

Analisis Butir Soal Tipe Multiple Choices Questions (MCQ) pada Ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMA

Rahmat Fajar¹, Junaidi Junaidi^{2*}

^{1,2}Universitas Negeri Padang

*Corresponding author, e-mail: ikesylvia@fis.unp.ac.id

Abstract

This study aims to determine the quality of the Odd Semester Final Examination Questions for Class XII Sociology Subjects at SMA Negeri 2 Padang Panjang in the 2021/2022 academic year in terms of analysis: 1) content and empirical validity, 2) reliability, 3) distinguishing characteristics, 4) difficulty level. And 5) distracting qualities. This research is a descriptive study using a quantitative approach because all the data or information obtained is manifested as numbers. The research subjects were students of class XII SMA Negeri 2 Padang Panjang, with a total population of 163 students. Data collection techniques were carried out using the documentation method to obtain data on questions, answer keys, and student test results. Based on the results of data analysis, the results of a review of aspects are obtained: 1) The validity of the content of the questions has a high category material aspect, a high category construction aspect, and a high category linguistic aspect. 2) Empirical validity, there are 7 items valid questions (17.5%) and 33 items (82.5%). 3) In terms of reliability, the question has low reliability of 0.19. 4) Discriminatory power, questions with less discriminating power 23 items (57.5%), questions with sufficient discriminating power 14 items (35%), good discriminating power three items (7.5%). 5) Difficulty level, 16 items (40%) are classified as easy, 12 items (30%) are classified as moderate, and 12 items (30%) are difficult. 6) The effectiveness of distractors, 29 items (72.5%) are eligible, and 11 items (27.5%) are not eligible. Thus it can be concluded that the quality of the items tested is not good.

Keywords: Analysis; Question quality; Sociology.

How to Cite: Fajar, R. & Junaidi, J. (2022). Analisis Butir Soal Tipe Multiple Choices Questions (MCQ) pada Ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMA. *Jurnal Sikola: Jurnal Kajian Pendidikan dan Pembelajaran*, 4(1), 23-34.



This is an open access article distributed under the Creative Commons Attribution Share-Alike 4.0 International License. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. © 2022 by author.

Pendahuluan

Evaluasi pembelajaran merupakan salah satu komponen dan tahap yang harus dilakukan oleh guru agar mengetahui keefektifan dari suatu pembelajaran yang sudah berlangsung. Evaluasi merupakan suatu kegiatan untuk memahami, menilai, dan mengukur potensi serta kemampuan siswa dalam mencapai tujuan pembelajaran sehingga dapat diperoleh suatu keputusan (Mania et al., 2020). Hasil yang didapatkan dari evaluasi bisa dijadikan umpan balik bagi guru untuk memperbaiki dan memaksimalkan program serta kegiatan dalam pembelajaran. Evaluasi dalam pendidikan berguna untuk mengukur dan menilai kualitas pendidikan agar tercapainya tujuan dari pembelajaran. Kegiatan evaluasi dilakukan untuk mendapatkan informasi yang menyangkut objek yang sedang dievaluasi. Guru harus melaksanakan evaluasi pembelajaran kepada siswa dengan tujuan agar mengetahui keberhasilan suatu pembelajaran yang telah dilaksanakan oleh siswanya. Untuk mengetahui keberhasilan tersebut guru perlu melakukan penilaian terhadap hasil belajar siswa.

Penilaian hasil belajar sangat berperan penting dalam suatu proses pembelajaran. Penilaian tidak hanya sebagai alat ukur ketercapaian siswa terhadap kompetensi yang telah ditetapkan namun juga memungkinkan guru menggunakan informasi kondisi siswa untuk memperbaiki suatu pembelajaran. (Kemendikbud, 2015). Ketepatan penilaian akan mempengaruhi upaya peningkatan kualitas pendidikan di sekolah, karena itu dapat dikatakan bahwa hasil penilaian merupakan salah satu hal yang mempengaruhi

kualitas pendidikan. Salah satu cara yang dilakukan untuk memperbaiki kualitas sekolah adalah dengan melakukan penilaian hasil belajar, baik itu soal esai maupun soal objektif. Soal yang dominan digunakan sekolah adalah soal pilihan ganda (*Multiple Choices Questions*). Soal yang diujikan untuk mengetahui kemampuan belajar siswa harus memenuhi standar yang ditentukan. Alat penilaian yang berkualitas harus memiliki lima karakteristik yaitu, validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan kualitas pengecoh (Pratiwi et al., 2013). Untuk mendapatkan kualitas soal yang baik perlu adanya keseimbangan dari dari seluruh karakteristik soal tersebut.

Tingkat kesukaran soal yang dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya. Persoalan yang penting dalam melakukan analisis tingkat kesukaran soal adalah penentuan proporsi dan kriteria soal yang termasuk mudah, sedang, dan sukar. Selanjutnya syarat soal yang bermutu adalah soal harus valid, dan reliabel. Valid maksudnya bahwa setiap alat ukur hanya mengukur satu dimensi/ aspek saja. Reliabel maksudnya bahwa setiap alat ukur harus dapat memberikan hasil pengukuran yang tepat dan cermat.

Kualitas kelengkapan tes (soal) bisa dilihat dari hasil yang diperoleh oleh siswa. Kalau siswa mendapatkan skor kurang baik, berarti tes yang disusun sangat sukar, sebaliknya bila semua siswa mendapatkan skor baik, dapat dikatakan kalau soal tes mudah bagi siswa. Guru dapat menghindari hal tersebut dengan melaksanakan analisis, salah satunya melaksanakan analisis soal. Analisis soal merupakan sesuatu prosedur yang sistematis, yang hendak membagikan informasi-data yang sangat spesial terhadap butir tes (Arikunto, 2015).

Presentase ketuntasan nilai ujian akhir semester ganjil mata pembelajaran sosiologi kelas XII tahun pelajaran 2021/2022 SMA Negeri se-Kota Padang Panjang dapat dilihat pada Tabel 1.

Tabel 1. Presentase Ketuntasan Nilai UAS Mata Pelajaran Sosiologi Semester Ganjil Kelas XII SMA Negeri se-Kota Padang Panjang Tahun Pelajaran 2021/2022

No	Nama Sekolah	Jumlah Siswa	KKM	Presentase Ketuntasan UAS
1.	SMAN 1 Sumatera Barat	57	80	8,77%
2.	SMAN 1 Padang Panjang	59	80	84,74%
3.	SMAN 2 Padang Panjang	163	76	1,23%
4.	SMAN 3 Padang Panjang	154	80	25,97%

Sumber: Data nilai siswa milik guru Sosiologi XII SMAN se-Kota Padang Panjang

Data yang penulis kumpulkan adalah data hasil belajar kelas XII karena merupakan kelas yang krusial dan kelas puncak di SMA, sehingga bentuk soal yang diberikan kepada siswa tersebut seharusnya soal yang berkualitas. Dari tabel tersebut terlihat bahwa rata-rata ketuntasan Ujian Akhir Semester paling rendah ada di SMAN 2 Padang Panjang yaitu 1,23%, sedangkan KKM di sekolah ini sudah paling rendah diantara sekolah lainnya. Ketika banyak siswa yang mendapatkan nilai rendah atau di bawah KKM (belum tuntas), berarti terdapat masalah yang harus dianalisis oleh guru. Ada beberapa kemungkinan berkaitan dengan hal di atas yakni: 1) soal yang disusun sangat sulit, 2) soal yang disusun belum mengacu kepada substansi atau materi yang telah diajarkan, 3) pembelajaran yang dilakukan oleh guru belum bisa dipahami oleh siswa dengan baik (Kunandar, 2015).

