



Analisis Butir Soal pada Materi Pasar dan Terbentuknya Harga Keseimbangan Mata Pelajaran Ekonomi Kelas X

Dinar Arfiansah¹, Fitri Rahmadhani², Olivia Junu Bentu³, Dian Rachmawati⁴
Fakultas Ekonomi dan Bisnis, Universitas Negeri Malang
Email: dinar.arfiansah.2304316@students.um.ac.id

A B S T R A C T

This study aims to analyze the quality of test items on the market and equilibrium price material for class X Economics through validity, reliability, difficulty level, and discriminating power tests. A quantitative descriptive approach was employed using a saturated sampling technique involving all class X students at SMA Negeri 3 Blitar. The instrument consisted of 20 multiple-choice questions and 5 essay questions, analyzed using IBM SPSS 25. The results indicate that the majority of test items met the validity criteria, although a few items were found to be invalid or could not be tested. In terms of consistency, the multiple-choice questions yielded high reliability, while the essay questions fell into the moderate category. The difficulty level aspect was dominated by the easy category, while the discriminating power of the multiple-choice questions varied and tended to be low, whereas the essay questions were in the sufficient-to-good category. These findings suggest that the instrument is reasonably adequate; however, further refinement is still needed to measure students' abilities more precisely. Periodic revision of the instrument is strongly recommended to enhance the accuracy and quality of economics learning evaluation in the future.

Keywords: item analysis, validity, reliability, difficulty level, discrimination index.

A B S T R A K

Penelitian ini bertujuan menganalisis kualitas butir soal pada materi pasar dan harga keseimbangan kelas X melalui pengujian validitas, reliabilitas, tingkat kesukaran, serta daya pembeda. Pendekatan kuantitatif deskriptif diterapkan dengan teknik sampling jenuh yang mencakup seluruh siswa kelas X di SMA Negeri 3 Blitar. Instrumen berupa 20 soal pilihan ganda dan 5 soal uraian dianalisis menggunakan IBM SPSS 25. Hasil analisis mengindikasikan bahwa mayoritas butir soal telah memenuhi kriteria valid, meskipun beberapa soal tidak valid atau tidak dapat diuji. Dari segi konsistensi, soal pilihan ganda memperoleh reliabilitas tinggi, sedangkan soal uraian pada kategori sedang. Aspek tingkat kesukaran didominasi kategori mudah, sementara daya pembeda soal pilihan ganda bervariasi dan cenderung rendah, sedangkan soal uraian berada pada kategori cukup hingga baik. Temuan ini menyiratkan bahwa instrumen tersebut sudah cukup layak, tetapi masih perlu penyempurnaan untuk mengukur kemampuan peserta didik secara lebih presisi. Revisi instrumen secara berkala sangat disarankan guna meningkatkan akurasi dan kualitas evaluasi pembelajaran ekonomi di masa mendatang.

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

PENDAHULUAN

Penilaian dalam ranah pendidikan kerap diidentikkan dengan konsep evaluasi. Evaluasi didefinisikan sebagai upaya sistematis untuk mengukur pencapaian hasil belajar sekaligus menyempurnakan proses pembelajaran. Secara spesifik, evaluasi merujuk pada proses terstruktur guna menilai sejauh mana tujuan instruksional telah tercapai oleh peserta didik (Kusumawati & Hadi, 2018). Di dalam konteks pendidikan, evaluasi memegang peran



krusial sebagai fondasi perumusan kebijakan, demi membentuk sumber daya manusia yang kompetitif di tengah era globalisasi. Dengan demikian, evaluasi atau penilaian merupakan elemen integral yang tak terpisahkan dari seluruh proses di lembaga pendidikan. Secara garis besar, evaluasi memiliki dua tujuan pokok, yakni membuktikan tingkat pemahaman peserta didik pasca-proses pembelajaran, serta menjadi rujukan untuk perbaikan proses tersebut.

Pelaksanaan evaluasi memerlukan alat atau teknik penilaian agar lebih terfokus dan terarah. Instrumen evaluasi pendidikan yang dimanfaatkan untuk pengumpulan data mencakup tes maupun nontes. Mengenai pengertian tes, Nurkencana dan Sunartana mendefinisikannya sebagai metode penilaian berupa tugas atau rangkaian tugas yang dikerjakan oleh siswa atau kelompok siswa, sehingga menghasilkan nilai yang merefleksikan perilaku atau prestasi belajar mereka (Magdalena et al., 2020). Salah satu format evaluasi yang paling lazim adalah tes tertulis, yang biasanya dirancang dalam bentuk butir soal untuk mengukur pemahaman peserta didik terhadap materi pelajaran (Arikunto, 2010). Khusus dalam pembelajaran Ekonomi di tingkat sekolah menengah atas (SMA), kualitas soal yang prima sangatlah esensial untuk menilai kemampuan berpikir kritis dan analitis siswa terhadap fenomena ekonomi yang rumit (Mulyasa, 2013). Meskipun demikian, masih kerap ditemui soal-soal dalam proses pembelajaran yang belum melalui analisis kualitas secara memadai, sehingga hasil penilaiannya berisiko kurang akurat dalam merepresentasikan kemampuan siswa secara autentik (S. M. Brookhart & Nitko, 2019).

