



JURNAL BASICEDU

Volume 8 Nomor 2 Tahun 2024 Halaman 1265 - 1277

Research & Learning in Elementary Education

<https://jbasic.org/index.php/basicedu>



Analisis Soal Tes Hasil Belajar Siswa Soal Berstandar Nasional Bahasa Indonesia Kelas 5 Sekolah Dasar

Melisa Pratiwi Sijabat^{1✉}, Krisdianti Hutabarat², Lestari Sitorus³, Syahrial⁴

Universitas Negeri Medan, Indonesia^{1,2,3,4}

E-mail: melisasijabat7@gmail.com¹, krisdiantihutabarat23@gmail.com², lestarisitorus63@gmail.com³,
syahrialpep@gmail.com⁴

Abstrak

Evaluasi merupakan hal yang penting karena dapat menilai sejauh mana keberhasilan pencapaian tujuan pembelajaran, sehingga hasilnya dapat digunakan sebagai dasar untuk langkah-langkah tindak lanjut. Penelitian ini dilakukan untuk menganalisis butir soal bahasa Indonesia yang berstandar nasional, yaitu UN dan USBN Bahasa Indonesia untuk membuat keputusan tentang setiap penilaian. Riset dilakukan dengan menggunakan metode penelitian survei. Berdasarkan hasil analisis butir soal bahasa Indonesia yang berstandar nasional, yaitu UN dan USBN maka dapat disimpulkan hasil analisis yaitu butir tes valid 16 butir dan 14 butir tidak valid, reliabilitas tes tinggi, 24 soal yang mesti dirombak dikarenakan indeks daya pembeda yang masih rendah, soal yang telah bisa diterima sebanyak 6 karena mempunyai indeks daya pembeda yang baik, distribusi tingkat kesulitan dari 30 soal yang digunakan adalah sebagai berikut: 27% mudah (8 soal), 73% sedang dan 0% sukar (0 soal) dan hasil uji tipuan dalam soal dengan menyeluruh terdapat 9 tipuan tidak berfungsi. Adapun Implikasi dari penelitian ini terhadap perkembangan keilmuan yaitu dalam penyusunan kembali soal yang tidak memenuhi standar.

Kata Kunci: Analisis Soal, Standar Nasional, Bahasa Indonesia, Sekolah Dasar.

Abstract

Evaluation is important because it can assess the extent of success in achieving learning objectives, so that the results can be used as a basis for follow-up steps. This research was conducted to analyze Indonesian language questions of national standard, namely the Indonesian National Examination and USBN to make decisions about each assessment. Research was conducted using survey research methods. Based on the results of the analysis of Indonesian language questions that are of national standard, namely the UN and USBN, it can be concluded that the results of the analysis are 16 valid test items and 14 invalid items, high test reliability, 24 questions that must be revised because the discriminating power index is still low, 6 questions were accepted because they had a good discriminating power index, the distribution of difficulty levels of the 30 questions used was as follows: 27% easy (8 questions), 73% medium and 0% difficult (0 questions) and the results of the trick test In the problem as a whole there are 9 tricks that don't work. The implications of this research for scientific development are in the rearrangement of questions that do not meet standards.

Keywords: *Question Analysis, National Standards, Indonesian, Elementary School.*

Copyright (c) 2024 Melisa Pratiwi Sijabat, Krisdianti Hutabarat, Lestari Sitorus, Syahria

✉ Corresponding author :

Email : melisasijabat7@gmail.com

DOI : <https://doi.org/10.31004/basicedu.v8i2.7354>

ISSN 2580-3735 (Media Cetak)

ISSN 2580-1147 (Media Online)