Berdasarkan permasalahan yang telah diuraikan di atas maka dalam artikel penelitian ini membahas mengenai analisis butir soal ujian akhir sekolah pada mata pelajaran sosiologi SMA khususnya yang telah diujikan pada siswa yang memiliki nilai ketuntasan paling rendah di Kota Padang Panjang yaitu SMAN 2 Kota Padang Panjang, mulai dari validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan kualitas pengecoh sehingga dapat diketahui kualitas dan kelayakan dari butir-butir soal tersebut.

Metode Penelitian

Penelitian ini termasuk jenis penelitian deskriptif dengan menggunakan pendekatan kuantitatif. Menurut Lurfi dan Ardi, penelitian deskriptif adalah penelitian yang mendeskripsikan suatu gejala, fakta, peristiwa atau kejadian yang sedang atau sudah terjadi (Ardi & Lufri, 2014). Tujuan dari penelitian deskriptif ini adalah untuk mendeskripsikan hasil analisis terhadap soal ujian akhir semester yang digunakan sebagai instrumen pengukuran hasil belajar siswa pada mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang tahun pelajaran 2021/2022.

Sesuai dengan masalah yang hendak diteliti, maka dalam penelitian ini yang dijadikan populasi penelitian adalah seluruh siswa yang mengikuti ujian akhir semester mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang tahun pelajaran 2021/2022. Alasan peneliti memilih kelas XII karena kelas

tersebut adalah kelas yang krusial sehingga soal yang diberikan harus soal yang berkualitas. Kelas XII IPS di SMAN 2 Kota Padang Panjang terdiri dari 5 kelas, XII IPS 1 berjumlah 34 siswa, XII IPS 2 berjumlah 35 siswa, XII IPS 3 berjumlah 32 siswa, XII IPS 4 berjumlah 33 siswa, dan XII IPS 5 berjumlah 29 siswa. Sehingga total populasi pada penelitian ini adalah 163 orang siswa.

Teknik pengambilan sampel dalam penelitian ini adalah teknik total sampling, artinya semua siswa yang menjadi populasi dijadikan sampel. Sampel penelitian ini berjumlah 163 orang siswa yang masing-masing telah mengerjakan soal pilihan ganda (*Multiple Choice Questions*) pada ujian semester mata pelajaran sosiologi yang berjumlah 40 butir yang terdiri dari 5 option. Artinya penulis melakukan telaah terhadap lembar jawaban siswa dan menyesuaikannya dengan kisi-kisi soal. Berikut kisi-kisi ujian akhir semester mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang tahun pelajaran 2021/2022.

Tabel 2. Kisi-kisi Ujian Akhir Semester Mata Pelajaran Sosiologi Kelas XII SMA Kota Padang Panjang Tahun Pelajaran 2021/2022

No.	Materi	Level Kognitif	Indikator Soal	No Soal
1.	Perubahan Sosial	L2	Mengidentifikasi bentuk-bentuk perubahan sosial	1
		L3	Mendeskripsikan teori-teori perubahan sosial	2
		L2	Mengklasifikasikan faktor penyebab perubahan sosial	3
		L2	Memberikan contoh kasus dampak perubahan sosial	4
2.	Globalisasi dan perubahan komunitas lokal	L2	Menjelaskan pengertian globalisasi	5
		L2	Mendeskripsikan proses globalisasi	6
		L3	Mengklasifikasikan ciri-ciri globalisasi	7
		L2	Mendeskripsikan globalisasi di Indonesia	10,11,12,13
		L3	Mendeskripsikan dampak positif globalisasi	14,15,16,17
		L2	Mendeskripsikan dampak negatif globalisasi	18,19,20,21
		L2	Mendeskripsikan globalisasi antara tantangan dan kesempatan	22, 23
		L2	Mendeskripsikan sikap dalam menghadapi tantangan globalisasi	24, 25
3	Ketimpangan Sosial	L2	Menjelaskan teori tentang ketimpangan sosial	26
		L2	Menjelaskan faktor penyebab ketimpangan sosial	27, 28, 29, 30
		L2	Mendeskripsikan bentuk-bentuk ketimpangan sosial	31, 32, 33, 34
		L2	Mendeskripsikan akibat ketimpangan sosial	35, 36
		L2	Mendeskripsikan pemecahan dan solusi ketimpangan sosial di Indonesia	37, 38
		L2	Mendeskripsikan upaya mengatasi ketimpangan sosial akibat perubahan sosial	39, 40

Variabel dalam penelitian ini adalah kualitas soal ujian akhir semester ganjil mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang tahun pelajaran 2021/2022 meliputi validitas isi, validitas empiris, reliabilitas, tingkat kesukaran, daya beda, dan fungsi pengecoh.

Jenis data dalam penelitian ini adalah data primer, yaitu data yang didapatkan langsung dari kisi-kisi soal, soal, kunci jawaban, dan lembar jawaban siswa yang diperoleh dari guru mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang. Sumber data penelitian adalah kisi-kisi, soal, kunci, dan lembar jawaban ujian akhir semester ganjil mata pelajaran sosiologi kelas XII di SMA Negeri 2 Kota Padang Panjang tahun pelajaran 2021/2022, serta lembar validasi isi soal yang diisi oleh ahli. Metode pengumpulan data yang digunakan dalam penelitian ini adalah dokumen. Proses pengumpulan data dokumen dilakukan dengan mengumpulkan 163 lembar jawaban ujian akhir semester sosiologi siswa dan melakukan telaah dan analisis butir soal, serta data validasi ahli terkait isi atau konten soal.

Penelitian ini menggunakan teknik analisis data secara kuantitatif. Analisis kuantitatif mencakup pengukuran validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Selain itu peneliti juga melakukan telaah validitas isi soal melalui ahli menggunakan lembar validasi soal. Analisis isi soal ditinjau dari aspek analisis materi, konstruksi, bahasa. Analisis materi, konstruksi, dan bahasa dalam

penelitian ini bertujuan untuk mengetahui validitas isi pada ujian akhir semester ganjil mata pelajaran sosiologi kelas XII di SMA Negeri 2 Kota Padang Panjang. Terdapat empat langkah dalam penghitungannya, yaitu menelaah butir soal menggunakan teknik panel, menentukan spesifikasi domain skala empat-point, menentukan model kesepakatan interrater, dan menghitung validitas isi.