Dalam mata Pelajaran Ekonomi, terkhususnya pada materi Pasar dan Harga Keseimbangan, kemampuan siswa untuk memahami konsep dasar seperti interaksi antara permintaan dan penawaran, mekanisme pembentukan harga, dan perubahan keseimbangan sangat penting untuk pemahaman ekonomi yang lebih kompleks di tingkat berikutnya. Konsep-konsep ini merupakan salah satu konsep utama yang mendasari analisis perilaku pasar dan pengambilan Keputusan ekonomi, sehingga instrument soal yang digunakan perlu dianalisis secara mendalam untuk memastikan bahwa soal tersebut mampu mengukur capaian kompetensi siswa secara tepat dan akurat.

Analisis butir soal (item analysis) menjadi tahapan krusial yang wajib dilaksanakan guna meningkatkan kualitas soal yang telah dirancang. Proses ini dimaksudkan untuk mengklasifikasikan butir soal menjadi kategori baik, kurang memadai, atau tidak layak pakai. Lewat analisis tersebut, informasi mengenai kelemahan soal dapat terungkap, sekaligus menyediakan panduan perbaikan (Farida & Musyarofah, 2021). Soal yang dipakai dalam evaluasi hasil belajar siswa harus berpedoman pada standar yang telah ditetapkan (Zein et al., 2013). Karenanya, instrumen penilaian diharuskan memenuhi sejumlah kriteria, seperti bukti validitas, reliabilitas, tingkat kesukaran, serta daya pembeda butir soal. Keempat elemen ini merupakan komponen analisis butir soal secara kuantitatif.

Validitas menyangkut apakah soal tes benar-benar mampu mengukur secara akurat apa yang dimaksudkan untuk diukur. Taherdoost (2016) menyatakan bahwa validitas merujuk pada tingkat ketepatan tes dalam mengukur apa yang seharusnya diukur. Penilaian validitas suatu alat ukur dapat dipahami sebagai pembangunan argumen berbasis bukti terkait sejauh mana instrumen tersebut efektif mengukur target pengukuran (Sullivan, 2011). Sementara itu, reliabilitas butir soal berkaitan dengan derajat kepercayaan terhadap hasil pengukuran. Reliabilitas mencerminkan kemampuan tes untuk menghasilkan pengukuran yang konsisten dan dapat diandalkan (Bajpai, 2014). Relasi antara validitas dan reliabilitas menegaskan bahwa validitas bersifat primer, dengan reliabilitas sebagai pendukung yang memperkuatnya (Sukiman, 2012). Mengenai tingkat kesukaran butir soal, Boopathiraj & Chellamani (2013) mendefinisikannya sebagai proporsi peserta tes yang berhasil menjawab dengan benar. Tingkat kesukaran ini ditentukan berdasarkan performa



siswa dalam menjawab, bukan sekadar asumsi penyusun soal, sebab soal yang dianggap mudah atau sulit oleh guru belum tentu demikian bagi siswa. Di samping itu, butir soal juga harus memiliki kemampuan membedakan siswa yang menguasai materi dari yang belum. Kemampuan ini dikenal sebagai daya beda atau diskriminasi. Kocdar et al. (2016) menjelaskan bahwa daya beda adalah kapabilitas item untuk memisahkan siswa berprestasi tinggi dari yang rendah. Dengan kata lain, soal berkualitas adalah yang dapat dijawab benar oleh siswa berpengetahuan, tetapi tidak oleh yang belum menguasai.

Dengan demikian, analisis terhadap validitas, reliabilitas, tingkat kesukaran, dan daya beda bertujuan untuk mengungkap kualitas butir soal secara mendalam. Selain itu, analisis ini memastikan kemampuan siswa dalam memahami konsep pasar dan harga keseimbangan secara autentik. Analisis butir soal juga membantu guru menyediakan umpan balik yang lebih presisi, merancang pembelajaran yang lebih adaptif, serta memfasilitasi pengambilan keputusan yang akurat oleh pendidik.

Oleh karena itu, penelitian ini bertujuan menganalisis butir soal pada materi pasar dan harga keseimbangan, yang sangat dibutuhkan para pendidik mengingat minimnya referensi instrumen evaluasi yang difokuskan pada materi tersebut—khususnya yang telah melalui pengujian analisis seperti validitas, reliabilitas, daya beda, dan tingkat kesukaran. Melalui penelitian ini, pendidik dan pihak sekolah, termasuk pengembang kurikulum, dapat memperoleh rujukan dalam menyusun serta menyempurnakan soal Ekonomi, terutama materi pasar dan harga keseimbangan, sehingga penilaian menjadi lebih akurat dan kredibel, sekaligus mendukung proses pembelajaran yang lebih responsif.

Berdasarkan uraian di atas, pertanyaan penelitian yang diajukan adalah: (1) Bagaimana tingkat validitas butir soal pada materi pasar dan harga keseimbangan kelas X? (2) Bagaimana reliabilitas instrumen soal yang digunakan? (3) Bagaimana tingkat kesukaran dan daya pembeda setiap butir soal? Artikel ini disusun dengan urutan sebagai berikut: bagian Metode menjelaskan rancangan penelitian, populasi, instrumen, dan teknik analisis; bagian Hasil dan Pembahasan memaparkan temuan uji validitas, reliabilitas, tingkat kesukaran, serta daya pembeda; dan bagian Kesimpulan merangkum temuan utama beserta implikasinya bagi pengembangan instrumen evaluasi pembelajaran ekonomi.

