

# Penggunaan Aplikasi Iteman Dalam Menganalisis Butir Soal Ulangan Akhir Semester Gasal TA 2023/2024 PPKn Kelas X Di SMAN 1 Prambanan Sleman

Afiyanto Nugroho <sup>a,1</sup>, Yayuk Hidayah <sup>b,2</sup>

<sup>1</sup> ([afiyantonugroho.2021@student.uny.ac.id](mailto:afiyantonugroho.2021@student.uny.ac.id)) Mahasiswa Departemen PKNH, Fishipol, UNY

<sup>2</sup> ([yayukhidayah@uny.ac.id](mailto:yayukhidayah@uny.ac.id)) Dosen Fakultas Ilmu Sosial, Hukum, dan Politik, UNY

<sup>a</sup> Mahasiswa (Departemen Pendidikan Kewarganegaraan dan Hukum FISHIPOL UNY), Yogyakarta Indonesia

<sup>b</sup> Dosen (Departemen Pendidikan Kewarganegaraan dan Hukum FISHIPOL UNY), Yogyakarta Indonesia

## ABSTRAK

Penelitian ini bertujuan untuk mengetahui dan mendeskripsikan kualitas butir soal Ulangan Akhir Semester Gasal Tahun Ajaran 2023/2024 Mata Pelajaran Pendidikan Kewarganegaraan Kelas X di SMA Negeri 1 Prambanan yang ditinjau dari validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Penelitian ini merupakan jenis penelitian evaluasi dengan pendekatan kuantitatif. Subjek penelitian ini adalah peserta didik kelas X di SMAN 1 Prambanan Sleman. Teknik pengambilan sampel menggunakan teknik purposive sampling dengan menggunakan rumus slovin sehingga diperoleh sampel sebanyak 72 peserta didik. Data diperoleh menggunakan metode dokumentasi yang berupa butir soal UAS, lembar jawab siswa, kunci jawaban, kisi-kisi soal. Data yang diperoleh dianalisis menggunakan aplikasi ITEMAN versi 4.3. Hasil penelitian menunjukkan bahwa: (1) Berdasarkan validitas berjumlah 32 soal atau 64% tergolong valid dan 18 soal atau 36% tergolong tidak valid (2) Berdasarkan reliabilitas, sebesar 0,733 yang tergolong tinggi. (3) Berdasarkan daya pembeda, berjumlah 5 soal atau 10% daya pembeda sangat jelek (negatif), 20 soal atau 40% daya pembeda jelek, 21 soal atau 42% daya pembeda cukup, 4 soal atau 8% daya pembeda baik. (4) Berdasarkan tingkat kesukaran, berjumlah 5 soal atau 10% tergolong sulit, 16 soal atau 32% tergolong sedang, 29 soal atau 58% tergolong mudah. (5) Berdasarkan efektivitas pengecoh, terdapat 9 soal atau 18% tergolong sangat baik, 11 soal atau 22% tergolong baik, 9 soal atau 18% tergolong cukup, 8 soal atau 16% tergolong jelek, dan 13 soal atau 26% tergolong sangat jelek. (6) Hasil keseluruhan analisis butir soal, berjumlah 9 soal atau 18% berkualitas sangat baik, 6 soal atau 12% berkualitas baik, 19 soal atau 38% berkualitas sedang, 10 soal atau 20% berkualitas tidak baik, 6 soal atau 12% berkualitas sangat tidak baik.

## ABSTRACT

*This study aims to determine and describe the quality of the items of the Final Examination of the Odd Semester of the 2023/2024 Academic Year of Class X Civics Education Subjects at SMA Negeri 1 Prambanan in terms of validity, reliability, differentiating power, difficulty level, and effectiveness of triggers. This research is a type of evaluation research with a quantitative approach. The subjects of this research were students of class X at SMAN 1 Prambanan Sleman. The sampling technique used purposive sampling technique using the slovin formula so that a sample of 72 students was obtained. Data obtained using the documentation method in the form of UAS items, student answer sheets, answer keys, question grids. The data obtained were analyzed using the ITEMAN version 4.3 application. The results showed that: (1) Based on validity, 32 questions or 64% are classified as valid and 18 questions or 36% are classified as invalid (2) Based on reliability, it is 0.733 which is classified as high. (3) Based on differentiating power, there are 5 questions or 10% very poor differentiating power (negative), 20 questions or 40% poor differentiating power, 21 questions or 42% sufficient differentiating power, 4 questions or 8% good differentiating power. (4) Based on the level of difficulty, there are 5 questions or 10% classified as difficult, 16 questions or 32% classified as medium, 29 questions or 58% classified as easy. (5) Based on the effectiveness of the checkers, there are 9 questions or 18% classified as very good, 11 questions or 22% classified as good, 9 questions or 18% classified as sufficient, 8 questions or 16% classified as bad, and 13 questions or 26% classified as very bad. (6) The overall results of item analysis, totaling 9 questions or 18% of very good quality, 6 questions or 12% of good quality, 19 questions or 38% of moderate quality, 10 questions or 20% of poor quality, 6 questions or 12% of very poor quality.*

## Sejarah Artikel

Diterima : 02-07-2025

Disetujui: 05-07-2025

## Kata kunci:

Analisis Butir Soal, PPKn, Iteman 4.3, SMAN 1 Prambanan

## Keywords:

Item Analysis, Civics Education, Iteman 4.3, SMAN 1 Prambanan

## Pendahuluan

Pendidikan adalah upaya sadar dan terencana untuk menciptakan lingkungan dan proses pembelajaran di mana siswa secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, dan keterampilan yang diperlukan untuk diri mereka sendiri, masyarakat, bangsa, dan negara. Tujuan dari pendidikan saat ini diharapkan dapat menghasilkan siswa yang mampu berkolaborasi dan berkomunikasi dengan baik, mahir menggunakan teknologi, berpikiran inovatif dan kreatif, dan mampu menyelesaikan masalah. Sarker & Ullah (2023), menyatakan bahwa cara yang paling efektif untuk mengukur kualitas pendidikan adalah dengan melihat dimensi proses seperti pembelajaran, konten mata pelajaran, keahlian guru, ketepatan konten kurikulum, media pembelajaran, dan penilaian siswa.

