



P-ISSN 2620-343X E-ISSN 2986-3074

Sabdasastra : Jurnal Pendidikan Bahasa Jawa

URL : <https://jurnal.uns.ac.id/sab/article/view/103398>

DOI: <https://doi.org/10.20961/sabpbj.v10i1.103398>

**STUDI EVALUATIF TERHADAP BUTIR SOAL UAS URAIAN
BAHASA JAWA KELAS VIII SMP DI KABUPATEN SUKOHARJO**

***EVALUATIVE STUDY OF JAVANESE LANGUAGE ESSAY
QUESTIONS IN THE EIGHTH GRADE FINAL EXAM IN SUKOHARJO
REGENCY***

Fadhilah Rizky Azari*, Astiana Ajeng Rahadini

Pendidikan Bahasa Jawa, Fakultas Keguruan dan Ilmu Pendidikan, Universitas
Sebelas Maret, Surakarta, Indonesia

**Corresponding author:* Fadhilah Rizky Azari dhilaazr@gmail.com

Submitted: 03/06/25

Accepted: 06/03/26

Published: 26/03/26

Abstrak

Pentingnya kegiatan evaluasi pembelajaran harus diimbangi dengan menggunakan instrumen evaluasi yang memiliki kualitas baik. Penelitian mengenai analisis butir soal dilaksanakan dengan tujuan untuk mengetahui kualitas butir soal yang digunakan berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan tingkat daya beda pada butir soal UAS uraian Bahasa Jawa kelas VIII SMP di Kabupaten Sukoharjo tahun ajaran 2023/2024. Penelitian menggunakan metode deskriptif kuantitatif dengan data berupa naskah soal, kisi-kisi, kunci jawaban, pedoman penskoran, dan 64 lembar jawaban siswa yang dipilih melalui *simple random sampling*. Analisis dilakukan menggunakan pendekatan Teori Tes Klasik, meliputi uji validitas (*Product Moment*), reliabilitas (*Alpha Cronbach*), tingkat kesukaran, dan daya pembeda. Hasil menunjukkan seluruh butir soal tidak valid (r hitung < r tabel pada taraf 0,05), meskipun instrumen memiliki reliabilitas sangat tinggi (0,990). Seluruh butir tergolong mudah dan sebagian besar memiliki daya beda rendah, bahkan negatif. Temuan ini menunjukkan instrumen belum mampu mengukur kompetensi siswa secara akurat maupun membedakan tingkat kemampuan siswa, sehingga perlu revisi menyeluruh sebelum digunakan sebagai instrumen evaluasi resmi.

Kata kunci: Analisis butir soal; Reliabilitas; Tingkat daya beda; Tingkat kesukaran; Validitas.

Abstract

The importance of learning evaluation must be supported by the use of high-quality assessment instruments. This study aims to analyze the quality of essay test items based on their validity, reliability, difficulty level, and discrimination index in the Grade VIII Javanese Language Final Semester Examination in Sukoharjo Regency for the 2023/2024 academic year. The research employed a quantitative descriptive method. The data consisted of test questions, test blueprints, answer keys, scoring rubrics, and 64 students' answer sheets selected through simple random sampling. The analysis was conducted using the Classical Test Theory approach, including validity testing (Product Moment correlation), reliability testing (Cronbach's Alpha), item difficulty analysis, and discrimination index analysis. The results revealed that all test items were invalid ($r\text{-count} < r\text{-table}$ at the 0.05 significance level), although the instrument demonstrated very high reliability (0.990). All items were categorized as easy, and most showed low or even negative discrimination indices. These findings indicate that the instrument has not accurately measured students' competencies nor effectively differentiated students' ability levels. Therefore, a comprehensive revision of the test construction is necessary before it is used as an official evaluation instrument.