METODE

Penelitian ini menerapkan pendekatan kuantitatif deskriptif guna menguraikan kualitas butir soal pada materi pasar dan pembentukan harga keseimbangan dalam pembelajaran ekonomi kelas X. Pendekatan tersebut dipilih karena data yang dianalisis berupa angka-angka yang diperoleh dari hasil tes peserta didik. Data tersebut kemudian diolah untuk mengidentifikasi karakteristik setiap butir soal. Dengan begitu, penelitian ini tidak hanya menyajikan temuan, tetapi juga memberikan gambaran objektif tentang mutu instrumen penilaian yang dipakai (Sugiyono, 2021).

Penelitian dilakukan di SMA Negeri 3 Blitar dengan melibatkan seluruh peserta didik kelas X yang mengikuti pembelajaran ekonomi pada materi pasar dan harga keseimbangan. Karena populasi tidak terlalu besar, teknik pengambilan sampel yang diterapkan adalah sampling jenuh atau total sampling, yakni menjadikan seluruh anggota populasi sebagai sampel. Teknik ini memastikan data yang dikumpul lebih komprehensif dan mencerminkan kondisi lapangan secara akurat (Arikunto, 2020).

Instrumen penelitian berupa tes hasil belajar yang mencakup 20 soal pilihan ganda dan 5 soal uraian. Penyusunan soal dimulai dari kisi-kisi yang disesuaikan dengan kompetensi dasar serta indikator pencapaian pembelajaran. Kisi-kisi ini berfungsi sebagai panduan agar setiap butir soal mewakili aspek kemampuan yang dimaksudkan. Penggunaan dua jenis soal juga bertujuan untuk memberikan keragaman dalam pengukuran kemampuan



peserta didik, baik pemahaman konsep maupun analisis sederhana terhadap isu ekonomi. Pengumpulan data dilaksanakan melalui pemberian tes kepada seluruh peserta didik sebagai sampel. Tes tersebut digelar dalam satu pertemuan sesuai jadwal yang telah ditetapkan. Jawaban peserta didik diperiksa dan diskor berdasarkan rubrik penskoran yang telah disusun sebelumnya. Skor tersebut kemudian menjadi dasar analisis butir soal.

Analisis data penelitian ini menekankan empat aspek pokok, yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Keempat aspek ini dijadikan tolok ukur kualitas instrumen. Uji validitas bertujuan untuk menilai sejauh mana setiap butir soal mengukur kompetensi yang ditargetkan, dengan membandingkan skor butir soal terhadap skor total. Butir soal dengan korelasi tinggi dianggap valid karena selaras dengan tes secara keseluruhan, sementara yang berkorelasi rendah memerlukan perbaikan karena kurang merepresentasikan kompetensi yang diukur (Widoyoko, 2022).

Uji reliabilitas selanjutnya mengukur konsistensi instrumen, di mana instrumen berkualitas akan menghasilkan output yang stabil dalam kondisi serupa. Reliabilitas dihitung menggunakan rumus Kuder-Richardson untuk soal pilihan ganda, sedangkan untuk soal uraian diterapkan pendekatan konsistensi internal. Nilai reliabilitas yang tinggi mencerminkan tingkat keandalan instrumen yang memadai (Azwar, 2021).

Analisis tingkat kesukaran bertujuan untuk mengetahui proporsi peserta didik yang berhasil menjawab benar setiap soal. Hasil perhitungan mengelompokkan butir soal menjadi kategori mudah, sedang, atau sulit. Komposisi tingkat kesukaran ini perlu diperhatikan, sebab soal yang terlalu mudah atau sulit cenderung kurang efektif dalam mengukur kemampuan peserta didik secara optimal (Putri & Lestari, 2023).

Analisis daya pembeda sebagai tahap akhir bertujuan untuk mengevaluasi kemampuan soal dalam membedakan peserta didik berdasarkan tingkat penguasaannya. Perhitungan dilakukan dengan membandingkan performa kelompok atas dan bawah. Soal dengan daya pembeda tinggi efektif mengidentifikasi perbedaan penguasaan materi, sedangkan yang rendah disarankan untuk direvisi atau diperbaiki (Rahman, 2022). Data yang diperoleh dari tes kemudian dilakukan analisis menggunakan metode statistik deskriptif. Analisis dalam penelitian ini dilakukan dengan menggunakan perangkat lunak uji statistik IBM SPSS versi 25 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 reliabilis

Pengujian reliabilitas digunakan untuk mengukur soal yang reliable atau konsistendengan 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
$P < 0,30$	Sukar
$0,30 < P < 0,70$	Sedang
$0,70 < P < 1,00$	Rendah

Sumber: Diolah dari Arikunto (2012)

4. Daya pembeda

Pengukuran daya pembeda dimanfaatkan untuk menilai kemampuan soal dalam memisahkan tingkat kemampuan kognitif siswa. Uji tingkat kesukaran dilakukan dengan merujuk pada klasifikasi nilai uji tingkat kesukaran seperti pada Tabel 4 berikut.

