

PENGUNAAN *ITEM AND TEST ANALYSIS* 4.3 DALAM MENGANALISIS BUTIR SOAL ULANGAN HARIAN PADA MATA PELAJARAN PANCASILA AND *CIVIC EDUCATION* KELAS VIII

Atika Cahya Fajriyati Nahdiyah^{1*}, Ach Chairy², Arif Surya Volta³

UIN Sunan Kalijaga Yogyakarta, Indonesia¹

IAI Agus Salim Metro Lampung, Indonesia²

Universitas Negeri Yogyakarta, Indonesia³

*Email Corresponding: 21204082034@student.uin-suka.ac.id

Abstrak

Dalam penelitian ini, terdapat beberapa permasalahan yang diidentifikasi, yaitu kurangnya pemahaman dan pengembangan soal oleh sebagian guru, analisis butir soal yang belum sesuai dengan prinsip, mekanisme, serta prosedur penilaian. Tujuan dari penelitian ini adalah untuk mengevaluasi kualitas soal Pendidikan Pancasila dan Kewarganegaraan di SMPN 2 Galur Kulon Progo dari beberapa aspek tertentu. Pengumpulan data dilakukan dengan teknik dokumentasi, termasuk lembar soal penilaian, kunci jawaban soal, dan lembar jawaban siswa. Jumlah siswa yang menjadi subjek penelitian sebanyak 161 siswa. Analisis data dilakukan menggunakan program Item and Analysis versi 4.3. Hasil dari penelitian ini adalah menunjukkan bahwa validitas menggunakan rumus Aiken's Validity, terdapat 18 butir soal dari 25 soal atau 72% soal dinyatakan validitas sedang atau valid dan 7 butir soal atau 28% soal dinyatakan validitas tinggi. Reliabilitas dari keseluruhan soal tersebut memiliki nilai sebesar 0,947 yaitu reliabilitas sangat tinggi. Indeks daya pembeda butir soal yaitu 72% berkategori baik, 12% berkategori sedang, 8% berkategori baik tetapi perlu direvisi, dan 8% berkategori buruk atau ditolak. Indeks tingkat kesukaran soal yaitu 16% dengan kriteria sukar, 56% dengan kriteria sedang, 24% dengan kriteria mudah dan 4% dengan kriteria soal sangat mudah. Fungsi *distractor* (pengecoh) dapat diketahui bahwa kualitas butir soal Pendidikan Pancasila dan Kewarganegaraan 28% sangat baik, 44% baik dengan perbaikan, dan 28% cukup baik dengan perbaikan. Rekomendasi perbaikan dapat dilakukan terutama pada soal-soal yang memiliki kategori sedang, perlu direvisi, atau dikategorikan buruk untuk meningkatkan kualitas keseluruhan instrumen penelitian.

Kata kunci: analisis butir soal, *item and test analysis*, pancasila and *civic education*

PENDAHULUAN

Keberhasilan pendidikan sangat bergantung pada kemampuan guru dalam melaksanakan kegiatan pembelajaran. Guru berperan penting dalam menciptakan lingkungan pembelajaran yang menyenangkan dan efektif. Diperlukan strategi pembelajaran inovatif dan model pembelajaran yang berbeda untuk meningkatkan kualitas pembelajaran. Hal ini menunjukkan pentingnya penggunaan metode pembelajaran yang menarik dan sesuai dengan perkembangan peserta didik. Evaluasi atau penilaian dianggap sebagai bagian terpenting dalam pendidikan dan pembelajaran. Proses ini membantu mengukur sejauh mana tujuan pembelajaran tercapai dan memberikan umpan balik yang berguna (Bowo, 2015).

Terdapat permasalahan dalam sistem penilaian pendidikan Indonesia, seperti fokus yang masih terlalu mendalam pada aspek kognitif peserta didik. Sistem ujian nasional (UAN) dan kelulusan ujian akhir sekolah disebutkan sebagai contoh instrumen yang terbatas dan belum melibatkan proses validasi yang memadai. Instrumen penilaian seringkali dibuat oleh guru tanpa melalui proses validasi

yang memadai. Hal ini dapat menyebabkan kesulitan bagi guru dalam menentukan unsur-unsur instrumen soal yang sesuai dengan standar kompetensi soal yang baik. Guru dihadapkan pada kesulitan dalam menyusun soal yang sesuai dengan standar kompetensi, yang menunjukkan perlunya pengembangan keterampilan guru dalam merancang instrumen penilaian yang valid dan relevan (Rotama dkk., 2020).