Sejalan dengan pendapat tersebut Gladushyna & Strietholt (2023), menyatakan bahwa metode yang paling populer untuk menilai kualitas pendidikan dengan mengukur pengetahuan dan keterampilan siswa menggunakan evaluasi pembelajaran. Dengan adanya penilaian atau evaluasi dapat memberikan dampak positif bagi keberhasilan peserta didik untuk dijadikan salah satu tolak ukur dalam rangkaian proses pembelajaran. Selain itu, penilaian menggunakan tes dapat dijadikan sebagai umpan balik atau feedback terhadap hasil belajar siswa guna memperbaiki kinerja pada saat merencanakan pembelajaran selanjutnya daripada hanya digunakan sebagai alat untuk memeriksa standar akhir pembelajaran (López-Hernández et al., 2023). Menurut Oktanin & Sukirno (2015), evaluasi pembelajaran merupakan salah satu langkah atau prosedur penting yang harus dilakukan oleh para pendidik untuk menilai seberapa baik siswa belajar.

Salah satu bentuk evaluasi adalah penyelenggaraan Ulangan Akhir Semester (UAS) yang bertujuan untuk mengukur pemahaman siswa terhadap materi yang telah diajarkan selama satu semester. Dalam konteks PPKn, soal-soal evaluasi harus dirancang untuk mengukur tidak hanya aspek kognitif siswa, tetapi juga aspek afektif dan psikomotor. Mata pelajaran Pendidikan Pancasila dan Kewarganegaraan (PPKn) di tingkat Sekolah Menengah Atas (SMA) memiliki peran strategis dalam membentuk karakter dan kompetensi kewarganegaraan siswa. Evaluasi pembelajaran dalam PPKn tidak hanya bertujuan untuk mengukur sejauh mana siswa memahami konsep-konsep kewarganegaraan, tetapi juga untuk menilai kemampuan mereka dalam menerapkan nilai-nilai tersebut dalam kehidupan sehari-hari. Oleh karena itu, kualitas soal-soal yang digunakan dalam evaluasi, seperti Ulangan Akhir Semester (UAS) sangat menentukan efektivitas mata pelajaran PPKn dalam mencapai tujuannya. Menurut Arikunto (2018), tes yang baik yaitu tes yang memiliki validitas, reliabilitas, objektivitas, praktikalitas, dan ekonomis.

Ketika sebuah tes digunakan untuk mencapai tujuan tertentu dan dapat memberikan hasil yang akurat, maka tes tersebut dianggap valid. Sebuah tes dianggap reliabel, jika diberikan kepada kelompok yang sama dalam beberapa kesempatan dan hasilnya menunjukkan konsisten (Lafrarchi, 2020). Sebuah tes dianggap objektif jika tidak dipengaruhi oleh faktor subjektif. Hal ini terutama berlaku untuk sistem yang menilai tes. Tes yang praktis adalah tes yang mudah diatur, dikelola, digunakan, dan diproses. Tes dianggap ekonomis jika tidak membutuhkan banyak biaya, waktu, dan tenaga. Hal ini berarti, guru harus menggunakan inovasi penilaian supaya memperoleh soal atau instrumen tes yang bermutu dengan menganalisis butir soal (Hardiana & Suyata, 2018). Untuk mendapatkan instrumen tes yang bermutu dapat diketahui melalui analisis butir soal. Analisis butir soal merupakan penelaahan soal ujian untuk menghasilkan perangkat soal yang berkualitas tinggi (Himawan & Suyata, 2022). Dari analisis butir soal ini dapat bertujuan untuk mengidentifikasi butir soal mana yang dapat masuk ke dalam bank soal, direvisi atau dibuang. Analisis butir soal ini dapat dihitung melalui beberapa aspek yaitu Validitas, Reliabilitas, Daya Pembeda, Tingkat Kesukaran, dan Efektivitas Pengecoh.

Berkaitan dengan analisis butir soal, Mardapi & Kartowagiran (2013), berpendapat bahwa dalam mengembangkan instrumen tes harus melalui beberapa tahapan yaitu merancang tes, melakukan uji coba tes, validitas, reliabilitas, daya pembeda, tingkat kesulitan, efektivitas pengecoh dan yang terakhir menginterpretasikan skor tes. Oleh karena itu, guru harus mampu menyusun tes dengan baik dan menganalisisnya terlebih dahulu sesuai dengan kompetensi dasar, sandar kompetensi, indikator dan tujuan pembelajaran. Analisis ini dilakukan setelah soal diberikan kepada peserta didik dan dijawab oleh peserta didik yang kemudian dianalisis untuk menilai kualitas butir soal. Dengan demikian, proses belajar yang dilaksanakan akan relevan dalam mengukur pemahaman siswa sesuai dengan tujuan pembelajaran.

Berdasarkan informasi yang diperoleh dari hasil pra observasi yang telah penulis lakukan pada tanggal 2 Mei 2024, yang berada di SMA N 1 Prambanan Sleman menunjukkan bahwa selama ini, guru PPKn belum pernah melakukan analisis butir soal terhadap soal yang diujikan kepada peserta didik. Guru PPKn di SMA N 1 Prambanan hanya membuat soal yang kemudian langsung diujikan kepada peserta didik sehingga belum ada langkah lanjutan mengenai analisis butir secara klasik menggunakan program komputer ITEMAN. Selain itu, fenomena di lapangan juga menunjukkan bahwa kebanyakan guru dalam membuat soal sering kali terlalu sulit dan terlalu mudah, padahal soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sulit. Hal tersebut disebabkan karena jam mengajar guru sangat padat sehingga kurangnya waktu untuk melakukan analisis butir soal, kurangnya pengetahuan dan pemahaman guru tentang analisis butir soal apalagi analisis butir soal yang ditinjau dari taraf kesukaran, daya pembeda, dan efektivitas distractor, masih sedikit guru yang melakukan analisis butir soal dengan bantuan komputer seperti Iteman, bigstep, dan lain-lain. Oleh karena itu, guru dituntut untuk melakukan analisis butir soal, karena analisis terhadap butir soal UAS sangatlah penting dilakukan untuk memperbaiki kualitas soal dan peningkatan mutu soal yang nantinya dapat menghasilkan butir-butir soal yang diketahui karakteristiknya (Matthews et al., 2022).