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
$D = \text{negatif}$	Sebaiknya diganti

Sumber: Diolah dari Arikunto (2012)

Berdasarkan keseluruhan hasil analisis tersebut, setiap butir soal kemudian diklasifikasikan ke dalam tiga kategori, yaitu layak digunakan, perlu direvisi, dan tidak layak digunakan. Klasifikasi ini menjadi dasar dalam menentukan tindak lanjut terhadap instrumen yang telah disusun, sehingga diharapkan dapat meningkatkan kualitas penilaian dalam pembelajaran ekonomi, khususnya pada materi pasar dan terbentuknya harga keseimbangan

HASIL DAN PEMBAHASAN

1. Uji Validitas Butiran Soal

Uji validitas menjadi tahapan esensial dalam analisis butir soal, yang dimaksudkan untuk memverifikasi bahwa setiap item tes secara tepat mengukur kompetensi yang ditargetkan (Siregar et al., 2024). Dalam penelitian ini, pengujian validitas dilakukan terhadap dua jenis instrumen, yakni 20 soal pilihan ganda dan 5 soal uraian, dengan menerapkan teknik korelasi antara skor tiap butir dan skor total. Nilai r tabel yang digunakan sebagai patokan adalah 0,2084, sehingga butir soal dikategorikan valid jika nilai r hitungnya lebih besar dari r tabel.

Tabel 5. Uji Validitas Pilihan Ganda

No. Butir Soal	r Hitung	r Tabel	Keterangan
1	.461**	2,084	valid
2	.302**	2,084	valid
3	.475**	2,084	valid



4	.447**	2,084	valid
5	.678**	2,084	valid
6	.273*	2,084	valid
7	.351**	2,084	valid
8	0.045	2,084	Tidak valid
9	.619**	2,084	valid
10	.616**	2,084	valid
11	.333**	2,084	valid
12	.c	2,084	Tidak bisa diuji
13	.427**	2,084	valid
14	0.1	2,084	Tidak valid
15	.430**	2,084	valid
16	.239*	2,084	valid
17	.480**	2,084	valid
18	.533**	2,084	valid
19	.c	2,084	Tidak bisa diuji
20	.503**	2,084	valid

Sumber : Output SPSS, diolah oleh peneliti (2026)

Tabel 6. Uji Validitas Uraian

No. Butir Soal	r Hitung	r Tabel	Keterangan
1	.662**	0,2084	Valid
2	.589**	0,2084	Valid
3	.569**	0,2084	Valid
4	.664**	0,2084	Valid
5	.685**	0,2084	Valid

Sumber : Output SPSS, diolah oleh peneliti (2026)

Hasil analisis terhadap soal pilihan ganda menunjukkan bahwa mayoritas butir telah memenuhi standar validitas. Dari 20 soal yang diuji, sebanyak 16 butir dikategorikan valid, yaitu nomor 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 15, 16, 17, 18, dan 20, karena nilai r hitungnya melebihi 0,2084. Fakta ini menandakan bahwa butir-butir tersebut efektif merepresentasikan indikator pembelajaran yang dimaksud. Namun, dua butir soal gagal memenuhi kriteria validitas, yakni nomor 8 ($r = 0,045$) dan nomor 14 ($r = 0,1$), sebab nilai korelasi mereka berada di bawah r tabel. Selain itu, dua butir lain, yaitu nomor 12 dan 19,



tidak dapat diuji validitasnya karena data tidak memenuhi persyaratan analisis (ditandai dengan nilai “.c”). Soal-soal yang tidak valid atau tidak dapat diuji ini memerlukan evaluasi ulang, mencakup kejelasan redaksi, kecocokan dengan indikator, serta tingkat kesulitan.

Pada instrumen soal uraian yang terdiri dari 5 butir, seluruh item dinyatakan valid. Hal ini dibuktikan oleh nilai r hitung yang berkisar antara 0,569 hingga 0,685, semuanya melebihi r tabel sebesar 0,2084. Dengan demikian, dapat disimpulkan bahwa soal uraian yang dirancang telah secara akurat mengukur kemampuan peserta didik sesuai tujuan pembelajaran.

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 proses pengembangan instrumen evaluasi di sekolah, guru perlu memastikan bahwa setiap butir soal yang digunakan benar-benar mampu mengukur kemampuan berpikir serta pemahaman siswa, tidak sekadar menguji aspek hafalan (Suprpto et al., 2020). Dengan tingkat validitas yang baik, kegiatan pembelajaran menjadi lebih bermakna karena guru memiliki instrumen yang tepat untuk memantau perkembangan siswa sekaligus memberikan umpan balik yang relevan.

Secara keseluruhan, dari 25 butir soal yang dianalisis, sebanyak 21 soal dinyatakan valid, 2 soal tidak valid, dan 2 soal tidak dapat diuji. Temuan ini mengindikasikan bahwa instrumen tes yang dipakai telah memiliki kualitas yang cukup baik untuk mengukur hasil belajar siswa. Validitas butir soal merupakan elemen krusial karena secara langsung memengaruhi akurasi hasil evaluasi. Butir yang valid efektif membedakan siswa yang telah menguasai materi dengan yang belum, sementara butir tidak valid berisiko menimbulkan data bias dan tidak merepresentasikan kemampuan siswa secara autentik.

