

WIDYA AKSARA

Jurnal Agama Hindu

Volume 30 Nomor 1 Maret 2025

ANALISIS VALIDITAS RELIABILITAS TINGKAT KESUKARAN DAYA PEMBEDA SOAL ASESMEN SUMATIF BAHASA JAWA SEKOLAH DASAR (SD) DI KABUPATEN KENDAL

Oleh:

Trimio

STAHN Jawa Dwipa Klaten Jawa Tengah

Email: elangtrio@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi kualitas soal Asesmen Sumatif mata pelajaran Bahasa Jawa untuk tingkat Sekolah Dasar di Kabupaten Kendal tahun ajaran 2023/2024. Fokus penelitian adalah menganalisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal-soal tersebut. Metode yang digunakan adalah kuantitatif dengan pendekatan deskriptif. Data diperoleh dari dokumen soal dan jawaban siswa, kemudian dianalisis menggunakan teknik analisis butir soal. Hasil penelitian menunjukkan bahwa soal-soal tersebut memiliki validitas tinggi, dengan mayoritas butir soal memenuhi kriteria yang ditetapkan. Reliabilitas soal juga termasuk dalam kategori baik, menunjukkan konsistensi dalam mengukur kemampuan siswa. Tingkat kesukaran soal sebagian besar berada pada kategori sedang, sehingga cocok untuk evaluasi pembelajaran. Daya pembeda soal juga cukup baik, dengan kemampuan membedakan siswa berkemampuan tinggi dan rendah. Berdasarkan temuan ini, dapat disimpulkan bahwa soal Asesmen Sumatif Bahasa Jawa di Kabupaten Kendal telah memenuhi standar kualitas yang baik. Namun, beberapa butir soal masih perlu direvisi untuk meningkatkan efektivitasnya. Penelitian ini merekomendasikan agar penyusun soal lebih memperhatikan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dalam menyusun soal evaluasi di masa depan.

Kata Kunci : asesmen, validitas, reliabilitas, tingkat kesukaran, daya pembeda

ABSTRACT

The aim of the research is to evaluate the quality of the summative assessment questions for the Javanese Language subject at the elementary school level in Kendal Regency in the academic year 2023/2024. The focus of the research is to analyze the validity, reliability, difficulty level, and the discriminating power of the questions. The method of the study belongs to quantitative with a descriptive approach. Data were obtained from question documents and student answer sheets, which were then analyzed using item analysis techniques. The results of the study show that the questions have high validity with the majority of the items meeting the established criteria. The reliability of the questions is also in a good category indicating consistency to measure students' abilities. Most of the questions have a moderate difficulty level which makes them suitable for learning evaluation. Additionally, the discriminating power of the questions is fairly good with the ability to differentiate between high-ability and low-ability students. Based on the research findings, it can be concluded that the Javanese Language Summative Assessment questions in Kendal Regency have met good quality

standards. However, some items still need to be revised to improve their effectiveness. This study recommends that question designers should pay more attention to aspects of validity, reliability, difficulty level as well as discriminating power to prepare evaluation questions in the future.

Keywords: assessment, validity, reliability, difficulty level, discriminating power

I. PENDAHULUAN

Asesmen sumatif merupakan salah satu aspek penting dalam proses pendidikan. Salah satu bentuk evaluasi yang sering digunakan di sekolah dasar (SD) adalah asesmen sumatif yang berfungsi sebagai alat untuk mengukur pencapaian kompetensi siswa dalam suatu mata pelajaran. Dalam konteks pendidikan di Kabupaten Kendal, mata pelajaran Bahasa Jawa memiliki peran penting dalam mempertahankan serta mengembangkan budaya dan kearifan lokal. Oleh karena itu, kualitas soal asesmen sumatif Bahasa Jawa di SD perlu dianalisis secara mendalam untuk memastikan efektivitasnya sebagai alat ukur yang valid dan reliabel.

Penyusunan soal yang baik harus memenuhi beberapa kriteria utama, seperti validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Validitas mengacu pada sejauh mana soal mampu mengukur apa yang seharusnya diukur, sedangkan reliabilitas berkaitan dengan konsistensi hasil yang diperoleh ketika soal digunakan dalam berbagai kondisi yang sama. Tingkat kesukaran soal harus disesuaikan dengan kemampuan siswa, sehingga tidak terlalu sulit atau terlalu mudah. Daya pembeda menjadi aspek penting lainnya karena menentukan kemampuan soal dalam membedakan siswa yang memiliki pemahaman tinggi dan rendah terhadap materi.