Dalam penelitian ini, untuk melakukan analisis butir soal menggunakan program ITEMAN. Menurut Yot-Domínguez & Marcelo (2017), program ITEMAN adalah salah satu teknologi digital yang dapat digunakan sebagai sarana untuk menunjang aktivitas pembelajaran dalam lingkup penilaian. Program ini dapat digunakan untuk menganalisis data file jawaban butir soal yang dihasilkan melalui manual entry data, menskor dan menganalisis data soal pilihan ganda dan skala Likert. ITEMAN merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik dimana dapat memberikan informasi tentang validitas setiap butir, standar error of measurement, mean, variance, standar deviasi, skew, kurtosis untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi skor. Berkaitan dengan penelitian mengenai analisis butir soal, ada beberapa penelitian relevan dan lebih dahulu dilakukan oleh (Komariah & Ruffi'i Ruffi'i, 2023), dengan penelitiannya yang berjudul *Analysis Of Multiple-Choice Questions In Class VII Mathematics At SMP Negeri 6 Bangkalan*. Hasil penelitian menunjukkan bahwa terdapat 20 soal berkategori valid, memiliki reliabilitas tinggi dengan skor 0,937, dan 20 soal memiliki daya pembeda yang baik. Persamaan penelitian ini, yaitu sama-sama menganalisis butir soal pilihan ganda. Perbedaannya, jika penelitian ini mengkaji soal matematika dengan SPSS 25.0 sementara dalam penelitian ini mengkaji soal PPKn dengan Iteman 4.3.

Kedua, penelitian Himawan & Nurgiyantoro (2022), yang berjudul *Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN*. Hasilnya menunjukkan seluruh soal yang dianalisis valid, memiliki reliabilitas yang sangat tinggi dengan skor Alpha 0.820, berdasarkan daya pembeda terdapat 20 soal layak untuk diujikan, 4 butir soal dengan kategori sedang, 8 soal dengan kategori cukup, dan 8 soal dengan kategori gugur atau tidak layak untuk diujikan, berdasarkan tingkat kesukaran terdapat 6 soal dengan kategori sangat mudah, 17 soal dengan kategori mudah, 8 butir soal dengan kategori

sedang, 9 soal dengan kategori sukar atau tinggi, berdasarkan efektivitas pengecoh terdapat 31 butir soal dengan pengecoh baik, 9 soal dengan pengecoh yang tidak berfungsi. Ketiga, penelitian yang dilakukan oleh Pradani & Efendi (2023), yang berjudul Analisis Butir Soal Ujian Sekolah Menggunakan Program IteMan (Analysis of School Exam Questions Using the IteMan Program). Hasilnya menunjukkan bahwa berdasarkan tingkat kesukaran terdapat 7 soal berkategori sulit, 20 soal berkategori sedang, dan 13 soal berkategori mudah, berdasarkan Indeks Daya Pembeda terdapat 4 soal berkategori sangat buruk, 7 soal berkategori buruk, 13 soal berkategori cukup, 14 soal berkategori baik, dan 2 butir soal berkategori sangat baik.

Berdasarkan berbagai penelitian yang sudah dilakukan, maka analisis butir soal ini sangat penting dilakukan untuk memberikan informasi yang akurat bagi guru dan pengambil kebijakan pendidikan dalam meningkatkan kualitas pendidikan di tingkat sekolah menengah atas dan memperbaiki kurikulum yang digunakan di sekolah tersebut. Menurut Rohmatdi et al., (2024), dengan melakukan analisis butir soal menggunakan aplikasi IteMan 4.3, guru dapat menentukan bahwa tes yang diberikan tidak hanya sekedar mengukur pemahaman siswa secara tepat, tetapi juga sebagai upaya perbaikan pembelajaran dan pengembangan dalam penyusunan soal di tes yang akan datang. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan pada evaluasi sumatif serta pendekatan terhadap dimensi berpikir tingkat tinggi (HOTS) yang menjadikan penelitian ini relevan untuk meningkatkan kualitas evaluasi pembelajaran PPKn secara komprehensif melalui analisis butir soal.

Berdasarkan uraian tersebut, maka penelitian ini berfokus pada analisis butir soal dalam pembelajaran Pendidikan Pancasila dan Kewarganegaraan dengan menggunakan program ITEMAN 4.3. Hal ini dikarenakan mengenai analisis butir soal di tingkat Sekolah Menengah Atas (SMA) menggunakan program IteMan masih jarang dilakukan sehingga perlu dilakukan penelitian yang memenuhi kebutuhan referensi, serta memunculkan penelitian yang relevan. Penelitian ini menghadirkan kebaruan, yaitu menyajikan sebuah analisis butir soal di jenjang Sekolah Menengah Atas (SMA) dengan memanfaatkan aplikasi iteman 4.3. Dengan menambahkan analisis berdasarkan validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh menjadikan pelengkap serta penyempurna pada penelitian sebelumnya. Melihat latar belakang permasalahan serta penelitian terdahulu betapa pentingnya evaluasi melalui analisis butir soal, menjadikan peneliti tertarik untuk melakukan penelitian yang berjudul : Penggunaan Aplikasi IteMan Dalam Menganalisis Butir Soal Ulangan Akhir Semester Gasal TA 2023/2024 PPKn Kelas X Di SMAN 1 Prambanan Sleman. Dengan adanya penelitian ini bertujuan untuk memberikan informasi bagi guru PPKn dalam meningkatkan kualitas butir soal yang layak dan tidak layak digunakan serta upaya perbaikan mutu soal yang nantinya dapat menghasilkan butir soal yang diketahui karakteristiknya sehingga mutu pendidikan di SMAN 1 Prambanan juga dapat meningkat sesuai dengan tujuan yang ditetapkan.

