

AN ANALYSIS OF THE QUALITY OF ODD SEMESTER FINAL EXAMINATION ITEMS IN INDONESIAN HISTORY SUBJECT FOR GRADE XII SOCIAL SCIENCE STUDENTS AT SMAN 7 KOTA SERANG ACADEMIC YEAR 2024/2025

Analisis Kualitas Butir Soal Penilaian Akhir Semester (PAS) Ganjil Mata Pelajaran Sejarah Indonesia Kelas XII IPS di SMAN 7 Kota Serang Tahun Ajaran 2024/2025

Shalama Qaulum Fadilla ^{1a} Arif Permana Putra ^{2b} Rikza Fauzan ^{3c}

¹²³Pendidikan Sejarah FKIP Universitas Sultan Ageng Tirtayasa, Jl. Raya Ciwaru No. 25 Serang Banten

^a2288210018@untirta.ac.id

^barif.permana@untirta.ac.id

^crikza.fauzan@untirta.ac.id

(*) Coresponding Author

2288210018@untirta.ac.id

How to Cite: Fadilla, S. Q., Putra, A. P., & Fauzan, R. (2026). An Analysis Of The Quality Of Odd Semester Final Examination Items In Indonesian History Subject For Grade XII Social Science Students at SMAN 7 Kota Serang Academic Year 2024/2025. doi: [10.36526/js.v3i2.7094](https://doi.org/10.36526/js.v3i2.7094)

Abstract

Received : 29-12-2025
 Revised : 08-05-2026
 Accepted : 01-06-2026

Keywords:

Item Analysis,
 Odd Semester Final Examination,
 Indonesian History

This study aims to analyze the quality of test items used in the Odd Semester Final Examination (Penilaian Akhir Semester/PAS) for the Indonesian History subject of Grade XII Social Science students at SMAN 7 Kota Serang in the 2024/2025 academic year. The analysis focuses on the appropriateness of the items in terms of content, construction, and language, as well as their psychometric characteristics, including validity, reliability, difficulty level, discrimination power, and distractor effectiveness. This research employed a descriptive quantitative approach. The subjects were Grade XII Social Science students, and data were collected through documentation of test items and students' answer sheets. The data were analyzed descriptively using an item review format and quantitatively using Microsoft Excel and Anates version 4. The results indicate that, overall, the quality of the test items was fairly good and aligned with the basic competencies. However, several items were found to be inconsistent with the specified indicators and required improvement in terms of construction and language. The item validity analysis showed that 60% of the items were valid, while 40% were invalid. The reliability coefficient was 0.52, indicating a moderate level of reliability. In terms of discrimination power, the items ranged from weak to very good, while the difficulty level analysis revealed that most items were categorized as moderate. Regarding distractor effectiveness, some items contained less functional distractors. This study highlights the importance of systematic item analysis to ensure the quality of assessment instruments and recommends continuous collaboration among teachers, schools, and the MGMP to improve learning evaluation practices.

PENDAHULUAN

Cita-cita dan tujuan bangsa Indonesia yang termuat dalam Pembukaan Undang-Undang Dasar (UUD) 1945 salah satunya yaitu, mencerdaskan kehidupan bangsa. Hal ini diwujudkan melalui pendidikan yang diselenggarakan di Indonesia. Pendidikan mempunyai peranan yang cukup penting untuk membentuk masyarakat Indonesia yang berkarakter dan berwawasan. Untuk mewujudkan hal tersebut banyak usaha yang dilakukan agar akses pendidikan di Indonesia dapat secara merata terlaksana. Sekolah-sekolah Negeri yang menjadi lembaga pendidikan formal dalam menyelenggarakan kegiatan belajar mengajar menunjukkan adanya usaha untuk menciptakan kualitas pendidikan di Indonesia yang lebih baik. Sebagai lembaga pendidikan formal, sekolah

menempati posisi penting dalam pelaksanaan kegiatan pembelajaran. Melalui proses pembelajaran yang terencana dan terarah, sekolah berupaya meningkatkan mutu pendidikan agar menghasilkan lulusan yang memiliki kompetensi akademik maupun non-akademik. Proses pendidikan juga memungkinkan siswa mengembangkan potensi dirinya secara menyeluruh, baik dari aspek intelektual, sosial, maupun kepribadian. (Sabariah, 2022).

Pendidikan pada hakikatnya merupakan sarana utama dalam pengembangan kualitas sumber daya manusia. Melalui pendidikan, para siswa memperoleh pengetahuan, keterampilan, serta nilai-nilai kehidupan yang diperlukan sebagai bekal menuju kedewasaan dan kematangan pribadi. (Nurdiana & Nursikin, 2025). Mengingat peran pendidikan yang sangat penting tersebut, maka diperlukan penyelenggaraan pendidikan yang dilakukan secara terencana, berkesinambungan, dan melibatkan berbagai pihak secara aktif guna menyiapkan generasi bangsa yang mampu berkontribusi bagi kemajuan masyarakat dan negara. (Susanto, 2024). Dengan demikian, pendidikan berperan sebagai sarana fundamental dalam upaya mencetak sumber daya manusia yang kompeten dan berdaya saing tinggi bagi kemajuan suatu negara. Semakin maju sebuah negara, maka semakin tinggi pula tingkat Sumber Daya Manusia (SDM) yang ada di negara tersebut.

Pendidikan yang berkualitas harus bisa memenuhi tujuan pembelajaran yang hendak dicapai dengan baik. Menurut Oemar Hamalik (2011) dalam bukunya yang berjudul "Proses Belajar Mengajar" tujuan pembelajaran menggambarkan capaian yang diharapkan dari siswa sebagai hasil dari keterlibatan mereka dalam proses belajar mengajar. Selain itu, tujuan pembelajaran bukan hanya sekedar penyampaian pengetahuan, konsep dan teori, serta informasi akademik melainkan melibatkan lebih dari sekedar pelaksanaan tes untuk siswa sebagai evaluasi capaian pembelajaran yang telah dilaksanakan. Diantara banyaknya dari tujuan pendidikan Nasional, salah satu yang diharapkan melalui pendidikan yang berkualitas yaitu para siswa kelak dapat membangun jiwa kebangsaan atau Nasionalisme dalam dirinya. Sejarah merupakan salah satu disiplin ilmu yang tidak hanya menyampaikan pengetahuan tentang masa lampau, namun juga turut membantu siswa memahami identitas Nasional dan nilai-nilai kebangsaan.

