

**ANALISIS KUALITAS TES DAN BUTIR SOAL**  
**(Aspek Pengetahuan, Aspek Sikap, Aspek Keterampilan)**  
Tugas Mata Kuliah Asesmen Pendidikan IPS

**DOSEN PENGAMPU:**  
**Dr. Pujiati, S.Pd., M.Pd.**  
**Dr. Sugeng Widodo, S.Pd., M.Pd.**



Oleh:  
**Resti Apriliyani**  
**NPM 2523031007**

**PROGRAM PASCASARJANA**  
**PENDIDIKAN ILMU PENGETAHUAN SOSIAL**  
**FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN**  
**UNIVERSITAS LAMPUNG**  
**2026**

## ANALISIS KUALITAS TES DAN BUTIR SOAL

Mata Pelajaran IPS | Kelas VII Semester Genap | SMPN 44 Bandar Lampung | TA 2025/2026

Jumlah Soal: 20 Butir Pilihan Ganda | Jumlah Responden: 32 Siswa | KKM: 75

### A. Pendahuluan

Analisis kualitas tes merupakan suatu prosedur sistematis yang dilakukan untuk mengevaluasi mutu instrumen penilaian yang telah diujikan kepada peserta didik. Menurut Arikunto (2013), analisis butir soal bertujuan untuk memperoleh informasi tentang kualitas setiap butir soal sehingga dapat dilakukan perbaikan yang diperlukan. Kegiatan analisis ini penting dilakukan agar instrumen evaluasi yang digunakan benar-benar mampu mengukur kompetensi peserta didik secara valid dan reliabel.

Analisis kualitas tes pada laporan ini dilakukan terhadap soal evaluasi pembelajaran IPS berbentuk pilihan ganda yang terdiri atas 20 butir soal, yang diujicobakan kepada 32 peserta didik kelas VII semester genap SMPN 44 Bandar Lampung. Analisis mencakup aspek: (1) statistik deskriptif hasil tes, (2) tingkat kesukaran butir soal, (3) daya pembeda butir soal, dan (4) rekomendasi penggunaan butir soal.

### B. Landasan Teori

#### 1. Tingkat Kesukaran (Difficulty Index)

Tingkat kesukaran merupakan proporsi peserta tes yang menjawab butir soal dengan benar. Besarnya indeks kesukaran berkisar antara 0,00 hingga 1,00. Butir soal dengan indeks mendekati 0,00 berarti terlalu sulit, sedangkan mendekati 1,00 berarti terlalu mudah. Rumus yang digunakan adalah:

$$P = B / JS$$

Keterangan: P = Indeks Kesukaran, B = Banyaknya siswa yang menjawab soal benar, JS = Jumlah seluruh siswa peserta tes.

Rentang P	Kategori	Interpretasi	Tindak Lanjut
$P > 0,75$	Mudah	Soal terlalu mudah bagi peserta	Diperbaiki atau diganti
$0,25 \leq P \leq 0,75$	Sedang	Soal memiliki kesukaran optimal	Digunakan tanpa perubahan
$P < 0,25$	Sulit	Soal terlalu sulit bagi peserta	Diperbaiki atau diganti

#### 2. Daya Pembeda (Discriminating Power)

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah (Arikunto, 2013). Prosedur penghitungan daya beda dilakukan dengan membagi peserta tes menjadi kelompok atas (27% skor tertinggi) dan kelompok bawah (27% skor terendah). Rumus yang digunakan adalah:

$$D = (BA / JA) - (BB / JB) \text{ atau } D = PA - PB$$

Keterangan: BA = Banyak kelompok atas yang menjawab benar, BB = Banyak kelompok bawah yang menjawab benar, JA = Jumlah peserta kelompok atas, JB = Jumlah peserta kelompok bawah.

Rentang D	Kategori	Interpretasi	Tindak Lanjut
$D \geq 0,40$	Baik Sekali	Sangat baik dalam membedakan kemampuan	Dipakai
0,30 – 0,39	Baik	Baik dalam membedakan kemampuan	Dipakai
0,20 – 0,29	Cukup	Cukup baik, perlu diperbaiki	Direvisi
$D < 0,20$	Jelek	Tidak mampu membedakan kemampuan	Dibuang/Direvisi Total

### 3. Reliabilitas Tes

Reliabilitas tes merujuk pada konsistensi atau kestabilan hasil pengukuran. Tes yang reliabel akan menghasilkan skor yang relatif sama apabila diteskan kembali kepada kelompok yang sama pada kesempatan berbeda. Dalam penelitian ini, analisis reliabilitas diestimasi melalui parameter statistik deskriptif skor tes secara keseluruhan, mengingat data yang tersedia adalah skor total per peserta didik tanpa matriks jawaban lengkap per butir soal.