## Metode

Penelitian ini merupakan jenis penelitian evaluasi dengan menggunakan pendekatan kuantitatif. Evaluasi dalam penelitian ini dilakukan untuk menganalisis kualitas butir soal Ulangan Akhir Semester (UAS) mata pelajaran PPKn Kelas X di SMA N 1 Prambanan Sleman menggunakan aplikasi ITEMAN versi 4.3. Populasi yang diteliti adalah seluruh siswa kelas X di SMA Negeri 1 Prambanan, Kabupaten Sleman yang terdiri dari tujuh kelas, yaitu tiga kelas IPS (X IPS 1, 2, 3) dan empat kelas IPA (X IPA 1, 2, 3, 4), dengan total jumlah siswa sebanyak 252 orang. Pengambilan sampel dalam penelitian ini dilakukan menggunakan metode Non-probability Sampling dengan teknik Purposive Sampling. Untuk menentukan jumlah sampel, peneliti menggunakan rumus teknik Slovin dengan tingkat kesalahan 10% atau 0,1 sehingga diperoleh sampel sebanyak 72 siswa. Sedangkan objek penelitian ini adalah soal ulangan akhir semester gasal mata pelajaran PPKn untuk kelas X yang berjumlah 50 butir soal pilihan ganda dan disusun oleh Guru Mata Pelajaran PPKn SMA N 1 Prambanan, Kabupaten Sleman. Teknik pengumpulan data melalui dokumentasi yang meliputi kisi-

kisi soal, naskah soal UAS mata pelajaran PPKn kelas X semester gasal tahun ajaran 2023/2024, kunci jawaban soal, lembar jawab siswa kelas X dan daftar nama siswa. Analisis data yang dilakukan terhadap butir soal Ulangan Akhir Semester Gasal mata pelajaran PPKn menggunakan Iteman 4.3 dengan mencari validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh.

### Hasil dan Pembahasan

Hasil yang diperoleh dari analisis butir soal Ulangan Akhir Semester Gasal Tahun Ajaran 2023/2024 Mata Pelajaran Pendidikan Kewarganegaraan Kelas X yang terdiri dari 50 butir soal dan diujikan kepada 72 siswa, didapatkan karakteristik butir soal secara keseluruhan sebagai berikut:

Dalam menguji validitas item, peneliti menggunakan rumus koefisien korelasi point biserial ( $Y_{pbi}$ ) dengan bantuan program ITEMAN versi 4.3. Hasil dari perhitungan tersebut kemudian dikonsultasikan dengan  $r_{tabel}$  pada taraf signifikan 5% dan disesuaikan dengan banyaknya peserta didik yang mengikuti tes. Jumlah peserta didik yang mengikuti tes sebanyak 72 siswa, jadi nilai standar  $r_{tabel}$  adalah sebesar 0,235. Apabila ( $Y_{pbi}$ ) >  $r_{tabel}$ , maka butir soal dinyatakan valid. Sebaliknya, apabila ( $Y_{pbi}$ ) < ( $r_{tabel}$ ), maka butir soal dinyatakan tidak valid.

Tabel 1

Ringkasan Statistik Hasil Analisis Validitas Butir Soal Iteman 4.3

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	50	34,375	5,095	12	43	0,687	0,198

Sumber : Data Primer Diolah, 2025

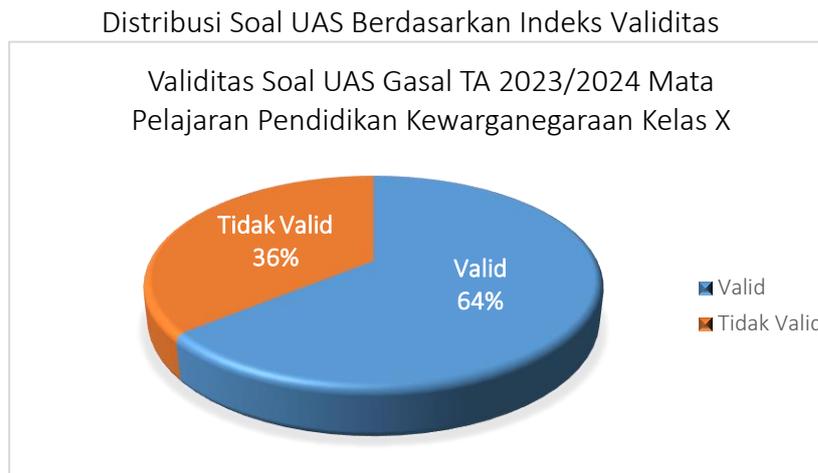
Berdasarkan tabel 1 menunjukkan bahwa hasil nilai rata-rata korelasi point biserial ( $r_{pbis}$ ) yaitu 0,198. Hal tersebut menunjukkan bahwa secara keseluruhan butir soal memiliki validitas yang sangat rendah. Dikarenakan standar nilai  $r_{tabel}$  yaitu 0,235 sebagaimana rumus korelasi point biserial  $0,198 < 0,235$ , maka butir soal dinyatakan tidak valid.

