



ANALISIS BUTIR SOAL IPS PADA MATERI PERUBAHAN SOSIAL DI SMP

Nilam Syaharani¹, Imay Mardiah², Silma Aisya Putri³, Pelita Sukma⁴, Sakilah⁵

Universitas Islam Negeri Sultan Syarif Kasim Riau^{1,2,3,4,5}

e-mail: nilamsyaharani053@gmail.com¹, imaymardiah07@gmail.com²,
silmaaisya536@gmail.com³, pelitasukma.nst@gmail.com⁴, Sakilah@uin-suska.ac.id⁵

Diterima: 30/5/2026; Direvisi: 10/6/2026; Diterbitkan: 18/6/2026

ABSTRAK

Evaluasi pembelajaran yang berkualitas sangat ditentukan oleh mutu instrumen tes yang digunakan, namun pada praktiknya masih banyak butir soal yang belum dianalisis secara sistematis sehingga akurasi pengukuran hasil belajar siswa kurang optimal. Penelitian ini bertujuan untuk menganalisis kualitas butir soal pada mata pelajaran Ilmu Pengetahuan Sosial (IPS) pada materi Perubahan Sosial dengan fokus pada validitas, reliabilitas, indeks kesukaran, dan daya pembeda. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan subjek seluruh siswa SMP Negeri 3 pinggir sebanyak 30 orang melalui teknik total sampling. Data dikumpulkan melalui dokumentasi berupa soal tes, kunci jawaban, dan lembar jawaban siswa, kemudian dianalisis menggunakan korelasi *product moment* untuk menguji validitas dan *Cronbach's Alpha* untuk reliabilitas, serta perhitungan indeks kesukaran dan daya pembeda. Hasil penelitian menunjukkan bahwa validitas berada pada kategori sedang (0,39), reliabilitas tergolong rendah (0,47), indeks kesukaran termasuk katogori mudah (0,74), dan daya pembeda menunjukkan kategori baik. Temuan ini mengindikasikan bahwa butir soal sudah layak digunakan sebagai instrumen evaluasi, namun masih memerlukan perbaikan terutama pada aspek validitas dan reliabilitas agar mampu mengukur kemampuan siswa secara lebih akurat dan konsisten.

Kata Kunci: *Analisis Butir Soal, Validitas, Reliabilitas.*

ABSTRACT

Quality learning evaluation is largely determined by the quality of the test instruments used, but in practice, many test ítems have not been systematically analyzed, resulting in suboptimal accuracy in measuring student learning outcomes. This study to analyze the quality of test ítems in Social Sciences (IPS) on Social Change, focusing on validity, reliability, difficulty index, and discriminatory power. The study used a descriptive qualitative approach with 30 students of SMP Negeri 3 Pinggir using a total sampling technique, data were collected through documentation in the form of test questions, answer keys, and student answer sheets. Then, they were analyzid using product momento correlation to test validity and Cronbach's Alpha for reliability, as well as calculation of difficulty index and discriminatory power. The results showed that validity was in the moderate category (0.39), reliability was low (0.47), difficulty index was in the easy category (0.74), and discriminatory power was in the good category. These findings indicate that the test ítems are suitable for use as evaluation instrument, but still require improvement, especially, to be able to measure student abilites more accurately and consentently.