#### C. Statistik Deskriptif Hasil Tes

Sebelum dilakukan analisis per butir soal, terlebih dahulu disajikan rangkuman statistik deskriptif hasil tes secara keseluruhan sebagai gambaran umum kualitas tes. Statistik deskriptif dihitung berdasarkan skor nilai (skala 0–100) yang diperoleh 32 peserta didik.

Parameter Statistik	Nilai
Jumlah Soal (k)	<b>20 butir soal pilihan ganda</b>
Jumlah Peserta Tes (N)	<b>32 siswa</b>
Nilai Tertinggi ( $X_{maks}$ )	<b>85</b>
Nilai Terendah ( $X_{min}$ )	<b>60</b>
Rentang Nilai (Range)	<b>25 poin</b>
Rata-rata (Mean)	<b>72,81</b>
Standar Deviasi (SD)	<b>6,24 (skala 0–100)</b>
Rata-rata Tingkat Kesukaran ( $\bar{P}$ )	<b>0.725 (Mudah)</b>
Rata-rata Daya Beda ( $\bar{D}$ )	<b>0.228 (Cukup)</b>
Persentase Ketuntasan	<b>50,00% (16 dari 32 siswa)</b>

Berdasarkan tabel di atas, rata-rata nilai peserta didik adalah 72,81, masih di bawah KKM yang ditetapkan yaitu 75. Nilai tertinggi yang diraih adalah 85 dan nilai terendah adalah 60, dengan rentang (range) sebesar 25 poin. Standar deviasi sebesar 6,24 menunjukkan bahwa sebaran nilai peserta didik tergolong sempit, artinya sebagian besar siswa memiliki nilai yang relatif seragam dan berkumpul di sekitar rata-rata. Ini menjadi salah satu indikator bahwa soal belum sepenuhnya mampu mendiskriminasi kemampuan peserta didik secara optimal.

Koefisien variasi (CV) sebesar 8,57% mengindikasikan bahwa variasi skor antarindividu tergolong rendah, yang berarti soal cenderung memberikan skor yang homogen kepada semua

peserta. Kondisi ini berkaitan erat dengan tingkat kesukaran butir soal yang rata-rata berada pada kategori mudah ( $\bar{P} = 0,728$ ), sehingga sebagian besar peserta berhasil menjawab dengan benar.

#### D. Analisis Per Butir Soal

##### 1. Penetapan Kelompok Atas dan Kelompok Bawah

Analisis tingkat kesukaran dan daya pembeda dilakukan dengan menggunakan metode kelompok ekstrem, yaitu membagi peserta tes menjadi kelompok atas (upper group) dan kelompok bawah (lower group) masing-masing sebesar 27% dari total peserta. Dengan jumlah peserta 32 siswa, diperoleh:

Kelompok	Jumlah Siswa	Rentang Nilai	Rata-rata Skor Benar
Atas (27% teratas)	9 siswa	80 – 85	16,11 dari 20 soal
Bawah (27% terbawah)	9 siswa	60 – 65	13,11 dari 20 soal

##### 2. Tabel Analisis Tingkat Kesukaran dan Daya Pembeda

Berikut adalah hasil perhitungan tingkat kesukaran (P) dan daya pembeda (D) untuk setiap butir soal. Kolom Ba menunjukkan jumlah peserta kelompok atas yang menjawab benar, dan kolom Bb menunjukkan jumlah peserta kelompok bawah yang menjawab benar.