Validitas dan reliabilitas merupakan dua aspek mendasar dalam penyusunan instrumen evaluasi. Menurut Sugiyono (2019), validitas suatu soal dapat diukur melalui berbagai cara, seperti validitas isi, validitas konstruk, dan validitas empiris. Soal yang valid harus sesuai dengan kompetensi yang diukur serta mencerminkan tujuan pembelajaran yang telah ditetapkan. Sementara itu, reliabilitas mengacu pada sejauh mana suatu instrumen memberikan hasil yang konsisten ketika digunakan berulang kali dalam kondisi yang sama (Arikunto, 2018). Soal yang reliabel akan memberikan hasil yang relatif stabil meskipun digunakan dalam waktu dan kelompok yang berbeda.

Tingkat kesukaran soal juga menjadi aspek yang harus diperhatikan dalam penyusunan soal asesmen sumatif. Menurut penelitian Tritantining (2013) menunjukkan bahwa bentuk-bentuk kesalahan berbahasa Jawa pada soal-soal ulangan akhir semester terdapat pada tataran linguistik fonologi, morfologi, sintaksis (frasa, klausa, kalimat), dan semantik. Penelitian Setiawan dan Wulandari (2021), menemukan bahwa soal yang terlalu mudah atau terlalu sulit tidak efektif dalam mengukur kemampuan siswa secara optimal. Oleh karena itu, tingkat kesukaran soal harus disesuaikan dengan kemampuan siswa yang diuji agar hasil evaluasi lebih akurat. Daya pembeda, di sisi lain, berfungsi untuk mengetahui sejauh mana suatu soal mampu membedakan siswa yang memiliki pemahaman tinggi dan rendah. Penelitian oleh Rahmawati (2022) menegaskan bahwa daya pembeda yang baik akan meningkatkan efektivitas soal dalam mengukur kemampuan siswa secara objektif.

Berdasarkan observasi awal, terdapat indikasi bahwa beberapa soal asesmen sumatif Bahasa Jawa di SD di Kabupaten Kendal belum sepenuhnya memenuhi standar kualitas yang baik. Beberapa guru melaporkan bahwa soal yang digunakan terkadang terlalu sulit bagi siswa atau justru terlalu mudah, sehingga kurang efektif dalam mengukur kompetensi siswa secara menyeluruh. Selain itu, masih ditemukan adanya soal yang tidak memiliki daya pembeda yang baik, yang berakibat pada kurang optimalnya proses evaluasi hasil belajar siswa.

Dengan mempertimbangkan berbagai teori dan penelitian terdahulu, analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal asesmen sumatif Bahasa Jawa di SD di Kabupaten Kendal menjadi sangat penting. Hasil dari penelitian ini diharapkan dapat

memberikan kontribusi bagi para pendidik dalam menyusun dan mengevaluasi soal yang lebih berkualitas, sehingga proses pembelajaran Bahasa Jawa di sekolah dasar dapat berlangsung secara lebih efektif dan bermakna.

Oleh karena itu, diperlukan analisis lebih lanjut untuk memastikan bahwa soal asesmen sumatif Bahasa Jawa yang digunakan di SD di Kabupaten Kendal memenuhi standar kualitas yang baik dengan menganalisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis deskriptif. Subjek penelitian adalah soal asesmen sumatif mata pelajaran Bahasa Jawa tingkat Sekolah Dasar (SD) di Kabupaten Kendal. Populasi penelitian sejumlah 53.191 siswa SD dari kelas I s.d. VI di Kabupaten Kendal. Teknik pengambilan sampel menggunakan rumus Isaac dan Michael (1995) yang sudah ditabelkan. Untuk populasi 53.191 dengan tingkat kepercayaan 95% dan tingkat error 5% sejumlah 346. Setelah ditentukan jumlah sampel, kemudian dilanjutkan dengan teknik *proportional random sampling* tiap-tiap kelas. Sampel kelas I sebanyak 62, kelas II 56, kelas III 59, kelas IV 54, kelas V 57, dan kelas VI 58 p siswa. Teknik pengumpulan data menggunakan studi dokumen yang berupa soal asesmen sumatif dan hasil pengerjaan siswa. Data yang dianalisis mencakup validitas *product moment*, reliabilitas *alpha*, tingkat kesukaran, dan daya pembeda dengan menggunakan *Statistical Product and Service Solutions* (SPSS) versi 25. Hasil analisis data disajikan dalam bentuk tabel dan deskripsi statistik sejauh mana soal asesmen sumatif Bahasa Jawa SD di Kabupaten Kendal telah memenuhi standar kualitas evaluasi yang baik.

