



Analisis Butir Soal Berbasis Level Kognitif pada Ujian Akhir Semester Mata Pelajaran PPKn Kelas VII SMP Negeri 2 Pemangkat

Henny Muldia

Universitas Tanjungpura, Kalimantan Barat, Indonesia

Email: hennymuldia09@gmail.com

Sulistyarini

Universitas Tanjungpura, Kalimantan Barat, Indonesia

Email: sulistyarini@fkip.untan.ac.id

Nuraini Asriati

Universitas Tanjungpura, Kalimantan Barat, Indonesia

Email: nuraini.asriati@fkip.untan.ac.id

**Corresponding Author*

Article History

Submitted : 2025-09-15

Accepted : 2025-12-02

Revised : 2025-11-02

Published : 2025-12-07

DOI: <https://doi.org/10.31571/jpkn.v9i2.9629>

Abstrak

Penelitian ini bertujuan untuk menganalisis kualitas item soal Ujian Akhir Semester Ganjil mata pelajaran PPKn di kelas VII SMP Negeri 2 Pemangkat TA 2024/2025. Analisisnya mencakup level kognitif soal dengan mengacu pada Taksonomi Bloom dan analisis empiris (validitas, reliabilitas, daya pembeda, dan tingkat kesukaran). Metode yang digunakan dalam penelitian ini adalah kuantitatif dengan pendekatan deskriptif. Hasil menunjukkan item soal memiliki kualitas cukup baik secara reliabilitas dan soal memiliki proporsi yang seimbang dari segi tingkat kesukarannya. Namun soal didominasi level kognitif rendah (LOTS) dan soal lemah pada dua aspek yaitu 40% butir soal ditemukan tidak valid serta 47% soal tidak memiliki daya pembeda. Berdasarkan analisis empiris terhadap 45 item soal yaitu 40 soal pilihan ganda dan 5 soal uraian, bahwa 27 item soal (60%) valid sedangkan 18 item soal (40%) tidak valid. Reliabilitas soalnya tinggi dengan nilai 0,717 untuk soal pilihan ganda, dan 0,651 untuk soal uraian. Tingkat kesukarannya bervariasi: 13 item (29%) mudah, 18 item (40%) sedang dan 14 item (31%) sukar. Sementara itu, daya pembedanya, 8 item (18%) baik, 16 item (35%) cukup baik, 17 item (38%) jelek, dan 4 item (9%) jelek sekali. Penelitian ini merekomendasikan perlunya peningkatan proporsi soal HOTS dan pelatihan penyusunan soal berstandar bagi guru.

Kata Kunci: Analisis Butir Soal, Level Kognitif, Ujian Akhir Semester, Taksonomi Bloom

Abstract

This study is designed to analyze the quality of the Odd Semester Final Examination questions for PPKn subjects in grade VII students of SMP Negeri 2 Pemangkat for the 2024/2025 academic year. The analysis includes the cognitive level of the question based on Bloom's Taxonomy and empirical analysis (namely validity, reliability, discrimination power, and difficulty level). The method used in this study is quantitative with a descriptive approach. The results showed that the question items has good quality in terms of reliability and the questions had a balanced proportion in terms of difficulty, but the questions were dominated by low cognitive level (LOTS) and weak questions in two aspects, namely 40% of the questions were found to be invalid and 47% of the questions did not have any differentiating power. Based on empirical analysis of the 45 items consisting of 40 multiple-choice and 5 essay questions, it was found 27 items (60%) were valid while 18 items (40%) as invalid. The reliability of the test items was high, with a coefficient of 0.717 for multiple-choice items and 0.651 for essay items. The difficulty level, 13 items (29%) as easy, 18 items (40%) as moderate, and 14 items (31%) as difficult. Meanwhile, the discrimination power, 8 items (18%) as good, 16 items (35%) as fair, 17 items (38%) as poor, and 4 items (9%) as very poor. This study recommends the need to increase the proportion of HOTS questions and standardized question preparation training for teachers.

Keyword: Item Analysis, Cognitive Level, Final Semester Examination, Bloom's Taxonomy



PENDAHULUAN

Peran penting pendidikan dalam meningkatkan mutu sumber daya manusia, di antaranya melalui kegiatan evaluasi hasil belajar. Evaluasi merupakan serangkaian kegiatan yang dilakukan guru untuk mengetahui keberhasilan pembelajaran sekaligus menjadi dasar pengambilan keputusan dalam perbaikan proses belajar (Arikunto, 2021). Dalam praktiknya, tes menjadi salah satu instrumen evaluasi yang efektif digunakan untuk mengukur kemampuan peserta didik (Suwanto & Zain, 2022). Oleh karena itu, kualitas item soal dalam tes, khususnya Ujian Akhir Semester (UAS), menjadi hal yang sangat menentukan dalam pencapaian tujuan pendidikan.

Berbagai penelitian sebelumnya menunjukkan bahwa kualitas soal yang digunakan guru masih belum optimal. Penelitian Zahrotunnaqiyah, Nulhakim, & Taufik (2023) menemukan bahwa komposisi soal lebih banyak berada pada level LOTS dengan persentase 50,8%, sementara soal HOTS hanya 7,5%. Hasil yang hampir serupa oleh Rahmadhani (2022) yang menyatakan bahwa soal LOTS mencapai 60%, MOTS 40%, dan HOTS 0%. Temuan tersebut menunjukkan bahwa guru masih cenderung membuat soal yang menekankan aspek berpikir rendah. Situasi ini bertentangan dengan kebutuhan kurikulum abad ke-21, yang mengutamakan pengembangan kemampuan berpikir tingkat tinggi. Soal-soal yang memuat ketiga level kognitif tersebut sangat baik diberikan kepada peserta didik sebagai alat untuk mengukur keterampilan berpikir peserta didik. Namun pada umumnya, tenaga pendidik di Indonesia rata-rata membuat soal dengan pembagian persentase tingkatan level kognitif LOTS, MOTS, dan HOTS yang tidak signifikan. Berdasarkan lembar kerja bimbingan teknis implementasi kurikulum 2013, menyatakan bahwa penyusunan soal HOTS minimal 10% dari jumlah soal keseluruhan (Wayan Budiyasa, 2019).