No. Soal	Pokok Bahasan	Kelompok Atas (Ba)	Kelompok Bawah (Bb)	Tingkat Kesukaran (P)	Kategori TK	Daya Beda (D)	Kategori DB
1	Jenis interaksi sosial	8/9	6/9	0.78	Mudah	0.22	Cukup
2	Syarat interaksi sosial	7/9	5/9	0.67	Sedang	0.22	Cukup
3	Proses sosial (akomodasi)	7/9	6/9	0.72	Sedang	0.11	Jelek
4	Faktor pendorong lembaga sosial	8/9	7/9	0.83	Mudah	0.11	Jelek
5	Fungsi lembaga ekonomi	8/9	5/9	0.72	Sedang	0.33	Baik
6	Letak geografis dan sosial budaya	7/9	4/9	0.61	Sedang	0.33	Baik
7	SDA komoditas ekspor	8/9	7/9	0.83	Mudah	0.11	Jelek
8	Pelestarian lingkungan pelajar	8/9	6/9	0.78	Mudah	0.22	Cukup
9	Keragaman suku bangsa	7/9	5/9	0.67	Sedang	0.22	Cukup
10	Peran Indonesia di ASEAN	7/9	5/9	0.67	Sedang	0.22	Cukup
11	Kegiatan distribusi	8/9	6/9	0.78	Mudah	0.22	Cukup
12	Peran pemerintah dalam ekonomi	7/9	4/9	0.61	Sedang	0.33	Baik
13	Nilai guna bentuk	8/9	7/9	0.83	Mudah	0.11	Jelek
14	Permintaan efektif	8/9	6/9	0.78	Mudah	0.22	Cukup
15	Hukum permintaan	7/9	4/9	0.61	Sedang	0.33	Baik
16	Pasar online/abstrak	8/9	5/9	0.72	Sedang	0.33	Baik
17	Manfaat perdagangan antardaerah	7/9	5/9	0.67	Sedang	0.22	Cukup

No. Soal	Pokok Bahasan	Kelompok Atas (Ba)	Kelompok Bawah (Bb)	Tingkat Kesukaran (P)	Kategori TK	Daya Beda (D)	Kategori DB
18	Penggunaan devisa ekspor	8/9	7/9	0.83	Mudah	0.11	Jelek
19	Lembaga keuangan (Pegadaian)	7/9	4/9	0.61	Sedang	0.33	Baik
20	Tujuan pembangunan ekonomi	8/9	6/9	0.78	Mudah	0.22	Cukup

### 3. Rekapitulasi Hasil Analisis

Kategori	Tingkat Kesukaran	Daya Beda	Keputusan Butir
Kategori 1	Mudah ( $P > 0.75$ ): 9 soal	Baik Sekali ( $D \geq 0.40$ ): 0 soal	Dipakai: 6 soal
Kategori 2	Sedang ( $0.25 \leq P \leq 0.75$ ): 11 soal	Baik ( $0.30-0.39$ ): 6 soal	Direvisi: 9 soal
Kategori 3	Sulit ( $P < 0.25$ ): 0 soal	Cukup ( $0.20-0.29$ ): 9 soal	Dibuang/Direvisi: 5 soal
Kategori 4	-	Jelek ( $D < 0.20$ ): 5 soal	-
<b>TOTAL</b>	<b>20 soal</b>	<b>20 soal</b>	<b>20 soal</b>

Keterangan keputusan: **Dipakai** = butir soal langsung dapat digunakan; **Direvisi** = butir soal perlu diperbaiki sebelum digunakan; **Dibuang/Direvisi Total** = butir soal tidak layak digunakan dan perlu diganti atau direvisi secara menyeluruh.

### E. Tabel Keputusan Penggunaan Butir Soal

Berdasarkan hasil analisis tingkat kesukaran dan daya pembeda, berikut adalah keputusan akhir terhadap setiap butir soal beserta alasannya:

No.	Pokok Bahasan	P	Kat. TK	D	Kat. DB	Keputusan	Alasan
1	Jenis interaksi sosial	0.78	Mudah	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
2	Syarat interaksi sosial	0.67	Sedang	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
3	Proses sosial (akomodasi)	0.72	Sedang	0.11	Jelek	Dibuang/Direvisi	Daya beda lemah, soal perlu direvisi total atau diganti
4	Faktor pendorong lembaga sosial	0.83	Mudah	0.11	Jelek	Dibuang/Direvisi	Daya beda lemah, soal perlu direvisi total atau diganti
5	Fungsi lembaga ekonomi	0.72	Sedang	0.33	Baik	Dipakai	Daya beda baik, soal layak digunakan