Berdasarkan hasil pengamatan yang dilakukan peneliti selama mahasiswa tersebut melaksanakan program praktik kependidikan (PK) UNY di SMPN 2 Galur Kulon Progo diperoleh informasi bahwa banyak siswa kelas VIII yang belum mencapai kriteria ketuntasan minimal (KKM) pada penilaian ulangan harian, dengan nilai di bawah 76. Bahkan, beberapa siswa perlu mengikuti remedial karena nilai mereka jauh di bawah standar. Kondisi rendahnya nilai siswa dirasakan karena kurang optimalnya guru dalam melakukan evaluasi pada proses pembelajaran. Guru belum memahami dan mengembangkan soal, serta belum melakukan analisis butir soal sesuai dengan prinsip, mekanisme, dan prosedur penilaian. Upaya perbaikan dan peningkatan kualitas pembelajaran, terutama pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan, dapat dilakukan dengan memberikan perhatian khusus pada pengembangan kompetensi guru, penyusunan soal yang baik, serta penerapan strategi evaluasi yang lebih efektif.

Analisis butir soal adalah proses yang melibatkan pengumpulan, ringkasan, dan penggunaan informasi dari tanggapan siswa. Tujuannya adalah membuat keputusan yang tepat terkait dengan setiap butir soal dalam sebuah penilaian. Tujuan penelitian adalah mempelajari dan meneliti setiap butir soal sebelum digunakan dalam ujian atau penilaian. Hal ini menunjukkan pentingnya mendeteksi potensi kelemahan atau kekurangan dalam soal sebelum diterapkan. Analisis dasar butir soal juga bertujuan untuk mengklarifikasi informasi diagnostik dari siswa, sehingga dapat diketahui sejauh mana pemahaman mereka terhadap materi pelajaran yang diajarkan oleh guru. Dengan melakukan analisis butir soal secara cermat, guru dapat memastikan bahwa tes yang diberikan tidak hanya mengukur pemahaman siswa secara akurat, tetapi juga memberikan informasi yang berharga untuk perbaikan pembelajaran di kelas. Proses ini dapat menjadi bagian integral dari upaya meningkatkan mutu pendidikan dan evaluasi pembelajaran (Adawiah & Ruchliyadi, 2019).

Para penulis soal dapat melakukan analisis kualitatif terhadap butir soal, baik dari segi isi (validitas isi) maupun bentuknya. Analisis kualitatif melibatkan pertimbangan terhadap kecocokan soal dengan materi yang diajarkan (validitas isi) dan konstruksi soal yang sesuai dengan tujuan pengukuran (validitas konstruk). Analisis kuantitatif melibatkan karakteristik statistik seperti kesulitan item, item diskriminasi, validitas item, dan reliabilitas. Kesulitan item mengukur seberapa sulit atau mudah soal tersebut, sementara item diskriminasi mengindikasikan sejauh mana soal dapat membedakan antara siswa yang memiliki kemampuan yang tinggi dan rendah. Penulis soal dapat melakukan perbaikan butir soal berdasarkan evaluasi atau kajian yang melibatkan analisis kualitatif, dan juga berdasarkan metode perbaikan empiris yang melibatkan data dan fakta terkait kinerja siswa terhadap soal.

Pernyataan tersebut menekankan bahwa metode terbaik adalah menggunakan keduanya, yaitu analisis kualitatif dan kuantitatif secara bersamaan atau dengan cara menggabungkannya. Penggabungan teknik ini dapat memberikan gambaran yang lebih komprehensif tentang kualitas butir soal. Analisis kualitatif dan kuantitatif memiliki kelebihan dan kekurangannya masing-masing. Pendekatan kualitatif dapat memberikan wawasan mendalam tentang relevansi dan konstruksi soal, sementara analisis kuantitatif dapat memberikan data numerik untuk mengukur kinerja dan karakteristik statistik soal. Pendekatan yang mencakup baik analisis kualitatif maupun kuantitatif dapat membantu memastikan bahwa butir soal tidak hanya valid dan reliabel, tetapi juga efektif dalam mengukur pemahaman siswa dengan cara yang bermakna dan sesuai dengan tujuan pengajaran (Saely & Shaleh, 2023).

Hasil wawancara dengan guru pendidikan pancasila dan kewarganegaraan di SMPN 2 Galur Kulon Progo menunjukkan adanya beberapa temuan penting terkait penyusunan soal dan analisis soal dalam mata pelajaran tersebut yaitu guru-guru pendidikan pancasila dan kewarganegaraan jarang melakukan analisis soal pada mata pelajaran tersebut. Hal ini mencerminkan kurangnya kesadaran atau kebiasaan dalam mengevaluasi dan memperbaiki kualitas soal. Fenomena di lapangan menunjukkan bahwa guru cenderung membuat soal yang terlalu sulit atau terlalu mudah. Kedua hal ini memiliki dampak negatif terhadap motivasi dan hasil belajar siswa. Kualitas pertanyaan yang baik adalah pertanyaan yang mampu memberikan tantangan sesuai dengan tingkat kemampuan siswa. Soal seharusnya tidak terlalu mudah atau sulit, melainkan sesuai dengan tingkat kesulitan yang mendukung pengembangan pemahaman siswa.