Keywords: *Item Analysis, Validity, Reliability.*



PENDAHULUAN

Evaluasi pembelajaran memegang peranan fundamental di dalam sistem operasional makro pendidikan karena berfungsi sebagai kompas utama untuk mengukur tingkat ketercapaian tujuan instruksional yang telah dirancang. Hasil dari penilaian ini secara teoretis mengemban misi sakral sebagai pijakan dasar bagi para pendidik dalam pengambilan keputusan strategis terkait perbaikan mutu proses belajar mengajar secara berkelanjutan (Agisna et al., 2023; Amaliana et al., 2025; Purnama, 2022). Dalam konteks mata pelajaran Ilmu Pengetahuan Sosial, esensi evaluasi tidak boleh hanya terjebak pada pengukuran ranah pengetahuan faktual yang bersifat hafalan kaku semata, melainkan wajib menyentuh kedalaman pemahaman konseptual siswa terhadap dinamika fenomena sosiokultural kemasyarakatan. Karakteristik materi yang sarat akan nilai kehidupan seperti perubahan sosial menuntut adanya ketersediaan instrumen tes yang memiliki kualitas standarisasi yang tinggi agar mampu merepresentasikan peta kompetensi anak secara akurat. Penyelenggaraan evaluasi yang kredibel dan terukur pada hakikatnya akan mendorong lahirnya iklim akademis yang sehat, memicu motivasi intrinsik siswa, serta membantu guru dalam memetakan serta mengklasifikasi tingkat kesiapan kognitif peserta didik dalam menghadapi tantangan era modernisasi global yang kian dinamis (Farida & Hastuti, 2026; Mawarni et al., 2023; Mtd et al., 2023; Ratnawati & Hasanah, 2021).

Namun demikian, terdapat jurang pemisah yang sangat lebar antara rumusan konseptual yang ideal tersebut dengan kenyataan objektif yang berlangsung di tataran praktis sekolah formal saat ini. Kondisi ideal menghendaki agar setiap instrumen lembar soal yang diujikan kepada peserta didik wajib melalui tahapan analisis butir soal yang sistematis dan komprehensif, mencakup parameter validitas, reliabilitas, tingkat kesukaran, serta daya pembeda. Sayangnya, fakta empiris di lapangan menyingkap sebuah fenomena yang memprihatinkan di mana mayoritas alat evaluasi yang digunakan guru belum pernah diuji secara klinis melalui pendekatan statistika modern. Banyak tenaga pendidik yang langsung menggulirkan paket soal ujian tanpa melakukan telaah awal teoretis maupun empiris, sehingga kualitas keilmiahannya sama sekali belum terjamin. Distorsi administrasi ini berdampak buruk pada terjadinya bias pengukuran akademis, di mana nilai akhir yang diperoleh siswa tidak lagi mencerminkan kapasitas intelektual mereka yang sesungguhnya. Kelesuan profesionalisme guru dalam mengelola evaluasi ini memicu kejenuhan massal dan mengaburkan arah perbaikan mutu kurikulum di tingkat persekolahan dasar (Bahri, 2023; Magdalena et al., 2023; Mtd et al., 2023; Nugrahini et al., 2023).

Potret distorsi metodologis evaluasi dan lemahnya standarisasi perangkat tes tersebut terkonfirmasi secara nyata melalui pengamatan awal yang dilakukan di SMP Negeri 3 Pinggir Kabupaten Bengkalis pada tahun ajaran 2026/2027. Di lembaga pendidikan menengah tersebut, tampak jelas bahwa subjek penelitian senantiasa menghadapi kendala kognitif yang pelik ketika berhadapan dengan soal-soal penalaran sosial pada materi perubahan sosial yang kompleks. Berdasarkan data observasi langsung, instrumen penilaian harian yang digunakan oleh guru geografi dan sosiologi di kelas tersebut terindikasi dibuat secara instan tanpa mengaitkan kesesuaian antara indikator soal dengan kompetensi dasar secara matang. Kondisi disonansi evaluatif ini diperparah oleh ketiadaan variasi pada tingkat kesukaran soal, di mana paket ujian cenderung didominasi oleh pertanyaan ingatan tingkat rendah yang monoton. Akibat dari lemahnya validitas alat ukur tersebut, perolehan nilai akademik siswa menjadi fluktuatif, memicu kecemasan akademis, serta gagal memberikan gambaran yang akurat mengenai letak kelemahan belajar subjek penelitian secara spesifik.