Mata pelajaran Sejarah Indonesia memiliki peran yang sangat penting dalam kurikulum pendidikan nasional karena melalui pembelajaran sejarah, siswa dapat dilatih untuk berpikir kritis dan kreatif, menumbuhkan rasa ingin tahu, memperoleh inspirasi, serta mengembangkan kemampuan dalam mencari, mengolah, dan menyampaikan informasi. Selain itu, pembelajaran sejarah juga berfungsi memperkuat semangat kebangsaan. (Purni, 2023). Maka dari itu, untuk mewujudkan tujuan pembelajaran sejarah yang diinginkan kepada siswa perlu adanya peran dan dukungan dari berbagai pihak. Salah satu yang sangat berperan penting dalam pelaksanaan pembelajaran Sejarah ialah guru sejarah. Guru sejarah memegang peranan yang sangat strategis karena menjadi pihak utama yang berhadapan langsung dengan siswa dalam proses pembelajaran sejarah di sekolah. Oleh karena itu, hasil belajar yang diperoleh dari pembelajaran sejarah perlu mendapatkan perhatian serius. Hal ini berguna untuk menentukan apakah tujuan dari pembelajaran sejarah itu sendiri berhasil atau tidak. Untuk memperoleh gambaran mengenai hasil belajar siswa perlu dilakukan pengukuran melalui serangkaian tes yang mencakup aspek pengetahuan dan keterampilan siswa.

Dalam sistem pendidikan di Indonesia, salah satu komponen terpentingnya yaitu kegiatan evaluasi. Istilah evaluasi ini mempunyai padanan kata dalam Bahasa Indonesia yaitu, penilaian. Pelaksanaan evaluasi diperlukan untuk mengetahui capaian hasil belajar siswa pada mata pelajaran sejarah. Hasil evaluasi tersebut menjadi informasi penting bagi guru dalam upaya meningkatkan kualitas pembelajaran sejarah di masa mendatang. Berdasarkan Undang-Undang Nomor 14 Tahun 2005 Pasal 1, pendidik profesional pada jenjang pendidikan formal, baik pendidikan anak usia dini, dasar, maupun menengah, memiliki tanggung jawab utama untuk mengajar, mendidik, membimbing, mengarahkan, melatih, menilai, dan melakukan evaluasi terhadap siswa. Dengan demikian, jelas bahwa salah satu tugas seorang guru yaitu mengevaluasi siswanya selain mengajar dan membimbing mereka. Menurut Idrus L (2019), evaluasi memiliki aspek utama sebagai proses

mengidentifikasi tingkat keberhasilan capaian belajar siswa atas materi atau bahan ajar yang telah disampaikan di kelas. Maksud dari hal tersebut ialah agar tujuan pembelajaran dapat disajikan secara nyata serta meyakinkan melalui evaluasi. Oleh karena itu, informasi mengenai pelaksanaan pembelajaran dan tingkat keberhasilan capaian belajar siswa diperlukan sebagai bahan evaluasi program pembelajaran ke depannya.

Dalam proses pembelajaran, penilaian merupakan komponen yang tidak terpisahkan dan memiliki peran strategis dalam menentukan ketercapaian tujuan pembelajaran. Penilaian tidak hanya berfungsi sebagai alat untuk mengukur hasil belajar siswa, tetapi juga sebagai dasar dalam pengambilan keputusan pendidikan, baik bagi guru, sekolah, maupun pemangku kepentingan lainnya. Melalui penilaian yang tepat, guru dapat memperoleh gambaran objektif mengenai tingkat penguasaan kompetensi siswa, efektivitas metode pembelajaran, serta kesesuaian materi yang diajarkan dengan capaian pembelajaran yang telah ditetapkan. Oleh karena itu, kualitas instrument penilaian menjadi faktor penting dalam menjamin akurasi hasil evaluasi. Instrumen penilaian yang disusun tanpa memperhatikan kaidah penyusunan soal yang baik berpotensi menghasilkan informasi yang bias dan kurang dapat dipertanggungjawabkan secara akademik, sehingga tujuan evaluasi pembelajaran tidak tercapai secara optimal.

Secara umum, evaluasi hasil belajar siswa dapat dilakukan melalui dua teknik penilaian, yaitu teknik tes dan non-tes. Instrumen tes digunakan sebagai alat ukur untuk mengetahui tingkat pemahaman siswa melalui serangkaian soal yang disusun secara terstruktur. Tes biasanya digunakan untuk menentukan suatu nilai. Amir Daien Indrakusuma (1993) menjelaskan bahwa tes memiliki sifat objektif serta dilakukan secara sistematis untuk memperoleh data yang relevan mengenai individu dengan cara yang efisien dan akurat. (Malawi dan Maruti, 2016).

Tes berfungsi sebagai alat untuk menilai hasil belajar siswa melalui pemberian sejumlah pertanyaan atau tugas yang harus diselesaikan. Dalam proses evaluasi pembelajaran, tes menjadi salah satu instrumen pengukuran yang paling umum digunakan di samping bentuk evaluasi lainnya. Sebagai alat ukur penilaian, tes atau soal sebaiknya melengkapi minimal 4 kriteria sebagai syarat tes yang baik dan berkualitas yaitu meliputi validitas, reliabilitas, tingkat kesukaran, serta daya pembeda soal. (Arifien, 2021). Dengan demikian, tujuan utama dari evaluasi yang berhasil baik bagi siswa maupun guru adalah untuk mengetahui perkembangan, hasil belajar, dan kemajuan siswa dalam satu semester. Agar evaluasi berjalan optimal, kualitas tes yang digunakan perlu diperhatikan dengan saksama.

Kualitas tes hasil belajar sangat dipengaruhi oleh kualitas setiap butir soalnya. Tes yang terdiri atas item-item berkualitas tinggi akan memberikan hasil yang lebih efektif, meskipun jumlah soalnya sedikit, dibandingkan dengan tes yang banyak namun memiliki item dengan kualitas rendah. Item-item yang berkualitas rendah tersebut tidak hanya menurunkan fungsi tes, akan tetapi dapat menghasilkan data pengukuran yang kurang akurat. Salah satu cara dalam menentukan kualitas tes hasil belajar siswa yaitu dengan dilakukannya analisis soal (*item analysis*). Utamanya analisis soal dapat dilaksanakan untuk tes objektif. (Surya et al., 2023). Namun demikian seringkali tes atau soal yang dijadikan instrumen dalam evaluasi tersebut tidak dilakukan analisis kualitas butir soal terlebih dahulu.

Analisis kualitas butir soal merupakan proses mengkaji setiap butir soal dalam sebuah tes yang dilakukan setelah pelaksanaan evaluasi. Kegiatan ini bertujuan untuk memastikan bahwa setiap butir soal memiliki kualitas yang memadai sehingga layak digunakan sebagai instrumen penilaian. (Arifien, 2021). Dalam proses evaluasi pembelajaran, guru perlu melakukan analisis terhadap butir soal yang telah digunakan. Melalui kegiatan ini, guru dapat menilai kualitas setiap butir soal dan mengelompokkannya ke dalam beberapa kategori, seperti sangat baik, baik, cukup, kurang baik, hingga tidak baik. Hasil analisis tersebut juga membantu guru menentukan butir soal mana yang dapat disimpan dalam bank soal untuk digunakan kembali, mana yang memerlukan revisi, serta butir soal yang harus dieliminasi/dibuang. (Arikunto, 2013).