Keywords: *Difficulty level; Discrimination index; Item analysis; Reliability; Validity.*

Sitasi: Azari, Fadhilah Rizky & Rahadini, Astiana Ajeng. (2026). Studi evaluatif terhadap butir soal UAS uraian bahasa Jawa kelas VIII SMP di Kabupaten Sukoharjo. *Sabdasastra : Jurnal Pendidikan Bahasa Jawa*, 10(1), 75–81. DOI: <https://doi.org/10.20961/sabpbj.v10i1.103398>

PENDAHULUAN

Evaluasi merupakan proses sistematis untuk menilai ketercapaian tujuan suatu kegiatan yang telah dilaksanakan (Anwar, 2021). Dalam konteks pendidikan, evaluasi menjadi komponen esensial untuk mengukur efektivitas pembelajaran dan ketercapaian kompetensi peserta didik dalam satu periode tertentu. Evaluasi dapat dilakukan pada tahap awal, proses, maupun akhir pembelajaran melalui berbagai bentuk tes dan penilaian (Phafiandita et al., 2022). Evaluasi yang dirancang secara tepat tidak hanya memberikan gambaran capaian hasil belajar siswa, tetapi juga menjadi dasar refleksi guru dalam menentukan strategi tindak lanjut pembelajaran (Santosa & Badawi, 2022). Dengan demikian, kualitas instrumen evaluasi sangat menentukan akurasi informasi yang diperoleh mengenai kemampuan siswa.

Secara umum, tes dalam evaluasi pembelajaran terbagi menjadi tes objektif dan tes subjektif. Tes subjektif, khususnya dalam bentuk soal uraian, dinilai lebih mampu mengungkap kemampuan berpikir kritis, penalaran, serta kemampuan siswa dalam mengorganisasi gagasan secara sistematis (Wiggins & McTighe, 2021). Namun demikian, berbagai penelitian menunjukkan bahwa soal uraian memiliki potensi kelemahan, antara lain cakupan materi yang terbatas, subjektivitas dalam penskoran (Sanusi & Aziez, 2021), serta rendahnya validitas dan reliabilitas apabila tidak disusun dengan kaidah yang tepat (Zamzania & Aristia, 2018). Kondisi tersebut menunjukkan bahwa penyusunan soal uraian memerlukan pengujian kualitas melalui analisis butir soal agar benar-benar berfungsi sebagai alat ukur yang akurat.

Analisis butir soal merupakan langkah penting untuk menjamin mutu instrumen evaluasi (Supandi & Farikhah, 2016). Melalui analisis tersebut, dapat

diketahui karakteristik butir soal berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda (Nurjanah, 2017). Validitas berfungsi memastikan bahwa instrumen benar-benar mengukur kompetensi yang seharusnya diukur (Kurniawan, 2015), sedangkan reliabilitas menunjukkan konsistensi hasil pengukuran (Sofiana, 2010).

Selain itu, tingkat kesukaran yang proporsional menjadi indikator penting dalam menentukan kualitas butir soal. Nur dan Palobo (2018) menyatakan bahwa butir soal dengan tingkat kesukaran yang seimbang akan menghasilkan pengukuran kemampuan siswa yang lebih akurat dan representatif. Di sisi lain, daya pembeda yang baik juga menjadi syarat utama agar suatu butir soal evaluasi mampu untuk mengklasifikasikan siswa berdasarkan tingkat kemampuannya. Sari et al. (2022) menegaskan bahwa soal dengan daya pembeda tinggi dapat membedakan secara jelas antara siswa berkemampuan tinggi dan rendah, sehingga hasil evaluasi lebih objektif dan tepat sasaran.

Meskipun berbagai penelitian telah mengkaji analisis butir soal, sebagian besar studi masih berfokus pada tes objektif, khususnya pilihan ganda, serta pada mata pelajaran umum seperti matematika dan IPA. Penelitian mengenai analisis butir soal uraian, terutama pada mata pelajaran Bahasa Jawa, masih relatif terbatas. Selain itu, kajian yang secara komprehensif menganalisis keempat aspek kualitas (validitas, reliabilitas, tingkat kesukaran, dan daya pembeda) pada soal uraian Ujian Akhir Semester Bahasa Jawa juga belum banyak ditemukan. Padahal, karakteristik mata pelajaran Bahasa Jawa yang menekankan aspek kebahasaan, kesastraan, dan kemampuan ekspresif menuntut instrumen evaluasi yang benar-benar teruji kualitasnya.