Kondisi darurat akademik terkait rendahnya mutu alat evaluasi tersebut menuntut dilakukannya sebuah intervensi ilmiah yang mendalam guna merestrukturisasi sistem penilaian berbasis materi sosiokultural yang rumit. Berbagai literatur ilmiah dalam beberapa tahun terakhir secara konsisten mengonfirmasi bahwa rendahnya tingkat reliabilitas instrumen tes di sekolah umumnya dipicu oleh buruknya penyusunan stimulus soal pada tahap awal konstruksi. Masalah menahun ini tidak boleh dibiarkan berlarut-larut karena materi perubahan sosial memiliki muatan cakupan yang sangat luas, dinamis, serta bersifat kontekstual yang membutuhkan instrumen evaluasi yang tepat. Melalui tindakan analisis butir soal secara matematis, guru dapat mengidentifikasi secara detail butir soal mana yang harus direvisi, dipertahankan, atau bahkan dibuang demi menyelamatkan validitas data. Upaya perbaikan yang terencana dan berbasis data empiris ini dipandang mampu menjembatani celah kosong sosiopedagogis, sehingga aktivitas pengukuran hasil belajar berubah fungsi dari sekadar formalitas pengisian rapor menjadi sarana refleksi kognitif yang humanis bagi anak (Farida & Hastuti, 2026; Mayuni et al., 2025; Yuza & Jannah, 2026).

Berpijak pada seluruh rangkaian latar belakang dan eskalasi problematika di atas, kajian ilmiah ini hadir dengan membawa nilai kebaruan serta inovasi berupa audit evaluasi terpadu. Nilai inovasi dari riset ini memfokuskan sasarannya pada analisis komprehensif yang mengintegrasikan uji validitas, reliabilitas, indeks kesukaran, dan daya pembeda butir soal IPS materi perubahan sosial dalam satu kesatuan utuh. Fokus subjek penelitian ini diarahkan secara spesifik pada performa kognitif serta lembar jawaban dari 30 siswa di SMP Negeri 3 Pinggir Kabupaten Bengkalis pada tahun ajaran 2026/2027 yang sedang mengadaptasi sistem kurikulum baru. Kebaruan kajian ini bersumber dari pemberian kontribusi praktis berupa rekomendasi paket perbaikan instrumen evaluasi yang orisinal, segar, dan dapat langsung diimplementasikan guru untuk mendongkrak efisiensi pengajaran. Melalui pembuktian empiris terhadap subjek penelitian di SMP Negeri 3 Pinggir Kabupaten Bengkalis pada tahun ajaran 2026/2027 ini, luaran teoretis riset diharapkan mampu menyelaraskan kesenjangan antara teori evaluasi dengan realitas operasional di dalam kelas.

METODE PENELITIAN

Penelitian deskriptif dengan pendekatan kuantitatif ini dirancang secara sistematis untuk menelaah mutu instrumen evaluasi pada mata pelajaran Ilmu Pengetahuan Sosial. Kegiatan riset operasional ini dilaksanakan di Sekolah Menengah Pertama Negeri 3 Pinggir, Kabupaten Bengkalis, Provinsi Riau. Populasi target dalam pengkajian ini melibatkan seluruh siswa di sekolah tersebut dengan penentuan sampel menggunakan teknik sampling total, sehingga diperoleh 30 siswa kelas IX sebagai responden aktif. Variabel utama yang diselidiki difokuskan pada tingkat kelayakan objektif dari naskah soal yang diujikan, khususnya pada materi perubahan sosial yang mencakup dimensi sosiokultural. Prosedur pelaksanaan riset diawali dengan pengumpulan berkas sekunder berupa lembar soal pilihan ganda dan uraian, kunci jawaban resmi, serta seluruh lembar jawaban milik siswa. Selanjutnya, data tersebut ditabulasi ke dalam matriks skor numerik untuk mempersiapkan tahap pengujian statistik. Langkah operasional ini murni dilakukan untuk memetakan kondisi riil perangkat tes di lapangan tanpa melibatkan intervensi perlakuan kelas ataupun manipulasi terhadap subjek.