Berkaitan dengan berbagai hal yang telah dipaparkan diatas perlu untuk dilakukan penelitian tentang analisis butir soal. Oleh karena itu, peneliti mencoba untuk meneliti soal-soal yang telah dibuat untuk peserta didik kelas VIII semester ganjil tahun pelajaran 2023/2024 di SMPN 2 Galur Kulon Progo oleh salah satu mahasiswa praktik kependidikan (PK) dari UNY pada mata pelajaran pendidikan pancasila dan kewarganegaraan pada penilaian harian kedua.

METODE

Metode harus secara jelas menggambarkan desain penelitian dan memberikan prosedur yang jelas yang dilakukan. Dengan kata lain, metode harus menjadi jawaban atas pertanyaan “Bagaimana Anda melakukannya?”. Paragraf pertama biasanya digunakan untuk menjelaskan desain penelitian dan prosedur secara rinci. Daftar metode dalam urutan yang sama ketika mereka akan muncul di hasil, dalam urutan logis di mana Anda melakukan penelitian.

Penelitian ini menggunakan pendekatan kuantitatif, yang fokus pada fenomena atau gejala dengan hubungan sebab-akibat (Sugiyono, 2016). Dalam hal ini, penelitian mengumpulkan data numerik untuk menganalisis kualitas butir soal. Pendekatan kuantitatif deskriptif digunakan untuk memberikan gambaran atau deskripsi tentang kualitas butir soal penilaian harian kedua pada mata pelajaran pendidikan pancasila dan kewarganegaraan. Data akan dianalisis secara statistik menggunakan program Item and Test Analysis (ITEMAN) versi 4.3. Penelitian dilakukan di SMPN 2 Galur Kulon Progo, Yogyakarta. Tempat ini merupakan lokasi yang relevan dengan tujuan penelitian untuk mengamati dan menganalisis kualitas soal pada mata pelajaran yang ditentukan. Waktu pelaksanaan penelitian mencakup bulan November sampai dengan Desember 2023. Periode ini dirancang agar mencakup penilaian harian kedua pada semester ganjil tahun pelajaran 2023/2024.

Subjek penelitian melibatkan siswa kelas VIII di SMPN 2 Galur, Kulon Progo, dengan pembagian ke dalam kelas-kelas tertentu (VIII A, VIII B, VIII C, VIII D, dan VIII F). Jumlah total subjek penelitian adalah 161 orang. Obyek penelitian adalah soal penilaian harian kedua mata pelajaran Pendidikan Pancasila dan Kewarganegaraan. Fokus penelitian adalah pada analisis kualitas butir soal tersebut. Data yang diperoleh akan dianalisis menggunakan statistik, dan khususnya dengan menggunakan program Item and Test Analysis (ITEMAN) versi 4.3. Ini mencerminkan pendekatan yang sistematis dalam menganalisis karakteristik butir soal. Penelitian ini diarahkan untuk memberikan pemahaman yang lebih mendalam mengenai kualitas butir soal pada mata pelajaran pendidikan pancasila dan kewarganegaraan di tingkat kelas VIII SMPN 2 Galur Kulon Progo, dengan pendekatan yang kuat pada aspek kuantitatif dan statistik. Dengan demikian, penelitian ini dapat memberikan kontribusi yang signifikan terhadap perbaikan dan pengembangan dalam penyusunan soal di masa depan.

HASIL DAN PEMBAHASAN

Pendekatan kuantitatif dengan menggunakan program Item and Test Analysis (ITEMAN) versi 4.3 dalam menganalisis soal penilaian harian kedua mata pelajaran Pendidikan Pancasila dan

Kewarganegaraan di kelas VIII SMPN 2 Galur, Kulon Progo, pada tahun pelajaran 2023/2024, melibatkan beberapa aspek penting yaitu validitas, reliabilitas, daya beda, tingkat kesukaran, dan fungsi distractor (pengecoh). Dengan melibatkan program ITEMAN versi 4.3, penelitian ini memiliki keunggulan dalam menggambarkan karakteristik butir soal secara lebih terperinci melalui analisis kuantitatif. Hasil dari analisis ini dapat memberikan panduan konkrit untuk perbaikan dan peningkatan kualitas soal pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan di SMPN 2 Galur, Kulon Progo, serta menjadi landasan untuk meningkatkan efektivitas pembelajaran.