Untuk mendapatkan hasil analisis validitas isi dari dua orang ahli penulis menggunakan model kesepakatan interrater. Model kesepakatan interrater merupakan metode statistik yang digunakan untuk menetapkan validitas isi tes secara keseluruhan. Selanjutnya, ketika kedua penelaah selesai menganalisis terhadap soal-soal tes, maka skala empat-pointnya dapat dibagi menjadi dua jenis, yaitu relevansi lemah (untuk peringkat 1 dan 2) dan relevansi kuat (untuk peringkat 3 dan 4). Penelaahan butir soal menggunakan teknik panel yang merupakan salah satu teknik yang dapat digunakan untuk menganalisis butir soal secara kualitatif. Teknik panel dilakukan oleh validator untuk menelaah soal setelah diberikan format penelaahan, kisi-kisi, butir, dan soal. Hasil analisis teknik panel selanjutnya dispesifikasi domainnya menggunakan skala empat-point. Spesifikasi domain menggunakan skala empat-point yang terdiri dari: (1) tidak relevan; (2) agak relevan; (3) relevan; dan (4) sangat relevan. Hasil spesifikasi tersebut digunakan sebagai data untuk mengisi tabel kesepakatan interrater.

Tabel 3. Model Kesepakatan Interrater 2 Ahli

Penelaah 2 \ Penelaah 1	Relevansi Lemah	Relevansi Kuat
	Relevansi Lemah	A
Relevansi Kuat	C	D

Sumber: Model Kesepakatan Interrater (Gregory, 2010)

Keterangan Tabel 2. yaitu: (1) kolom A merupakan kesepakatan soal relevansi lemah antara kedua ahli; (2) kolom B merupakan soal relevansi kuat menurut penelaah 1, namun penelaah 2 menilai relevansi lemah; (3) kolom C merupakan soal relevansi kuat menurut penelaah 2, namun penelaah 1 menilai relevansi lemah; serta (4) kolom D merupakan kesepakatan soal relevansi kuat kedua ahli. Data kesepakatan interrater selanjutnya dimasukkan kedalam rumus validitas isi. Berikut rumus validitas isi menurut (Gregory, 2010).

$$\text{Validitas isi} = \frac{D}{A + B + C + D}$$

Hasil penghitungan validitas isi selanjutnya dikategorikan berdasarkan kriteria validitas isi. Berikut kriteria validitas isi yang disajikan pada tabel dibawah ini (Gregory, 2010).

Tabel 4. Interpretasi Validitas Isi

Kriteria Validitas Isi	Interpretasi
0,80-1,00	Tinggi
0,40-0,79	Sedang
0,00-0,39	Rendah

Sumber: Suharsimi Arikunto (Arikunto, 2015)

Validitas empiris adalah ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian yang tak terpisah dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut. Menurut Arikunto indeks korelasi point biserial (Ypbi) yang diperoleh dari hasil perhitungan dikonsultasikan dengan r tabel pada taraf signifikansi 5% sesuai jumlah siswa yang diteliti. Apabila y pbi > r tabel maka butir soal tersebut valid. (Arikunto, 2015). Rumus yang akan digunakan untuk menghitung validitas adalah (y pbi).

Interpretasi tingkat kesukaran telah lengkap dengan tafsirannya, sehingga peneliti hanya membaca data yang telah diolah oleh ANATES ver.4.0.9. Kategori tingkat kesukaran dapat dilihat pada dibawah ini.

Tabel 5. Interpretasi Tingkat Kesukaran

Batasan	Interpretasi
0% < TK ≤ 30%	Soal sukar
31% < TK ≤ 70%	Soal sedang
71% < TK ≤ 100%	Soal mudah

Sumber : Suharsimi Arikunto (Arikunto, 2015)

Interpretasi daya pembeda masih belum memiliki tafsiran, hanya ada nilai indeks saja. Menurut Arikunto, kategori daya pembeda dapat dilihat pada Tabel 6.

Tabel 6. Interpretasi Daya Pembeda

Batasan	Interpretasi
$0,00 < DP \leq 0,20$	Jelek
$0,21 < DP \leq 0,40$	Cukup
$0,41 < DP \leq 0,70$	Baik
$0,71 < DP \leq 1,00$	Baik sekali

Sumber: Suharsimi Arikunto (Arikunto, 2015)

Interpretasi kualitas pengecoh telah lengkap dengan tafsirannya, sehingga penulis dalam penelitian ini membaca data yang sudah diolah oleh ANATES ver.4.0.9 dan menginterpretasikannya. Pedoman interpretasi keberfungsiam pengecoh dapat dilihat pada berikut.

Tabel 7. Interpretasi Fungsi Pengecoh

Batasan	Interpretasi
++	Sangat Baik
+	Baik
-	Kurang Baik
--	Buruk
---	Sangat Buruk

Sumber: Suharsimi Arikunto (Arikunto, 2015)

Selanjutnya pedoman interpretasi reliabilitas dapat dilihat pada Tabel 8. penafsiran nilai reliabilitas

Tabel 8. Interpretasi Nilai Reliabilitas

Koefisien Reliabilitas	Tingkat Reliabilitas
0,800-1,000	Sangat tinggi
0,600-0,799	Tinggi
0,400-0,599	Cukup
0,200-0,399	Rendah
0,000-0,200	Sangat Rendah

Sumber : Suharsimi Arikunto (Arikunto, 2015)

Analisis kuantitatif adalah analisis pada butir soal yang telah diuji cobakan, analisis dilakukan dengan pendekatan Teori tes klasik (*Classical Test Theory*), disebut juga sebagai "*true score theory*". Teori ini dimulai dari asumsi bahwa efek sistematis antara tanggapan peserta ujian hanya disebabkan oleh variasi dalam kemampuan minat. Semua sumber variasi potensial lainnya yang ada dalam bahan pengujian seperti kondisi eksternal atau kondisi internal peserta ujian diasumsikan konstan melalui standarisasi yang ketat atau memiliki efek yang nonsistematis atau acak secara alami (Vincent & Shanmugam, 2020).

Teori klasik mengasumsikan bahwa setiap individu memiliki skor sebenarnya yang akan diperoleh jika tidak ada kesalahan dalam pengukuran. Namun, karena alat ukur kadangkala tidak sempurna, skor yang diamati untuk setiap orang mungkin berbeda dari kemampuan sejati individu (Magno, 2009). Teori tes klasik menggunakan standar deviasi kesalahan sebagai ukuran dasar kesalahan. Biasanya ini disebut kesalahan standar pengukuran. Dalam praktiknya, standar deviasi dari validitas dan reliabilitas tes digunakan untuk memperkirakan kesalahan standar pengukuran. Semakin besar kesalahan standar pengukuran, semakin tidak pasti keakuratan atribut yang diukur. Sebaliknya, kesalahan standar pengukuran yang kecil menunjukkan bahwa skor individu mungkin mendekati skor sebenarnya (Magno, 2009).