Jurnal Basicedu Vol 8 No 2 Tahun 2024
p-ISSN 2580-3735 e-ISSN 2580-1147

PENDAHULUAN

Dalam proses belajar, evaluasi termasuk bagian yang tak terpisahkan (Fiska et al., 2021). Evaluasi meliputi dua proses utama, yaitu pengukuran yang melibatkan perbandingan dari sesuatu yang diteliti menggunakan alat yang sudah ditata dengan sistematis, serta penilaian yang termasuk nilai interpretasi dari perhitungan tersebut (Elviana, 2020). Evaluasi dianggap penting karena dapat menilai sejauh mana keberhasilan pencapaian tujuan pembelajaran, sehingga hasilnya dapat digunakan sebagai dasar untuk langkah-langkah tindak lanjut (Lesta Ariany et al., 2018). Melalui evaluasi, diharapkan mutu belajar mengajar yang diberikan terhadap pengajar dapat selalu berkembang, yang nantinya akan berpengaruh terhadap peningkatan keahlian pelajar. Perkembangan kemampuan siswa berkontribusi pada peningkatan kualitas pendidikan secara keseluruhan, karena kualitas pendidikan sejalan dengan kemampuan siswa (Khasanah et al., 2023).

Kemampuan peserta didik bisa diperoleh melalui hasil pengujian, yang dapat mampu diaplikasikan dengan menggunakan instrumen pengukuran berbentuk tes atau non-tes. Instrumen pengukuran yang benar nantinya mendapatkan data yang berkualitas. Pengajar bisa menilai keterampilan pelajar secara akurat apabila menggunakan instrumen yang berkualitas. Sebuah tes dianggap benar jika terdapat 5 kriteria, yakni: validitas, reliabilitas, objektivitas, praktikabilitas, serta ekonomis. menurut pandangan Arikunto tersebut, kriteria setidaknya untuk instrumen pengukuran yang benar ialah validitas serta reliabilitas. Terlepas dari validitas serta reliabilitas, sebuah tes dianggap benar apabila memiliki daya pembeda yang baik, tingkat kesulitan yang sesuai, dan analisis distraktor (untuk soal opsi ganda) yang memuaskan (Amelia, 2017a).

Salah satu langkah yang diambil oleh pemerintahan guna mengevaluasi penguasaan kompetensi yang sudah ditentukan dalam pendidikan dasar ialah adanya pelaksanaan UN dan USBN. USBN merupakan aktifitas penghitungan hasil kemampuan siswa yang dilaksanakan oleh Satuan Pendidikan yang merujuk terhadap Standar Kompetensi Lulusan. USBN bertujuan guna memberikan pengakuan atas prestasi belajar. Kegiatan ini memberikan informasi mengenai tingkat ketuntasan atau pencapaian selama tahap belajar dalam kurun waktu enam tahun bagi satuan pendidikan jenjang SD. tahapan belajar dilaksanakan agar menggapai maksud yaitu hasil belajar. USBN dilaksanakan oleh Badan Standarisasi Nasional Pendidikan (BSNP) yang memiliki sifat mandiri serta profesional. Selain USBN ada juga UN untuk mengevaluasi penguasaan kompetensi yang telah ditetapkan dalam pendidikan dasar UN merupakan bentuk penilaian pendidikan yang digunakan sebagai wadah guna mengevaluasi pencapaian standar nasional pendidikan. UN menunjukkan informasi yang penting pada pemutusan ketetapan terhadap penguasa ketentuan pendidikan pada negara. berdasarkan kemendikbud No. 66 thn 2013 terkait Standar Penilaian Pendidikan, UN termasuk agenda peninjauan kualitas peserta didik yang dilaksanakan dengan nasional guna meninjau capaian Standar Nasional Pendidikan. Menurut (Eko Putra Widoyoko, 2009), UN adalah bentuk tes yang dipakai guna pengambilan keputusan terkait dengan individu yang diuji, seperti sertifikasi, kelulusan, seleksi, dan lain sebagainya. Pada tingkat SD, mapel yang diujikan yaitu matematika, Bahasa Indonesia, serta IPA. Hasil ujian nasional bisa digunakan guna beragam keperluan, termasuk untuk pembandingan pada upaya pembimbingan serta penyajian bantuan terhadap satuan pendidikan untuk menambah kualitas Pendidikan (Latifah et al., 2024).