Validitas

Validitas adalah standar yang menunjukkan tingkat ketepatan suatu instrumen dalam mengukur apa yang akan diukur. Sebuah data dapat dikatakan valid apabila sesuai dengan keadaan senyatanya.

Tabel 1. Butir Soal berdasarkan Indeks Validitas

No	Validitas	No. Butir Soal	Jumlah	Persentase
1	Rendah ($<0,4$)	-	-	-
2	Sedang ($\leq 0,8$)	1, 3, 4, 5, 6, 8, 10, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24 25	18	72%
3	Tinggi ($>0,8$)	2, 7, 9, 11, 12, 13, 18	7	28%

Analisis validitas soal menggunakan rumus Aiken's Validity pada penilaian harian kedua mata pelajaran Pendidikan Pancasila dan Kewarganegaraan kelas VIII semester ganjil tahun pelajaran 2022/2023 di SMPN 2 Galur Kulon Progo memberikan hasil yang positif. Penelitian menggunakan rumus Aiken's Validity untuk mengukur validitas butir soal. Aiken's Validity adalah metode yang digunakan untuk mengevaluasi tingkat validitas suatu instrumen pengukuran, dalam hal ini, validitas butir soal. Dari hasil pengujian validitas, diketahui bahwa 72% dari total butir soal dinyatakan memiliki validitas sedang atau valid, sementara 28% dinyatakan memiliki validitas tinggi. Hal ini mengindikasikan bahwa sebagian besar butir soal telah berhasil mengukur apa yang seharusnya diukur, sesuai dengan teori Sudijono (Sudijono, 2011). Butir soal yang dinyatakan validitas sedang atau tinggi dapat dianggap telah menjalankan fungsinya dengan baik, yaitu mengukur dengan tepat apa yang seharusnya diukur. Ini mendukung kualitas butir soal sebagai alat ukur pemahaman siswa terhadap materi Pendidikan Pancasila dan Kewarganegaraan.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa soal penilaian harian kedua pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan kelas VIII semester ganjil tahun pelajaran 2023/2024 di SMPN 2 Galur Kulon Progo memiliki kualitas yang baik dari segi validitasnya. Meskipun sebagian besar butir soal dinyatakan valid, butir soal yang tidak valid perlu diperbaiki. Rekomendasi perbaikan dapat diberikan agar soal-soal yang akan diujikan kepada peserta didik memiliki validitas yang baik. Dengan demikian, hasil analisis validitas ini dapat menjadi dasar untuk meningkatkan dan menyempurnakan soal-soal yang akan digunakan dalam penilaian harian kedua pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan di SMPN 2 Galur Kulon Progo.

Reliabilitas

Reliabilitas merupakan ukuran sejauh mana suatu instrumen pengukuran, dalam hal ini, tes pilihan ganda, dapat diandalkan dan memberikan hasil yang konsisten. Proses pengujian reliabilitas menggunakan program Item and Analysis (ITEMAN) versi 4.3, dan hasil perhitungan kemudian diinterpretasikan dengan memanfaatkan nilai koefisien reliabilitas (r_{11}). Berikut adalah interpretasi hasil reliabilitas berdasarkan patokan yang digunakan:

1. Jika nilai r_{11} sama dengan atau lebih besar daripada 0,70, maka tes hasil belajar yang diuji reliabilitasnya dapat dianggap memiliki reliabilitas yang tinggi (reliable). Ini menunjukkan bahwa instrumen tersebut dapat diandalkan dan memberikan hasil yang konsisten.

2. Jika nilai r_{11} lebih kecil daripada 0,70, maka tes hasil belajar yang diuji reliabilitasnya dianggap belum memiliki reliabilitas yang tinggi (unreliable). Ini mengindikasikan bahwa instrumen tersebut mungkin tidak memberikan hasil yang konsisten (Sudijono, 2011).

Tabel 2. Kriteria Reliabilitas (Guilford, 1956)

Nilai Siswa	Tingkat Kemampuan Berpikir Tingkat Tinggi
$0,00 < r_{11} \leq 0,20$	Reliabilitas sangat rendah
$0,20 < r_{11} \leq 0,40$	Reliabilitas rendah
$0,40 < r_{11} \leq 0,60$	Reliabilitas sedang
$0,60 < r_{11} \leq 0,80$	Reliabilitas tinggi
$0,80 < r_{11} \leq 1,00$	Reliabilitas sangat tinggi

Tabel hasil perhitungan reliabilitas dengan bantuan program *Item and Analysis* (ITEMAN) versi 4.3 disajikan pada Tabel 3.