Berdasarkan kondisi tersebut, terdapat kebutuhan mendesak untuk melakukan analisis mendalam terhadap kualitas butir soal uraian pada Ujian Akhir Semester Bahasa Jawa. Penelitian ini diharapkan dapat mengisi kekosongan kajian empiris terkait mutu instrumen evaluasi pada mata pelajaran Bahasa Jawa sekaligus memberikan kontribusi praktis bagi guru dalam menyusun soal uraian yang lebih valid, reliabel, dan mampu mengukur kompetensi siswa secara akurat.

METODE PENELITIAN

Penelitian ini menggunakan metode deskriptif kuantitatif untuk menganalisis kualitas butir soal uraian UAS Bahasa Jawa kelas VIII SMP tahun pelajaran 2023/2024 di Kabupaten Sukoharjo. Data diperoleh melalui dokumentasi berupa naskah soal, kisi-kisi, kunci jawaban, pedoman penskoran, dan 64 lembar jawaban siswa yang dipilih dengan teknik simple random sampling karena populasi relatif homogen.

Instrumen penelitian berupa lembar analisis butir soal berbasis Teori Tes Klasik. Analisis data dilakukan secara kuantitatif dengan menghitung validitas (Product Moment), reliabilitas (Alpha Cronbach), tingkat kesukaran (proporsi skor rata-rata terhadap skor maksimum), dan daya beda (perbandingan kelompok atas dan bawah). Keempat aspek tersebut dianalisis mengacu pada rumus analisis butir

soal menurut Nurgiyantoro (2016: 226). Validasi data dilakukan melalui triangulasi dokumen dan pengecekan ulang hasil perhitungan.

HASIL PENELITIAN DAN PEMBAHASAN

Validitas

Hasil analisis menunjukkan bahwa kelima butir soal uraian memiliki koefisien korelasi di bawah r tabel pada taraf signifikansi 0,05 sehingga seluruhnya dinyatakan tidak valid. Artinya, butir soal belum mampu mengukur kompetensi yang seharusnya diukur secara tepat. Temuan ini berbeda dengan hasil penelitian Ida dan Musyarofah (2021) yang melaporkan bahwa “sebagian besar butir soal yang dianalisis telah memenuhi kriteria validitas dengan kategori sedang hingga tinggi setelah melalui uji korelasi.” Demikian pula, penelitian Nurjanah (2017) menyatakan bahwa lebih dari 60% butir soal yang dianalisis berada pada kategori valid dan layak digunakan setelah dilakukan uji empiris.

Perbedaan hasil tersebut menunjukkan bahwa soal uraian UAS Bahasa Jawa dalam penelitian ini memiliki kelemahan mendasar pada aspek konstruksi dan kesesuaian indikator. Jika pada penelitian sebelumnya validitas tercapai karena keselarasan antara kisi-kisi, indikator, dan rumusan soal, maka dalam penelitian ini ketidaksesuaian tersebut diduga menjadi penyebab rendahnya koefisien validitas.

Reliabilitas

Hasil pengukuran reliabilitas dalam penelitian ini menghasilkan koefisien reliabilitas sebesar 0,990 yang menunjukkan kategori sangat tinggi. Hal ini sejalan dengan pendapat Solichin (2017) yang menyatakan bahwa instrumen dengan koefisien reliabilitas di atas 0,70 tergolong sangat baik dan memiliki konsistensi internal yang kuat. Penelitian Sanusi dan Aziez (2021) juga menemukan bahwa instrumen dengan reliabilitas tinggi cenderung menghasilkan pengukuran yang stabil meskipun digunakan pada waktu berbeda.

Namun, berbeda dengan penelitian-penelitian tersebut yang menunjukkan reliabilitas tinggi disertai validitas yang memadai, penelitian ini justru menemukan reliabilitas sangat tinggi, tetapi seluruh butir tidak valid. Kondisi ini memperkuat teori bahwa reliabilitas tidak menjamin validitas. Tingginya reliabilitas dalam penelitian ini diduga karena bentuk soal yang struktural dan homogen sehingga variasi skor siswa relatif kecil, bukan karena kualitas konstruk yang baik.