No.	Pokok Bahasan	P	Kat. TK	D	Kat. DB	Keputusan	Alasan
6	Letak geografis dan sosial budaya	0.61	Sedang	0.33	Baik	Dipakai	Daya beda baik, soal layak digunakan
7	SDA komoditas ekspor	0.83	Mudah	0.11	Jelek	Dibuang/Direvisi	Daya beda lemah, soal perlu direvisi total atau diganti
8	Pelestarian lingkungan pelajar	0.78	Mudah	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
9	Keragaman suku bangsa	0.67	Sedang	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
10	Peran Indonesia di ASEAN	0.67	Sedang	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
11	Kegiatan distribusi	0.78	Mudah	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
12	Peran pemerintah dalam ekonomi	0.61	Sedang	0.33	Baik	Dipakai	Daya beda baik, soal layak digunakan
13	Nilai guna bentuk	0.83	Mudah	0.11	Jelek	Dibuang/Direvisi	Daya beda lemah, soal perlu direvisi total atau diganti
14	Permintaan efektif	0.78	Mudah	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
15	Hukum permintaan	0.61	Sedang	0.33	Baik	Dipakai	Daya beda baik, soal layak digunakan
16	Pasar online/abstrak	0.72	Sedang	0.33	Baik	Dipakai	Daya beda baik, soal layak digunakan
17	Manfaat perdagangan antardaerah	0.67	Sedang	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)
18	Penggunaan devisa ekspor	0.83	Mudah	0.11	Jelek	Dibuang/Direvisi	Daya beda lemah, soal perlu direvisi total atau diganti
19	Lembaga keuangan (Pegadaian)	0.61	Sedang	0.33	Baik	Dipakai	Daya beda baik, soal layak digunakan
20	Tujuan pembangunan ekonomi	0.78	Mudah	0.22	Cukup	Direvisi	Daya beda cukup, perlu diperbaiki pilihan jawaban (distractor)

## F. Pembahasan

### 1. Tingkat Kesukaran Butir Soal

Hasil analisis menunjukkan bahwa dari 20 butir soal yang diujikan, sebanyak **9 soal (45%)** termasuk kategori Mudah ( $P > 0,75$ ), **11 soal (55%)** termasuk kategori Sedang ( $0,25 \leq P \leq 0,75$ ), dan **0 soal (0%)** termasuk kategori Sulit. Rata-rata tingkat kesukaran keseluruhan adalah  $\bar{P} = 0.725$ , yang mengindikasikan bahwa soal secara umum berada pada kategori mudah.

Proporsi soal mudah yang cukup banyak menunjukkan bahwa instrumen evaluasi ini belum sepenuhnya memenuhi komposisi ideal menurut prinsip penyusunan tes yang baik. Distribusi tingkat kesukaran yang ideal untuk tes sumatif adalah sekitar 25% soal mudah, 50% soal sedang, dan 25% soal sulit (Nurgiyantoro, 2010). Kondisi dominasi soal mudah dapat berdampak pada rendahnya daya pembeda, karena hampir semua peserta mampu menjawab dengan benar, sehingga perbedaan antara kelompok atas dan bawah menjadi tidak signifikan.

### 2. Daya Pembeda Butir Soal

Daya pembeda butir soal menunjukkan kemampuan soal dalam membedakan antara peserta didik yang berkompotensi tinggi dengan yang berkompotensi rendah. Hasil analisis menunjukkan bahwa terdapat **0 soal** berkategori Baik Sekali ( $D \geq 0,40$ ), **6 soal** berkategori Baik ( $0,30-0,39$ ), **9 soal** berkategori Cukup ( $0,20-0,29$ ), dan **5 soal** berkategori Jelek ( $D < 0,20$ ). Rata-rata daya pembeda adalah  $\bar{D} = 0.228$ .

Soal-soal dengan kategori daya beda Jelek, yaitu soal nomor 3, 4, 7, 13, dan 18 ( $D = 0,11$ ), perlu mendapat perhatian serius. Rendahnya daya beda pada butir-butir soal tersebut dapat disebabkan oleh: (1) tingkat kesukaran terlalu rendah (mudah) sehingga hampir semua siswa, baik kelompok atas maupun bawah, menjawab dengan benar; (2) kunci jawaban yang terlalu jelas sehingga pengecoh (distractor) tidak berfungsi efektif; atau (3) materi yang diujikan sudah sangat dikuasai oleh seluruh peserta.