Penilaian Akhir Semester (PAS) adalah bentuk evaluasi rutin yang dilaksanakan oleh sekolah pada akhir semester ganjil untuk menilai proses pembelajaran yang telah berlangsung

antara guru dan siswa selama satu semester. (Kiswati, 2020). Hasil PAS sering dijadikan acuan utama dalam menentukan nilai akhir, kelulusan, serta tindak lanjut pembelajaran pada semester berikutnya. Soal-soal PAS seharusnya mampu mengukur kompetensi siswa secara valid dan reliabel sesuai dengan indikator yang telah ditetapkan.

Namun, dalam praktiknya, penyusunan soal PAS sering kali belum disertai dengan evaluasi kualitas butir soal secara sistematis. Kondisi tersebut dapat menyebabkan soal yang digunakan tidak sepenuhnya mencerminkan kemampuan siswa yang sebenarnya, baik karena tingkat kesukaran yang tidak proporsional, daya pembeda yang rendah, maupun pengecoh yang tidak berfungsi secara optimal. Oleh sebab itu, analisis butir soal menjadi langkah penting untuk menjamin kualitas instrumen penilaian yang digunakan dalam PAS.

Soal yang akan dilakukan analisis pada penelitian kali ini yaitu, soal Penilaian Akhir Semester (PAS) Ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025. Penilaian Akhir Semester (PAS) merupakan salah satu bentuk evaluasi yang dilaksanakan di SMAN 7 Kota Serang. Penilaian Akhir Semester (PAS) adalah instrumen yang dipakai untuk menilai pencapaian siswa selama satu semester. Penelitian ini berfokus pada analisis soal Penilaian Akhir Semester (PAS) tahun ajaran 2024/2025 pada mata pelajaran Sejarah Indonesia kelas XII IPS di semester ganjil. Berdasarkan hasil wawancara bersama guru sejarah di SMAN 7 Kota Serang yaitu, Isti Duwiyanti, S.Pd. soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS tahun ajaran 2024/2025 berbentuk soal pilihan ganda (PG) yang berjumlah 20 soal. Adapun dalam penyusunan soal tersebut guru mengacu pada sumber referensi seperti buku paket Sejarah Indonesia kelas XII yang digunakan selama pembelajaran maupun lembar kerja siswa yang dibuat dari berbagai sumber penerbit buku. Soal pada tes tersebut disusun berdasarkan bahan ajar atau materi yang sudah disampaikan di kelas selama satu semester. Soal-soal ini mengukur kemampuan siswa dalam aspek kognitif.

Hasil Penilaian Akhir Semester (PAS) ganjil pada mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 menunjukkan hasil belajar siswa yang masih rendah. Nilai siswa banyak yang masih di bawah Kriteria Ketuntasan Minimal (KKM). Di samping itu, soal yang dipakai belum dilakukannya analisis kualitas butir soal yang mendalam oleh guru. Adanya keterbatasan waktu bagi guru dimana waktu yang tersedia untuk menganalisis butir soal terbatas, sehingga analisis tidak bisa dilakukan secara mendalam. Tak hanya itu, adanya *deadline* untuk pengumpulan data atau penyelesaian analisis bisa membuat proses analisis terburu-buru sehingga menghasilkan data yang kurang akurat. Dengan demikian, melalui penelitian ini diharapkan dapat membantu guru untuk memberikan informasi mengenai data analisis yang lebih akurat sehingga dapat memberikan masukan yang bermanfaat bagi guru untuk perbaikan dalam penyusunan soal Penilaian Akhir Semester (PAS) pada mata pelajaran Sejarah Indonesia yang lebih baik di masa mendatang.

METODE

Penelitian ini merupakan jenis penelitian deskriptif kuantitatif. Penelitian kuantitatif adalah penelitian yang menghasilkan data berupa angka-angka atau pernyataan yang dinilai atau dianalisis secara statistik. Sementara itu, penelitian deskriptif kuantitatif bertujuan untuk menggambarkan suatu fenomena baik melalui uraian kalimat maupun data numerik, yang terjadi pada masa kini ataupun masa lampau. (Sukmadinata, 2012).

Penelitian dilaksanakan di SMAN 7 Kota Serang pada tahun ajaran 2024/2025. Subjek penelitian adalah siswa kelas XII IPS yang mengikuti Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia yang berjumlah 90 siswa. Objek penelitian berupa perangkat evaluasi PAS yang digunakan oleh guru, meliputi soal PAS berbentuk pilihan ganda yang berjumlah 20 butir soal, kunci jawaban, lembar jawaban siswa, kisi-kisi soal, Rencana Pelaksanaan Pembelajaran (RPP), serta silabus yang digunakan pada semester tersebut. Perangkat tersebut dianalisis untuk mengetahui kesesuaian antara soal yang disusun dengan kompetensi dasar dan indikator pembelajaran.

Sumber data terdiri atas data primer dan data sekunder. Data primer diperoleh dari hasil jawaban siswa pada soal PAS ganjil mata pelajaran Sejarah Indonesia kelas XII IPS. Data sekunder diperoleh dari dokumen pendukung berupa kisi-kisi soal, RPP, silabus, serta dokumen administrasi pembelajaran lainnya yang relevan dengan pelaksanaan evaluasi. Teknik pengumpulan data yang digunakan adalah dokumentasi, yaitu dengan mengumpulkan dan menelaah dokumen-dokumen yang berkaitan dengan pelaksanaan PAS.

Teknik analisis data dalam penelitian ini menggunakan teknik analisis kualitatif dan kuantitatif. Analisis kualitatif dilakukan dengan menelaah butir soal berdasarkan aspek materi, konstruksi, dan bahasa menggunakan format penelaahan soal. Analisis ini bertujuan untuk menilai kesesuaian isi soal dengan indikator pembelajaran, kejelasan rumusan soal dan pilihan jawaban, serta ketepatan penggunaan bahasa sesuai kaidah penulisan soal yang baik dan benar. (Kusaeri, 2014). Selanjutnya, analisis kuantitatif dilakukan untuk mengetahui karakteristik statistik butir soal. Analisis ini mencakup pengujian validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor/pengecoh soal. Pengolahan data kuantitatif dilakukan dengan bantuan perangkat lunak *Microsoft Excel* dan program *Anates versi 4*. Hasil analisis kuantitatif digunakan untuk mengelompokkan butir soal ke dalam kategori tertentu sesuai dengan kriteria yang telah ditetapkan. Melalui tahapan analisis tersebut, penelitian ini diharapkan dapat memberikan gambaran menyeluruh mengenai kualitas soal PAS ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang. Hasil penelitian ini diharapkan dapat menjadi bahan evaluasi bagi guru dalam menyusun instrumen penilaian yang lebih berkualitas serta sebagai dasar perbaikan dalam pelaksanaan evaluasi pembelajaran di sekolah.