Selain itu, penelitian lain menunjukkan bahwa guru sering kali tidak melakukan analisis item soal setelah tes dilaksanakan. Menurut Muchlizani, Mania, & Rasyid (2023) pada dasarnya, analisis terhadap item soal memiliki peranan penting dalam mengidentifikasi kualitas instrumen, melakukan perbaikan terhadap item soal yang tidak memadai, dan merancang bank soal yang lebih bermutu. Namun, fakta di lapangan menunjukkan guru lebih sering mengambil soal dari buku, internet, atau mendaur ulang soal tahun sebelumnya tanpa analisis mendalam (Ambarwati & Ismiyati, 2022). Kondisi ini menyebabkan kualitas evaluasi pembelajaran menjadi kurang efektif. Permasalahan ini kian kompleks ketika guru menghadapi keterbatasan waktu, tenaga, dan kurangnya pelatihan terkait analisis item soal.

Berdasarkan temuan penelitian yang telah dilakukan tersebut, terdapat dua masalah pokok yang teridentifikasi, yaitu dominasi pertanyaan dengan tingkat kognitif rendah dan kurangnya kemampuan pendidik dalam menganalisis soal ujian. Faktanya, kegiatan analisis soal memiliki

peran penting dalam memastikan validitas, reliabilitas, tingkat kesukaran, dan daya beda setiap soal (Widodo, 2021). Dengan demikian, penelitian ini dimaksudkan untuk menganalisis kualitas soal Ujian Akhir Semester mata pelajaran PPKn untuk peserta didik kelas VII di SMP Negeri 2 Pemangkat dengan menganalisis validitas, reliabilitas, tingkat kesulitan, dan daya beda soal-soal tersebut guna memperoleh pemahaman yang lebih menyeluruh tentang mutu instrumen penilaian tersebut.

METODE

Penelitian ini menggunakan pendekatan metode deskriptif kuantitatif dengan maksud untuk menganalisis kualitas item soal Ujian Akhir Semester pada mata pelajaran PPKn bagi peserta didik kelas VII di SMP Negeri 2 Pemangkat TA 2024/2025. Subjek dalam penelitian ini adalah peserta didik kelas VII Tahun Ajaran 2024/2025 pada semester ganjil yakni kelas VII A (27 peserta didik) dan VII B (28 peserta didik). Sedangkan objek yang dianalisis dalam penelitian ini mencakup keseluruhan instrumen ujian yang meliputi 40 item soal berbentuk pilihan ganda dan 5 soal dalam bentuk uraian termasuk jawaban peserta didik, sehingga seluruh populasi dijadikan sampel penelitian. Data diperoleh melalui teknik dokumentasi berupa lembar soal, kunci jawaban, serta hasil pekerjaan peserta didik. Analisis data dilakukan dengan menggunakan aplikasi ITEMAN versi 4.3 untuk menguji validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, kemudian hasilnya dipersentasekan dan diinterpretasikan secara deskriptif.

Prosedur penelitian yang dilakukan dalam penelitian ini mencakup beberapa tahapan, yakni 1) pengumpulan data dengan mengumpulkan semua soal (PG dan uraian) yang digunakan pada ujian PPKn kelas VII di SMP Negeri 2 Pemangkat; 2) mengklasifikasikan atau mengelompokkan tingkat kognitif untuk masing-masing kemampuan kognitif berdasarkan Taksonomi Bloom revisi; 3) menghitung jumlah soal dan persentase untuk masing-masing tingkat kognitif; 4) melakukan analisis kualitas butir soal (validitas, reliabilitas, tingkat kesukaran, daya pembeda) menggunakan Aplikasi ITEMAN versi 4.3 sedangkan pada soal uraian diakumulasikan menggunakan Microsoft excel; 5) penarikan kesimpulan berdasarkan hasil analisis.

HASIL DAN PEMBAHASAN

Level Kognitif

Merujuk pada hasil yang diperoleh dari analisis data terhadap soal Ujian Akhir Semester Ganjil mata Pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat berdasarkan tingkatan kognitif dalam Taksonomi Bloom menunjukkan distribusi sebagai berikut: dari total 45 item soal yang dianalisis, sebanyak 21 item soal (47%) berada pada level Mengingat (C1), yang

meliputi nomor 1,2,3,4,5,6,7,8,11,12, 14,21,28,31,35,38,39 (PG) dan 1,2,4,5 (uraian). Sementara itu, 19 item soal (42%) berada pada level Memahami (C2), yang meliputi nomor 9,10,13,15,16,17,19,20,22,23,24,25,26,29,30,32,33,37 (PG) dan nomor 3 (uraian). Adapun soal yang berada pada level Menerapkan (C3) berjumlah 4 item (9%) yakni nomor 18,27,34,36. Sedangkan hanya terdapat 1 item soal yang berada pada level Menganalisis (C4)1 yaitu nomor 40.