Analisis kuantitatif menurut pendekatan teori tes klasik menghasilkan karakteristik butir yang meliputi tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Selain itu, dengan analisis kuantitatif pendekatan teori klasik juga dapat diketahui reliabilitas soal tes, dan validitas soal. Untuk melihat tingkat kesukaran, daya pembeda, dan efektivitas pengecoh dilakukan analisis setiap butir tes, sedangkan reliabilitas dan validitas empiris dapat dilihat dengan cara menganalisis soal tes secara keseluruhan (Kartowagiran, 2012).

Hasil dan Pembahasan

Validitas Soal

Validitas isi

Analisis materi, konstruksi, dan bahasa bertujuan untuk mencari validitas isi pada soal ujian akhir semester ganjil mata pelajaran sosiologi kelas XII SMA Negeri 2 Padang Panjang tahun 2022. Teknik ini dilakukan oleh validator yang telah diberi satu set kisi-kisi soal, soal beserta kunci jawabannya, format analisis, pedoman analisis, dan selanjutnya para penelaah menganalisis secara tersendiri (Departemen Pendidikan Nasional, 2008).

Hasil analisis dari kedua validator tersebut diinterpretasikan menggunakan skala empat-point. Hasil skala empat-point selanjutnya dikategorikan menjadi dua kelompok yaitu relevansi lemah dan relevansi kuat. Skala empat point yang sudah dikategorikan, digunakan sebagai data untuk mengisi model kesepakatan *interrater*.

Data hasil kesepakatan *interrater* dimasukkan kedalam rumus validitas isi menurut Gregory (2010). Berikut perhitungan indeks validitas isi ditinjau dari aspek materi, konstruksi, dan bahasa:

Validitas isi aspek materi:

$$\text{Validitas isi} = \frac{27}{5+9+3+23} = \frac{27}{40} = 0,57 \text{ (Kategori Sedang)}$$

Validitas isi aspek konstruksi:

$$\text{Validitas isi} = \frac{29}{0+12+1+27} = \frac{29}{40} = 0,72 \text{ (Kategori Sedang)}$$

Validitas isi aspek bahasa:

$$\text{Validitas isi} = \frac{29}{0+10+1+29} = \frac{29}{40} = 0,67 \text{ (Kategori Sedang)}$$

Hasil perhitungan validitas isi kemudian di kelompokkan berdasarkan kriteria validitas isi. Berdasarkan kriteria tersebut, kategori tinggi memiliki rentang 0,40-0,79 sehingga validitas isi pada soal ujian akhir semester ganjil mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang tahun 2022 pada aspek materi berkategori sedang, pada aspek konstruksi berkategori sedang, dan pada aspek bahasa berkategori sedang.

Validitas Empiris

Menurut Arikunto Indeks korelasi *point biserial* (Y_{pbi}) yang diperoleh dari hasil perhitungan dikonsultasikan dengan r tabel pada taraf signifikansi 5% sesuai jumlah peserta didik yang diteliti. Apabila $y_{pbi} > r_{tabel}$ maka butir soal tersebut valid. (Arikunto, 2015). Rumus yang akan digunakan untuk menghitung validitas adalah y_{pbi} .

Hasil analisis validitas empiris pada butir soal adalah 7 butir (17,5%) yang valid. Butir soal yang tidak valid adalah 33 butir, berarti dari keseluruhan soal yang dianalisis 82,5% soal tidak valid. Kemudian apabila 40 soal tersebut didistribusikan berdasarkan indeks validitasnya maka akan didapat hasil pada Tabel 9.

Tabel 9. Distribusi Validitas Empiris Ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMAN 2 Padang Panjang

Indeks Validitas	No Butir Soal	Jumlah	Presentase
Jika nilai indeks diatas atau sama 0,304 maka soal dinyatakan valid	18, 19, 26, 27, 29, 31, 35	7	17,5%
Jika nilai indeks kurang dari 0,304 maka soal dinyatakan tidak valid	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 28, 30, 32, 33, 34, 36, 37, 38, 39, 40	33	82,5%

Sumber: Diolah dari Data Primer (2022)

Suatu teknik evaluasi dapat dikatakan mempunyai validitas yang tinggi (valid) jika teknik evaluasi tersebut dapat sepenuhnya mengukur kemampuan tertentu yang diharapkan, sehingga dapat dikatakan bahwa alat tersebut mampu dan mengukur kemampuan siswa dalam mencapai kompetensi yang diharapkan pada materi yang diajarkan. Skor pada butir soal dapat menyebabkan skor total menjadi tinggi atau rendah, karena itu sebuah item yang memiliki validitas tinggi jika skor pada item mempunyai kesejajaran dengan skor total (Rahmani, 2015).

Nilai validitas pada soal ujian akhir semester hanya mencapai 17,5% dari keseluruhan, hal ini menandakan bahwa masih banyak soal yang belum mampu mengukur ketercapaian belajar siswa, sehingga dapat ditarik kesimpulan bahwa soal ini masih belum memenuhi kriteria validitas. Menurut Scarvia B. Anderson (dalam Arikunto) mengungkapkan bahwa sebuah tes bisa saja reliabel tetapi tidak valid, namun sebuah tes yang valid biasanya reliabel. Soal yang baik jika soal tersebut tergolong valid sehingga mampu mengukur kemampuan siswa. Soal yang valid adalah soal yang telah memiliki korelasi positif dengan r hitung lebih besar dari r tabel (Arikunto, 2015).

Berdasarkan uraian di atas dapat disimpulkan bahwa soal-soal ujian akhir semester ganjil mata pelajaran sosiologi kelas XII SMAN 2 Padang Panjang belum memiliki kualitas yang baik dari segi validitas empirisnya. Untuk selanjutnya guru perlu melakukan analisis validitas soal pada sampel tertentu sebelum melaksanakan ujian. Butir soal yang valid dapat disimpan di bank soal. Sebaliknya, butir soal yang tidak valid perlu diperbaiki dengan menyesuaikan indikator dan meningkatkan penguasaan teknik tentang penyusunan butir soal.

Tingkat Kesukaran

Perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah (Arifin, 2013). Soal yang terlalu mudah tidak mampu membuat siswa berkeinginan untuk memecahkannya, sebaliknya soal yang terlalu sulit akan menyebabkan siswa menjadi putus asa dan menyerah serta menurunkan tingkat semangat siswa karena diluar kemampuannya (Daryanto, 2014).