Tabel 3. Hasil Pengujian Reliabilitas Butir Soal

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0,947	1,834	0,900	0,928	0,910	0,947	0,962	0,953

Dengan memanfaatkan interpretasi nilai koefisien reliabilitas dan indeks kriteria reliabilitas, penelitian ini dapat memberikan pemahaman tentang sejauh mana tes pilihan ganda pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan kelas VIII semester ganjil tahun pelajaran 2023/2024 di SMPN 2 Galur Kulon Progo dapat diandalkan. Evaluasi reliabilitas ini penting untuk memastikan bahwa instrumen pengukuran memberikan hasil yang konsisten dan dapat dipercaya.

Berdasarkan perhitungan reliabilitas dengan bantuan program *Item and Analysis* (ITEMAN) versi 4.3 mengacu pada kriteria di atas, “Alpha (KR-20)” menunjukkan reliabilitas dari tes ini. Telah diketahui bahwa dari 161 peserta didik yang mengerjakan 25 butir soal pilihan ganda telah menunjukkan bahwa angka reliabilitas dari keseluruhan soal tersebut yaitu sebesar 0,947. Hal ini menunjukkan bahwa soal penilaian harian kedua pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan kelas VIII SMPN 2 Galur Kulon Progo Tahun Pelajaran 2023/2024 memiliki tingkat Reliabilitas yang sangat tinggi.

Berdasarkan uraian di atas dapat peneliti simpulkan bahwa soal penilaian harian kedua pada mata pelajaran Pendidikan Pancasila dan Kewarganegaraan kelas VIII SMPN 2 Galur Kulon Progo merupakan soal yang baik apabila dilihat dari segi reliabilitasnya. Instrumen tes dikatakan memiliki tingkat reliabilitas yang tinggi apabila dapat dengan ajeg memberikan data yang sesuai dengan kenyataan. Reliabilitas soal yang tinggi atau sangat tinggi akan memberikan hasil yang konsisten dalam mengukur suatu hal yang akan diukur. Soal dengan reliabilitas rendah tidak akan dapat memberikan data yang valid (Aeni & Suyato, 2020)

Daya Pembeda

Daya pembeda (*discrimination index*) adalah sebuah ukuran yang mengindikasikan sejauh mana suatu butir soal mampu membedakan antara peserta didik yang memiliki kemampuan tinggi dengan yang memiliki kemampuan rendah terkait dengan materi yang diukur oleh soal tersebut. Koefisien korelasi point-biserial digunakan untuk mengukur daya pembeda butir soal. Proses perhitungan daya pembeda menggunakan program *Item and Test Analysis* (ITEMAN) versi 4.3, dan hasilnya diinterpretasikan untuk menilai seberapa baik butir soal tersebut dalam membedakan antara peserta didik yang memiliki kemampuan tinggi dengan yang memiliki kemampuan rendah. Pada umumnya, interpretasi daya pembeda melalui koefisien korelasi point-biserial dapat mengikuti panduan berikut:

Tabel 4. Kriteria Daya Pembeda

Kriteria Daya Pembeda	
Kriteria	Skala
0,40-1,00	Baik
0,30-0,39	Sedang (Tidak Perlu direvisi)
0,20-0,29	Cukup Baik (Perlu direvisi)
Negatif-0,19	Buruk (Ditolak)

Hasil dari analisis daya pembeda dari soal penilaian harian kedua mata pelajaran Pendidikan Pancasila dan Kewarganegaraan tahun pelajaran 2022/2023 terlihat pada tabel berikut :

Tabel 5. *Summary Statistic*

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	25	14,025	8,002	3	25	0,561	0,624

“Mean Rpbis” menunjukkan daya pembeda dari tes ini. Nilai 0,624 berarti secara keseluruhan soal ini memiliki daya pembeda yang baik. Distribusi daya pembeda soal dari butir soal tersebut dapat dilihat pada Tabel 6.