II. PEMBAHASAN

1. Analisis Validitas

Validitas sebuah soal dinyatakan baik jika memiliki koefisien validitas yang tinggi, yang biasanya diukur dengan angka korelasi. Arikunto (2018) dan Sugiyono (2019) menyatakan bahwa soal dikategorikan valid jika memiliki koefisien validitas dalam rentang 0,40 hingga 1,00, dengan klasifikasi sangat tinggi (0,70 – 1,00), tinggi (0,40 – 0,69), cukup (0,20 – 0,39), rendah (0,00 – 0,19), dan tidak valid (<0,00). Butir soal dinyatakan valid apabila $r_{hitung} > r_{tabel}$ atau $\text{sig 2-tailed} < 0,05$

Berdasarkan analisis data validitas *product moment* menggunakan SPSS versi 25 terhadap soal asesmen sumatif Bahasa Jawa tahun pelajaran 2023/2024, diperoleh sebagai berikut.

Tabel 1 Rata-rata Validitas Soal Asesmen Sumatif Bahasa Jawa

No.	Kelas-N	Jumlah Soal	Valid – Tidak Valid	Rata-rata r_{hitung}	$r_{tabel} - N$	Kategori
1.	I – 62	30	28-2	0,41	0,254	tinggi
2.	II – 56	30	29-1	0,52	0,266	tinggi
3.	III – 59	35	34-1	0,58	0,254	tinggi
4.	IV – 54	50	47-3	0,43	0,266	tinggi
5.	V – 57	50	49-1	0,62	0,266	tinggi
6.	VI - 58	40	40	0,74	0,254	sangat tinggi
Rata-rata				0,55		tinggi

Sumber: Data Penelitian (diolah)

Berdasarkan data tersebut diperoleh rata-rata validitas butir soal asesmen sumatif Bahasa Jawa tiap kelas dalam kategori tinggi bahkan ada yang sangat tinggi. Secara keseluruhan rata-rata validitas butir soal sebesar 0,55 kategori tinggi. Hal ini berarti secara umum soal asesmen sumatif Bahasa Jawa tepat mengukur kemampuan siswa. Soal yang disusun sudah mencerminkan materi yang relevan dengan tujuan evaluasi. Hal tersebut selaras dengan pendapat Allen dan Yen (1979) menegaskan bahwa semakin tinggi korelasi

antara hasil tes dengan konstruk atau kriteria yang relevan, semakin tinggi pula validitas tes tersebut. Anastasi dan Urbina (1997) juga menekankan pentingnya instrumen tidak hanya mengukur hafalan, tetapi juga keterampilan dan pemahaman mendalam.

Untuk menguji validitas secara kuantitatif, metode statistik seperti Product Moment Pearson sering digunakan dalam analisis korelasi antara butir soal dengan skor total. Selain itu, metode *Content Validity Ratio* (CVR) dari Lawshe digunakan untuk mengukur validitas isi berdasarkan pendapat panel ahli. Pendekatan ini didukung oleh Crocker dan Algina (2008), yang menyatakan bahwa kombinasi antara analisis statistik dan expert judgment sangat penting dalam menilai kualitas instrumen tes. Dengan memahami konsep validitas secara komprehensif, pengembang tes dapat memastikan bahwa soal yang digunakan benar-benar mampu mengukur apa yang seharusnya diukur dengan akurat dan dapat dipercaya.