Hasil analisis tingkat kesukaran soal ujian akhir semester ganjil mata pelajaran sosiologi XII SMA Negeri 2 Kota Padang Panjang diperoleh data soal yang berkategori mudah berjumlah 16 butir soal, kategori sedang ada 12 butir soal, dan kategori sukar ada 12 butir soal. Distribusi tingkat kesukaran soal dapat dilihat pada Tabel 10.

Tabel 10. Distribusi Tingkat Kesukaran Ujian Akhir Semester ganjil Mata pelajaran Sosiologi XII SMA Negeri 2 Kota Padang Panjang

Kategori	Nomor Soal	Presentase
Mudah	2, 3, 4, 5, 6, 8, 10, 11, 15, 18, 21, 27, 30, 32, 34, 38	40%
Sedang	1, 7, 14, 19, 20, 21, 23, 26, 28, 29, 31, 33	30%
Sukar	9, 12, 13, 16, 17, 24, 25, 35, 36, 37, 39, 40	30%

Sumber: Diolah dari Data Primer (2022)

Berdasarkan hasil analisis soal, terdapat 16 (40%) soal ujian akhir semester genap ini tergolong mudah, 12 soal (30%) sedang, dan 12 soal (30%) sukar. Dengan demikian soal ini belum tergolong baik karena belum memenuhi standar tingkat kesukaran butir soal yaitu sukar 25%, sedang 50%, dan mudah 25% (Rahmawati, 2014).

Dengan demikian soal sebelum diujikan perlu dianalisis dan diperbaiki. Menurut A. Muri Yusuf bahwa tindak lanjut yang perlu dilakukan dalam memperbaiki tingkat kesukaran soal yaitu: a) Butir soal dengan tingkat kesukarannya baik (dalam artian cukup atau sedang) butir soal bisa dimasukkan dalam bank soal dan dikeluarkan lagi untuk tes selanjutnya; b) Butir soal dengan tingkat kesukaran sukar, ada tiga kemungkinan tindak lanjut yang dilakukan yaitu butir soal dibuang, diteliti ulang atau direvisi kembali, dan butir soal dapat digunakan untuk tes yang sangat ketat dalam artian sebagian besar peserta tidak akan lulus dalam tes tersebut; c) Butir soal dengan tingkat kesukaran mudah, juga ada 3 kemungkinan tindak lanjut yang dapat dilakukan yaitu butir soal dibuang, diteliti ulang atau direvisi kembali, dan butir soal dapat digunakan untuk tes yang sangat longgar dalam artian sebagian besar peserta lulus dalam tes tersebut, sehingga tes tersebut boleh dikatakan hanya sebagai formalitas saja (Yusuf, 2015).

Daya Pembeda

Sebuah soal tes dikatakan berkualitas tinggi jika mampu membedakan antara siswa yang pandai (berkompetensi tinggi) dengan siswa yang kurang pandai (kompetensi rendah) (Haryanto, 2015). Sama halnya menurut Arikunto (Arikunto, 2015), pembeda soal adalah daya soal dalam menyeleksi tingkat kemampuan siswa (pandai dan kurang pandai). Analisis daya pembeda dilakukan guna memahami kemahiran soal dalam membedakan tingkat kemampuan siswa yang tergolong tinggi dan rendah dari segi akademik. Analisis daya pembeda diperoleh dari perhitungan indeks diskriminasi soal tes.

Terdapat lima klasifikasi daya beda soal yaitu kategori sangat jelek, jelek, cukup, baik, dan baik sekali. Menurut Prasetyo semakin tinggi nilai daya beda soal (bernilai positif) maka semakin baik soal tersebut,

meskipun memiliki nilai positif, soal yang sebaiknya digunakan adalah soal yang memiliki daya beda cukup, baik, dan baik sekali (Prasetyo, 2012).

Hasil analisis daya pembeda, diperoleh 23 butir soal yang berkategori jelek, 14 butir soal kategori cukup, 3 butir soal baik, dan tidak ada terdapat soal yang berkategori baik sekali. Berdasarkan analisis daya pembeda berarti ada 57,5% soal yang termasuk kategori jelek, 35% soal yang termasuk kategori cukup, dan 7,5% soal termasuk kategori baik. Distribusi daya pembeda dapat dilihat pada Tabel 11.

Tabel 11. Distribusi Daya Pembeda ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMA Negeri 2 Kota Padang Panjang

Kategori	Nomor Soal	Presentase
Jelek	1, 2, 3, 5, 6, 10, 11, 12, 13, 15, 16, 17, 21, 22, 23, 24, 25, 30, 33, 34, 37, 39, 40	57,5%
Sedang	4, 7, 8, 9, 14, 18, 20, 27, 28, 29, 32, 35, 36, 38	35%
Baik	19, 26, 31	7,5%
Baik Sekali	-	-

Sumber: Diolah dari Data Primer (2022)

Dari analisis daya pembeda, soal yang memiliki daya pembeda jelek artinya soal tersebut banyak dijawab benar oleh siswa yang berkemampuan rendah dan banyak dijawab salah oleh siswa yang berkemampuan tinggi (Friatma et al., 2017). Faktor daya pembeda soal jelek salah satu faktornya disebabkan karena menjawab soal dengan cara menebak saja. Hal ini juga sesuai dengan pendapat Daryanto bahwa butir soal jelek karena lebih banyak dijawab benar oleh kelompok bawah dibandingkan dengan jawaban benar dari kelompok atas (Daryanto, 2014). Ini berarti bahwa untuk menjawab soal dengan benar, dapat dilakukan dengan menebak. Soal seperti ini seharusnya tidak digunakan kembali untuk ujian selanjutnya. Apabila akan digunakan, maka soal perlu direvisi. Ini dilakukan agar soal dapat berfungsi sesuai dengan tujuan yang diinginkan.

Kualitas Pengecoh

Pada soal bentuk pilihan ganda memiliki alternatif jawaban (opsi) yang merupakan pengecoh. Butir soal yang baik, pengecohnya akan dipilih secara merata oleh siswa yang menjawab salah. Sebaliknya, butir soal yang kurang baik, pengecohnya akan dipilih secara tidak merata. Suatu pilihan jawaban (pengecoh) dapat dikatakan berfungsi apabila pengecoh; 1) paling tidak dipilih oleh 5% peserta tes/siswa, 2) lebih banyak dipilih oleh kelompok siswa yang belum paham materi. Kategori fungsi pengecoh yang baik adalah adanya simbol + dan ++ yang menandakan bahwa fungsi pengecoh berfungsi (Depdiknas, 2010). Distribusi hasil analisis kualitas pengecoh dapat dilihat pada Tabel 12.