Instrumen pengumpulan data dalam penelitian ini sepenuhnya mengandalkan teknik dokumentasi terhadap pemenuhan lembar jawaban dari 30 peserta didik. Pengujian mutu butir soal secara kuantitatif dianalisis menggunakan 4 parameter statistik, yaitu uji validitas, reliabilitas, indeks kesukaran, dan daya pembeda. Pengujian dilakukan secara komputersasi, di



mana validitas dihitung menggunakan rumus korelasi *product moment pearson* dan reliabilitas diukur lewat koefisien *cronbach alpha*. Formulasi matematika juga diterapkan untuk menghitung indeks kesukaran guna mengelompokkan soal ke dalam kategori mudah, sedang, atau sulit, serta rumus daya pembeda untuk melihat kemampuan soal dalam memisahkan kelompok siswa berkemampuan tinggi dan rendah. Seluruh rangkaian analisis data numerik ini dikerjakan secara terpadu guna menghasilkan output deskriptif yang akurat mengenai standarisasi alat evaluasi. Prosedur metodologi ini dipaparkan secara jelas dan ringkas untuk menjelaskan teknik pengauditan naskah ujian tanpa mencantumkan teori umum, tambahan gambar eksternal, ataupun rincian angka hasil akhir penelitian.

HASIL DAN PEMBAHASAN

Hasil

Hasil penelitian disajikan berdasarkan analisis terhadap kualitas butir soal yang meliputi validitas, reliabilitas, indek kesukaran dan daya pembeda. Data diperoleh dari hasil pengerjaan soal oleh 30 siswa yang kemudian dianalisis secara kuantitatif.

Tabel 1. Hasil Analisis Butir Soal

No.	Aspek Analisis	Nilai	Kategori
1.	Validitas	0,39	Sedang
2.	Reliabilitas	0,47	Rendah
3.	Indeks Kesukaran	0,74	Mudah
4.	Daya Pembeda PG	7,375	Baik
5.	Daya Pembeda Essay	3,125	Baik

Hasil menunjukkan validitas berada pada kategori sedang, reliabilitas rendah, indek kesukaran mudah, dan daya pembeda baik. Berdasarkan tabel.1 dapat diketahui bahwa uji validitas soal berada pada kategori sedang dengan nilai 0,39. Hal ini, menunjukkan bahwa sebagian besar butir soal sudah cukup mampu mengukur kompetensi yang diharapkan, meskipun masih terdapat beberapa soal yang perlu diperbaiki. Nilai reliabilitas sebesar 0,47 menunjukkan bahwa tingkat konsistensi instrumen masih rendah. Artinya, hasil pengukuran belum sepenuhnya stabil jika dilakukan pengujian ulang pada kondisi yang berbeda. Indek kesukaran sebesar 0,74 mengindikasikan bahwa soal termasuk dalam kategori mudah. Sebagian besar siswa mampu menjawab soal dengan benar sehingga tingkat kesulitan soal relatif rendah sementara itu, daya pembeda soal menunjukkan hasil yang baik, baik pada soal pilihan ganda (7,375) maupun essay (3,125). Hal ini, menunjukkan bahwa terdapat soal yang mampu membedakan antara siswa yang memiliki kemampuan tinggi dan kemampuan rendah. Secara keseluruhan, hasil penelitian ini menunjukkan bahwasannya kualitas butir soal sudah cukup baik, namun masih memerlukan perbaikan pada beberapa aspek agar lebih optimal,

Pembahasan

Analisis mendalam mengenai karakteristik psikometrik instrumen evaluasi di Sekolah Menengah Pertama Negeri 3 Pinggir menunjukkan bahwa kualitas perangkat tes berada pada level yang cukup memadai namun memerlukan rekonstruksi struktural. Pengujian kuantitatif terhadap performa 30 siswa menghasilkan nilai validitas sebesar 0,39 yang menempatkan instrumen dalam kriteria sedang. Capaian empiris ini mengindikasikan bahwa sebagian besar konstruksi pertanyaan sudah mampu merepresentasikan indikator kompetensi dasar yang diharapkan dalam kurikulum ilmu pengetahuan sosial. Meskipun demikian, angka 0,39 tersebut juga memberikan sinyal bahwa keselarasan antara tujuan instruksional dengan stimulus



kebahasaan di dalam naskah ujian belum sepenuhnya tercapai secara ideal. Ketidaksempurnaan ini biasanya dipicu oleh bias redaksional atau ketidakjelasan kalimat yang dapat mendistorsi pemahaman siswa terhadap konsep materi sosial yang bersifat kontekstual. Oleh sebab itu, guru perlu melakukan modifikasi teknis pada kalimat stimulan agar instrumen ujian mampu memotret peta kemampuan akademik siswa secara lebih presisi (Aulia et al., 2024; Handayani et al., 2023; Listeani et al., 2021; Rahmawati et al., 2021).