2. Analisis Reliabilitas

Reliabilitas suatu soal menunjukkan sejauh mana instrumen tersebut dapat memberikan hasil yang konsisten dan dapat dipercaya dalam berbagai kondisi pengukuran. Arikunto (2018) menyatakan bahwa reliabilitas yang baik ditandai dengan hasil tes yang stabil ketika diberikan kepada kelompok yang sama dalam waktu berbeda, dengan kategori reliabilitas sangat tinggi (0,80 – 1,00), tinggi (0,60 – 0,79), cukup (0,40 – 0,59), rendah (0,20 – 0,39), dan sangat rendah (<0,20). Azwar (2018) menambahkan bahwa reliabilitas bukan hanya sekadar angka statistik, tetapi juga harus didukung dengan perancangan soal yang baik, di mana setiap butir soal harus memiliki keseragaman dalam mengukur aspek yang diinginkan serta tidak menimbulkan bias pada kelompok tertentu. Sementara itu, Sugiyono (2017) menekankan bahwa reliabilitas dapat ditingkatkan dengan memperbanyak jumlah butir soal yang berkualitas serta melakukan uji coba pada sampel yang representatif agar hasil pengukurannya lebih akurat dan dapat dipercaya.

Berdasarkan analisis data reliabilitas *alpha Cronbach* menggunakan SPSS versi 25 terhadap soal asesmen sumatif Bahasa Jawa tahun pelajaran 2023/2024, diperoleh angka sebagai berikut.

Tabel 2 Reliabilitas Soal Asesmen Sumatif Bahasa Jawa

No.	Kelas-N	Koefisien reliabilitas	Kategori
1.	I – 62	0,60	tinggi
2.	II – 56	0,62	tinggi
3.	III – 59	0,61	tinggi
4.	IV – 54	0,60	cukup
5.	V – 57	0,72	tinggi
6.	VI - 58	0,81	sangat tinggi
	Rata-rata	0,66	tinggi

Mengacu data tersebut diketahui bahwa secara keseluruhan soal asesmen sumatif Bahasa Jawa dari kelas I s.d. VI memperoleh rata-rata koefisien reliabilitas sebesar 0,66 dalam kategori tinggi. Hal ini berarti soal-soal asesmen sumatif Bahasa Jawa yang digunakan untuk memperoleh informasi penguasaan materi dapat dipercaya sebagai alat pengumpulan data dan mampu mengungkapkan informasi yang sebenarnya di sekolah. Hal tersebut selaras dengan pendapat Nunnally dan Bernstein (1994), bahwa instrumen dengan reliabilitas tinggi akan memberikan hasil yang stabil dalam berbagai kondisi pengukuran. Selain itu, Feldt dan Brennan (1989) menyatakan bahwa tingkat reliabilitas juga dipengaruhi oleh jumlah butir soal dalam tes, di mana semakin banyak butir yang berkualitas, semakin tinggi reliabilitasnya. Crocker dan Algina (2008) menambahkan bahwa pemilihan metode pengujian reliabilitas harus disesuaikan dengan jenis instrumen agar hasilnya dapat diinterpretasikan secara akurat. Jika reliabilitas suatu soal rendah, maka

hasil pengukuran tidak dapat dipercaya karena adanya faktor ketidakkonsistenan dalam penyajian soal atau jawaban peserta tes.

3. Analisis Tingkat Kesukaran

Tingkat kesukaran soal adalah aspek penting dalam evaluasi pendidikan, karena berfungsi untuk mengukur seberapa mudah atau sulit suatu soal bagi siswa. Menurut Nitko (1996), tingkat kesukaran soal adalah peluang bagi peserta tes untuk menjawab benar suatu soal pada tingkat kemampuan tertentu, yang biasanya dinyatakan dalam bentuk indeks dengan rentang 0,00 hingga 1,00. Indeks 0,00 menunjukkan soal yang sangat sulit, sedangkan indeks 1,00 menunjukkan soal yang sangat mudah.

Berdasarkan analisis data menggunakan SPSS terhadap soal asesmen sumatif Bahasa Jawa tahun pelajaran 2023/2024, diperoleh sebagai berikut.