Tabel 12. Distribusi Kualitas Pengecoh Ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMA Negeri 2 Kota Padang Panjang

Kategori	Nomor Soal
Berkualitas	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 16, 18, 19, 20, 22, 24, 27, 28, 29, 30, 31, 32, 33, 35, 37, 38, 39, 40
Tidak Berkualitas	4, 12, 14, 15, 17, 21, 23, 25, 26, 35, 36

Sumber: Diolah dari Data Primer (2022)

Berdasarkan hasil analisis, kualitas opsi pengecoh yang tidak berkualitas adalah 27,5% (11 soal) dan yang berkualitas adalah 72,5% (29 soal). Persentase tersebut diperoleh dengan membandingkan opsi pengecoh yang masuk kriteria berkualitas/tidak berkualitas dengan keseluruhan opsi pengecoh. Berdasarkan hasil analisis tersebut dapat disimpulkan bahwa kualitas opsi pengecoh pada soal ini sangat bagus, karena lebih banyak soal yang opsi pengecohnya berfungsi.

Soal yang baik adalah soal yang sudah memenuhi semua kriteria yang ada, berdasarkan hasil analisis yang dilakukan soal-soal yang telah diujikan tersebut ternyata masih ada yang belum memenuhi dua kriteria yaitu kriteria: validitas, dan fungsi pengecoh. Maka dapat disimpulkan bahwa soal ujian akhir semester genap mata pelajaran sosiologi kelas XII tahun pelajaran 2021/2022 di SMAN 2 Padang Panjang belum memenuhi kriteria baik.

Reliabilitas

Reliabilitas adalah tingkat atau derajat konsistensi dari suatu instrumen. Reliabilitas berkaitan dengan pertanyaan, apakah suatu tes teliti dan dapat dipercaya sesuai dengan kriteria yang telah ditetapkan. Suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda (Amalia & Widayati, 2012), yang dipentingkan di sini adalah

ketelitiannya sejauh mana tes atau alat tersebut dapat dipercaya kebenarannya. Rentangan reliabilitas yang baik berkisar antara 0,400-1,000 (Purwanto, 2012).

Berdasarkan hasil analisis reliabilitas, nilai reliabilitas yang diperoleh adalah sebesar 0,19. Nilai reliabilitas pada rentangan 0,00-0,20 artinya memiliki reliabilitas yang sangat rendah (Depdiknas, 2010). Dapat diambil kesimpulan tes ini belum memenuhi persyaratan reliabilitas, sehingga tes ini belum layak diujicobakan di lain waktu ataupun tempat pada materi yang sama (Reski & Sylvia, 2021). Rendahnya koefisien reliabilitas soal karena terbatasnya jumlah butir soal yang dibuat oleh guru, sehingga guru dapat menambahkan jumlah soal yang valid. Suatu instrumen tes yang memiliki validitas yang baik pada setiap butirnya juga akan memiliki tingkat reliabilitas yang tinggi juga, hal ini selaras dengan yang disampaikan oleh Suharsimi Arikunto (Arikunto, 2015).

Kualitas Butir Soal

Hasil analisis data dari keseluruhan ditinjau dari aspek validasi, reliabilitas, daya pembeda, tingkat kesukaran, dan fungsi pengecoh dapat dilihat pada Tabel 13.

Tabel 13. Analisis Data Dilihat Dari Semua Aspek

No	Aspek yang dianalisis	Keterangan
1	Validitas Empiris	7 Soal (Valid)
2	Daya Pembeda	17 (Baik)
3	Tingkat Kesukaran	12 (Sedang)
4	Kualitas Pengecoh	72,5% (berkualitas)
5	Reliabilitas	0,19 (Sangat Rendah)

Sumber: Diolah dari Data Primer (2022)

Tabel di atas memperlihatkan hasil analisis terhadap 40 soal yang diujikan pada ujian akhir semester pada pembelajaran sosiologi di SMAN 2 Kota Padang Panjang. Hasil penelaah ahli terhadap validitas isi yang diujikan pada soal yaitu aspek materi, konstruksi, dan pada aspek bahasa berkategori sedang dengan kesepakatan interrater dua orang ahli berada pada rentang 0,52–0,72. Sedangkan hasil analisis validitas empiris memberikan hasil pengukuran hanya 7 soal yang valid dari 40 buah soal. Artinya soal yang diujikan ini belum dapat mengukur apa yang seharusnya diukur.

Selanjutnya pada tingkat kesukaran soal, hanya 12 soal (30%) dari 40 soal yang memiliki tingkat kesukaran sedang, sedangkan 16 (40%) soal ujian akhir semester ini berada pada kategori mudah, dan 12 soal (30 %) sukar. Dengan demikian soal ini belum tergolong baik karena belum memenuhi standar tingkat kesukaran butir soal yaitu sukar 25%, sedang 50%, dan mudah 25%. Berdasarkan data diatas, maka soal dengan tingkat kesukaran mudah dan sukar perlu diteliti dan direvisi kembali secara terperinci ataupun dibuang jika soal dinyatakan tidak baik. Sedangkan untuk butir soal dengan indeks tingkat kesukaran sedang bisa disimpan di bank soal dan bisa dijadikan soal tes selanjutnya (Arifin, 2013). Hal ini sependapat dengan Arikunto yang menyatakan soal yang baik merupakan soal yang tidak sangat mudah dan tidak sangat sukar karena soal yang sangat mudah tidak bisa membangkitkan kemampuan peserta didik untuk memecahkan soal yang ada. Sedangkan soal yang sangat sukar mematahkan semangat peserta didik untuk mencoba kembali karena karena menganggap dirinya tidak mampu (Arikunto, 2015).

Analisis daya pembeda dilakukan guna memahami kemahiran soal dalam membedakan tingkat kemampuan peserta didik yang tergolong tinggi dan rendah dari segi akademik. Hasil analisis daya pembeda, diperoleh 23 butir soal yang berkategori jelek, artinya ada 23 soal yang tidak bisa membedakan peserta didik yang pandai dan kurang pandai, 14 butir soal kategori cukup, artinya 14 butir soal tersebut hampir bisa membedakan kemampuan peserta didik yang pandai dan kurang pandai, 3 butir soal baik, artinya hanya 3 buah soal yang sedikit bisa membedakan peserta didik yang pandai dan kurang pandai dan tidak ada terdapat soal yang berkategori baik sekali, artinya tidak ada soal yang mampu benar-benar membedakan siswa menjadi dua golongan yaitu siswa yang memiliki kemampuan tinggi dan rendah. Artinya hanya 3 soal atau 7,5% yang hampir dapat membedakan klasifikasi kemampuan siswa.