HASIL DAN PEMBAHASAN

Hasil

Soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 yang berjumlah 20 butir soal pilihan ganda akan diteliti melalui analisis kuantitatif menggunakan *Microsoft Excel* dan program *Anates Ver.4* serta analisis secara kualitatif menggunakan format penelaahan soal.

Tabel 1. Hasil Penelaahan Soal Pilihan Ganda

No.	Aspek Penilaian	Jumlah Soal			
		Sesuai	%	Tidak Sesuai	%
A.	Materi				
1.	Butir soal sesuai dengan indikator	16	80%	4	20%
2.	Isi materi benar secara keilmuan	20	100%	0	0%
3.	Hanya ada satu kunci jawaban benar	20	100%	0	0%
4.	Isi materi sesuai kelas/jenjang pendidikan	20	100%	0	0%
5.	Butir pengecoh berfungsi dengan baik	20	100%	0	0%
B.	Konstruksi				
1.	Pokok soal jelas	19	95%	1	5%
2.	Pokok soal tidak mengarah ke jawaban benar	20	100%	0	0%
3.	Pilihan jawaban dirumuskan dengan jelas	20	100%	0	0%
4.	Pilihan jawaban homogeny	20	100%	0	0%
5.	Tidak ada bentuk negatif ganda	20	100%	0	0%
6.	Panjang pilihan jawaban kurang lebih sama	20	100%	0	0%
7.	Antar butir soal tidak bergantung satu sama lain	20	100%	0	0%
8.	Pilihan dalam bentuk angka/waktu diurutkan	20	100%	0	0%
C.	Bahasa				
1.	Bahasa komunikatif	18	90%	2	10%
2.	Kalimat gramatikal	20	100%	0	0%
3.	Kalimat tidak bermakna ganda	19	95%	1	5%
4.	Kosakata baku umum/netral	20	100%	0	0%

Analisis kualitatif dilakukan dengan menelaah butir soal berdasarkan tiga aspek, yaitu materi, konstruksi, dan bahasa. Berdasarkan hasil penelaahan, sebagian besar butir soal telah

sesuai dengan kompetensi dasar dan indikator pembelajaran yang ditetapkan dalam kisi-kisi. Dari aspek materi, mayoritas soal telah mengukur materi yang relevan dengan tujuan pembelajaran Sejarah Indonesia kelas XII IPS. Namun, masih ditemukan beberapa butir soal yang belum sepenuhnya sesuai atau selaras dengan indikator pembelajaran, sehingga perlu dilakukan perbaikan agar soal lebih representatif dalam mengukur kompetensi siswa.

Dari aspek konstruksi, sebagian besar soal telah disusun sesuai dengan kaidah penulisan soal pilihan ganda yang baik, seperti kejelasan pokok soal dan homogenitas pilihan jawaban. Meskipun demikian, terdapat beberapa butir soal yang memiliki kelemahan dalam perumusan stem dan alternatif jawaban, sehingga berpotensi menimbulkan penafsiran ganda. Pada aspek bahasa, sebagian besar soal telah menggunakan bahasa yang komunikatif dan sesuai dengan kaidah bahasa Indonesia yang baik dan benar. Akan tetapi, masih ditemukan beberapa soal dengan penggunaan kalimat yang kurang efektif dan tanda baca yang kurang tepat.

Tabel 2. Distribusi Butir Soal berdasarkan Validitas Item

No.	Indeks Validitas	Hasil Analisis	
		Butir Soal	Persen
1	≥0,207 (Valid)	3, 5, 6, 7, 8, 9,	60%
		12, 13, 14,	
		15, 17, 20	
2	≤0,207 (Tidak Valid)	1, 2, 4,	40%
		10, 11, 16,	
		18, 19	

Analisis kuantitatif dilakukan untuk mengetahui karakteristik statistik butir soal. Hasil analisis validitas menunjukkan bahwa sebagian butir soal termasuk dalam kategori valid, sementara beberapa soal lainnya dinyatakan tidak valid. Soal-soal yang valid menunjukkan kemampuan yang baik dalam mengukur kompetensi yang diharapkan, sedangkan soal yang tidak valid perlu direvisi atau diganti agar dapat digunakan sebagai instrumen penilaian yang lebih akurat.

Tabel 3. Distribusi Butir Soal berdasarkan Reliabilitas

Indeks Reliabilitas	Hasil Analisis
0,800 - 1,000 (Sangat tinggi, reliabel)	0,52
0,600 - 0,799 (Tinggi, reliabel)	
0,400 - 0,599 (Cukup, reliabel)	
0,200 - 0,399 (Rendah, tidak reliabel)	
0,000 - 0,199 (Sangat rendah, tidak reliabel)	

Hasil analisis reliabilitas menunjukkan bahwa instrumen soal PAS memiliki tingkat reliabilitas sedang/cukup. Hal ini mengindikasikan bahwa soal secara umum cukup konsisten dalam mengukur kemampuan siswa, meskipun masih terdapat ruang untuk peningkatan kualitas agar hasil pengukuran menjadi lebih stabil dan dapat diandalkan. Reliabilitas yang sedang menunjukkan bahwa instrumen penilaian belum sepenuhnya optimal dan memerlukan perbaikan dalam penyusunan butir soal.

Tabel 4. Distribusi Butir Soal berdasarkan Tingkat Kesukaran

No.	Indeks Tingkat Kesukaran	Analisis Soal	
		Butir Soal	Persen
1	0,00 - 0,30 (sukar/sulit)	7, 10, 11, 16	20%
2	0,31 - 0,70 (sedang)	1, 2, 3, 4, 5, 6, 8, 13, 14, 18, 19, 20	60%
3	0,71 - 1,00 (mudah)	9, 12, 15, 17	20%

Berdasarkan analisis tingkat kesukaran, sebagian besar butir soal berada pada kategori sedang. Hal ini menunjukkan bahwa soal yang digunakan telah cukup proporsional dalam mengukur kemampuan siswa, tidak terlalu mudah dan tidak terlalu sulit. Namun, masih ditemukan beberapa soal yang termasuk dalam kategori mudah dan sukar. Keberadaan soal dengan tingkat kesukaran yang kurang proporsional dapat memengaruhi ketepatan hasil pengukuran kemampuan siswa.