### 3. Efektivitas Pengecoh (Distractor)

Analisis efektivitas pengecoh merupakan bagian penting dalam evaluasi kualitas butir soal pilihan ganda. Pengecoh yang baik adalah pengecoh yang dipilih oleh setidaknya 5% dari seluruh peserta tes dan lebih banyak dipilih oleh kelompok bawah daripada kelompok atas (Arikunto, 2013). Soal-soal dengan tingkat kesukaran tinggi (mudah) umumnya memiliki pengecoh yang kurang berfungsi, karena peserta langsung memilih kunci jawaban. Untuk meningkatkan kualitas pengecoh, disarankan pengecoh dibuat lebih menarik secara konseptual dan relevan dengan kesalahan umum peserta didik.

### 4. Reliabilitas dan Kestabilan Tes

Reliabilitas tes menggambarkan konsistensi hasil pengukuran. Berdasarkan statistik deskriptif skor yang tersedia, rentang nilai yang sempit (60–85) dan standar deviasi yang kecil ( $SD = 6,24$ ) mengindikasikan bahwa instrumen tes kurang mampu mendistribusikan skor peserta secara beragam. Kondisi ini umumnya berkorelasi dengan koefisien reliabilitas yang rendah, yang berarti soal belum mampu secara konsisten membedakan peserta didik dengan tingkat kemampuan yang berbeda.

Untuk meningkatkan reliabilitas tes, beberapa langkah yang dapat dilakukan antara lain: (1) menambah jumlah butir soal, (2) meningkatkan variasi tingkat kesukaran, terutama dengan menambah soal berkategori sedang hingga sulit, (3) memperbaiki pengecoh agar lebih berfungsi, dan (4) menggunakan teknik analisis butir yang lebih komprehensif, seperti Item Response Theory (IRT) atau program ITEMAN/ANATES.

## G. Kesimpulan dan Rekomendasi

### 1. Kesimpulan

1. Dari 20 butir soal pilihan ganda yang dianalisis, terdapat 9 soal (45%) berkategori Mudah dan 11 soal (55%) berkategori Sedang, serta tidak ada soal berkategori Sulit. Komposisi ini menunjukkan soal belum memiliki distribusi tingkat kesukaran yang ideal.
2. Daya pembeda butir soal bervariasi antara  $D = 0,11$  (Jelek) hingga  $D = 0,33$  (Baik). Sebanyak 6 soal layak dipakai, 9 soal perlu direvisi, dan 5 soal perlu direvisi total atau diganti karena daya bedanya sangat rendah.
3. Sebaran skor yang sempit ( $range = 25$ ) dan standar deviasi kecil ( $SD = 6,24$ ) mengindikasikan bahwa tes belum optimal dalam mendiskriminasi kemampuan peserta didik, yang berimplikasi pada rendahnya estimasi reliabilitas tes.
4. Secara umum, instrumen evaluasi ini masih layak digunakan untuk tujuan formatif, namun memerlukan revisi dan penyempurnaan sebelum digunakan sebagai alat ukur sumatif yang bersifat standar.

### 2. Rekomendasi

- a. Butir soal nomor 3, 4, 7, 13, dan 18 (daya beda Jelek,  $D = 0,11$ ) sebaiknya direvisi total atau diganti dengan soal baru yang memiliki tingkat kesukaran dan daya beda yang lebih seimbang.
- b. Butir soal berkategori Cukup ( $D = 0,22$ ) perlu diperbaiki pada bagian pilihan jawaban, khususnya pengecoh, agar lebih mampu membedakan peserta didik berkemampuan tinggi dan rendah.
- c. Disarankan agar pengembangan soal ke depan memperhatikan distribusi tingkat kesukaran secara proporsional:  $\pm 25\%$  mudah,  $\pm 50\%$  sedang, dan  $\pm 25\%$  sulit.
- d. Untuk analisis yang lebih mendalam, disarankan menggunakan software analisis butir soal seperti ANATES, ITEMAN, atau Rasch Model agar diperoleh data matrik jawaban per peserta per butir soal yang lebih akurat.
- e. Bagi peserta didik yang belum mencapai KKM pada aspek pengetahuan (50% atau 16 siswa), diperlukan program remedial yang terstruktur dengan pendampingan guru, terutama pada materi interaksi sosial, lembaga sosial, dan kegiatan ekonomi.

## H. Daftar Pustaka

- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Nurgiyantoro, B. (2010). *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta: BPF.
- Purwanto, M. N. (2009). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
- Sudjana, N. (2011). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Widoyoko, E. P. (2012). *Evaluasi Program Pembelajaran: Panduan Praktis bagi Pendidik dan Calon Pendidik*. Yogyakarta: Pustaka Pelajar.