Kualitas pengecoh pada opsi jawaban soal tes adalah kemampuan soal dalam mengecoh peserta didik menjawab soal tes yang diberikan. Berdasarkan pendapat Arikunto, bahwa pengecoh dinyatakan baik jika pengecoh tersebut dipilih sekurang-kurangnya 5% dari sejumlah peserta tes (Arikunto, 2015). Berdasarkan hasil analisis, kualitas opsi pengecoh yang tidak berkualitas adalah 27,5% (11 soal) dan yang berkualitas adalah 72,5% (29 soal). Membuat pengecoh yang baik pada tes pilihan ganda sulit, karena pengecoh yang kurang baik akan mengakibatkan rendahnya daya pembeda, dan apabila satu atau dua pengecoh tidak berfungsi baik maka mengakibatkan tingkat kesukaran juga rendah. Hal inilah yang terjadi pada 40 soal yang diujikan pada ujian akhir semester pada pembelajaran sosiologi di SMAN 2 Kota Padang Panjang. Guru fokus pada merancang option soal dengan baik yang memperlihatkan bukti 72,5% atau 29 soal memiliki

kualitas yang baik pada option pengecoh, namun gagal dalam membuat soal yang memiliki kemampuan mengukur hal yang harus diukur (indikator ketercapaian kompetensi), tingkat kesukaran soal dan daya beda yang baik, sehingga reliabilitas soal tersebut rendah (Mardiani & Sylvia, 2019). Artinya soal tersebut belum mampu mengukur secara konsisten tidak berubah jika dilaksanakan pengukuran kembali terhadap subjek yang sama ataupun subjek yang berbeda (Mardapi, 2012).

Sebuah tes yang baik, akan bisa mengungkapkan keadaan sebenarnya dari siswa, dan tes yang tidak baik tidak akan bisa mengungkap apa kemampuan sebenarnya siswa. Sebuah tes yang baik harus valid dan reliabel. Dalam pandangan Samuel Messick dalam Linn, validitas merupakan penilaian menyeluruh dimana bukti empiris dan logika teori mendukung pengambilan keputusan serta tindakan berdasarkan skor tes atau model-model penilaian yang lain. Validitas sebuah tes dapat dilakukan dalam berbagai bentuk seperti *content validity*, *criterion validity* dan *construct-related validity* (Linn & Miller, 1993). Meskipun idealnya validasi dapat dilakukan dengan memakai semua bentuk validitas tes tersebut, tetapi pengembang tes dapat memilih bentuk validasi dengan melihat tujuan pengembangan tes. Selain valid, alat ukur yang baik juga harus reliabel. Dalam pandangan Aiken sebuah tes dikatakan reliabel jika skor yang diperoleh oleh peserta relatif sama meskipun dilakukan pengukuran berulang-ulang. Untuk memperoleh skor yang sama, maka tidak boleh ada kesalahan pengukuran. Dengan demikian, keandalan sebuah alat ukur dapat dilihat dari dua petunjuk yaitu kesalahan baku pengukuran dan koefisien reliabilitas (Linn & Miller, 1993).

Memahami kualitas butir soal bisa dilihat dari hasil analisis validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh butir soal. Kualitas butir soal terbagi menjadi 5 kategori yaitu sangat baik, baik, sedang, tidak baik, dan sangat tidak baik menurut Skala Likert. Butir soal dengan kualitas sangat baik harus berada di bank soal dengan menjaga kerahasiaan soal tersebut sehingga bisa dimanfaatkan kembali pada saat ujian mendatang. Butir soal dengan kualitas baik dan sedang belum bisa bergabung dengan bank soal, karena harus direvisi kembali sesuai dengan indikatornya. Serta butir soal dengan kualitas tidak baik dan sangat tidak baik lebih bagus dihapus karena memerlukan revisi yang lebih terperinci. Data analisis 40 soal ujian akhir semester ini 7 soal dapat dimasukkan ke dalam bank soal, sebanyak 9 butir soal membutuhkan revisi dan 22 butir soal dihapuskan. Hasil analisis ini dapat dilihat penjabarannya dalam tabel 13 di bawah ini.

Tabel 14. Hasil Rekapitulasi Analisis Soal Yang Dapat Dipakai Dalam Ujian Akhir Semester Ganjil Mata Pelajaran Sosiologi Kelas XII SMA Negeri 2 Kota Padang Panjang

Kriteria Kualitas Soal	Nomor Soal	Presentase
Sangat Baik	19, 27, 29, 31	10%
Baik	7, 18, 20, 26, 28	12,5%
Sedang	1, 8, 9, 14, 22, 32, 33, 35, 38	22,5%
Tidak Baik	2, 3, 4, 5, 6, 10, 11, 13, 16, 23, 24, 30, 34, 36, 37, 39, 40	42,5%
Sangat Tidak Baik	12, 15, 17, 21, 25	12,5%

Sumber: Diolah dari Data Primer (2022)

Penyebab kegagalan butir soal bisa dilihat kembali berdasarkan aspek validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh tiap butir soal. Penyebab kegagalan 33 butir soal ujian akhir semester mata pelajaran sosiologi kelas XII di SMAN 2 Kota Padang Panjang adalah : *Pertama*, soal ini tidak mampu menguji materi yang sudah dibelajarkan sesuai dengan indikator pencapaian kompetensi (tujuan yang dicapai), validitas butir soal belum mampu mendukung skor total tes. *Kedua*, soal ini memiliki kualitas sangat mudah. Tingkat kesukaran, soal pilihan ganda yang tergolong mudah belum bisa menaksir kemampuan peserta test. *Ketiga*, soal yang dipakai belum bisa menyeleksi antara kelompok atas dengan bawah, artinya soal ini tidak mampu mengkategorikan siswa menjadi dua kelompok pintar dan kurang. Hal inilah yang memberikan sumbangan terhadap hasil belajar siswa, yang menyebabkan tingkat ketuntasan siswa dalam ujian akhir semester genap tahun 2021/2022 hanya 1,23%.

Menurut teori tes klasik suatu alat ukur berupa tes dikatakan baik apabila memiliki banyak butir yang termasuk dalam kategori baik, dan harga indeks reliabilitas tinggi. Teori tes klasik menekankan pada skor mentah dari satu ujian yang dihasilkan. Skor mentah menunjukkan kemampuan seseorang. Dari skor mentah ini maka berbagai analisis dan interpretasi bisa dihasilkan sesuai dengan keperluan studi yang dilakukan (Sumintono, B., & Widhiarso, 2015). Teori tes klasik memiliki keterbatasan yang mendasar yaitu hasil estimasi parameter tergantung pada karakteristik peserta ujian (*group dependent*). Hal ini berimplikasi pada tingkat kesukaran soal akan menjadi rendah jika tes diujikan pada kelompok peserta tes berkemampuan tinggi dan sebaliknya jika tes diujikan pada peserta dengan kemampuan rendah maka tingkat kesukaran tes itu akan tinggi. Kedua hasil estimasi kemampuan peserta tergantung pada karakteristik butir soal (*item dependent*). Keterbatasan ini menyebabkan estimasi kemampuan peserta akan rendah jika soal yang diberikan

berada di atas kemampuannya. Sebaliknya, estimasi kemampuan peserta akan tinggi bila soal yang diujikan berada dibawah tingkat kemampuannya (Saifuddin, 2002).