Tabel 1. Distribusi Item Soal Menurut Klasifikasi Level Kognitif

Level Kognitif	Dimensi Proses Berpikir	No.Item Soal		Jumlah	Presentase
		Pilihan Ganda	Uraian		
LOTS	C1	1,2,3,4,5,6,7,8,11,12,14,21,28,31,35,38,39	1,2,4,5	21	47%
	C2	9,10,13,15,16,17,19,20,22,23,24,25,26,29,30,32,33,37	3	19	42%
MOTS	C3	18,27,34,36	-	4	9%
HOTS	C4	40	-	1	2%

Berdasarkan tabel yang disajikan, item soal pilihan ganda pada soal UAS memuat 3 level kognitif, yaitu LOTS dengan 21 item soal (47 %) dimensi proses berpikir mengingat (C1) dan 19 item soal (42%) dimensi proses berpikir memahami (C2), MOTS dengan 4 item soal (9%) dimensi proses berpikir menerapkan (C3), dan HOTS dengan 1 item soal (2 %) dimensi proses berpikir menganalisis (C4).

Pembuatan soal hendaknya disesuaikan dengan tingkat perkembangan kognitif peserta didik. Secara umum, pembelajaran di SMP sudah mulai diarahkan ke penalaran dan pemecahan masalah. Artinya untuk soal-soal dijenjang SMP sudah mulai diarahkan ke C3 dan ditingkatkan ke C4-C5. Sedangkan temuan dari penelitian ini yang mengindikasikan bahwa pada soal Ujian Akhir Semester Ganjil mata Pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat terdapat soal dengan kategori C1 dan C2 lebih banyak dibandingkan soal dengan kategori C3.

Sebaran persentase soal yang baik pada setiap jenjang kognitif kognitif taksonomi Bloom menunjukkan proporsi sebagai berikut: 5% untuk level C1, 10% untuk level C2, 45% untuk level C3, 25% untuk level C4, 10% untuk level C5, dan 5 % untuk level C6 (Helmawati, 2019). Berdasarkan temuan analisis, instrumen soal UAS menunjukkan keterbatasan dan bahkan ketiadaan item soal yang berada pada level kognitif C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta). Kondisi ini bertentangan dengan kebutuhan pendidikan saat ini, mengingat pemerintah telah mengganti Ujian Nasional (UN) dengan Asasemen Kemampuan Minimum (AKM). Instrumen soal AKM dirancang mengacu pada model soal *Programme for International Student Assasement* (PISA) yang menekankan kemampuan berpikir Tingkat tinggi atau *High Order Thinking Skill* (HOTS) dengan cakupan level kognitif C4 hingga C6

(Febrila et al, 2021). Peserta didik yang terbiasa dengan soal C1/C2 di sekolah akan mengalami “*shock*” saat dihadapkan pada soal AKM karena mereka tidak pernah dilatih untuk melakukan penalaran tingkat tinggi. Hal inilah yang menyebabkan rendahnya skor PISA nasional dan tidak tercapainya tujuan reformasi pendidikan melalui AKM.

Melalui hasil pemetaan item soal UAS yang digunakan dalam penelitian ini menunjukkan mayoritas soal didominasi oleh kategori LOTS (*Low Order Thinking Skill*) dengan jumlah 40 item soal, sementara kategori MOTS (*Middle Order Thinking Skill*) hanya terdapat 4 item soal, dan kategori HOTS (*High Order Thinking Skill*) sebanyak 1 item soal. Sementara itu, aspek menganalisis, mengevaluasi dan mencipta nyaris tidak termuat dalam soal Ujian Akhir Semester Ganjil mata Pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat. Hal tersebut menandakan bahwa soal UAS tidak sesuai dengan proporsi ideal level kognitif suatu soal.

Temuan ini mengidentifikasi bahwa instrumen soal UAS sebagian besar berupa soal LOTS khususnya pada aspek Mengingat (C1) dan Memahami (C2) yang hanya melatih peserta didik untuk mengingat seperti menanyakan tanggal suatu peristiwa, nama tokoh, atau bunyi pasal-pasal UUD 1945. Selain itu juga peserta didik hanya diminta memahami konsep dasar dan memberikan contoh dalam kehidupan sehari-hari. Peserta didik belum memperoleh soal-soal yang bersifat aplikatif atau permasalahan kontekstual dalam kehidupan nyata, bahkan belum dibekali latihan-latihan soal untuk melakukan pembuktian atau menarik kesimpulan berdasarkan informasi yang tersedia serta peserta didik belum diminta untuk memberikan penilaian terhadap suatu fenomena.

Kekurangan soal pada level C4 ke atas mengindikasikan soal UAS belum cukup mendorong peserta didik untuk melakukan analisis mendalam, pengambilan keputusan, atau berpikir kritis yang sangat dibutuhkan untuk menghadapi tantangan pembelajaran abad 21. Temuan ini sangat bertentangan dengan fokus pendidikan modern saat ini yakni tuntutan pembelajaran secara mendalam dan bermakna (*Deep Learning*) yang dalam konteks pedagoginya bukanlah sekedar menghafal melainkan pemahaman konseptual yang mendalam dan pemaknaan. Dalam Taksonomi Bloom, *Deep Learning* beroperasi di level kognitif tinggi: Menerapkan (C3) dalam konteks baru, Menganalisis (C4), Mengevaluasi (C5), dan Mencipta (C6).

Validitas

Validitas merupakan suatu ukuran yang menunjukkan tingkat kevalidan atau kesahihan suatu instrumen. Hasil dari penelitian menunjukkan bahwa validitas item soal Ujian Akhir Semester Ganjil pada pembelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat memiliki indeks validitas dengan kategori “valid” berjumlah 27 item (60%), kategori “tidak valid” berjumlah 18 item (40%).