Meskipun mayoritas butir soal telah memenuhi kriteria validitas, revisi tetap diperlukan pada soal yang belum valid atau tidak dapat diuji. Perbaikan ini dapat dilakukan melalui penyempurnaan redaksi, penyesuaian dengan indikator pembelajaran, serta penilaian ulang tingkat kesulitan dan daya pembeda. Akibatnya, instrumen evaluasi yang dihasilkan akan semakin berkualitas dan memberikan gambaran yang lebih presisi tentang pencapaian belajar peserta didik. Dalam konteks pembelajaran Ekonomi, validitas butir soal memegang peran strategis karena materi pelajaran erat kaitannya dengan kehidupan sehari-hari siswa (Welsandt et al., 2024). Karenanya, butir soal valid tidak hanya mencerminkan pencapaian kognitif, tetapi juga merangsang refleksi kritis terhadap isu-isu ekonomi di sekitar, seperti kelangkaan sumber daya, pengambilan keputusan rumah tangga, serta pengelolaan kebutuhan dan keinginan. Dengan demikian, evaluasi bukan sekadar alat



pengukur, melainkan komponen integral dari proses pembelajaran (Elliott & Balasubramanyam, 2016). Hasil uji validitas ini dapat menjadi fondasi kokoh untuk pengambilan keputusan perbaikan pembelajaran, baik dalam strategi pengajaran, penguatan materi, maupun pengembangan soal masa depan yang lebih kontekstual, adil, dan mendukung literasi ekonomi siswa secara komprehensif (Black & Wiliam, 1998; Wiliam, 2011).

2. Uji Realibilitas

Uji reliabilitas pada instrumen penelitian bertujuan untuk menilai sejauh mana kuesioner yang digunakan dalam pengumpulan data telah mencapai tingkat konsistensi yang memadai (Dewi & Sudaryanto, 2020). Pada penelitian ini, pengujian reliabilitas dilakukan dengan menggunakan koefisien Alpha Cronbach. Menurut Putri (dalam Dewi & Sudaryanto, 2020), apabila suatu variabel memiliki nilai Alpha Cronbach lebih dari 0,60, maka variabel tersebut dapat dinyatakan reliabel atau konsisten dalam melakukan pengukuran.

Berdasarkan data dan tabel, terdapat dua kategori soal yang diuji. Kategori pertama dengan soal pilihan ganda berjumlah 20 butir soal yang dimana nilai Cronbach's Alpha menunjukkan sebesar 0.702 yang termasuk didalam kategori realibilitas tinggi. Artinya, soal-soal pada kategori pilihan ganda ini konsisten dalam mengukur konstruk yang sama dan dapat diandalkan untuk proses evaluasi pembelajaran. Sementara itu, untuk kategori soal uraian yang terdiri dari 5 butir dengan nilai Cronbach's Alpha sebesar 0,619—yang termasuk dalam kategori reliabilitas sedang—menandakan bahwa instrumen tersebut masih dapat digunakan, meskipun tingkat konsistensi antar butir belum optimal. Karenanya, perbaikan lanjutan diperlukan, khususnya pada kejelasan soal, kesesuaian dengan indikator pembelajaran, serta variasi tingkat kognitif, guna meningkatkan reliabilitasnya.

Tabel 7. Uji Relibilitas

Cronbach's Alpha	N of Items	Keterangan
0.702	20	Relibilitas tinggi
0.619	5	Realibilitas Sedang

Sumber : Output SPSS, diolah oleh peneliti (2026)

Nilai reliabilitas sebesar 0,702 pada 20 butir soal menandakan bahwa instrumen tersebut memiliki tingkat konsistensi yang baik dalam mengukur kemampuan siswa. Fakta ini mengimplikasikan bahwa hasil pengukuran yang diperoleh relatif stabil dan dapat dipercaya, terutama dalam mengevaluasi pemahaman peserta didik terhadap materi Ekonomi. Tingkat konsistensi yang tinggi ini esensial karena menjamin bahwa hasil evaluasi secara objektif merefleksikan kemampuan siswa, tanpa dipengaruhi faktor kebetulan atau inkonsistensi instrumen.

Di sisi lain, nilai reliabilitas sebesar 0,619 pada 5 butir soal menunjukkan bahwa instrumen berada pada kategori sedang. Meskipun masih dapat digunakan dalam proses penilaian, tingkat konsistensinya belum sepenuhnya optimal. Oleh karena itu, butir-butir soal tersebut perlu ditinjau kembali, terutama dari segi kejelasan bahasa, kesesuaian dengan indikator pembelajaran, serta keseimbangan tingkat kesulitan. Perbaikan ini esensial agar instrumen mampu menghasilkan data yang lebih akurat dan konsisten.

Dalam pengembangan instrumen evaluasi, reliabilitas merupakan aspek penting yang perlu diperhatikan selain validitas. Instrumen dengan reliabilitas yang memadai akan menghasilkan data yang lebih kredibel, sehingga dapat menjadi fondasi pengambilan keputusan pembelajaran, seperti penyempurnaan metode pengajaran, penilaian tingkat penguasaan materi siswa, serta pengembangan butir soal yang lebih unggul di masa depan.



Karenanya, guru dianjurkan untuk tidak hanya terpaku pada uji validitas, melainkan melengkapi pengembangan instrumen dengan uji reliabilitas demi menjamin kualitas evaluasi secara menyeluruh (Earle & Earle, 2020; Govindasamy et al., 2024; Thompson et al., 2024).

3. Uji Tingkat Kesukaran Soal

Tingkat kesukaran suatu butir soal menunjukkan sejauh mana siswa mampu menjawab soal dengan benar, yang dihitung menggunakan rumus $P = \text{Mean} / \text{Skor Maksimum}$. Pada soal pilihan ganda ini, skor maksimum adalah 3 (rentang nilai 0–3). Berdasarkan data yang telah diolah, analisis dilakukan terhadap 20 butir soal yang diujikan kepada 89 responden.