Tabel 3 Rata-rata Indeks Kesukaraan Soal Asesmen Sumatif Bahasa Jawa

No.	Kelas	Rata-rata Indeks Tingkat Kesukaran	Keterangan
1.	I	0,52	Sedang
2.	II	0,51	Sedang
3.	III	0,48	Sedang
4.	IV	0,47	Sedang
5.	V	0,58	Sedang
6.	VI	0,69	Sedang
	Rata-rata	0,54	sedang

Sumber: Data Penelitian (diolah)

Dari data tersebut diketahui bahwa secara keseluruhan soal asesmen sumatif Bahasa Jawa yang disusun oleh Dinas Pendidikan Kabupaten Kendal, diperoleh rata-rata 0,54 termasuk kategori sedang, yang artinya efektif digunakan sebagai alat untuk mengukur kemampuan pendidikan. Hal tersebut selaras dengan pendapat Nitko (1996) bahwa soal dengan tingkat kesukaran sedang dianggap paling efektif untuk mengevaluasi kemampuan siswa, karena tidak terlalu mudah sehingga semua siswa dapat menjawabnya dengan benar, dan tidak terlalu sulit sehingga hanya sedikit siswa yang dapat menjawabnya. Dengan demikian, soal kategori sedang dapat membedakan antara siswa yang memiliki pemahaman yang baik dan yang kurang. Lebih lanjut Nitko mengklasifikasikan tingkat kesukaran soal menjadi tiga kategori: soal sukar (indeks 0,00–0,30), soal sedang (indeks 0,31–0,70), dan soal mudah (indeks 0,71–1,00).

Pentingnya mempertimbangkan tingkat kesukaran soal juga ditegaskan oleh Arifin (2012), yang menyatakan bahwa butir soal yang baik adalah soal dengan pengecoh yang dipilih peserta tes secara merata. Hal ini menunjukkan bahwa soal tersebut memiliki tingkat kesukaran yang sesuai dan mampu membedakan kemampuan siswa secara efektif. Oleh karena itu, dalam penyusunan instrumen evaluasi, pendidik harus memastikan bahwa soal-soal yang disusun memiliki tingkat kesukaran yang bervariasi dan seimbang untuk mendapatkan gambaran yang akurat mengenai kemampuan siswa.

4. Analisis Daya Pembeda

Daya pembeda soal merupakan aspek penting dalam evaluasi instrumen tes, karena menunjukkan sejauh mana suatu soal dapat membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Berdasarkan analisis data menggunakan SPSS terhadap soal asesmen sumatif Bahasa Jawa tahun pelajaran 2023/2024, diperoleh sebagai berikut.

Tabel 4. Rata-rata Indeks Kesukaraan Soal Asesmen Sumatif Bahasa Jawa

No.	Kelas	Rata-rata Indeks Daya Pembeda	Keterangan
1.	I	0,58	baik
2.	II	0,49	baik
3.	III	0,59	baik
4.	IV	0,42	baik
5.	V	0,62	baik
6.	VI	0,74	sangat baik
Rata-rata		0,49	baik

Sumber: Data Penelitian (diolah)

Dari data tersebut diketahui bahwa secara keseluruhan soal asesmen sumatif Bahasa Jawa yang disusun oleh Dinas Pendidikan Kabupaten Kendal, diperoleh rata-rata skor 0,49 termasuk kategori baik, yang artinya memiliki daya pembeda yang baik. Hal tersebut selaras dengan pendapat Ebel dan Frisbie (1991), bahwa indeks daya pembeda berkisar antara -1,00 hingga 1,00. Soal dengan daya pembeda negatif (di bawah 0,00) menunjukkan bahwa siswa berkemampuan rendah lebih banyak menjawab benar dibandingkan siswa berkemampuan tinggi, sehingga dianggap bermasalah dan harus direvisi atau dibuang. Soal dengan daya pembeda 0,00 – 0,19 dikategorikan sebagai jelek, karena tidak cukup membedakan siswa berkemampuan tinggi dan rendah. Soal dengan daya pembeda 0,20 – 0,39 dikategorikan cukup, sementara soal dengan daya pembeda 0,40 – 0,69 dianggap baik. Soal dengan daya pembeda 0,70 – 1,00 dikategorikan sebagai sangat baik, karena mampu membedakan siswa dengan efektif.

Daya pembeda yang baik memungkinkan guru atau penyusun tes untuk menilai apakah soal tersebut efektif dalam mengukur perbedaan kemampuan siswa. Soal dengan daya pembeda yang tinggi akan lebih banyak dijawab benar oleh siswa berkemampuan tinggi dan dijawab salah oleh siswa berkemampuan rendah. Sebaliknya, soal yang memiliki daya pembeda rendah atau negatif perlu direvisi atau dihilangkan dari tes, karena tidak memberikan informasi yang akurat tentang kemampuan siswa.