Aspek reliabilitas instrumen mencatatkan nilai sebesar 0,47 yang menempatkan perangkat evaluasi ini dalam kategori konsistensi yang rendah. Secara metodologis, perolehan indeks sebesar 0,47 ini menandakan bahwa stabilitas hasil pengukuran yang dihasilkan oleh perangkat tes tersebut belum sepenuhnya ajek apabila diujikan kembali pada situasi yang berbeda. Ketidakstabilan capaian skor ini dapat dipengaruhi oleh beberapa faktor internal, seperti keterbatasan jumlah butir pertanyaan yang didistribusikan kepada responden atau adanya ketidakseimbangan proporsi bobot materi. Selain itu, rendahnya angka keterandalan ini mengindikasikan bahwa paket soal yang digunakan belum melewati tahapan uji coba awal serta revisi mendalam secara berkelanjutan sebelum diaplikasikan dalam penilaian formal. Implikasi dari temuan ini menuntut para pengajar untuk memperluas cakupan sampel soal serta melakukan penyaringan ketat terhadap elemen pertanyaan yang memiliki tingkat varians skor yang terlalu lebar demi mendongkrak derajat keterpercayaan instrumen (Himawan & Nurgiyantoro, 2022; Murniatie et al., 2025; Trilestari et al., 2025).

Karakteristik tingkat kesukaran soal dalam studi ini menunjukkan angka sebesar 0,74 yang mengindikasikan bahwa paket pertanyaan tersebut masuk dalam kriteria mudah. Berdasarkan parameter teoritis evaluasi pendidikan, besaran indeks 0,74 mencerminkan bahwa mayoritas siswa dari total 30 partisipan dapat menyelesaikan tugas ujian dengan tingkat keberhasilan yang tinggi. Pola distribusi yang condong pada tingkat kesulitan rendah ini berpotensi mereduksi fungsi instrumen dalam memetakan kemampuan berpikir tingkat tinggi atau *higher order thinking skills* siswa. Lemahnya tantangan akademis dalam lembar evaluasi ini membuat guru kehilangan potret objektif mengenai kedalaman pemahaman analisis kritis siswa terhadap materi ilmu pengetahuan sosial. Guna menciptakan sistem penilaian yang komprehensif, format soal masa depan harus dikembangkan dengan skema yang lebih proporsional yang memadukan derajat kesulitan rendah, menengah, dan tinggi secara seimbang agar mampu menstimulasi kemampuan kognitif tingkat lanjut (Riskiani & Suryadi, 2026; Sudianto & Kisno, 2021; Syarifah et al., 2020; Valen & Satria, 2021).

Meskipun aspek konsistensi dan tingkat kesulitan masih memerlukan perbaikan, instrumen ini menunjukkan performa yang sangat baik pada fungsi pemisahan kompetensi objek didik. Parameter daya pembeda untuk tipe soal pilihan ganda berhasil mencatatkan nilai sebesar 7,375 sedangkan untuk tipe soal esai meraih skor sebesar 3,125. Angka 7,375 dan 3,125 ini menjadi bukti kuat bahwa seluruh komponen pertanyaan memiliki efisiensi yang tinggi dalam memisahkan kelompok siswa berkemampuan akademik atas dengan kelompok bawah. Eksistensi daya diferensiasi yang tinggi merupakan indikator krusial yang menunjukkan bahwa butir soal tidak sekadar mudah dijawab secara acak, melainkan memiliki validitas internal yang fungsional dalam mendeteksi variasi kecerdasan siswa. Keunggulan teknis ini sangat menguntungkan bagi guru dalam melakukan pemetaan diagnostik guna merancang program remedial maupun pengayaan yang tepat sasaran bagi setiap individu di dalam kelas (Fatikasari et al., 2026; Habibi et al., 2024; Prismanata, 2025).