Tabel 2

Distribusi Soal UAS Berdasarkan Indeks Validitas

No.	Validitas	No. Butir Soal	Jumlah	Persentase
1.	> 0,235 (Valid)	3, 5, 9, 10, 11, 12, 15, 16, 17, 18, 20, 22, 23, 25, 26, 28, 29, 30, 34, 35, 36, 37, 39, 40, 42, 43, 44, 45, 46, 48, 49, 50	32	64%
2.	< 0,235 (Tidak Valid)	1, 2, 4, 6, 7, 8, 13, 14, 19, 21, 24, 27, 31, 32, 33, 38, 41, 47	18	36%

Gambar 1 Sumber : Data Primer Diolah, 2025



Sumber : Data Primer Diolah, 2025

Sedangkan berdasarkan tabel 2 diagram tersebut menunjukkan bahwa hasil analisis validitas butir soal Ulangan Akhir Semester Gasal Tahun Ajaran 2023/2024 Mata Pelajaran Pendidikan Kewarganegaraan Kelas X berdasarkan Item per soal terdapat 32 butir soal atau 64% dinyatakan valid dan 18 butir soal atau 36% dinyatakan tidak valid. Hasil butir soal yang valid ini mengindikasikan bahwa sebagian besar butir soal dapat mengukur apa yang hendak diukur, sebagaimana dengan teori (Sudijono, 2013). Hasil butir soal yang valid sebesar 64 % dapat menjalankan fungsinya dengan baik sehingga dapat dijadikan alat ukur pemahaman siswa dalam menguasai materi pembelajaran. Di samping itu, untuk butir soal yang tidak valid sebesar 36% dapat disebabkan oleh beberapa faktor, yakni faktor dari jawaban siswa, faktor instrumen evaluasi, faktor administrasi evaluasi dan penskoran (Arifin, 2014). maka soal tersebut perlu dibuang atau direvisi terlebih dahulu dengan menyesuaikan indikator pencapaian kompetensi dalam silabus dan meningkatkan penguasaan teknik dalam menyusun butir soal. Berikutnya adalah hasil analisis reliabilitas butir soal menggunakan iteman 4.3.

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	0,733	2,632	0,619	0,608	0,510	0,765	0,756	0,675

Sumber : Data Primer Diolah, 2025

Berdasarkan tabel 3 menunjukkan hasil nilai reliabilitas Alpha (KR-20) yaitu 0,733 dengan standar error pengukuran sebesar 2,632. Artinya, Soal Ulangan Akhir Semester Gasal Tahun Ajaran 2023/2024 Mata Pelajaran Pendidikan Kewarganegaraan Kelas X memiliki nilai indeks reliabilitas yang tinggi. Nilai  $r_{11}$  yang didapat yaitu  $0,733 > 0,70$ , maka secara keseluruhan item soal tersebut dinyatakan reliabel. Sebagaimana teori yang ungkapkan oleh Azwar (2016), butir soal yang memiliki reliabilitas yang tinggi dapat memberikan hasil yang serupa saat diujikan kepada kelompok yang sama

dan waktu yang berbeda. Gagasan tersebut diperkuat oleh Sudijono (2013), yang menyatakan bahwa butir soal yang memiliki reliabilitas tinggi dapat menghasilkan data yang relatif sama walaupun diujikan dalam kesempatan yang berbeda.

Selanjutnya adalah hasil analisis daya pembeda butir soal menggunakan iteman 4.3. Hasil dari analisis daya pembeda dapat dilihat pada tabel 1. Diketahui bahwa nilai Mean Rpbis sebesar 0,198. Hal ini menunjukkan bahwa secara keseluruhan butir soal ini memiliki daya pembeda yang jelek sebagaimana  $0,198 < 0,20$ .

Tabel 3  
Distribusi Soal UAS Berdasarkan Indeks Daya Pembeda

No.	Daya Pembeda	Butir Soal	Jumlah	Persentase
1.	Tanda Negatif	7, 13, 21, 32, 38	5	10%
2.	< 0,20 (Jelek)	1, 2, 3, 4, 5, 6, 8, 9, 11, 14, 15, 19, 22, 24, 27, 31, 33, 41, 47, 49,	20	40%
3.	0,20 – 0,39 (Cukup)	10, 12, 16, 17, 20, 23, 26, 28, 29, 34, 36, 37, 39, 40, 42, 43, 44, 45, 46, 48, 50	21	42%
4.	0,40 – 0,69 (Baik)	18, 25, 30, 35	4	8%
5.	0,70 – 1,00 (Sangat Baik)	–	0	0%

Sumber : Data Primer Diolah, 2025

Gambar 2  
Distribusi Soal UAS Berdasarkan Indeks Daya Pembeda



Sumber : Data Primer Diolah, 2025

Berdasarkan tabel 3 dan gambar 2 tersebut menunjukkan hasil analisis daya pembeda menggunakan program iteman versi 4.3, dari 50 butir soal yang diujikan kepada 72 siswa, dapat diketahui bahwa terdapat 5 butir soal atau 10% memiliki daya pembeda yang sangat jelek (negatif), 20 butir soal atau 40% memiliki daya pembeda yang jelek, 21 butir soal atau 42% memiliki daya pembeda yang cukup, 4 butir soal atau 8% memiliki daya pembeda yang baik. Indeks daya pembeda yang positif (baik) menunjukkan bahwa soal tersebut efektif dalam membedakan siswa yang telah memahami materi dengan siswa yang belum memahami materi. Sebaliknya, Indeks daya pembeda negatif atau jelek menunjukkan bahwa kurangnya kemampuan item dalam membedakan antara siswa yang memiliki kemampuan tinggi dan rendah sehingga tidak dapat digunakan dalam evaluasi. Sebagaimana teori Sudijono (2013), menyatakan bahwa hasil uji analisis daya pembeda ini sangat penting dilakukan sebab dalam menyusun butir soal harus mampu memberikan perbedaan kemampuan diantara peserta tes.

Hasil analisis selanjutnya yaitu tingkat kesukaran. Berdasarkan tabel 1 hasil analisis tingkat kesulitan butir soal menggunakan aplikasi Iteman 4.3, diketahui bahwa nilai proportional correct (Mean P) sebesar 0,687. Hal ini menunjukkan bahwa secara keseluruhan butir soal tersebut memiliki tingkat kesulitan yang tergolong sedang.