Tabel 8. Uji Tingkat Kesukaran Pilihan Ganda

No. Butiran Soal	N		Mean	Tingkat Kesukaran ($P = \text{Mean}/3$)	Keterangan
	Valid	Missing			
1	89	0	2.9	0.96	Mudah
2	89	0	2.9	0.96	Mudah
3	89	0	2.12	0.70	Sedang
4	89	0	2.7	0.90	Mudah
5	89	0	2.83	0.94	Mudah
6	89	0	0.4	0.13	Sukar
7	89	0	2.66	0.88	Mudah
8	89	0	2.97	0.99	Mudah
9	89	0	2.7	0.90	Mudah
10	89	0	2.73	0.91	Mudah
11	89	0	2.66	0.88	Mudah
12	89	0	3	1.00	Mudah
13	89	0	2.97	0.99	Mudah
14	89	0	2.97	0.99	Mudah
15	89	0	2.9	0.96	Mudah
16	89	0	2.9	0.96	Mudah
17	89	0	2.87	0.95	Mudah
18	89	0	2.66	0.88	Mudah
19	89	0	3	1	Mudah
20	89	0	2.83	0.94	Mudah

Sumber: Output SPSS, diolah oleh peneliti (2026)

Tabel 9. Uji Tingkat Kesukaran Uraian

No. Butiran Soal	N		Mean	Tingkat Kesukaran ($P = \text{Mean}/8$)	Keterangan
	Valid	Missing			
1	89	0	6.6	0.82	Mudah
2	89	0	6.99	0.87	Mudah
3	89	0	7.03	0.87	Mudah



4	89	0	6.7	0.83	Mudah
5	89	0	6.96	0.87	Mudah

Sumber: Output SPSS, diolah oleh peneliti (2026)

Hasil perhitungan menunjukkan bahwa sebanyak 18 dari 20 butir soal memiliki nilai P di atas 0,70, sehingga tergolong dalam kategori mudah. Hal ini menandakan bahwa sebagian besar siswa mampu menjawab soal dengan tingkat keberhasilan yang tinggi. Nilai P tertinggi terdapat pada soal nomor 12 dan 19 dengan nilai 1,00, yang berarti hampir seluruh siswa menjawab kedua soal tersebut dengan benar. Selain itu, beberapa butir lain seperti nomor 8, 13, dan 14 juga memiliki nilai yang sangat tinggi (0,99), sehingga semakin menegaskan dominasi kategori mudah.

Sementara itu, satu butir soal termasuk kategori sedang, yakni nomor 3 dengan nilai P sebesar 0,70, yang berada pada batas rentang $0,30 \leq P \leq 0,70$. Di sisi lain, satu butir soal tergolong sukar, yaitu nomor 6 dengan nilai P sebesar 0,13 ($P < 0,30$), menandakan bahwa hanya sebagian kecil siswa yang berhasil menjawabnya dengan benar.

Secara umum, dapat disimpulkan bahwa tingkat kesukaran soal masih didominasi oleh kategori mudah. Hal ini membuktikan bahwa soal-soal yang diberikan cenderung kurang menantang bagi siswa, meskipun di sisi lain juga dapat mengindikasikan bahwa siswa telah memahami materi dengan cukup baik. Namun, dari perspektif evaluasi pembelajaran, komposisi tingkat kesukaran yang tidak seimbang menjadi kelemahan karena belum sepenuhnya merepresentasikan variasi kemampuan siswa secara komprehensif.

Kondisi ini menunjukkan instrumen evaluasi ini belum sepenuhnya optimal dalam membedakan kemampuan siswa, khususnya pada kelompok dengan kemampuan menengah dan rendah. Oleh karena itu, dalam penyusunan soal selanjutnya disarankan untuk memperhatikan proporsi tingkat kesukaran agar lebih seimbang, misalnya dengan menambah jumlah soal kategori sedang dan sukar. Dengan demikian, hasil evaluasi tidak hanya menunjukkan tingkat keberhasilan siswa secara umum, tetapi juga mampu mengukur kemampuan berpikir pada berbagai tingkat kognitif secara lebih akurat dan komprehensif.

4. Uji Daya Pembeda soal

Uji daya pembeda dilaksanakan untuk mengukur sejauh mana suatu butir soal efektif membedakan siswa dengan penguasaan materi tinggi dari yang rendah. Dalam analisis ini, indikator yang dipakai adalah nilai Corrected Item-Total Correlation (CITC), di mana semakin tinggi nilai CITC, semakin superior kemampuan butir soal dalam memisahkan kemampuan siswa. Secara garis besar, butir dengan $CITC \geq 0,40$ diklasifikasikan sebagai “baik”, rentang 0,20–0,39 sebagai “cukup”, sedangkan $< 0,20$ sebagai “jelek”.

Tabel 10. Uji Daya Pembeda Pilihan Ganda

No. Butir Soal	<i>Corrected Item-Total Correlation</i>	Keterangan
1	0.403	Baik
2	0.262	Cukup
3	0.252	Cukup
4	0.333	Cukup
5	0.605	Baik
6	0.107	Jelek
7	0.156	Jelek
8	-0.006	Jelek
9	0.469	Baik



10	0.517	Baik
11	0.196	Jelek
12	0	Jelek
13	0.397	Cukup
14	0.051	Jelek
15	0.367	Cukup
16	0.125	Jelek
17	0.413	Baik
18	0.429	Baik
19	0	Jelek
20	0.372	Cukup

Sumber: Output SPSS, diolah oleh peneliti (2026)