Haladyna dan Downing (1989) menekankan bahwa daya pembeda yang tinggi akan meningkatkan validitas tes dan membantu guru dalam mengidentifikasi kesulitan belajar siswa. Untuk meningkatkan daya pembeda, diperlukan perbaikan dalam perumusan soal, pemilihan pengecoh yang efektif, serta uji coba soal sebelum diterapkan secara luas. Soal dengan daya pembeda yang baik akan menghasilkan tes yang lebih akurat dan adil dalam mengukur kemampuan siswa, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan pendidikan yang lebih tepat.

5. Pembahasan

Dalam penyusunan soal asesmen sumatif Bahasa Jawa, validitas, reliabilitas, tingkat kesukaraan, dan daya pembeda merupakan komponen yang saling berkaitan dan harus dianalisis secara terintegrasi agar menghasilkan soal yang berkualitas. Validitas berperan sebagai dasar utama dalam menentukan apakah soal yang disusun benar-benar mengukur kompetensi siswa dalam bahasa Jawa, seperti pemahaman teks, penggunaan aksara Jawa, atau keterampilan berkomunikasi dalam bahasa tersebut. Namun, validitas tidak akan berarti banyak jika soal yang telah dirancang tidak memiliki reliabilitas yang baik. Soal yang valid tetapi tidak reliabel akan menghasilkan hasil ujian yang inkonsisten, sehingga tidak dapat digunakan sebagai alat ukur yang dapat diandalkan. Oleh karena itu, validitas dan reliabilitas harus diperhitungkan secara bersamaan untuk memastikan bahwa soal tidak hanya sesuai dengan tujuan pembelajaran tetapi juga memberikan hasil yang stabil dan konsisten.

Selain validitas dan reliabilitas, analisis tingkat kesukaran dan daya pembeda juga harus dilakukan dalam satu kesatuan agar soal dapat mengukur kemampuan siswa dengan optimal. Jika suatu soal memiliki validitas tinggi tetapi tingkat kesukarannya terlalu tinggi atau terlalu rendah, maka fungsinya sebagai alat evaluasi menjadi kurang efektif. Tingkat kesukaran yang ideal harus disusun secara proporsional, yaitu dengan kombinasi soal mudah, sedang, dan sulit agar siswa dapat menunjukkan kemampuannya secara bertahap. Di sisi lain, daya pembeda memastikan bahwa soal mampu mengidentifikasi perbedaan tingkat pemahaman siswa. Soal yang terlalu mudah atau terlalu sulit akan memiliki daya pembeda yang rendah, sehingga tidak dapat secara efektif membedakan siswa dengan penguasaan materi yang baik dan kurang baik. Oleh karena itu, analisis tingkat kesukaran dan daya pembeda harus dilakukan secara simultan agar setiap butir soal memberikan informasi yang akurat tentang kemampuan siswa.

Hasil penelitian menunjukkan bahwa soal asesmen sumatif Bahasa Jawa di Kabupaten Kendal memiliki validitas tinggi, reliabilitas baik, tingkat kesukaran sedang, dan daya pembeda yang cukup baik. Temuan ini sejalan dengan pendapat Linn dan Gronlund (2000) yang menekankan pentingnya keempat aspek tersebut dalam penyusunan soal evaluasi. Selain itu, Black dan Wiliam (1998) dalam "*Inside the Black Box*" menekankan bahwa *assessment* harus digunakan sebagai alat untuk meningkatkan pembelajaran. Mereka menyoroti pentingnya soal yang valid dan reliabel untuk memberikan umpan balik yang akurat, tingkat kesukaran yang menantang tetapi tidak terlalu sulit, serta daya pembeda yang baik untuk mengidentifikasi kebutuhan siswa. Dengan memenuhi kriteria ini, *assessment* tidak hanya mengukur kemampuan siswa tetapi juga membantu guru dalam merancang pembelajaran yang lebih efektif dan mendukung perkembangan siswa secara holistik.