Tabel 5. Distribusi Butir Soal berdasarkan Daya Pembeda

No.	Indeks Daya Pembeda	Analisis Soal	
		Butir Soal	Persen
1	< 0,20 (Lemah)	1, 2, 4, 10, 11, 18, 19	35%
2	0,20 - 0,40 (Cukup)	9, 16, 20	15%
3	0,40 - 0,70 (Baik)	3, 7, 8, 12, 15, 17	30%
4	0,70 - 1,00 (Sangat baik)	5, 6, 13, 14	20%

Hasil analisis daya pembeda menunjukkan variasi kemampuan soal dalam membedakan siswa yang memiliki kemampuan tinggi dan rendah. Sebagian butir soal memiliki daya pembeda yang baik hingga sangat baik, sementara beberapa soal lainnya memiliki daya pembeda yang rendah. Soal dengan daya pembeda yang baik menunjukkan efektivitas dalam mengidentifikasi perbedaan kemampuan siswa, sedangkan soal dengan daya pembeda rendah kurang efektif dan perlu dilakukan revisi.

Tabel 6. Distribusi Butir Soal berdasarkan Efektivitas Pengecoh

No.	Indeks Pengecoh	Analisis Soal	
		Butir Soal	Persen
1	76% - 125% (Sangat baik)	9, 12, 15	15%
2	51% - 75% (Baik)	1, 2, 5, 13, 14, 17	30%
3	26% - 50% (Kurang baik)	3, 4, 6, 8, 16, 18, 19, 20	40%
4	0% - 25% (Jelek)	7, 10, 11	15%
5	>200% (Sangat jelek)	-	0%

Analisis efektivitas distraktor/pengecoh soal menunjukkan bahwa tidak semua alternatif jawaban berfungsi dengan baik. Beberapa pengecoh masih kurang dipilih oleh siswa, sehingga tidak berfungsi secara optimal dalam mengecoh siswa yang belum menguasai materi. Pengecoh yang tidak efektif menunjukkan perlunya perbaikan dalam penyusunan pilihan jawaban agar seluruh alternatif dapat berfungsi sesuai dengan tujuan penilaian.

Pembahasan

Analisis Kualitatif

Soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025, analisis dilakukan dengan mengelompokkan soal berdasarkan kesesuaiannya ditinjau dari aspek materi, konstruksi, dan bahasa melalui penggunaan format penelaahan soal. (Kusaeri dan Suprananto, 2012). Proses ini bertujuan untuk memastikan bahwa butir soal yang dianalisis telah memenuhi ketentuan dari segi materi, konstruksi, serta penggunaan bahasa. Berdasarkan hasil penelaahan dari aspek materi, secara umum butir soal telah sesuai atau selaras dengan kompetensi dasar yang tercantum dalam Silabus dan RPP. Namun demikian, masih ditemukan beberapa soal yang memerlukan perbaikan. Terdapat 4 butir soal yang belum memenuhi kriteria penelaahan soal dari segi materi yaitu pada nomor 5, 7, 12, dan 20, karena indikator yang digunakan tidak sesuai dengan indikator yang tercantum dalam kisi-kisi soal. Kemudian dari segi konstruksi yaitu pada soal nomor 20 memiliki pokok soal yang kurang jelas, pokok soal belum dirumuskan dengan baik yang dapat menimbulkan potensi multitafsir sehingga soal tersebut perlu direvisi. Selanjutnya dari segi bahasa yaitu terdapat pada soal nomor 1 dan 20 menggunakan bahasa yang tidak komunikatif dan bermakna ganda.

Analisis Kuantitatif

Validitas merupakan derajat ketepatan suatu instrumen dalam mengukur apa yang seharusnya diukur (Arikunto, 2013). Apabila tes dapat memberikan informasi secara akurat mengenai apa yang sedang dinilai atau diukur, maka instrumen tersebut dikatakan memiliki tingkat validitas yang tinggi. (Purwanto, 2010). Pengujian validitas pada penelitian ini dilakukan dengan menggunakan teknik *point biserial* dan *product moment*. Nilai koefisien validitas yang diperoleh dari hasil perhitungan selanjutnya dibandingkan dengan nilai r_{tabel} pada taraf signifikansi 5%.

Subjek penelitian adalah siswa yang mengikuti PAS ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025, dengan jumlah responden sebanyak 90 siswa. Pada taraf signifikansi 5% dengan derajat kebebasan ($n - 2$) sebesar 88, diperoleh nilai r_{tabel} sebesar 0,207. Kriteria pengambilan keputusan dalam uji validitas ini adalah apabila nilai $r_{hitung} > r_{tabel}$ maka butir soal dinyatakan valid. Sebaliknya, apabila nilai $r_{hitung} < r_{tabel}$ maka butir soal dinyatakan tidak valid. Berdasarkan hasil analisis terhadap 20 butir soal PAS ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025, diketahui bahwa terdapat 12 butir soal yang memenuhi kriteria valid, sedangkan 8 butir soal lainnya dinyatakan tidak valid.

Reliabilitas menunjukkan tingkat konsistensi suatu instrumen penilaian apabila digunakan dalam kondisi yang relatif sama (Sudjana, 2017). Pengujian reliabilitas dalam penelitian ini dilakukan dengan menggunakan rumus Kuder-Richardson 20 (K-R. 20). Hasil perhitungan reliabilitas kemudian diinterpretasikan berdasarkan kriteria yang berlaku, yaitu instrumen dinyatakan memiliki reliabilitas tinggi apabila koefisien reliabilitas bernilai $> 0,70$. (Sudijono, 2015). Berdasarkan hasil analisis terhadap butir soal, diperoleh nilai koefisien reliabilitas sebesar 0,52. Dengan demikian, soal PAS ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 termasuk dalam kategori reliabilitas sedang/cukup.

Tingkat kesukaran soal merupakan indikator penting dalam menentukan kualitas instrumen penilaian. Soal yang baik seharusnya memiliki tingkat kesukaran yang seimbang, tidak terlalu mudah dan tidak terlalu sulit (Arifin, 2016). Hasil penghitungan tingkat kesukaran butir soal dikelompokkan ke dalam tiga kategori, yaitu: butir soal dengan nilai $P 0,00 - 0,30$ termasuk dalam kategori sukar/sulit, nilai $P 0,31 - 0,70$ diklasifikasikan sebagai sedang, sedangkan butir soal dengan nilai $P 0,71 - 1,00$ berada pada kategori mudah.