Tabel 6. *Frequency Distribution for the Rpbis*

Score	Frequency	Item
-0,1 to 0,0	1	10
0,0 to 0,1	0	-
0,1 to 0,2	1	6
0,2 to 0,3	2	15, 24
0,3 to 0,4	3	8, 9, 16
0,4 to 0,5	0	-
0,5 to 0,6	1	4
0,6 to 0,7	3	2, 3, 18
0,7 to 0,8	6	1, 5, 7, 20, 21, 22,
0,8 to 0,9	8	11, 12, 13, 14, 17, 19, 23, 25
0,9 to 1,0	0	-

Dari 25 soal yang dianalisis, sebanyak 18 soal (72%) memiliki daya pembeda baik, 3 soal (12%) memiliki daya pembeda sedang, 2 soal (8%) memiliki daya pembeda cukup baik tetapi perlu direvisi, dan 2 soal (8%) memiliki daya pembeda buruk atau ditolak. Butir soal nomor 6 dan 10 memiliki daya pembeda buruk atau ditolak. Menurut Chatterji (2003), indeks diskriminasi negatif menunjukkan bahwa subjek tidak dapat didiskriminasi berdasarkan tujuan yang ditetapkan. Indeks daya pembeda negatif dapat disebabkan oleh skrip atau konten objek yang membingungkan, kesalahan kunci, dan menunjukkan bahwa produk tidak memenuhi tujuan kualifikasi produk yang dihasilkan.

Butir soal yang baik adalah yang memiliki indeks daya beda positif, menunjukkan bahwa siswa yang menjawab benar memiliki skor ujian yang relatif lebih tinggi. Sebaliknya, butir soal dengan indeks daya beda negatif kurang baik karena tidak dapat membedakan antara kemampuan tinggi dan rendah. Kesimpulan dari pendapat di atas adalah bahwa butir soal yang baik memiliki indeks daya beda positif, sementara yang negatif menunjukkan kurangnya kemampuan dalam membedakan

antara peserta didik yang memiliki kemampuan tinggi dan rendah. Faktor-faktor yang dapat mempengaruhi indeks daya beda negatif termasuk penulisan pertanyaan yang kurang jelas, kesalahan kunci jawaban, dan isi/kompetensi yang diukur kurang jelas (Adawiah & Ruchliyadi, 2019).

Tingkat Kesukaran

Tingkat kesukaran mencerminkan seberapa sulit atau mudah suatu soal dianggap oleh peserta didik. Secara umum, tingkat kesukaran diukur berdasarkan peluang peserta didik untuk menjawab benar soal tersebut pada tingkat kemampuan tertentu. Jika sebuah soal memiliki tingkat kesukaran yang sedang, hal itu berarti soal tersebut dirancang agar tidak terlalu mudah maupun terlalu sulit. Artinya, sebagian besar peserta didik pada tingkat kemampuan yang diharapkan memiliki peluang yang sebanding untuk menjawab dengan benar. Tingkat kesukaran yang sedang dianggap ideal karena dapat memberikan informasi yang lebih akurat tentang kemampuan peserta didik.

Penting untuk menciptakan variasi dalam tingkat kesukaran soal di suatu tes, sehingga dapat mengukur sejauh mana peserta didik menguasai materi dengan cermat. Hal ini memungkinkan untuk mendapatkan gambaran yang lebih lengkap tentang tingkat pemahaman dan kemampuan peserta didik dalam menghadapi materi pelajaran. Kategori soal berdasarkan indeks tingkat kesukaran dengan program *Item and Test Analysis* (ITEMAN) versi 4.3 yaitu

Tabel 7. Kriteria Tingkat Kesukaran

Kriteria Tingkat Kesukaran	
Kriteria	Skala
0,00-0,20	Sangat Sukar
0,21-0,40	Sukar
0,41-0,60	Sedang
0,61-0,80	Mudah
0,81-1,00	Sangat Mudah

Tingkat Kesukaran soal penilaian harian kedua mata pelajaran Pendidikan Pancasila dan Kewarganegaraan kelas VIII SMPN 2 Galur Kulon Progo pada semester ganjil tahun pelajaran 2022/2023 dapat terlihat pada tabel berikut ini.

Tabel 8. Hasil Kriteria Tingkat Kesukaran per Butir Soal

Kriteria	Butir Soal	Jumlah
Sangat Sukar 0,00-0,20	-	-
Sukar 0,21-0,40	3, 9, 15, 18	4
Sedang 0,41-0,60	1, 5, 7, 11, 12, 13, 14, 17, 19, 20, 21, 22, 23, 25	14
Mudah 0,61-0,80	2, 4, 6, 8, 10, 16	6
Sangat Mudah 0,81-1,00	24	1

Tabel 8, memperlihatkan bahwa terdapat 16% butir soal yang memiliki kriteria sukar (0,21-0,40), terdapat 56% butir soal yang memiliki kriteria sedang (0,41-0,60), terdapat 24% butir soal yang memiliki kriteria mudah (0,61-0,80) dan 4% butir soal yang memiliki kriteria soal sangat mudah. Distribusi tingkat kesukaran dari butir soal dapat dilihat pada Tabel 9.