Secara akumulatif, potret komprehensif dari hasil evaluasi ini menegaskan bahwa kualitas instrumen ujian ilmu pengetahuan sosial sudah berada pada jalur yang benar namun



belum menyentuh level performa optimal. Kelemahan pada nilai validitas 0,39 dan reliabilitas 0,47 menjadi ruang keterbatasan utama yang harus segera diselesaikan melalui tindakan revisi butir soal secara total. Implikasi praktis dari riset ini mewajibkan pihak manajemen sekolah untuk mengadakan pelatihan penyusunan kisi-kisi soal berbasis digital yang terstandarisasi bagi para guru secara periodik. Keterbatasan penelitian ini terletak pada kecilnya ukuran sampel yang hanya melibatkan 30 siswa pada satu instansi, sehingga generalisasi tingkat keandalan soal belum dapat diproyeksikan pada skala daerah yang lebih luas. Oleh karena itu, direkomendasikan bagi pengembangan riset selanjutnya untuk memperluas cakupan subjek uji coba serta mengintegrasikan analisis semiotika bahasa untuk menyempurnakan kualitas evaluasi pembelajaran.

KESIMPULAN

Berdasarkan tujuan penelitian untuk menganalisis kualitas butir soal mata pelajaran Ilmu Pengetahuan Sosial (IPS) pada materi perubahan sosial, hasil penelitian ini menunjukkan bahwa instrumen yang digunakan telah memenuhi sebagian kriteria kualitas, namun belum sepenuhnya optimal. Validitas yang berada pada kategori sedang mengindikasikan bahwa butir soal cukup mampu merepresentasikan kompetensi yang diukur, meskipun masih memerlukan penyempurnaan dalam penyusunan indikator dan redaksi soal. Reliabilitas yang rendah menunjukkan bahwa konsistensi instrumen perlu ditingkatkan melalui revisi dan pengujian ulang agar hasil evaluasi lebih stabil dan dapat dipercaya. Disisi lain, indeks kesukaran yang cenderung udah serta daya pembeda yang baik menunjukkan bahwa soal mampu dijawab oleh sebagian besar siswa sekaligus tetap dapat membedakan tingkat kemampuan mereka. Secara substansif, temuan ini menegaskan bahwa instrumen evaluasi tidak hanya ditentukan oleh satu aspek saja, tetapi merupakan hasil keterpaduan dari antara uji validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Oleh karena itu, pengembangan instrumen evaluasi kedepannya perlu dilakukan secara sistematis dengan memperhatikan keseimbangan tingkat kesulitan dan peningkatan kualitas konstruk soal. Penelitian ini juga membuka peluang untuk pengembangan lebih lanjut seperti penggunaan model analisis modern (misalnya *Model Rasch*), penerapan pada jumlah sampel yang lebih besar, serta pengembangan soal berbasis Ilmu Pengetahuan Sosial (IPS) yang terstandar. Dengan demikian hasil penelitian ini tidak hanya berkontribusi pada peningkatan kualitas evaluasi pembelajaran, tetapi juga dapat menjadi dasar bagi pengembangan instrumen yang lebih akurat dan aplikatif di masa mendatang.

DAFTAR PUSTAKA

- Agisna, R., Jauhari, Z. A., Zuar, M. S., Sholihin, M., & Khusnul, A. (2023). Evaluasi pembelajaran. *Social Science Academic*, 1(2), 353–362. <https://doi.org/10.37680/ssa.v1i2.3582>
- Amaliana, A., Arbeni, W. W., Azzahira, F., Fadilla, A. F., & Rabita, A. S. (2025). Perencanaan instrumen tes dan evaluasi hasil belajar. *PEMA: Jurnal Pendidikan dan Pengabdian Kepada Masyarakat*, 5(3), 657–665. <https://doi.org/10.56832/pema.v5i3.2272>
- Aulia, R., Sulistiawan, M. J., Wagiran, W., & Naryatmojo, D. L. (2024). Rekonstruksi asesmen formatif modul ajar teks deskripsi bagi siswa SMP di era merdeka belajar. *Jurnal Onoma: Pendidikan Bahasa dan Sastra*, 10(1), 257–267. <https://doi.org/10.30605/onoma.v10i1.3188>