Tabel 4

Distribusi Soal UAS Berdasarkan Indeks Tingkat Kesukaran

No.	Tingkat Kesukaran	Butir Soal	Jumlah	Persentase
1.	0,00 – 0,30 (Sulit)	1, 4, 7, 21, 38	5	10%
2.	0,31 – 0,70 (Sedang)	6, 8, 12, 13, 14, 16, 20, 27, 34, 40, 42, 43, 44, 46, 47, 50	16	32%
3.	0,71 – 1,00 (Mudah)	2, 3, 5, 9, 10, 11, 15, 17, 18, 19, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 35, 36, 37, 39, 41, 45, 48, 49	29	58%

Sumber : Data Primer Diolah, 2025

Gambar 3

Distribusi Soal UAS Berdasarkan Indeks Tingkat Kesukaran



Sumber : Data Primer Diolah, 2025

Sedangkan analisis tingkat kesulitan berdasarkan item per masing-masing soal, terdapat 5 butir soal atau 10% termasuk dalam kriteria sulit, 16 butir soal atau 32% tergolong sedang, dan 29 butir soal atau 58% termasuk dalam kategori mudah. Butir soal dalam kategori sedang dinyatakan sebagai butir soal yang baik sebab soal tersebut tidak terlalu sukar dan tidak terlalu mudah. Hal ini sesuai dengan pendapat Sudjana (2019), menyatakan bahwa butir soal yang baik adalah butir soal yang tidak terlalu mudah dan tidak terlalu sulit, yang mana butir soal tersebut berada pada derajat kesukaran sedang. Tingkat kesukaran sedang dianggap ideal karena dapat menggambarkan informasi yang akurat tentang kemampuan peserta didik. Instrumen soal yang mudah dapat membuat siswa menyepelekan soal karena tidak terangsang untuk memecahkan permasalahan di dalam soal. Sebaliknya instrumen soal yang sulit membuat siswa merasa putus asa atau mudah menyerah karena di luar batas kemampuannya.

Hasil analisis selanjutnya yaitu efektivitas pengecoh. Hasil analisis keefektifan pola sebaran jawaban menggunakan bantuan Iteman versi 4.3 dapat dilihat pada tabel sebagai berikut.

Tabel 5

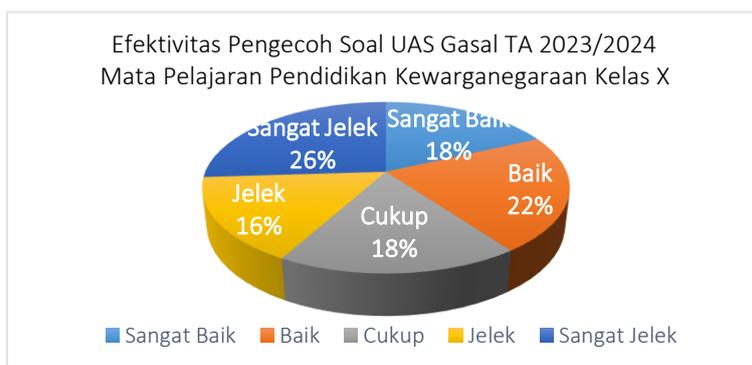
Distribusi Soal UAS Berdasarkan Indeks Efektivitas Pengecoh

No.	Efektivitas Pengecoh	Butir Soal	Jumlah	Persentase
1.	Sangat Baik	4, 8, 20, 21, 27, 34, 38, 40, 46	9	18%
2.	Baik	1, 7, 10, 12, 14, 15, 16, 36, 44, 45, 50	11	22%
3.	Cukup	5, 6, 13, 19, 22, 28, 29, 42, 47	9	18%
4.	Jelek	24, 25, 26, 30, 32, 37, 41, 43	8	16%
5.	Sangat Jelek	2, 3, 9, 11, 17, 18, 23, 31, 33, 35, 39, 48, 49	13	26%

Sumber : Data Primer Diolah, 2025

Gambar 4

Distribusi Soal UAS Berdasarkan Indeks Efektivitas Pengecoh



Sumber : Data Primer Diolah, 2025

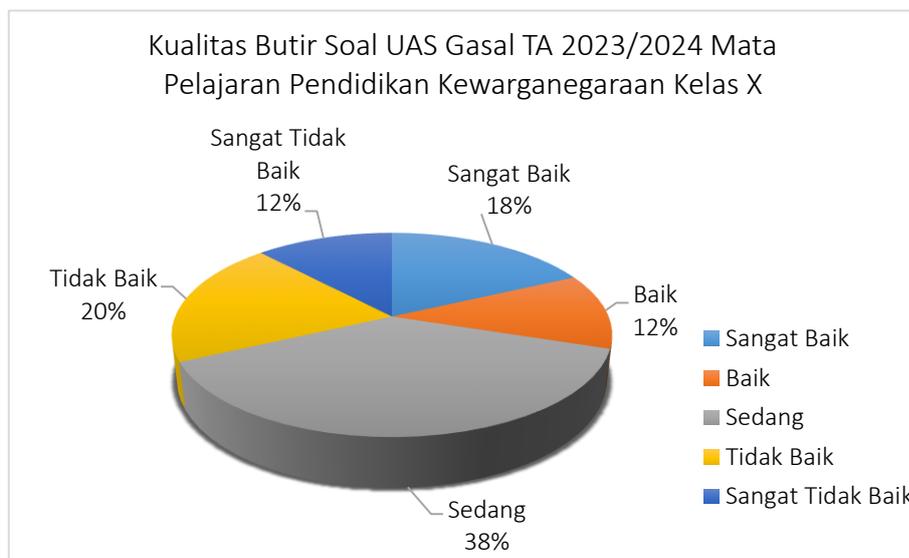
Berdasarkan tabel 5 dan gambar 4 menunjukkan bahwa hasil analisis efektivitas pengecoh menggunakan aplikasi Itean versi 4.3, dapat diketahui bahwa terdapat 9 butir soal atau 18% memiliki efektivitas pengecoh yang sangat baik, 11 butir soal atau 22% memiliki efektivitas pengecoh yang baik, 9 butir soal atau 18% memiliki efektivitas pengecoh yang cukup, 8 butir soal atau 16% memiliki efektivitas pengecoh yang jelek, dan 13 butir soal atau 26% memiliki efektivitas pengecoh yang sangat jelek. Efektivitas pengecoh dapat berfungsi dengan baik apabila opsi pengecoh dipilih secara merata oleh peserta tes. Namun, jika salah satu atau lebih dari pengecoh tidak dipilih oleh peserta tes maka pengecoh tersebut tidak berfungsi dengan baik (Himawan & Nurgiyantoro, 2022).