Berdasarkan analisis tingkat kesukaran terhadap seluruh butir soal, diperoleh hasil bahwa sebanyak 4 butir soal (20%), yaitu nomor 7, 10, 11, dan 16, termasuk dalam kategori sukar/sulit. Selanjutnya, 12 butir soal (60%), yaitu nomor 1, 2, 3, 4, 5, 6, 8, 13, 14, 18, 19, dan 20, berada pada kategori sedang. Adapun 4 butir soal (20%) lainnya, yaitu nomor 9, 12, 15, 17, tergolong

dalam kategori mudah. Hasil penelitian ini menunjukkan bahwa sebagian besar soal berada pada kategori sedang, yang sesuai dengan pendapat Sudjana (2017) bahwa soal dengan tingkat kesukaran sedang lebih efektif dalam mengukur kemampuan siswa secara optimal.

Daryanto (2018) mengungkapkan bahwa daya pembeda menunjukkan sejauh mana suatu butir soal dapat membedakan antara peserta didik yang memiliki kemampuan tinggi dan berkemampuan rendah. Dalam penghitungan daya pembeda, peserta tes dikelompokkan berdasarkan tingkat pencapaian hasil belajar, yaitu kelompok berkemampuan tinggi dan kelompok berkemampuan rendah. Penelitian ini melibatkan 90 siswa sehingga termasuk dalam kategori kelompok kecil. Oleh karena itu, peserta dibagi secara seimbang menjadi dua bagian, masing-masing sebagai kelompok atas (J_A) dan kelompok bawah (J_B). Selanjutnya, daya pembeda setiap butir soal dihitung menggunakan rumus yang telah ditetapkan.

Berdasarkan hasil analisis indeks daya pembeda butir soal, diketahui bahwa terdapat 7 butir soal (35%), yaitu nomor 1, 2, 4, 10, 11, 18, dan 19 yang termasuk dalam kategori daya pembeda lemah. Sebanyak 3 butir soal (15%), yaitu nomor 9, 16, dan 20 berada pada kategori cukup. Selanjutnya, terdapat 6 butir soal (30%), yaitu nomor 3, 7, 8, 12, 15, dan 17 yang masuk dalam kategori baik, serta 4 butir soal (20%), yaitu nomor 5, 6, 13, dan 14 yang termasuk dalam kategori sangat baik.

Pengecoh merupakan alternatif jawaban yang berfungsi untuk mengelabui atau mengecoh siswa yang belum menguasai materi. Pengecoh yang efektif adalah pengecoh yang dipilih oleh sejumlah siswa dan memiliki daya tarik yang relatif seimbang (Arifin, 2016). Hasil perhitungan efektivitas pada setiap pengecoh ditafsirkan berdasarkan kriteria Indeks Pengecoh (IP). Pengecoh dinyatakan sangat baik apabila memiliki nilai IP antara 76% - 125%. Pengecoh dengan nilai IP 51% - 75% atau 126% - 150% termasuk dalam kategori baik. Selanjutnya, pengecoh dikategorikan kurang baik apabila nilai IP berada pada rentang 26% - 50% atau 151% - 175%. Pengecoh dengan nilai IP 0% - 25% atau 176% - 200% diklasifikasikan sebagai jelek, sedangkan nilai IP yang melebihi 200% menunjukkan bahwa pengecoh tersebut termasuk dalam kategori sangat jelek.

Berdasarkan hasil analisis efektivitas pengecoh menunjukkan bahwa 3 soal (15%) yaitu pada nomor 9, 12, 15 berkategori sangat baik, 6 soal (30%) pada nomor 1, 2, 5, 13, 14, 17 berkategori baik, 8 soal (40%) nomor 3, 4, 6, 8, 16, 18, 19, 20 berkategori kurang baik, dan 3 soal (15%) pada nomor 7, 10, 11 berkategori jelek.

Kualitas Butir Soal

Penentuan kualitas butir soal dalam penelitian ini dilakukan dengan menggunakan kriteria yang disesuaikan dari skala *Likert*. Oktanin dan Sukirno (2015) menjelaskan bahwa penilaian kualitas butir soal didasarkan pada pemenuhan empat indikator utama, yaitu validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, dengan kriteria sebagai berikut :

- a) Butir soal dikategorikan sangat baik apabila telah memenuhi keempat indikator kualitas soal. Soal dengan kriteria ini layak disimpan dalam bank soal dan dapat digunakan kembali pada kesempatan selanjutnya.
- b) Butir soal termasuk dalam kategori baik apabila memenuhi tiga dari empat indikator kualitas. Soal tersebut masih dapat dimasukkan ke dalam bank soal, namun perlu dilakukan perbaikan agar seluruh indikator dapat terpenuhi secara optimal.
- c) Butir soal diklasifikasikan sedang apabila hanya memenuhi dua dari empat indikator kualitas. Oleh karena itu, butir soal perlu direvisi terlebih dahulu sebelum dapat disimpan dan dimanfaatkan kembali dalam bank soal.
- d) Butir soal dinyatakan tidak baik jika hanya memenuhi satu dari empat indikator kualitas. Soal dengan kondisi ini memerlukan revisi yang lebih mendalam, sehingga lebih disarankan untuk tidak digunakan kembali atau lebih baik dibuang.
- e) Butir soal dikategorikan sangat tidak baik apabila tidak memenuhi seluruh indikator kualitas yang telah ditetapkan. Soal tersebut tidak layak dimasukkan ke dalam bank soal dan sebaiknya dieliminasi atau dibuang.
- f) Selain syarat berbasis butir, maka soal secara keseluruhan harus reliabel.

Berdasarkan hasil analisis butir soal maka diketahui kualitas butir soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 sebagai berikut :

Tabel 7. Kualitas Butir Soal Pilihan Ganda

Nomor Soal	Validitas	Daya Pembeda	Tingkat Kesukaran	Efektivitas Pengecoh	Keterangan
1	Tidak Valid	Lemah	Sedang	Baik	Soal Revisi
2	Tidak Valid	Lemah	Sedang	Baik	Soal Revisi
3	Valid	Baik	Sedang	Kurang Baik	Soal Baik
4	Tidak Valid	Lemah	Sedang	Kurang Baik	Soal Buang
5	Valid	Sangat Baik	Sedang	Baik	Soal Sangat Baik
6	Valid	Sangat Baik	Sedang	Kurang Baik	Soal Baik
7	Valid	Baik	Sukar/Sulit	Jelek	Soal Revisi
8	Valid	Baik	Sedang	Kurang Baik	Soal Baik
9	Valid	Cukup	Mudah	Sangat Baik	Soal Revisi
10	Tidak Valid	Lemah	Sukar/Sulit	Jelek	Soal Buang
11	Tidak Valid	Lemah	Sukar/Sulit	Jelek	Soal Buang
12	Valid	Baik	Mudah	Sangat Baik	Soal Baik
13	Valid	Sangat Baik	Sedang	Baik	Soal Sangat Baik
14	Valid	Sangat Baik	Sedang	Baik	Soal Sangat Baik
15	Valid	Baik	Mudah	Sangat Baik	Soal Baik
16	Tidak Valid	Cukup	Sukar/Sulit	Kurang Baik	Soal Buang
17	Valid	Baik	Mudah	Baik	Soal Baik
18	Tidak Valid	Lemah	Sedang	Kurang Baik	Soal Buang
19	Tidak Valid	Lemah	Sedang	Kurang Baik	Soal Buang
20	Valid	Cukup	Sedang	Kurang Baik	Soal Revisi