- Bahri, M. S. (2023). Problematika evaluasi pembelajaran dalam mencapai tujuan pendidikan di masa merdeka belajar. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(4), 2871–2880. <https://doi.org/10.54371/jiip.v6i4.1954>
- Farida, A., & Hastuti, M. S. A. (2026). Transformasi pedagogi abad 21: Mewujudkan ekosistem pendidikan masa depan yang adaptif dan kontekstual. *MANAJERIAL: Jurnal Inovasi Manajemen dan Supervisi Pendidikan*, 6(2), 492–502. <https://doi.org/10.51878/manajerial.v6i2.10770>
- Fatikasari, F., Yuwono, A., & Sukoyo, J. (2026). Efektivitas model pembelajaran berdiferensiasi pada mata pelajaran bahasa Jawa dalam materi sastra piwulang. *LEARNING: Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran*, 6(1), 29–39. <https://doi.org/10.51878/learning.v6i1.8903>
- Habibi, A., Herianto, E., & Istiningsih, S. (2024). Penerapan tes diagnostik digital sebagai pemantik pembelajaran berdiferensiasi dalam implementasi kurikulum merdeka. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 7(3), 2964–2968. <https://doi.org/10.54371/jiip.v7i3.4163>
- Handayani, L., Rusilowati, A., Mindyarto, B. N., Astuti, B., Sutikno, S., Prayitno, W. S. W., Kusuma, G. S., Maulana, H. C., Salsabila, A. N., & Vivadi, M. R. (2023). Peningkatan keterampilan menyusun stimulus asesmen literasi numerasi guru IPA Kabupaten Banjarnegara. *E-Dimas: Jurnal Pengabdian kepada Masyarakat*, 14(4), 842–850. <https://doi.org/10.26877/e-dimas.v14i4.17344>
- 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 ITEMAN. *KEMBARA: Journal of Scientific Language Literature and Teaching*, 8(1), 160–180. <https://doi.org/10.22219/kembara.v8i1.20530>
- Listeani, F. Y., Safi'i, I., & Ibrahim, N. (2021). Kelayakan instrumen evaluasi buku sekolah elektronik (BSE) bahasa Indonesia siswa SMA kelas XI. *DIGLOSLIA: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 4(3), 375–386. <https://doi.org/10.30872/diglosia.v4i3.216>
- Magdalena, I., Hidayati, N., Dewi, R. H., Septiara, S. W., & Maulida, Z. (2023). Pentingnya evaluasi dalam proses pembelajaran dan akibat memanipulasinya. *MASALIQ*, 3(5), 810–823. <https://doi.org/10.58578/masaliq.v3i5.1379>
- Mawarni, W. T., Alfiansyah, M., Chaniago, F. Z., & Azzacky, F. (2023). Urgensi evaluasi pendidikan dalam proses pembelajaran menurut Tafsir Al-Misbah Q.S Al-Hasyr ayat 18-19. *ANTHOR: Education and Learning Journal*, 2(4), 494–499. <https://doi.org/10.31004/anthor.v2i4.197>
- Mayuni, N. P. W., Yudana, I. M., & Sulindawati, N. L. G. E. (2025). Optimalisasi rapor pendidikan untuk peningkatan kualitas pembelajaran di sekolah penggerak. *Journal of Education Action Research*, 9(2), 298–306. <https://doi.org/10.23887/jeaar.v9i2.91955>
- Mtd, N. P., Butarbutar, M. I., Sinulingga, S. A. B., Marpaung, J. R., & Harahap, R. M. (2023). Pentingnya evaluasi dalam pembelajaran dan akibat memanipulasinya. *Dewantara: Jurnal Pendidikan Sosial Humaniora*, 2(1), 249–261. <https://doi.org/10.3060/dewantara.v2i1.722>
- Murniatie, I. U., Wahyudi, D. A., Siswiyanti, F., Hidayah, L., Wicaksono, H., & Sinarsih, S. (2025). Kualitas soal ujian akhir semester pada mata pelajaran bahasa Indonesia di