Tingkat kesukaran

Tabel 1. Tingkat kesukaran

No	Butir soal	Tingkat kesukaran soal
1.	1	1,00
2.	2	0,93
3.	3	0,76
4.	4	0,88

Hasil analisis menunjukkan bahwa seluruh butir soal berada pada kategori mudah (0,74–1,00). Temuan ini tidak sejalan dengan rekomendasi **Putriani et al. (2020)** yang menyatakan bahwa komposisi ideal butir soal sebaiknya mencakup variasi mudah, sedang, dan sukar agar mampu mengukur kemampuan siswa secara komprehensif. Penelitian **Alti et al. (2023)** juga menegaskan bahwa instrumen dengan proporsi tingkat kesukaran yang seimbang akan menghasilkan distribusi skor yang lebih representatif.

Berbeda dengan hasil penelitian tersebut, dominasi soal kategori mudah dalam penelitian ini menyebabkan distribusi nilai siswa cenderung tinggi dan homogen. Hal ini berdampak pada rendahnya kemampuan instrumen dalam memetakan perbedaan capaian belajar siswa.

Tingkat daya beda

Tabel 2. Tingkat daya beda

No	Butir soal	Tingkat daya beda soal
1.	1	0,00
2.	2	-0,07
3.	3	0,49
4.	4	0,07
5.	5	0,15

Hasil analisis menunjukkan empat butir soal memiliki daya beda di bawah 0,20 bahkan terdapat nilai negatif, sedangkan hanya satu butir berada pada kategori baik (0,49). Menurut **Muluki (2020)**, butir soal dengan daya beda di bawah 0,20 tergolong jelek dan sebaiknya direvisi atau dibuang karena tidak mampu membedakan kemampuan siswa secara efektif. Penelitian **Naibaho et al. (2021)** juga menemukan bahwa butir soal yang baik umumnya memiliki daya beda minimal 0,30 agar dapat mengklasifikasikan siswa kelompok atas dan bawah secara jelas.

Dibandingkan dengan hasil penelitian tersebut, rendahnya daya beda dalam penelitian ini menunjukkan bahwa sebagian besar butir soal belum memenuhi fungsi diskriminatifnya. Kondisi ini selaras dengan tingkat kesukaran yang terlalu mudah, sehingga hampir seluruh siswa memperoleh skor tinggi dan perbedaan kemampuan tidak tampak signifikan.

KESIMPULAN

Berdasarkan hasil analisis, seluruh butir soal uraian UAS Bahasa Jawa dinyatakan tidak valid meskipun memiliki reliabilitas sangat tinggi (0,990). Selain itu, seluruh butir berada pada kategori mudah dan sebagian besar memiliki daya beda rendah. Temuan ini menunjukkan bahwa instrumen evaluasi belum mampu mengukur kompetensi siswa secara tepat maupun membedakan kemampuan siswa secara optimal. Tingginya reliabilitas yang tidak diikuti validitas menegaskan bahwa konsistensi skor tidak menjamin ketepatan konstruk pengukuran.

Implikasinya, soal uraian yang digunakan belum efektif sebagai alat evaluasi hasil belajar dan berpotensi menghasilkan pemetaan capaian siswa yang kurang akurat. Oleh karena itu, guru perlu melakukan revisi menyeluruh terhadap konstruksi soal dengan memperhatikan kesesuaian indikator, variasi tingkat kesukaran, serta peningkatan daya beda. Disarankan pula agar penyusunan soal didahului dengan telaah kisi-kisi yang cermat dan diikuti uji coba serta analisis butir soal sebelum digunakan sebagai instrumen evaluasi resmi.