Tabel 9. *Frequency Distribution for the P values*

Score	Frequency
0,0 to 0,1	0
0,1 to 0,2	0
0,2 to 0,3	0
0,3 to 0,4	3
0,4 to 0,5	7
0,5 to 0,6	8
0,6 to 0,7	2
0,7 to 0,8	4
0,8 to 0,9	1
0,9 to 1,0	0

Fungsi Distractor (Pengecoh)

Suatu butir soal dapat dikategorikan sebagai soal yang baik apabila pengecoh dapat berfungsi dengan baik. Keefektifan pengecoh dapat dilihat dari pola sebaran jawaban yang diperoleh dengan menghitung banyaknya peserta didik yang memilih jawaban a,b,c, atau d. Dari pola sebaran jawaban tersebut maka akan dapat dilihat apakah pengecoh dapat berfungsi dengan baik atau tidak. Sebuah pengecoh dapat dikatakan berfungsi dengan baik apabila paling sedikit dipilih oleh 5% pengikut tes. Prop menunjukkan keberfungsian distraktor.

Distraktor A, B, C dan D dengan nilai <5% berarti distraktor tersebut tidak baik (ditolak) maka perlu diganti, apabila jawaban semua pengecoh telah berfungsi dengan baik (>5%) maka soal tersebut masuk kedalam kategori diterima tanpa direvisi. Dengan kriteria ini, dapat diharapkan bahwa butir soal yang baik tidak hanya mengukur pengetahuan peserta didik, tetapi juga mampu membedakan antara peserta yang paham dan yang tidak. Proses evaluasi ini merupakan langkah penting dalam memastikan kualitas tes dan memberikan informasi yang akurat tentang pemahaman peserta didik terhadap materi pelajaran.

Tabel 10. Fungsi Distractor (Pengecoh)

No	Pilihan (Prop. Endorsing)				Keterangan
	A	B	C	D	
1	0,056 (N =9)	0,050 (N =8)	0,584 (N =94)	0,081 (N =13)	Sangat Baik
2	0,043 (N =7)	0,025 (N =4)	0,230 (N =37)	0,702 (N =113)	Cukup Baik, tetapi opsi A dan B perlu diperbaiki
3	0,404 (N =65)	0,000 (N =0)	0,168 (N =27)	0,429 (N =69)	Baik tetapi opsi B, perlu diperbaiki
4	0,745 (N =120)	0,006 (N =1)	0,205 (N =33)	0,043 (N =7)	Cukup Baik, tetapi opsi B dan D perlu diperbaiki
5	0,304 (N =49)	0,571 (N =92)	0,081 (N =13)	0,043 (N =7)	Baik, tetapi opsi D perlu diperbaiki
6	0,031 (N =5)	0,783 (N =126)	0,075 (N =12)	0,112 (N =18)	Baik, tetapi opsi A perlu diperbaiki
7	0,106	0,087	0,317	0,491	Sangat Baik

No	Pilihan (Prop. Endorsing)				Keterangan
	A	B	C	D	
	(N =17)	(N =14)	(N =51)	(N =79)	
8	0,081 (N =13)	0,012 (N = 2)	0,627 (N =101)	0,280 (N =45)	Baik, tetapi opsi B perlu diperbaiki
9	0,398 (N =64)	0,211 (N =34)	0,205 (N =33)	0,174 (N =28)	Sangat Baik
10	0,037 (N =6)	0,708 (N =114)	0,075 (N =12)	0,180 (N =29)	Baik, tetapi opsi A perlu diperbaiki
11	0,484 (N =78)	0,348 (N =56)	0,130 (N =21)	0,037 (N =6)	Baik, tetapi opsi D perlu diperbaiki
12	0,398 (N =64)	0,031 (N =5)	0,528 (N =85)	0,043 (N =7)	Cukup Baik, tetapi opsi B dan D perlu diperbaiki
13	0,565 (N =91)	0,031 (N =5)	0,385 (N =62)	0,019 (N =3)	Cukup Baik, tetapi opsi B dan D perlu diperbaiki
14	0,429 (N =69)	0,019 (N =3)	0,068 (N =11)	0,484 (N =78)	Baik, tetapi opsi B perlu diperbaiki
15	0,130 (N =21)	0,360 (N =58)	0,168 (N =27)	0,342 (N =55)	Sangat Baik
16	0,075 (N =12)	0,671 (N =108)	0,217 (N =35)	0,037 (N =6)	Baik, tetapi opsi D perlu diperbaiki
17	0,087 (N =14)	0,292 (N =47)	0,137 (N =22)	0,484 (N =78)	Sangat Baik
18	0,149 (N =24)	0,398 (N =64)	0,068 (N =11)	0,385 (N =62)	Sangat Baik
19	0,012 (N =2)	0,317 (N =51)	0,571 (N =92)	0,099 (N =16)	Baik, tetapi opsi A perlu diperbaiki
20	0,012 (N =2)	0,025 (N =4)	0,379 (N =61)	0,584 (N =94)	Cukup Baik, tetapi opsi A dan B perlu diperbaiki
21	0,081 (N =13)	0,460 (N =74)	0,087 (N =14)	0,373 (N =60)	Sangat Baik
22	0,043 (N =7)	0,224 (N =36)	0,236 (N =38)	0,497 (N =80)	Baik, tetapi opsi A perlu diperbaiki
23	0,019 (N =3)	0,565 (N =91)	0,006 (N =1)	0,410 (N =66)	Cukup Baik, tetapi opsi A dan C perlu diperbaiki
24	0,068 (N =11)	0,820 (N =132)	0,025 (N =4)	0,087 (N =14)	Baik, tetapi opsi C perlu diperbaiki
25	0,540 (N =87)	0,422 (N =68)	0,012 (N =2)	0,025 (N =4)	Cukup Baik, tetapi opsi C dan D perlu diperbaiki