Kesimpulan

Hasil analisis soal dilihat dari beberapa aspek terkait dengan kualitas soal sebagai berikut, dilihat dari asepek: 1) berdasarkan validita isi, soal tersebut memiliki aspek materi dengan kategori sedang, aspek konstruksi dengan kategori sedang, dan aspek bahasa dengan kategori sedang. 2) berdasarkan validitas empiris, terdapat 7 (17,5%) soal yang valid, dan 33 (82,5%) soal yang tidak valid. Hasil tersebut menunjukkan bahwa validitas empiris yang dimiliki soal belum baik. 3) berdasarkan reliabilitas, soal tersebut memiliki reliabilitas rendah yaitu 0,19. Hasil tersebut menunjukkan bawah reliabilitas butir pada soal dikatakan sangat rendah atau tidak reliabel. 4) berdasarkan daya pembeda, 23 (57,5%) soal yang tergolong jelek, 14 (35%) soal tergolong sedang, dan 3 (7,5%) soal tergolong baik. 5) berdasarkan tingkat kesukaran, terdapat 16 (40%) soal ujian akhir semester genap ini tergolong mudah, 12 soal (30%) sedang, dan 12 soal (30 %) sukar. 6) berdasarkan kualitas pengecoh, terdapat 27,5% opsi pengecoh yang tidak berkualitas dan 72,5% yang berkualitas. Dengan demikian dapat ditarik kesimpulan bahwa soal ujian akhir semester ganjil mata pelajaran sosiologi kelas XII SMA Negeri 2 Kota Padang Panjang tahun pelajaran 2021/2022 belum memenuhi kriteria baik.

Daftar Rujukan

- Amalia, A. N., & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas Xii Sma Mata Pelajaran Ekonomi Akuntansi Di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1). <https://doi.org/10.21831/jpai.v10i1.919>
- Ardi, A. & Lufri, L. (2014). *Metodologi Penelitian: Penelitian Kuantitatif, PTK dan Penelitian Pengembangan*. Padang: UNP Press.
- Arifin, Z. (2013). *Evaluasi Pembelajaran*. Jakarta: Remaja Rosdakarya.
- Arikunto, S. (2015). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Daryanto, D. (2014). *Evaluasi Pendidikan*. Jakarta: Rineka Cipta.
- Departemen Pendidikan Nasional. (2008). Peraturan Pemerintah Republik Indonesia Nomor 74 Tahun 2008. <https://doi.org/peraturan.bpk.go.id/Home/Details/4892/pp-no-74-tahun-2008>
- Depdiknas. (2010). *Petunjuk Teknis Analisis Butir Soal di SMA*. Jakarta: Direktorat Pembina SMA.
- Friatma, A., Syamsurizal, S. & Helendra, H. (2017). Analisis Kualitas Soal Ujian Akhir Semester Genap Mata Pelajaran Biologi Kelas XI IPA SMA Negeri Wilayah Selatan Kabupaten Solok Tahun Pelajaran 2015/2016. *Bioeducation Journal*, 1(2), 50–67. <https://doi.org/10.24036/bioedu.v1i2.43>
- Gregory, J. R. (2010). *Tes Psikologi Sejarah, Prinsip, dan Aplikasi Edisi Keenam Jilid 1*. Jakarta: Erlangga.
- Haryanto, I. B. (2015). *Asesmen Pembelajaran*. Jakarta: Remaja Rosdakarya.
- Kartowagiran, B. (2012). Penulisan Butir Soal. <http://staff.uny.ac.id/sites/default/files/penelitian/prof-dr-badrun-kartowagiran-mpd/penulisan-dan-analisis-butir-soal.pdf>
- Kemendikbud. (2015). Panduan Penilaian Untuk Sekolah Menengah Pertama (SMP). 32–35. <https://doi.org/repositori.kemdikbud.go.id/18051/1/1>.
- Kunandar, K. (2015). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013)*. Edisi Revisi. Jakarta: Raja Wali Press.
- Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11. <https://doi.org/psycnet.apa.org/record/2010-11917-001>
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis Butir Soal Ujian Akhir Sekolah. *Al Asma : Journal of Islamic Education*, 2(2), 274. <https://doi.org/10.24252/asma.v2i2.16569>
- Mardapi, D. (2012). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Jakarta: Mitra Cendikia.
- Mardiani, M., & Sylvia, I. (2019). Pengembangan Tes Diagnostik Pilihan Objektif Tiga Tingkat Guna Mendeteksi Miskonsepsi Ciri-Ciri Sosiologi. *Jurnal Sikola: Jurnal Kajian Pendidikan dan Pembelajaran*, 1(1), 69–79. <https://doi.org/10.24036/sikola.v1i1.1>
- Prasetyo, AN., S. & B. (2012). Analisis Butir Soal Ulangan Akhir Gasal Biologi SMA Kelas X Tahun Pelajaran 2012/2013 di KabupatenPasuruan dengan Menggunakan Program ITEMAN dan Kesesuaiannya dengan Tuntutan KTSP. Universitas Negeri Malang.
- Pratiwi, D., Zakso, A., & Yusuf, M. (2013). Analisis Kualitas Soal Semester Ganjil Pelajaran Sosiologi Kelas XI SMA. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 2(9), 1-10.
- Purwanto, N. (2012). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Jakarta: Remaja Rosdakarya Offset.

-
- Rahmani, M. (2015). Analisis Kualitas Butir Soal Buatan Guru Biologi Kelas X SMA Negeri 1 Tanah Pinoh. *JPPK*, 2(1), 3–16. <https://doi.org/10.26418/jppk.v4i2.8970>
- Rahmawati, S. (2014). *Penilaian Dalam Kurikulum 2013*. Yogyakarta: Andi.
- Reski, P., & Sylvia, I. (2021). Rancangan Soal Open-Ended untuk Meningkatkan Kemampuan Penalaran Peserta Didik dalam Pembelajaran Sosiologi di SMA N 1 Sungai Aur Kabupaten Pasaman Barat. *Jurnal Sikola: Jurnal Kajian Pendidikan dan Pembelajaran*, 2(3), 147–161.
- Linn, R. & Miller, N. E. G. (1993). *Educational Measurement (Third Edit)*. American National Council on Measurement in Education.
- Saifuddin, A. (2002). *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Andi Publisher.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Model Rasch untuk Penelitian I Lmu-Ilmu Sosial*. Jakarta: Trim Komunikata Publishing House.
- Vincent, W., & Shanmugam, S. K. S. (2020). The Role of Classical Test Theory to Determine the Quality of Classroom Teaching Test Items. *Pedagogia: Jurnal Pendidikan*, 9(1), 5–34.
- Yusuf, A. M. (2015). *Asesmen dan Evaluasi Pendidikan*. Jakarta: Prenada Media Grup.