Dengan mempertimbangkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda secara bersamaan, penyusunan soal asesmen sumatif Bahasa Jawa SD di Kabupaten Kendal tahun ajaran 2023/2024 dapat menghasilkan instrumen evaluasi yang berkualitas tinggi. Jika salah satu aspek ini diabaikan, maka kualitas soal akan menurun, sehingga hasil ujian tidak mencerminkan kemampuan siswa secara objektif. Misalnya, soal dengan reliabilitas tinggi tetapi validitas rendah akan memberikan hasil yang konsisten tetapi tidak mengukur kemampuan yang seharusnya diuji. Begitu pula, soal dengan validitas tinggi tetapi daya pembeda rendah tidak akan efektif dalam mengklasifikasikan siswa berdasarkan tingkat pemahamannya. Oleh karena itu, pendekatan yang terintegrasi dalam menganalisis keempat aspek ini sangat penting agar soal asesmen sumatif Bahasa Jawa dapat menjadi alat evaluasi yang tidak hanya adil dan akurat tetapi juga memberikan umpan balik yang bermanfaat bagi siswa dan guru dalam meningkatkan kualitas pembelajaran bahasa Jawa.

III. PENUTUP

Berdasarkan analisis hasil penelitian dapat disimpulkan bahwa soal asesmen sumatif Bahasa Jawa SD di Kabupaten Kendal tahun ajaran 2023/2024 telah memenuhi standar kualitas yang baik. Hal tersebut dapat dilihat dari hasil analisis validitas (tinggi), reliabilitas (tinggi), tingkat kesukaran (sedang), dan daya pembeda (baik). Keempat hasil analisis tersebut berada dalam kategori yang memadai, Namun, beberapa butir soal masih perlu direvisi untuk meningkatkan kualitasnya. Penelitian ini merekomendasikan agar penyusun soal lebih memperhatikan aspek-aspek tersebut dalam penyusunan soal evaluasi di masa depan.

DAFTAR PUSTAKA

Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole.

- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama.
- Arikunto, S. (2018). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, S. (2018). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80(2), 139-148. Bloomington, IN: Phi Delta Kappa International.
- Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Mason, OH: Cengage Learning.
- Dinas Pendidikan dan Kebudayaan Provinsi Jawa Tengah. (2022). *Pedoman Kurikulum Muatan Lokal Bahasa Jawa Jenjang Pendidikan Dasar dan Pendidikan Menengah di Provinsi Jawa Tengah*. Semarang: Dinas Pendidikan dan Kebudayaan Provinsi Jawa Tengah.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Feldt, L. S., & Brennan, R. L. (1989). *Reliability*. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 105–146). New York: Macmillan.
- Haladyna, T. M., & Downing, S. M. (Eds.). (2006). *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Isaac, S., & Michael, W. B. (1995). *Handbook in Research and Evaluation: A Collection of Principles, Methods, and Strategies Useful in the Planning, Design, and Evaluation of Studies in Education and the Behavioral Sciences* (3rd ed.). San Diego, CA: EdITS Publishers.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching*. Upper Saddle River, NJ: Pearson.
- Nitko, A. J. (1996). *Educational Assessment of Students* (2nd ed.). Upper Saddle River, NJ: Merrill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.
- Rahmawati, L. (2022). *Efektivitas Pendekatan STEM (Science, Technology, Engineering, and Mathematics) terhadap Kemampuan Berpikir Kreatif dan Berpikir Kritis Matematis: Studi Meta-Analisis*. Tesis Magister, Universitas Pendidikan Indonesia.
- Setiawan, T. Y., Destrinelli, D., & Wulandari, B. A. (2022). *Keterampilan Berfikir Kritis pada Pembelajaran IPA Menggunakan Model Pembelajaran RADEC di Sekolah Dasar: Systematic Literature Review*. *Justek: Jurnal Sains dan Teknologi*, 5(2), 133–141.
- Sugiyono. (2019). *Metode Penelitian Pendidikan: Kuantitatif, Kualitatif, Kombinasi, R&D dan Penelitian Pendidikan*. Bandung: Alfabeta.
- Tritantining, A. M. (2013). Kesalahan Berbahasa pada Soal-Soal Ulangan Akhir Semester Gasal Mata Pelajaran Bahasa Jawa SD Negeri Se-Kabupaten Pati Tahun Ajaran 2012/2013. *Piwulang: Jurnal Pendidikan Bahasa Jawa*, 1(1).