Tabel 5  
Distribusi Soal UAS Berdasarkan Kualitas Butir Soal

No.	Kualitas Butir Soal	Butir Soal	Jumlah	Persentase
1.	Sangat Baik	12, 16, 20, 34, 40, 42, 44, 46, 50	9	18%
2.	Baik	10, 28, 29, 36, 43, 45	6	12%
3.	Sedang	5, 6, 8, 13, 14, 15, 17, 18, 22, 23, 25, 26, 27, 30, 35, 37, 39, 47, 48	19	38%
4.	Tidak Baik	1, 3, 4, 7, 9, 11, 19, 21, 38, 49	10	20%
5.	Sangat Tidak Baik	2, 24, 31, 32, 33, 41	6	12%

Sumber : Data Primer Diolah, 2025

Gambar 5  
Distribusi Soal UAS Berdasarkan Kualitas Butir Soal



Sumber : Data Primer Diolah, 2025

Setelah mengetahui hasil uji analisis validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh, selanjutnya peneliti menentukan hasil analisis secara keseluruhan dan mengklasifikasikan kualitas butir soal berdasarkan *Skala Likert* yang terbagi dalam lima kategori, yaitu sangat baik, baik, sedang, tidak baik, dan sangat tidak baik. Jika soal memenuhi 4 kriteria maka soal tersebut sangat baik dan soal dapat disimpan. Apabila soal hanya memenuhi 3 kriteria maka soal tersebut baik dan dapat digunakan kembali tetapi perlu direvisi. Apabila soal hanya memenuhi 2 kriteria maka soal ini tergolong cukup dan tidak dapat disimpan ke dalam bank soal sehingga perlu diperbaiki. Jika butir soal hanya memenuhi 1 atau tidak memenuhi semua kriteria maka soal tergolong tidak baik atau sangat tidak baik maka soal tersebut perlu diganti dengan soal baru.

Berdasarkan hasil secara keseluruhan menggunakan aplikasi Iteaman versi 4.3 pada Soal Ulangan Akhir Semester Gasal Tahun Ajaran 2023/2024 Mata Pelajaran Pendidikan Kewarganegaraan Kelas X yang terdiri dari 50 butir soal dan diujikan kepada 72 peserta didik berdasarkan validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh, diketahui bahwa terdapat 9 butir soal atau 18% berkualitas sangat baik, 6 butir soal atau 12% berkualitas baik, 19 butir soal atau 38% berkualitas sedang, 10 butir soal atau 20% berkualitas tidak baik, dan 6 butir soal atau 12% berkualitas sangat tidak baik (ditolak). Kualitas butir soal yang tidak baik dan sangat tidak baik ini disebabkan karena tidak memenuhi kriteria dari salah satu standar kualitas butir soal yang ditinjau dari validitas, daya pembeda, tingkat kesukaran, dan pola sebaran jawaban.

Tabel 6  
Penyebab Kegagalan Butir Soal UAS PPKn

No.	Penyebab Kegagalan	Butir Soal	Jumlah	Persentase
1.	Validitas (Tidak Valid)	1, 2, 4, 6, 7, 8, 13, 14, 19, 21, 24, 27, 31, 32, 33, 38, 41, 47	18	36%
2.	Daya Pembeda (Jelek atau Sangat Jelek/ Negatif)	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 19, 21, 22, 24, 27, 31, 32, 33, 38, 41, 47, 49	25	50%
3.	Tingkat Kesukaran (Mudah atau Sulit)	1, 2, 3, 4, 5, 7, 9, 10, 11, 15, 17, 18, 19, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 39, 41, 45, 48, 49	34	68%
4.	Efektivitas Pengecoh (Jelek / Sangat Jelek)	2, 3, 9, 11, 17, 18, 23, 24, 25, 26, 30, 31, 32, 33, 35, 37, 39, 41, 43, 48, 49	21	42%

Sumber : Data Primer Diolah, 2025

Berdasarkan tabel 6 tersebut dapat diketahui bahwa penyebab kegagalan butir soal yang terbesar adalah tingkat kesukaran. Hasil tersebut menandakan bahwa butir soal tersebut terlalu sulit dan terlalu mudah. Penyebab kegagalan yang kedua yakni daya pembeda. Hasil ini mengindikasikan bahwa butir soal tersebut tidak dapat membedakan peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. Penyebab kegagalan yang ketiga yaitu efektivitas pengecoh. Hasil ini membuktikan bahwa opsi pengecoh tidak dipilih secara merata oleh peserta tes dan tidak memiliki daya tarik bagi peserta tes yang kurang menguasai materi. Penyebab kegagalan yang keempat ialah validitas. Hasil tersebut menunjukkan bahwa butir soal tidak memiliki kesesuaian dan kesejajaran arah dengan skor totalnya.