Tabel 11. Uji Daya Pembeda Uraian

No. Butir Soal	<i>Corrected Item-Total Correlation</i>	Keterangan
1	0.35	Cukup
2	0.325	Cukup
3	0.309	Cukup
4	0.424	Baik
5	0.496	Baik

Sumber: Output SPSS, diolah oleh peneliti (2026)

Hasil analisis terhadap 20 butir soal pilihan ganda menunjukkan variasi kualitas daya pembeda yang cukup beragam. Sebanyak 6 butir termasuk kategori “baik”, yakni nomor 1 (0,403), 5 (0,605), 9 (0,469), 10 (0,517), 17 (0,413), dan 18 (0,429), yang dinilai efektif membedakan siswa berprestasi tinggi dari rendah. Selain itu, 7 butir berada pada kategori “cukup”, yaitu nomor 2 (0,262), 3 (0,252), 4 (0,333), 13 (0,397), 15 (0,367), dan 20 (0,372); meskipun belum optimal, butir-butir ini masih dapat dimanfaatkan dengan perbaikan tertentu.

Namun demikian, terdapat 7 butir soal yang tergolong dalam kategori “jelek”, yaitu nomor 6 (0,107), 7 (0,156), 8 (-0,006), 11 (0,196), 12 (0), 14 (0,051), 16 (0,125), dan 19 (0). Nilai CITC yang rendah bahkan negatif menunjukkan bahwa butir-butir tersebut tidak mampu membedakan kemampuan siswa secara efektif, bahkan berpotensi menghasilkan hasil evaluasi yang bias. Sejalan dengan pendapat Nunnally & Bernstein (1994), butir soal dengan nilai diskriminasi di bawah 0,20 sebaiknya tidak digunakan dalam instrumen akhir karena tidak memberikan kontribusi yang signifikan dalam membedakan performa siswa (Triola & Holloway, 2011).

Selanjutnya, pada analisis daya pembeda soal uraian yang berjumlah 5 butir, diperoleh hasil yang relatif lebih baik. Dua butir soal, yaitu nomor 4 (0,424) dan 5 (0,496), termasuk dalam kategori “baik”, sedangkan tiga butir lainnya, yaitu nomor 1 (0,350), 2 (0,325), dan 3 (0,309), berada pada kategori “cukup”. Fakta ini mengindikasikan bahwa secara keseluruhan, soal uraian telah memiliki daya pembeda yang cukup baik untuk membedakan tingkat pemahaman siswa, walaupun masih ada peluang penyempurnaan.

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.

Hasil analisis daya pembeda mengungkap bahwa kualitas butir soal dalam instrumen tidak seragam. Butir dengan daya pembeda tinggi dapat menjadi rujukan untuk penyusunan soal berikutnya karena lebih akurat mengukur kemampuan berpikir siswa. Sebaliknya, butir dengan daya pembeda rendah atau negatif memerlukan evaluasi ulang, termasuk redaksi, kesesuaian indikator pembelajaran, serta tingkat kesukaran yang mungkin terlalu mudah atau sulit.

Dengan demikian, meskipun sebagian butir telah menunjukkan kualitas yang baik, perbaikan tetap diperlukan, khususnya pada butir dengan kategori “jelek”. Revisi dapat dilakukan melalui perbaikan kejelasan pertanyaan, penyesuaian dengan indikator pembelajaran, serta peningkatan kualitas distraktor pada soal pilihan ganda. Melalui upaya tersebut, diharapkan instrumen evaluasi yang digunakan menjadi lebih akurat, objektif, dan mampu memberikan gambaran yang lebih tepat mengenai kemampuan peserta didik.

KESIMPULAN

Berdasarkan hasil penelitian, kualitas butir soal pada materi pasar dan harga keseimbangan untuk kelas X secara keseluruhan tergolong cukup baik, walaupun beberapa butir masih memerlukan perbaikan. Dari 25 butir soal yang dianalisis, 21 butir dinyatakan valid, 2 butir tidak valid, dan 2 butir tidak dapat diuji. Uji reliabilitas mengindikasikan koefisien sebesar 0,702 untuk soal pilihan ganda (kategori tinggi), serta 0,619 untuk soal uraian (kategori sedang), yang menunjukkan instrumen tes cukup konsisten dan layak untuk evaluasi pembelajaran.

Dari aspek tingkat kesukaran, mayoritas butir soal termasuk kategori mudah (18 soal), diikuti 1 soal sedang dan 1 soal sukar. Sementara itu, uji daya pembeda mengungkap variasi kualitas soal pilihan ganda—dengan 6 butir baik, 7 cukup, dan 7 jelek—sedangkan soal uraian didominasi kategori cukup dan baik. Dengan demikian, instrumen penelitian ini sudah mampu mengukur hasil belajar peserta didik dengan cukup baik, tetapi masih perlu revisi terutama pada butir soal yang tidak valid, terlalu mudah, dan memiliki daya pembeda rendah agar kualitas evaluasi menjadi lebih akurat, efektif, dan komprehensif

DAFTAR RUJUKAN

- Afandi, I. (2025). Educational Evaluation: Definition, Function, and General Requirements as a Pillar for Improving the Quality of Learning. *Responsive: Jurnal Manajemen Pendidikan Islam*, 1(2), 125-130.
- Arikunto, S. (2020). *Dasar-dasar evaluasi pendidikan* (Edisi revisi). Jakarta: Bumi Aksara.



