- Welsandt, N. C. J., Fortunati, F., Winther, E., & Abs, H. J. (2024). Constructing and validating authentic assessments: the case of a new technology-based assessment of economic literacy. *Empirical Research in Vocational Education and Training*, 16(1), 1–27. <https://doi.org/10.1186/S40461-024-00158-0/TABLES/5>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/J.STUEDUC.2011.03.001>