## Simpulan

Berdasarkan hasil dan pembahasan tersebut dapat disimpulkan bahwa Soal Ulangan Akhir Semester Gasal Tahun Ajaran 2023/2024 Mata Pelajaran Pendidikan Kewarganegaraan Kelas X yang terdiri dari 50 butir soal dan diujikan kepada 72 siswa menunjukkan butir soal tersebut tergolong cukup baik karena 19 dari 50 soal berkualitas sedang. Hal tersebut mengindikasikan bahwa butir soal tersebut belum sepenuhnya memiliki kualitas yang sangat baik karena ada beberapa soal yang belum memenuhi standar kualitas butir soal yang ditinjau dari validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Butir soal yang dapat disimpan dalam bank soal dan dapat digunakan kembali pada tes yang akan datang hanya berjumlah 9 soal (tergolong sangat baik). Sementara butir soal yang masih perlu diperbaiki berjumlah 25 soal (tergolong baik dan sedang). Sedangkan butir soal yang tidak dapat digunakan lagi atau dibuang berjumlah 16 soal (tergolong tidak baik dan sangat tidak baik). Penelitian ini diharapkan dapat menjadi acuan bagi guru PPKn, sekolah, atau lembaga pendidikan lainnya, betapa pentingnya dalam meningkatkan kualitas evaluasi pembelajaran PPKn secara komprehensif melalui analisis butir soal, dimana tes yang diberikan tidak hanya sekedar mengukur pemahaman siswa secara tepat, tetapi juga sebagai upaya pengembangan dalam penyusunan soal yang nantinya dapat menghasilkan butir soal yang diketahui karakteristiknya dan upaya memperbaiki kurikulum untuk meningkatkan mutu pendidikan di SMAN 1 Prambanan.

## Referensi

- Arifin, Z. (2014). *Evaluasi Pembelajaran*. Pt Remaja Rosdakarya.
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan, Edisi 3*. Bumi Aksara.
- Azwar, S. (2016). *Tes Prestasi: Fungsi Dan Pengembangan Pengukuran Prestasi Belajar*. Pustaka Belajar.
- Gladushyna, O., & Strietholt, R. (2023). Measuring Education: Do We Need A Plethora Of Assessment Studies Or Just A Single Score? *International Journal Of Educational Research Open*, 5, 100281. <https://doi.org/10.1016/j.ijedro.2023.100281>
- Hardiana, M. T. A. N., & Suyata, P. (2018). The Effectiveness Of Vak (Visual, Auditory, Kinesthetic) Model In Learning Of Summary Writing. *International Journal Of Research And Review*, 5, 43–49.
- Himawan, R., & Nurgiyantoro, B. (2022). Analisis Butir Soal Latihan Penilaian Akhir Semester Ganjil Mata Pelajaran Bahasa Indonesia Kelas Viii Smpn 1 Bambanglipuro Bantul Menggunakan Program Iteaman. *Kembara: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 8, 160–180.
- Himawan, R., & Suyata, P. (2022). Developing Hots Questions: Evaluating Persuasive Speech Text Learning In Grade Ix Of Junior High School. *Gramatika Stkip Pgri Sumatera Barat*, 8(1). <https://doi.org/10.22202/jg.2022.v8i1.4810>
- Komariah, K., & Rufi'i Rufi'i. (2023). Analysis Of Multiple-Choice Questions In Class Vii Mathematics At Smp Negeri 6 Bangkalan. *Jurnal Review Pendidikan Dan Pengajaran*, 6, 1912–1919.
- Lafarchi, N. (2020). Assessing Islamic Religious Education Curriculum In Flemish Public Secondary Schools. *Religions*, 11(3), 110. <https://doi.org/10.3390/Rel11030110>

- López-Hernández, A., Buckingham, L. R., & Strotmann, B. (2023). Enhancing Learning-Oriented Assessment Through Co-Teaching In Higher Education. *Studies In Educational Evaluation*, 79, 101307. <https://doi.org/10.1016/j.stueduc.2023.101307>
- Mardapi, D., & Kartowagiran, B. (2013). Pengembangan Instrumen Pengukur Hasil Belajar Nirbias Dan Terskala Baku. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(2), 326–341. <https://doi.org/10.21831/pep.v15i2.1100>
- Matthews, R. A., Pineault, L., & Hong, Y. (2022). Correction To: Normalizing The Use Of Single-Item Measures: Validation Of The Single-Item Compendium For Organizational Psychology. *Journal Of Business And Psychology*, 37(4), 873–873. <https://doi.org/10.1007/s10869-022-09816-0>
- Oktanin, W. S., & Sukirno, S. (2015). Analisis Butir Soal Ujian Akhir Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 13(1). <https://doi.org/10.21831/jpai.v13i1.5183>
- Pradani, R. A., & Efendi, A. (2023). Analisis Butir Soal Ujian Sekolah Menggunakan Program IteMan (Analysis Of School Exam Questions Using The IteMan Program). *Indonesian Language Education And Literature*, 8(2), 275. <https://doi.org/10.24235/ileal.v8i2.11002>
- Rohmatdi, A., Martono, M., Zafrullah, Z., & Safitri, R. (2024). Optimizing Question Quality In Junior High School Exams: Classical Test Theory Evaluation With IteMan 4.3. *Al-Ishlah: Jurnal Pendidikan*, 16(4), 4363–4375. <https://doi.org/10.35445/alishlah.v16i4.5671>
- Sarker, M. F., & Ullah, M. S. (2023). A Review Of Quality Assessment Criteria In Secondary Education With The Impact Of The Covid-19 Pandemic. *Social Sciences & Humanities Open*, 8(1), 100740. <https://doi.org/10.1016/j.ssaho.2023.100740>
- Sudijono, A. (2013). *Pengantar Evaluasi Pendidikan*. Pt. Raja Grafindo Persada.
- Sudjana, N. (2019). *Penilaian Hasil Proses Belajar Mengajar*. Remaja Rosdakarya.
- Yot-Domínguez, C., & Marcelo, C. (2017). University Students' Self-Regulated Learning Using Digital Technologies. *International Journal Of Educational Technology In Higher Education*, 14(1), 38. <https://doi.org/10.1186/s41239-017-0076-8>