DAFTAR PUSTAKA

- Alti, R. P., Zulyuri, Z., & Violita, V. (2023). Analisis kualitas butir soal ujian tengah semester mata pelajaran biologi kelas X di MAN 1 Solok Selatan. *Jurnal Metaedukasi: Jurnal Ilmiah Pendidikan*, 4(2), 70–75.
- Anwar, K. (2021). Urgensi evaluasi dalam proses pembelajaran. *Rausyan Fikr: Jurnal Pemikiran Dan Pencerahan*, 17(1), 108–118.
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan reliabilitas dalam analisis butir soal. *Al-Mu'Arrib: Journal of Arabic Education*, 1(1), 34–44.
- Kurniawan, T. (2015). Analisis butir soal ulangan akhir semester gasal mata pelajaran IPS Sekolah Dasar (Analysis of Odd Semester Final Test Items in Elementary School of Social Studies Subjects). *Journal of Elementary Education*, 4(1), 2.
- Muluki, A. (2020). Analisis kualitas butir tes semester ganjil mata pelajaran IPA kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86.
- Naibaho, M.A. U., Taroreh, M. L. G., & Kapantow, G. H. M. (2021). Kajian keberlanjutan program Yayasan Pembangunan Berkelanjutan Sulawesi Utara

- (Ypbsu) pasca penutupan tambang Pt. Newmont Minahasa Raya (Pt. Nmr) di Kecamatan Ratatotok Minahasa Tenggara. *Agri-Sosioekonomi*, 17(2), 177.
- Nur, A. S., & Palobo, M. (2018). Pelatihan analisis butir soal berbasis komputerisasi pada guru SD. *MATAPPA: Jurnal Pengabdian Kepada Masyarakat*, 1(1), 5.
- Nurgiyantoro, B. (2017). Penilaian pembelajaran bahasa berbasis kompetensi. Yogyakarta: BPFE-Yogyakarta.
- Nurjanah, N. (2017). Analisis butir soal pilihan ganda dari aspek kebahasaan. *Faktor Jurnal Ilmiah Kependidikan*, 2(1), 69–78.
- Phafiandita, A. N., Permadani, A., Pradani, A. S., & Wahyudi, M. I. (2022). Urgensi evaluasi pembelajaran di Kelas. *JIRA: Jurnal Inovasi Dan Riset Akademik*, 3(2), 111–121.
- Putriani, D., Turahmah, F., Sunarti, N. R., Ismarliana, P., & Walid, A. (2020). Analisis butir soal UAS Biologi 2018/2019 kelas X dan XI SMAN 11 Kota Bengkulu. *Journal of Biology Learning*, 2(1), 1–7.
- Santosa, S., & Badawi, J. A. (2022). Analisis butir soal pilihan ganda tema pertumbuhan dan perkembangan makhluk hidup Kelas III Madrasah Ibtidaiyah. *Jurnal Basicedu*, 6(2), 1678–1686.
- Sanusi, R. N. A., & Aziez, F. (2021). Analisis butir soal tes objektif dan subjektif untuk keterampilan membaca pemahaman pada kelas VII SMP N 3 Kalibagor. *Metafora: Jurnal Pembelajaran Bahasa Dan Sastra*, 8(1), 99.
- Sari, V. N. I., Utomo, A. P. Y., & Sumarwati. (2022). Kualitas soal bahasa indonesia di SMP Muhammadiyah 1 Pontianak: Analisis butir soal. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 11(2), 112–119.
- Sofiana, S. (2010). *Analisis butir soal ulangan kenaikan kelas mata pelajaran kimia kelas X SMA Negeri 8 Surakarta tahun ajaran 2009/2010*. (Skripsi) Universitas Sebelas Maret.
- Solichin, M. (2017). Analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Dirāsāt: Jurnal Manajemen & Pendidikan Islam*, 2(2), 192–213.
- Supandi, S., & Farikhah, L. (2016). Analisis butir soal matematika pada instrumen uji coba materi segitiga. *JIPMat*, 1(1), 71–78.
- Wiggins, G., & McTighe, J. (2021). *Understanding by design*. ASCD.
- Zamzania, A. W. H., & Aristia, R. (2018). Jenis-jenis instrumen dalam evaluasi pembelajaran. *Universitas Muhammadiyah Sidoarjo*, 1–13.