Tabel 2. Distribusi Item Soal Menurut Indeks Validitas

No.Item Soal		Jumlah	Kriteria	Presentase
Pilihan Ganda	Uraian			
4,7,8,10,15,16,18,20,22,23,24,26,27,28,30,32,33,34,36,37,39	1,2,3,4,5	27	Valid	60%
1,2,3,5,6,9,11,12,13,14,17,19,21,29,31,35,38,40	-	18	Tidak Valid	40%

Berdasarkan hasil pengolahan data menunjukkan bahwa dalam kategori soal pilihan ganda, terdapat 22 soal yang memenuhi kriteria valid sementara 18 soal lainnya tidak valid. Sementara itu, untuk kategori soal uraian, semua 5 item soalnya yang dianalisis (100%) menunjukkan kevalidan.

Dari paparan diatas, dapat ditarik kesimpulan bahwa soal UAS dalam penelitian ini terklasifikasi sebagai soal yang memiliki kualitas baik, mengingat total item soal yang memenuhi kriteria validitas mencapai 27 item atau setara dengan 60% dari keseluruhan 45 item soal dengan rincian kategori soal yang valid yaitu validitas dengan kategori tinggi terdapat 6 item soal (13%) dari 45 item soal yakni nomor 10, 16, 24, 26, 34, dan 36. Selanjutnya, validitas dengan kategori cukup terdapat 10 item soal (22%) dari 45 item soal yakni nomor 20,22,23,27,33,39 (PG) dan untuk soal uraian yang meliputi nomor 2,3,4, dan 5. Validitas dengan kategori rendah terdapat 14 (31%) item soal dari 45 item soal yakni nomor 1,2,4,7,9,15,17,18,25,28,30,32,37 dan soal uraian nomor 1. Validitas dengan kategori sangat rendah terdapat 15 item soal (33%) dari 45 item soal yakni nomor 3,5,6, 8,11,12,13,14,19,21,29,31,35,38,40.

Validitas sebuah soal akan tinggi jika skor yang didapatkan pada soal tersebut berkorelasi dengan skor total (Arikunto 2021). Dari pendapat tersebut, dapat dikatakan bahwa soal yang valid memiliki konstruksi yang baik dan mencakup materi yang relevan dengan tujuan pengukurannya. Jika suatu instrumen pengukuran memiliki validitas yang tinggi, itu menandakan bahwa instrumen tersebut efektif dalam mencapai tujuan pengukuran yang ditetapkan dan hasil pengukuran memberikan informasi yang sesuai dengan tujuan awal. Namun jika instrumen pengukuran memiliki validitas rendah, hal itu menandakan bahwa tujuan pengukuran tidak tercapai seperti yang diharapkan (Hartono et al. 2024)

Fakta menunjukkan hanya 35% soal yang mencapai tingkat validitas cukup hingga tinggi sementara terdapat 64% soal memiliki validitas rendah dan dan sangat rendah. Rendahnya validitas tersebut menunjukkan bahwa instrumen soal tidak berhasil mengukur kompetensi yang diharapkan, sehingga dapat memberikan representasi yang kurang akurat mengenai kemampuan peserta didik. Item-item soal dengan validitas rendah dan sangat rendah memerlukan perbaikan. Upaya perbaikan yang dapat dilakukan adalah menyelaraskan

formulasi soal dengan indikator kompetensi yang ditargetkan, mengingat ketidaksesuaian ini menyebabkan instrumen evaluasi tidak mampu mengukur aspek yang dimaksud. Selain itu, perlu juga dilakukan penyesuaian materi dengan konteks pembelajaran dan kemampuan tingkat kognitif yang dimiliki peserta didik (Ambarwati, 2021).

Berdasarkan paparan diatas, dapat ditarik kesimpulan bahwa soal Ujian Akhir Semester mata Pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat dari segi validitas termasuk baik namun perlu beberapa perbaikan pada item soal yang memiliki validitas rendah maupun sangat rendah. Secara Umum, tindak langkah selanjutnya yang bisa ditempuh ialah item soal yang valid bisa segera diarsipkan dan dimasukkan pada bank soal. Adapun, item soal yang tidak valid bisa direvisi melalui sinkronisasi teknik penyusunan instrumen dengan substansi pokok bahasan yang berfungsi sebagai target pengukuran.

Reliabilitas

Reliabilitas merupakan ketepatan suatu instrumen dalam mengukur apa yang seharusnya diukur, mencakup ketelitian hasil pengukuran serta konsistensi hasil apabila pengukuran dilakukan berulang kali.

Tabel 3. Distribusi Item Soal Menurut Indeks Reliabilitas

Jenis Tes	Alpha	Kriteria
Pilihan Ganda	0,717	Tinggi
Essay	0,651	Tinggi

Berdasarkan hasil pengolahan data menggunakan aplikasi diperoleh nilai indeks reliabilitas untuk soal pilihan ganda mencapai 0,717. Mengacu pada standar penilaian yang berlaku, tingkat reliabilitas tersebut terklasifikasi tinggi. Adapun untuk soal uraian, hasil analisis menunjukkan indeks reliabilitas sebesar 0,651 tergolong kategori tinggi. Dengan demikian dapat ditarik kesimpulan bahwa instrumen soal Ujian Akhir Semester Mata Pelajaran PPKn Kelas VII SMP Negeri 2 Pemangkat baik soal pilihan ganda maupun uraian memiliki reliabilitas yang tinggi.