Efektivitas *distractor* (pengecoh) soal tes ialah bagaimana kemampuan pengecoh soal itu berfungsi untuk mengecoh peserta didik yang kurang cakap memilih alternatif jawaban tersebut. Penulisan soal bentuk pilihan ganda harus memiliki keefektifitasan pengecoh, artinya jangan sampai jawaban menjadi sebuah hadiah untuk peserta didik, tetapi jawaban tersebut dapat menunjukkan kemampuan yang sesungguhnya terkait dengan siapa yang memiliki pengetahuan, kurang memiliki pengetahuan atau bingung dengan materi yang disampaikan (Chatterji, 2003).

Tabel 10, memperlihatkan bahwa terdapat 28% soal yang memiliki kategori fungsi *distractor* (pengecoh) sangat baik, 44% soal dengan kategori fungsi *distractor* (pengecoh) baik dengan perbaikan, dan 28% soal dengan kategori fungsi *distractor* (pengecoh) cukup baik dengan perbaikan.

KESIMPULAN

Secara keseluruhan, hasil analisis menunjukkan bahwa tes Pendidikan Pancasila dan Kewarganegaraan ini memiliki validitas yang baik, reliabilitas sangat tinggi, serta indikator daya pembeda dan tingkat kesukaran yang beragam. Evaluasi terhadap fungsi *distractor* juga memberikan gambaran komprehensif tentang kualitas butir soal dalam mengukur pemahaman peserta didik. Langkah-langkah perbaikan dapat diarahkan pada butir soal yang perlu direvisi dan peningkatan *distractor* agar mencapai kualitas yang lebih optimal.

REFERENSI

- Adawiah, R., & Ruchliyadi, D. A. (2019). Analisis Kualitas Soal Buatan Guru Pendidikan Pancasila dan Kewarganegaraan (PPKn). *Jurnal Pendidikan Kewarganegaraan*, 9(2), 58. <https://doi.org/10.20527/kewarganegaraan.v9i2.7554>
- Aeni, N., & Suyato. (2020). Analisis Butir Soal Mata Pelajaran PPKn Kelas IX SMP 1 Karangjambu Purbalingga Tahun 2019/2020. *Kewarganegaraan, Jurnal Pendidikan Volume, Hukum*, 9(5), 491–501.
- Bowo, A. N. A. (2015). *Cerita Cinta Belajar Mengajar*. Deepublish.
- Chatterji, M. (2003). *Designing and Using Tools for Educational Assessment*. USA: Pearson Education, Inc.
- Guilford, J. P. (1956). *Fundamental Statistics in Psychology and Education*. New York: Mc Graw-Hill Book Co. Inc.
- Rotama, A. D., Budiutomo, T. W., & Bowo, A. N. A. (2020). ANALISIS BUTIR SOAL PENILAIAN TENGAH SEMESTER MATA PELAJARAN PPKn KELAS VII DI SMP MUHAMMADIYAH 7 YOGYAKARTA. *Academy of Education Journal*, 11(01), 24–35. <https://doi.org/10.47200/aoej.v11i01.314>
- Saely, E., & Shaleh. (2023). ANALISI BUTIR SOAL PENILIAN AKHIR SEMESTER (PAS) PADA MATA PELAJARAN PPKn MENGGUNAKAN APLIKASI WINSTEP DAN ITEM AND TES ANALYSIS (ITEMAN) 4.3. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 4(1), 88–100.
- Sudijono, A. (2011). *Pengantar Evaluasi Pendidikan*. Yogyakarta: PT Rajagrafindo Persada.
- Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: PT Alfabet.