Hasil persentase dari kualitas butir soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 secara keseluruhan adalah sebagai berikut :

Tabel 8. Distribusi persentase kualitas butir soal ditinjau dari validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh

No.	Kriteria	Butir Soal	Jumlah	Persen
1	Berkualitas (Diterima)	3, 5, 6, 8, 12, 13, 14, 15, 17	9	45%
2	Kurang Berkualitas (Direvisi)	1, 2, 7, 9, 20	5	25%
3	Tidak Berkualitas (Ditolak)	4, 10, 11, 16, 18, 19	6	30%

Berdasarkan hasil analisis dapat diketahui bahwa soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 terdapat 9 soal dalam kategori berkualitas yaitu soal nomor 3, 5, 6, 8, 12, 13, 14, 15, 17. Kemudian 5 soal kategori kurang berkualitas dan perlu revisi yaitu pada nomor 1, 2, 7, 9, 20. Selanjutnya 6 soal berkategori tidak berkualitas yaitu soal nomor 4, 10, 11, 16, 18, 19. Soal-soal yang berkualitas dapat disimpan di bank soal sehingga dapat digunakan kembali. Sementara itu, soal-soal yang kurang berkualitas perlu direvisi terlebih dahulu agar memenuhi empat kriteria kemudian dapat disimpan di bank soal. Adapun soal-soal yang tidak berkualitas sebaiknya dibuang atau tidak dipakai kembali.

Dapat disimpulkan dalam penelitian ini masih ada beberapa butir soal yang perlu direvisi dari hasil analisis kualitatif dalam segi materi, konstruksi, dan bahasa. Pada analisis kuantitatif terdapat soal yang perlu direvisi dan yang sebaiknya dibuang atau tidak dipakai kembali. Sedangkan soal-soal yang berkualitas dapat disimpan di bank soal sehingga dapat digunakan kembali. Untuk soal yang kurang berkualitas sebaiknya direvisi terlebih dahulu dengan melihat atau memperbaiki kriteria soal yang tidak terpenuhi. Sementara itu, untuk soal yang masuk kategori tidak berkualitas sebaiknya dibuang atau diganti dengan soal yang lebih berkualitas.

PENUTUP

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025. Analisis kualitas butir soal dilakukan untuk mengetahui sejauh mana instrumen evaluasi yang digunakan telah memenuhi kaidah penilaian yang baik dan mampu mengukur capaian kompetensi siswa secara objektif. Kualitas instrumen evaluasi menjadi aspek penting dalam proses pembelajaran karena hasil penilaian berfungsi sebagai dasar pengambilan keputusan dalam perbaikan pembelajaran, penentuan ketuntasan belajar, serta pemetaan kemampuan siswa. Berdasarkan hasil analisis data yang telah dilakukan, diperoleh beberapa simpulan yang menjawab rumusan masalah dan tujuan penelitian sebagai berikut :

- 1) Kualitas soal Penilaian Akhir Semester (PAS) ganjil mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tahun ajaran 2024/2025 memiliki kualitas yang cukup baik karena sudah sesuai dengan Kompetensi Dasar (KD). Perlu sedikit perbaikan dalam segi materi, konstruksi, dan bahasa pada beberapa soal.
- 2) Ditinjau dari aspek validitas, sebagian besar butir soal PAS mata pelajaran Sejarah Indonesia kelas XII IPS berada pada kategori valid, yang menunjukkan bahwa butir soal tersebut telah mampu mengukur kompetensi yang seharusnya diukur sesuai dengan indikator pembelajaran. Meskipun demikian, masih ditemukan sejumlah butir soal yang tidak valid, sehingga perlu dilakukan revisi atau penggantian agar kesesuaian antara tujuan pembelajaran, indikator, dan butir soal dapat lebih optimal.
- 3) Ditinjau dari aspek reliabilitas, instrumen PAS menunjukkan tingkat reliabilitas yang tergolong sedang/cukup. Hal ini menandakan bahwa tes memiliki konsistensi internal yang memadai dan dapat digunakan sebagai alat ukur yang relatif stabil. Hal ini menunjukkan bahwa hasil penilaian dapat dipercaya untuk menggambarkan kemampuan siswa secara umum, meskipun tetap diperlukan penyempurnaan pada butir-butir tertentu agar konsistensi tes dapat semakin ditingkatkan.
- 4) Ditinjau dari tingkat kesukaran, butir soal PAS berada pada kategori beragam, mulai dari mudah, sedang, hingga sulit/sukar. Sebagian besar butir soal berada pada tingkat kesukaran sedang, yang merupakan kategori ideal dalam penyusunan tes. Namun, masih terdapat butir soal yang terlalu mudah atau terlalu sukar, sehingga komposisi tingkat kesukaran perlu diperhatikan agar sebaran soal lebih proporsional dan mampu membedakan kemampuan siswa secara lebih akurat.
- 5) Ditinjau dari daya pembeda, sebagian butir soal telah memiliki daya pembeda yang baik dan cukup, yang berarti mampu membedakan antara siswa yang berkemampuan tinggi dan rendah. Akan tetapi, masih ditemukan butir soal dengan daya pembeda rendah, sehingga kurang efektif dalam mengidentifikasi perbedaan kemampuan siswa. Oleh karena itu, butir soal dengan daya pembeda rendah perlu direvisi atau diganti agar fungsi evaluasi dapat berjalan secara optimal.
- 6) Ditinjau dari efektivitas pengecoh, tidak seluruh alternatif jawaban berfungsi secara efektif. Beberapa pengecoh belum mampu menarik perhatian siswa yang memiliki kemampuan rendah, sehingga pemilihan jawaban cenderung terpusat pada opsi tertentu. Kondisi ini menunjukkan perlunya perbaikan dalam perumusan pengecoh agar setiap alternatif jawaban memiliki peluang yang seimbang untuk dipilih dan dapat meningkatkan kualitas butir soal secara keseluruhan.

Secara umum, hasil penelitian menunjukkan bahwa kualitas butir soal PAS mata pelajaran Sejarah Indonesia kelas XII IPS di SMAN 7 Kota Serang tergolong cukup baik, baik ditinjau dari aspek kuantitatif maupun kualitatif, namun masih memerlukan perbaikan pada beberapa indikator dan kaidah penelaahan soal agar instrumen evaluasi yang digunakan benar-benar optimal.