Merujuk pada hasil analisis data, diperoleh informasi bahwa soal pilihan ganda memiliki tingkat reliabilitas sebesar 0,717 yang tergolong tinggi. Hal ini menunjukkan bahwa soal pilihan ganda tersebut reliabel. Adapun pada soal uraian, diketahui tingkat reliabilitas berada di angka 0,651. Mengacu pada pandangan Anas Sudjiono (dalam Rahman & Nasryah, 2019), suatu tes dinyatakan reliabel apabila memiliki koefisien $\geq 0,70$. Berdasarkan kriteria tersebut, soal pilihan ganda dapat dikatakan reliabel dengan kategori tinggi, sementara soal uraian belum memenuhi standar reliabilitas karena nilai 0,651 berada dibawah ambang batas 0,7 menurut standar yang dikemukakan Anas Sudjiono.

Hal ini mengarah pada kesimpulan bahwa soal Ujian Akhir Semester Ganjil pada mata pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat merupakan soal yang memiliki

nilai reliabilitas yang tinggi. Sejalan dengan pendapat Arikunto (2021) suatu instrumen dikatakan memiliki reliabilitas tinggi apabila dapat memberikan hasil yang tetap atau konsisten jika diuji berkali-kali dalam kondisi yang relatif sama. Namun, hasil analisis item soal pada Ujian Akhir Semester Ganjil mata pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat pada mengindikasikan yakni walaupun instrumen tes soal pilihan ganda mempunyai derajat konsistensi yang tinggi, tetapi soal uraian secara standar belum dapat dikatakan reliabel meskipun nilainya mendekati batas minimal reliabel.

Dari temuan diatas, peneliti merujuk pada pendapat Kusaeri dan Suprananto dalam Ambarwati (2022) aspek-aspek yang mempengaruhi reliabilitas meliputi: 1) Kian banyak item soal dalam tes, kian konsisten tes tersebut. 2) Kian lama durasi tes, kian ajeg tes tersebut. 3) Kian terbatas tingkat kesukaran item soal, kian tinggi konsistensinya. 4) item-item soal yang sangat berkorelasi dapat mengurangi konsistensi. 5) Kian objektif dalam pemberian skor, kian tinggi konsistensinya.

Tingkat Kesukaran

Tingkat kesukaran item soal merupakan probabilitas untuk memberikan jawaban yang tepat dalam sebuah tes dengan taraf penguasaan tertentu yang lazimnya diwujudkan melalui indikator.

Tabel 4. Distribusi Item Soal Menurut Indeks Tingkat Kesukaran

No.Item Soal		Jumlah	Kriteria	Presentase
Pilihan Ganda	Uraian			
4,5,6,7,11,12,13,14,19,21,31,38,40	4	14	Sukar	31%
1,3,8,9,15,17,20,22,25,28,29,30,32,35,36	1,3	18	Sedang	40%
2,10,16,18,23,24,26,27,33,34,37,39	5	13	Mudah	29%

Berdasarkan hasil pengolahan data menggunakan aplikasi diperoleh tingkat kesukaran dari total 45 item soal yang dianalisis menunjukkan bahwa soal dengan kategori sukar sebanyak 14 item (31%), soal yang berkategori sedang sebanyak 18 item (40%), serta soal yang diaktegorikan mudah berjumlah 13 item (29%). Hasil tersebut memperlihatkan bahwa item soal UAS dalam penelitian ini didominasi dengan tingkat kesukaran yang tergolong “sedang”. Menurut peneliti, soal UAS tersebut menunjukkan sebaran tingkat kesukaran yang proporsional. Hal ini terlihat dari hasil analisis yang mengindikasikan bahwa item-item soal terdistribusi secara merata dalam tiga kategori tingkat kesukaran yakni mudah, sedang, dan sukar. Dengan kata lain, komposisi soal menunjukkan keseimbangan yang baik antara kategori “mudah” dan “sukar”. Pada penelitian ini diperoleh hasil yakni 40% soal sedang, 31% soal mudah, dan 29% soal sulit.

Sebagaimana sudah dikemukakan instrumen bisa dinilai bermutu jikalau tingkat kesukaran butir soal mempunyai interval “menengah” maknanya instrumen jangan terlampau

sulit serta jangan terlampau mudah (Arikunto, 2021). Instrumen tes yang terlampau mudah cenderung gagal menstimulasi peserta didik untuk berpikir kritis. Kontrasnya, instrumen tes yang terlampau sulit berisiko menurunkan semangat mereka dan berpotensi kehilangan motivasi dalam menyelesaikan soal karena berada diluar batas kemampuan yang dimilikinya.

Berdasarkan hasil tersebut peneliti merujuk pada pendapat yang memberikan persentasi proporsi soal seimbang dengan pertimbangan 3-4-3. Artinya komposisi soal terdiri dari 30% berkategori mudah, 40% berkategori sedang, dan 30% berkategori sukar (Yadnyawati, 2019). Dengan demikian, soal Ujian Akhir Semester mata Pelajaran PPKn di SMP Negeri 2 Pemangkat mengindikasi bahwa sebaran soal mempunyai kualitas yang baik dari aspek tingkat kesukaran yakni instrumen tes mempunyai rasio tingkat kesukaran yang proporsional maknanya instrumen tes dalam penelitian ini tidak sepenuhnya sulit atau sepenuhnya mudah. (Putriani et al. 2020).

Suatu item soal terklasifikasi mudah, dapat diprediksi bahwa 1) pilihan pengecoh pada item tersebut tidak efektif; 2) mayoritas peserta didik sanggup menjawab pertanyaan dengan tepat maknanya yakni materi yang diteskan sudah dikuasai oleh sebagian besar peserta didik. (UIN Sumatera Utara, 2021). Pada penelitian ini, soal dengan tingkat kesukaran mudah sebesar 29% , faktor yang menyebabkan mudahnya soal adalah mayoritas peserta didik sudah memahami materi yang ditanyakan serta penggunaan instruksi yang terlalu jelas sehingga jawaban dapat diprediksi oleh peserta didik.