- sekolah menengah atas. *Indonesian Language Education and Literature*, 11(1), 163–163. <https://doi.org/10.24235/ileal.v11i1.20473>
- Nugrahini, Y., Rakhmawati, I., & Asriyanti, F. D. (2023). Pendampingan dan pelatihan membuat media pembelajaran dengan aplikasi POWTOON pada guru-guru sekolah dasar di lingkungan Kabupaten Tulungagung. *Journal of Human and Education (JAHE)*, 3(4), 178–184. <https://doi.org/10.31004/jh.v3i4.447>
- Prismanata, Y. (2025). Instrumen tes diagnostik berbasis multiple intelligences dalam implementasi pembelajaran berdiferensiasi: Sebuah tinjauan literatur sistematis. *Journal of Instructional Technology*, 6(2), 60–60. <https://doi.org/10.20527/j-instech.v6i2.16292>
- Purnama, B. J. (2022). Workshop teknik kelompok sebagai strategi efektif meningkatkan kompetensi guru dalam penyusunan instrumen penilaian. *Ideguru: Jurnal Karya Ilmiah Guru*, 7(3), 308–316. <https://doi.org/10.51169/ideguru.v7i3.445>
- Rahmawati, I., Suryana, Y., & Hidayat, S. (2021). Analisis kesesuaian soal penilaian tengah semester IPA dengan kaidah penyusunan soal pada aspek bahasa di sekolah dasar. *EDUKATIF: JURNAL ILMU PENDIDIKAN*, 3(6), 3636–3646. <https://doi.org/10.31004/edukatif.v3i6.975>
- Riskiani, R., & Suryadi, Y. (2026). Analisis penerapan problem based learning dalam menstimulus berpikir kritis IPS di SMP. *SOCIAL: Jurnal Inovasi Pendidikan IPS*, 6(2), 669–680. <https://doi.org/10.51878/social.v6i2.9916>
- Ratnawati, R., & Hasanah, M. (2021). Peran guru dalam meningkatkan motivasi belajar siswa selama pandemi Covid-19. *PAEDAGOGY: Jurnal Ilmu Pendidikan dan Psikologi*, 1(1), 60–69. <https://doi.org/10.51878/paedagogy.v1i1.339>
- Sudianto, S., & Kisno, K. (2021). Potret kesiapan guru sekolah dasar dan manajemen sekolah dalam menghadapi asesmen nasional. *Jurnal Akuntabilitas Manajemen Pendidikan*, 9(1), 85–97. <https://doi.org/10.21831/jamp.v9i1.39260>
- Syarifah, L. L., Yenni, Y., & Dewi, W. K. (2020). Analisis soal-soal pada buku ajar matematika siswa kelas XI ditinjau dari aspek kognitif. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 4(2), 1259–1272. <https://doi.org/10.31004/cendekia.v4i2.335>
- Trilestari, R. A., Ardana, N. W., Wulandari, M. D., Desideria, M., Putri, I. W. S., Ramadhan, R., Hakim, L., & Pratiwi, V. (2025). Menilai mutu instrumen penelitian: Analisis butir soal lingkup kerja akuntansi fase E berbasis aplikasi Anates. *ALACRITY: Journal of Education*, 1223–1235. <https://doi.org/10.52121/alacrity.v5i3.923>
- Valen, A., & Satria, T. G. (2021). Analisis tingkat kesulitan soal PAS (penilaian akhir semester) mata pelajaran IPS di sekolah dasar. *Jurnal Basicedu*, 5(4), 2199–2208. <https://doi.org/10.31004/basicedu.v5i4.1149>
- Yuza, S. F., & Jannah, M. (2026). Kajian meta-analisis: Indikator prioritas rapor pendidikan dalam kurikulum merdeka. *LEARNING: Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran*, 6(1), 548–563. <https://doi.org/10.51878/learning.v6i1.9216>