- Arikunto, S. (2010). *Prosedur penelitian: Suatu pendekatan praktik* (Edisi revisi). Rineka Cipta.
- Ahmad, K. A., Kadir, N. L. A., Zakaria, M. Z., Zamri, M. L., Rahman, S., & Mustafa, N. M. Y. (2021). Distractor Efficiency of an Intermediate English Proficiency Course Final Examination Paper. 11(2), 103–112.
- Azwar, S. (2021). *Reliabilitas dan validitas* (Edisi 4). Yogyakarta: Pustaka Pelajar.
- Bajpai, S. (2014). Reliability and validity assessment in educational research. *Research Journal of Education*, 1(1), 1–8.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189–193.
- Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students* (8th ed.). Pearson.
- Buuren, S. van, & Eekhout, I. (2023). Validity. 65–68. <https://doi.org/10.1201/9781003216315-7>
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.
- Earle, M., & Earle, L. (2020). Assessment, reliability and validity in education: A practical guide. *Journal of Educational Research*, 18(1), 45–58.
- Elliott, C., & Balasubramanyam, V. (2016). Assessment as integral component of learning. *International Journal of Educational Evaluation*, 4(2), 101–112.
- Finch, W. H., French, B. F., Immekus, J. C., & Dai, S. (2025). Reliability. *Educational and Psychological Measurement*, 51–87.
- Goel, N., Deshmukh, K., Patel, B. C., & Chacko, S. (2021). Tools and Rubrics for Assessment of Learning Outcomes. 211–254
- Emeliazola, & Zakir, S. (2024). Konsep Evaluasi Dalam Pembelajaran. *Jurnal Ilmiah Research and Development Student, Rawamangun/Jurnal Ilmiah Research and Development Student*, 2(2), 53–66.
- Karnia, R. (2024). Importance of Reliability and Validity in Research. *Psychology and Behavioral Sciences*, 13(6), 137–141.
- Kumar, R. V. (2024). Cronbach's Alpha: Genesis, Issues and Alternatives. <https://doi.org/10.1177/ijim.241234970>
- Kocdar, S., Karadeniz, A., Peytcheva-Forsyth, R., & Stoeva, V. (2016). Cheating and plagiarism in e-assessment: Students' perspectives. *Open Praxis*, 10(3), 221–235. <https://doi.org/10.5944/openpraxis.10.3.873>
- Laila, L., Nabila, A., & Widyanti, E. (2024). Konsep dasar evaluasi pembelajaran. *Jurnal Manajemen Dan Pendidikan Agama Islam*, 2(5), 252-262.
- Mulyasa, E. (2013). *Pengembangan dan implementasi kurikulum 2013*. Remaja Rosdakarya.



- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Putri, L. A., & Lestari, S. (2023). Analisis tingkat kesukaran soal dalam evaluasi pembelajaran ekonomi. *Jurnal Kajian Pendidikan*, 11(2), 120–128.
- Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi*, 13(3), 273–284.
- Rahayu, C., Cahyani, B. H., & Khosiyono, B. H. C. (2024). Analysis of mathematics exam practice questions for 6th grade elementary students. *Union*, 12(1), 26–34
- Rahman, F. (2022). Analisis daya pembeda dan tingkat kesukaran soal pada pembelajaran ekonomi SMA. *Jurnal Pendidikan Ekonomi Indonesia*, 10(1), 45–53.
- Rizkiyah, A. M. (2025). ANALISIS KUALITAS TES DAN BUTIR SOAL UJIAN MADRASAH MATA PELAJARAN BAHASA ARAB KELAS IX. *Jurnal Psikososial dan Pendidikan*, 1(2), 533-543.
- Sadie, P. (2023). Problem Analysis. 33–41. https://doi.org/10.1007/978-3-031-25817-6_4
- Saputri, H. A. S., & Larasati, N. J. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986-2995.
- Sari, M. P. (2022). Uji validitas dan reliabilitas instrumen tes hasil belajar siswa. *Jurnal Pendidikan dan Pembelajaran*, 9(1), 33–41.
- Son, A. L. (2019). Instrumentasi kemampuan pemecahan masalah matematis: analisis reliabilitas, validitas, tingkat kesukaran dan daya beda butir soal. *Gema wiralodra*, 10(1), 41-52.
- Sugiyono. (2021). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta.
- Sukiman. (2012). Pengembangan sistem evaluasi. *Insan Madani*.
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, 3(2), 119–120. <https://doi.org/10.4300/JGME-D-11-00075.1>
- Taherdoost, H. (2016). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, 5(3), 28–36.
- Udeh, C. A. (2025). Research methodology. 25–28. https://doi.org/10.70593/978-93-49307-40-7_3
- Validity. (2023). 35–44. <https://doi.org/10.1016/b978-0-12-818630-5.10005-3>
- Widoyoko, E. P. (2022). *Teknik penyusunan instrumen penelitian*. Yogyakarta: Pustaka Pelajar.
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Zakaria, F., Sasoko, W. H., & Utami, E. (2024). Analysis of Question Items Using the Differentiating Power Method. *International Journal of Informatics, Information System and Computer Engineering*, 6(1), 63–69.



Zahfa, F., Fauzi, A., Yansyah, E. A., Fadhillah, F., Syahdilla, J., Zahfa, F., Fauzi, A., Yansyah, E. A., Fadhillah, F., & Syahdilla, J. (2025). Validity and Reliability of Educational Evaluation Tests. *Jurnal Riset Ilmu Pendidikan*, 5(1), 1–5.