Selanjutnya, apabila suatu item soal masuk dalam kategori sukar, penyebabnya ialah 1) kunci jawaban berpotensi keliru. 2) item pertanyaannya menyediakan lebih dari satu opsi jawaban yang dapat dinilai benar. 3) pokok bahasan yang diteskan belum diajarkan atau proses pembelajaran belum selesai, sehingga peserta didik belum menguasai pencapaian kompetensi dasar secara optimal. 4) materi yang diuji tidak sesuai dengan jenis soal yang digunakan (misalnya, merangkum cerita tetapi pada format instrumen pilihan ganda). 5) proposisi atau susunan kata pada instrumen terlampau rumit serta berbelit-belit. (UIN Sumatera Utara, 2021).

Pada penelitian ini, faktor yang menyebabkan kesukaran soal sebesar 31% adalah terdapat beberapa soal dengan pernyataan atau kalimat soal terlalu kompleks dan panjang. Selain itu, soal terlalu banyak menekankan hafalan secara spesifik dan detail seperti tanggal, nama tokoh, sejarah, pasal, dll. Maka hal ini, bisa membingungkan peserta didik karena sulit untuk diingat dan biasanya untuk pilihan jawabannya terlalu mirip dan membuat peserta didik memilih jawaban yang kurang tepat.

Tindak lanjut menurut Sudjiono dalam (Muluki et al. 2020) ialah memasukkan item soal yang berkategori sedang kedalam bank soal sehingga bisa dipakai lagi sebagai alat penilaian di masa mendatang. Sementara itu, item soal yang terklasifikasi mudah atau sulit memerlukan

tindakan lebih lanjut berupa dilakukan revisi, penelaahan ulang, pelacakan, dan penelusuran mendalam agar dapat mengidentifikasi faktor-faktor yang menyebabkan butir item itu sulit atau mudah. Dengan demikian, soal-soal tersebut bisa diperbaiki atau digunakan lagi dalam ujian selanjutnya.

Daya Pembeda

Tabel 5. Distribusi Item Soal Menurut Indeks Daya Pembeda

No.Item Soal		Jumlah	Kriteria	Presentase
Pilihan Ganda	Uraian			
		-	Sangat Baik	
10,16,23,24,26,33,34,36	-	8	Baik	18%
4,7,15,18,20,22,27,28,30,32,37,38,39	1,2,3	16	Cukup	35%
1,2,3,5,8,9,11,12,13,,14,17,19,25,29,40	4,5	17	Jelek	38%
6,21,31,35	-	4	Jelek Sekali	9%

Berdasarkan hasil pengolahan data diperoleh informasi mengenai daya beda dari total 45 item soal yang dianalisis menunjukkan bahwa 8 item soal (18%) menunjukkan daya beda baik, 16 item soal (35%) menunjukkan daya pembeda pada kategori cukup baik, 17 item soal (38%) menunjukkan daya beda pada kategori jelek, serta 4 item soal (9%) menunjukkan daya beda pada kategori Jelek Sekali.

Sebuah instrumen tes dikatakan tidak baik dan tidak memiliki daya pembeda apabila berhasil dijawab dengan benar oleh peserta didik dari kelompok atas dan juga kelompok bawah. Kondisi serupa juga terjadi apabila sebagian besar peserta didik dari kedua kelompok tersebut memberikan jawaban yang salah, maka soal itu sama tidak baiknya, karena tidak memiliki daya beda. (Arikunto, 2021). Sebagai contoh peneliti menemukan sebanyak 6 item soal yaitu soal nomor 3, 21,31, 35 dan 38 (PG) dan nomor 4 (uraian). Hal tersebut menggambarkan kondisi di mana peserta didik yang berada pada level kemampuan bawah justru menunjukkan hasil yang lebih baik dalam menjawab item-item tes dibandingkan peserta didik pada level kemampuan atas.

Maka dari hasil analisis tersebut ditemukan bahwa item soal tersebut gagal dalam membedakan antara peserta didik yang pandai dan peserta didik yang kurang pandai. Hal ini bertentangan dengan pendapat diatas, dimana seharusnya hanya peserta didik kelompok atas yang mampu menjawab dengan benar, namun pada kasus ini justru sebagian peserta didik kelompok bawah mampu memberikan lebih banyak jawaban benar dibandingkan kelompok atas.

Berdasarkan pandangan dari Departemen Pendidikan Nasional Republik Indonesia dalam Ambarwati (2022), kegagalan suatu item soal dalam mengukur perbedaan kemampuan peserta didik dipicu oleh berbagai faktor antara lain 1) ketidaktepatan kunci jawaban; 2) terdapat lebih dari satu jawaban yang dapat dianggap benar pada item soal tersebut; 3) ketidakjelasan item

soal mengenai aspek yang hendak diukur; 4) alternatif jawaban tidak efektif; 5) instrumen yang ditekankan terlampau rumit alhasil mendorong mayoritas peserta didik menggunakan strategi menebak dalam menentukan pilihan jawaban; 6) adanya persepsi dari mayoritas peserta didik yang memahami materi bahwa terdapat kesalahan informasi dalam item soalnya.