Sebagai prospek pengembangan, hasil penelitian ini dapat dimanfaatkan oleh guru sebagai bahan refleksi dalam menyusun dan mengevaluasi instrumen penilaian, khususnya pada mata pelajaran Sejarah Indonesia. Guru diharapkan dapat menggunakan hasil analisis butir soal sebagai dasar untuk melakukan revisi soal sebelum digunakan kembali pada periode penilaian berikutnya. Selain itu, sekolah dapat menjadikan penelitian ini sebagai referensi dalam meningkatkan mutu evaluasi pembelajaran secara berkelanjutan. Bagi peneliti selanjutnya, penelitian ini dapat dikembangkan dengan menganalisis kualitas butir soal pada jenjang atau mata pelajaran yang berbeda, serta menggunakan pendekatan atau perangkat analisis yang lebih beragam. Penelitian lanjutan juga dapat mengkaji hubungan antara kualitas butir soal dengan hasil belajar siswa, sehingga diperoleh gambaran yang lebih komprehensif mengenai efektivitas evaluasi dalam proses pembelajaran.

DAFTAR PUSTAKA

- Aman. (2019). Final examination test instruments for history subject in Yogyakarta, Indonesia: A quality analysis. *Universal Journal of Educational Research*, 7(12), 2857–2866. <https://doi.org/10.13189/ujer.2019.071237>.
- Arifien, M., & Wijayanto, P. A. (2021). Analisis Kualitas Butir Soal Ujian Tengah Semester Geografi SMA Islam Al Azhar 14 Kota Semarang. *Geodika: Jurnal Kajian Ilmu Dan Pendidikan Geografi*, 5(2), 206–215. <https://doi.org/10.29408/geodika.v5i2.3696>.
- Arifin, Z. (2016). *Evaluasi pembelajaran*. Remaja Rosdakarya.
- Arikunto, S. (2013). *Dasar-dasar evaluasi pendidikan*. Bumi Aksara.
- Chen, Y., Li, X., Liu, J., & Ying, Z. (2021). Item response theory: A statistical framework for educational and psychological measurement. *Journal of Educational and Behavioral Statistics*.
- Daryanto. (2018). *Evaluasi pendidikan*. Rineka Cipta.
- Faridah, A. (2021). Karakteristik butir soal penilaian akhir semester mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang. *Fajar Historia: Jurnal Ilmu Sejarah dan Pendidikan*, 5(2), 148–163. <https://doi.org/10.29408/fhs.v5i2.4609>.
- Hamalik, O. (2011). *Proses belajar mengajar*. Bumi Aksara.
- Hanifah, F. (2023). Analysis of final semester assessment question items for Class XI Indonesian History subjects. *Indonesian Journal of History Education*, 8(2). <https://doi.org/10.15294/ijhe.v8i2.68079>.
- Kiswati. (2020). *Pedoman praktis penilaian kurikulum 2013*. Qahar Publisher.
- Hayat, B. (2021). Klasika: Program analisis item dan tes dengan pendekatan klasik. *JP3I: Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, 10(1), 1–11. <https://doi.org/10.15408/jp3i.v10i1.20551>.
- Idrus, L. (2019). Evaluasi dalam proses pembelajaran. *Jurnal Evaluasi Pendidikan*, 9(2), 920–935. <https://doi.org/https://doi.org/10.35673/ajmpi.v9i2.427>.
- Kusaeri, & Suprananto. (2012). Pengukuran dan penilaian pendidikan. Graha Ilmu.
- Kusaeri. (2014). *Acuan dan teknik penilaian proses dan hasil belajar dalam kurikulum 2013*. Ar-Ruzz Media.
- Malawi, I., & Maruti, E. S. (2016). *Evaluasi pendidikan*. Media Grafika.
- Martínez, P. S. R., Barbosa-Pereira, D., Valgas-Costa, M., & Mansur-Alves, M. (2020). Item analysis of the Child Neuropsychological Assessment Test (TENI): Classical test theory and item response theory. *Applied Neuropsychology: Child*. <https://doi.org/10.1080/21622965.2020.1846128>.

- Nurdiana, A., & Nursikin, M. (2025). Konsep pendidikan nilai dalam filsafat pendidikan menurut Al-Ghazali dan Ki Hadjar Dewantara. *Afeksi: Jurnal Penelitian Dan Evaluasi Pendidikan*, 6(5), 970–976. <https://doi.org/https://doi.org/10.59698/afeksi.v6i5.542>.
- Purwanto, M. N. (2010). *Prinsip-prinsip dan teknik evaluasi pengajaran*. Remaja Rosdakarya.
- Sudijono, A. (2015). *Pengantar evaluasi pendidikan*. Rajawali Pers.
- Sukmadinata, N. S. (2012). *Metode penelitian pendidikan*. Remaja Rosdakarya.
- Triono, D., Sarno, R., & Sungkono, K. R. (2020). Item analysis for examination test in the postgraduate students' selection with Classical Test Theory and Rasch Measurement Model. In *Proceedings of the 2020 International Seminar on Application for Technology of Information and Communication* (pp. 523–529). IEEE. <https://doi.org/10.1109/iSemantic50169.2020.9234204>.
- Oktanin, W. S., & Sukirno, S. (2015). Analisis butir soal ujian akhir mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, XIII(1), 35–44. <https://doi.org/https://doi.org/10.21831/jpai.v13i1.5183>.
- Purni, T. (2023). Pentingnya pendidikan sejarah sebagai penguat pendidikan yang berkarakter. *KRINOK: Jurnal Pendidikan Sejarah & Sejarah FKIP UNJA*, 2(1), 190–197. <https://doi.org/https://doi.org/10.22437/krinok.v2i1.24723>.
- Sabariah, S. (2022). Manajemen sekolah dalam meningkatkan mutu pendidikan. *Edukatif: Jurnal Ilmu Pendidikan*, 4(1), 116–122. <https://doi.org/10.31004/edukatif.v4i1.1764>.
- Surya, A. D., Sumarno, S., & Muhtarom, M. (2023). Analisis kualitas instrumen tes hasil belajar IPAS materi wujud zat dan perubahannya. *FONDATIA: Jurnal Pendidikan Dasar*, 7(2), 271–282. <https://doi.org/10.36088/fondata.v7i2.3190>.
- Susanto, M. A. (2024). Konsep kepribadian guru menurut imam Al-Ghazali dalam kitab Ihyā' Ulumuddin. *Jurnal Al-Muharrir: Jurnal Pendidikan Agama Islam*, 1(1), 15–22. <https://doi.org/https://doi.org/10.59829/p0knje61>.