Merujuk pada pendapat diatas, maka peneliti memprediksi salah satu faktor yang dapat menyebabkan soal UAS dalam penelitian ini tidak memiliki daya pembeda sebesar 47% karena item-item soal yang dimaksud diatas termasuk soal dengan tingkat kesukaran yang sulit. Jadi jawaban peserta didik hampir semua jawabannya salah baik yang pandai maupun yang tidak pandai seperti pada soal nomor 31 yang dimana hanya 5 peserta didik (1 orang kelompok atas dan 4 orang kelompok bawah) yang merespons secara akurat dari 56 murid. Selain itu peneliti juga menemukan banyaknya soal yang tidak valid menyebabkan item soal penelitian ini berdaya pembeda kurang baik serta kompetensi yang diukur tidak jelas karena sebelum membuat soal, guru tidak merancang kisi-kisi soal.

Secara keseluruhan, hasil analisis diperoleh bahwa daya pembeda yang terdapat disoal tersebut didominasi pada kategori jelek artinya soal tidak memiliki daya pembeda. Dengan demikian soal Ujian Akhir Semester mata Pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat tidak dapat mengukur kemampuan peserta didik secara akurat karena soal tersebut tidak mampu membedakan antara peserta didik yang berkemampuan tinggi atau pandai dengan peserta didik yang berkemampuan rendah atau kurang pandai. Item soal yang tidak memiliki daya pembeda sebaiknya direvisi atau diganti dengan item soal yang berbeda (Arikunto, 2021).

Kualitas Butir Soal

Temuan analisis terhadap butir-butir soal pada aspek validitas, reliabilitas, tingkat kesukaran dan daya pembeda digunakan sebagai dasar untuk menentukan klasifikasi kualitas butir soal ke dalam kategori baik, cukup, tidak baik, dan sangat tidak baik.

Tabel 6. Distribusi Kualitas Item Soal

Kualitas Item Soal	No.Item Soal		Jumlah	Presentase
	Pilihan Ganda	Uraian		
Baik (3)	15,20,22,28,30,32,36	4	9	20%
Sedang (2)	4,7,8,10,16,18,23,24,26,27,33,34,37,39	1,3	15	33%
Tidak Baik (1)	1,3,9,17,25,29	5	8	18%
Sangat tidak baik (0)	2,5,6,11,12,13,14,19,21,31,35,38,40		13	29%

Berdasarkan hasil uji menggunakan aplikasi ITEMAN versi 4.3 pada soal Ujian Akhir Semester Ganjil mata Pelajaran PPKn kelas VII SMP Negeri 2 Pemangkat mengindikasikan yakni dari totalitas 45 item soal yang dianalisis, sejumlah 9 item soal (20 %) dikategorikan memiliki kualitas baik meliputi nomor item 15,20,22,28,30,32,36 (PG) dan 1,3 (uraian).

Selanjutnya, 15 item soal (33%) berkategori sedang yaitu nomor item 4,7,8,10,16,18,23,24,26,27,33,34,37,39 (PG) dan 2 (uraian). Adapun 8 item soal (18%) berkategori kurang baik, dengan nomor item 1,3,9,17,25,29 (PG) dan 4,5 (uraian). Sisanya, sebanyak 13 item soal atau 29% terklasifikasi sangat tidak baik, mencakup nomor item 2,5,6,11,12,13,14,19,21,31,35,38,40 (PG).

Item soal yang tergolong memiliki kualitas “baik” dan sedang masih perlu untuk dilakukan revisi atau perbaikan sebelum nantinya akan dimasukkan ke dalam bank soal. Perbaikan soal disesuaikan terlebih dahulu dengan indikator keagalannya. Sementara itu, item soal yang masuk klasifikasi “Tidak baik” dan “sangat tidak baik” sebaiknya dieliminasi mengingat item soal tersebut membutuhkan perbaikan secara menyeluruh.

Merujuk pada hasil yang diperoleh dari analisis data, kesimpulan yang didapat yakni soal pada Ujian Akhir Semester ganjil mata Pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat menunjukkan pada soal pilihan ganda sebanyak 21 item soal butuh dikoreksi ataupun direvisi, 19 item soal lainnya sebaiknya dibuang atau tidak digunakan lagi. Sementara itu, untuk kategori soal uraian, terdapat 3 item soal yang perlu ditingkatkan dengan perbaikan atau revisi, dan 2 item soal yang tidak layak digunakan lagi. Faktor-faktor yang menyebabkan kegagalan item soal bisa dikenali dari penelaahan terhadap aspek validitas, tingkat kesukaran, dan daya pembeda pada setiap instrument tes.

SIMPULAN

Dari temuan riset dan pembahasan analisis item soal berbasis level kognitif dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda pada soal Ujian Akhir Semester Ganjil mata pelajaran PPKn kelas VII di SMP Negeri 2 Pemangkat Tahun Ajaran 2024/2025, diperoleh kesimpulan bahwa soal pilihan ganda dan uraian secara keseluruhan berkualitas cukup baik meskipun masih didominasi oleh soal level LOTS serta kelemahan pada kemampuan soal membedakan kemampuan peserta didik. Instrumen Ujian Akhir Semester PPKn di SMP Negeri 2 Pemangkat memiliki dua aspek yang masih lemah yakni segi validitas dan daya pembeda. Banyaknya soal yang tidak valid sebesar 40% menunjukkan instrumen memiliki kegagalan fungsi ukur sedangkan dari segi daya pembeda mayoritas soal 47% gagal membedakan antara siswa yang berkemampuan tinggi dan rendah. Sedangkan dari segi reliabilitas soal memiliki konsistensi yang tinggi artinya instrumen soal cukup konsisten dan dapat diandalkan. Kemudian sebaran tingkat kesulitan soal yang proporsional yakni 31% terklasifikasi sukar, 40% terklasifikasi sedang, dan 29% terklasifikasi mudah. Keterbatasan dalam penelitian ini hanya meneliti satu sekolah sehingga membatasi generalisasi dan juga tidak ada validasi antar-penilai dalam mengkategorisasikan level kognitif Taksonomi Bloom.

DAFTAR PUSTAKA

- Ambarwati, Y. F., & Ismiyati, I. (2022). Analisis Item Soal Pilihan Ganda Ulangan Akhir Semester Genap Mata Pelajaran Kearsipan. *Measurement In Educational Research (Meter)*, 1(2), 64–75. <https://doi.org/10.33292/meter.v1i2.144>
- Arikunto, S. (2021). *Dasar-Dasar Evaluasi Pendidikan Edisi 3*. Bumi Aksara.
- Burwash, S. C., Snover, R., & Krueger, R. (2016). Up Bloom's pyramid with slices of Fink's pie: Mapping an occupational therapy curriculum. *The Open Journal of Occupational Therapy*, 4(4). <https://doi.org/10.15453/2168-6408.1235>
- Febrila, L. G., Yensy N.A., Susanta, A. (2021). Analisis Tingkat Kognitif dan Kualitas Soal Penilaian Akhir Semester Ganjil Kelas VII SMP IT Darul Fikri Arga Makmur Tahun Ajaran 2020/2021. *Jurnal Penelitian Pembelajaran Matematika Sekolah*, 5, 281–195. <https://doi.org/10.33369/jp2ms.v.ii.281-295>
- Hartono, I. D. I., Tenriawaru, A. B., & Ningsih, K. (2024). Analisis Item Soal Penilaian Sumatif IPA Menggunakan Anates. *Jurnal Kajian Pembelajaran dan Keilmuan*, 8, 162–171. <https://doi.org/10.26418/jurnalkpk.v8i2.78282>
- Heng, C. S. (2015). Framework of Assessment for the Evaluation of Thinking Skills of Tertiary Level Students. *Advances in Language and Literary Studies*, 6(5). <https://doi.org/10.7575/aiac.all.v.6n.5p.67>
- Helmawati. (2019). *Pembelajaran dan Penilaian Berbasis HOTS*. Bandung: PT. Remaja Rosdakarya.
- Kashyap, S. (2015). Item Analysis Of Multiple Choice Questions. *International Journal of Current Research*, 7(12). <http://www.journalcra.com>
- Linnette D'sar, J., & Liza Visbal-Dionaldo2, M. (2017). Analysis of Multiple-Choice Questions: Item Difficulty, Discrimination Index and Distractor Efficiency. *International Journal of Nursing Education*, 9(3).
- Muchlizani, N., Mania, S., & Rasyid, M. N. A. (2023). Analisis Kualitas Item Soal Ujian Akhir Semester Mata Pelajaran Akidah Akhlak Kelas V MI Radhiatul Adawiyah Makassar. XII (1), 224–240.
- Muluki, A., Bundu, P., & Sukmawati. (2020). Analisis Kualitas Item Tes Semester Ganjil Mata Pelajaran IPA Kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86–96.
- Mulyani, S., Krismonita, M., & Yamtinah, S. (2022). Analisis Item Soal dan Kecukupan HOTS Soal Ujian Akhir Semester Mata Pelajaran Kimia SMK Kelas X. *PAEDAGOGIA*, 25(2), 162. <https://doi.org/10.20961/paedagogia.v25i2.60913>
- OECD. (2022). PISA 2022 Results Factsheets Indonesia PUBE. *OECD Publishing*. Diambil dari <https://oecdch.art/a40de1dbaf/C108>.
- Putriani, D., Turahmah, F., Sunarti, N.R., Ismarliana, P., & Walid, A. (2020). Analisis Item Soal UAS Biologi 2018/2019 Kelas X dan XI SMAN 11 Kota Bengkulu. *Journal of Biology Learning*, 2(1), 1-7. <https://doi.org/10.32585/.v2i1.559>
- Rahmadhani. (2022). Analisis Level Soal dan Level Kognitif Soal Ujian Akhir Semester Mata Pelajaran Fisika di SMAS Babul Maghfirah. Universitas Islam Negeri Ar-Raniry Darussalam, Banda Aceh.
- Rahman, A. A., & Nasryah, E. C. (2019). *Evaluasi Pembelajaran*. Meulaboh: Uwais Inspirasi Indonesia.

- Ramadhani, S. (2022). Analisis Item Soal HOTS (Higher Order Thingking Skills) Mata Pelajaran PAI di SMA Negeri 1 Peukan Aceh Besar. Universitas Islam Negeri Ar-Raniry, Banda Aceh.
- Riinawati. (2021). *Pengantar Evaluasi Pendidikan*. Banjarmasin : Thema Publishing.
- Shalikhah, N. D., Purnanto, A. W., & Nugroho, I. (2021). Soal Higher Order Thingking Skills (HOTS) Matematika Pada Buku Tematik Terpadu Kurikulum 2013. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(2), 701. <https://doi.org/10.24127/ajpm.v10i2.3442>
- Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dn R & D (2 ed). Bandung: Alfabeta.
- Widodo, H. (2021). *Evaluasi Pendidikan*. Yogyakarta: UAD PRESS.
- Yadnyawati, I. A.D (2019). *Evaluasi Pembelajaran*. Denpasar. UNHI Press
- Zahrotunnaqiyah E, Nulhakim, M., & Taufik, A. N. (2023). Analisis Higher Order Thinking Skills (HOTS) Pada Soal Penilaian Akhir Semester Mata Pelajaran IPA Kelas VII Di Kabupaten Serang. *Jurnal Pendidikan MIPA*, 13(1), 67–72. <https://doi.org/10.37630/jpm.v13i1.810>