Berdasarkan peranan tes menjadi alat ukur, keberhasilan suatu tes bisa dinilai dari kemampuannya guna menunjukkan informasi berdasarkan kondisi mestinya pada objek yang diukur. Maka dari itu, saat hendak diaplikasikan, tes hasil belajar perlu diamati secara menyeluruh agar bisa menunjukkan hasil yang maksimal (Aulia Ulfa Dewi, 2016).

Aktivitas mengamati satuan soal termasuk tahapan pengelompokan, peringkasan, serta pemakaian informasi pada jawaban siswa agar mengambil langkah terkait seluruh penilaian. pengamatan satuan soal bertujuan untuk memberikan alasan terkait butiran tak berjalan sebagaimana mestinya serta tingkat efektifnya butir tersebut (Ebel & Frisbie, 2016).

Sebelumnya sudah terdapat penelitian tentang analisis hasil ujian nasional pelajaran matematika tingkat Madrasah Tsanawiyah (MTs) di Kabupaten Sidoarjo yang memfokuskan pada data hasil ujian nasional yang banyak dikuasai dan tidak dikuasai oleh siswa (8). (Latifah et al., 2024).

Berbeda dengan penelitian sebelumnya yang hanya menganalisis tingkat kesukaran dan penguasaan kompetensi dasar pada ujian sekolah berstandar nasional matematika. Penelitian yang dilakukan ini memberikan penambahan analisis validitas, analisis daya beda, analisis reliabilitas dan analisis pengecoh untuk menganalisa setiap butir soal dengan baik. Soal yang dianalisa adalah soal ujian berstandar nasional yaitu soal UN dan soal USBN yang belum pernah dilakukan sebelumnya.

Analisis utama untuk analisis butir soal pilihan ganda adalah indeks kesukaran butir (p), dan indeks daya beda butir (D). Selain itu juga ada sebuah langkah untuk menganalisis keefektifan pengecoh dalam butir pilihan ganda, atau sering disebut sebagai analisis distraktor (Ratni et al., 2023). Sependapat dengan ahli lain, (Shermis & Di Vesta, 2011). berfungsi dengan baik dan apakah butir layak untuk digunakan atau butir buruk untuk dibuang ada dua langkah untuk menentukan yaitu kesukaran dan daya beda. Analisis butir soal dalam penelitian ini dilakukan secara kuantitatif yaitu analisis berdasarkan data empirik dengan pendekatan analisis secara klasik terkait dengan ciri-ciri statistiknya yaitu tingkat kesukaran, indeks daya beda, efektivitas distraktor dan uji realibilitas. Berdasarkan uraian di atas, maka pada penelitian ini sangat perlu dilakukan untuk menganalisis aspek-aspek tingkat validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh, sehingga hasilnya akan dimanfaatkan dalam memberikan rekomendasi bagi soal/tes yang digunakan pada saat USBN di tingkat SD berikutnya.

Permasalahan sebagai fokus pada observasi ini ialah menganalisis lembaran soal bahasa Indonesia yang berstandar nasional, yaitu UN dan USBN Bahasa Indonesia guna mengambil langkah terkait seluruh penilaian. Kegunaan pada observasi ini ialah antara lain: Pertama, untuk mahasiswa, observasi ini menjadi sarana untuk mengaplikasikan atau menerapkan ilmu pengetahuan dalam menganalisis suatu butir tes. Kedua, bagi pendidik, penelitian ini bisa sebagai sumber guna mengadakan evaluasi belajar mengajar, khususnya dalam menganalisis aspek-aspek seperti validitas, reliabilitas, tingkat kesulitan, daya pembandingan, serta distraktor dari lembaran tes.

METODE

Variabel yang diteliti dalam observasi ini ialah soal berstandar nasional Bahasa Indonesia yang diujikan pada peserta didik kelas 5 pada SDN 101774 Sampali. Populasi penelitian merupakan peserta didik kelas 5^B. Sampel yang diambil dari peserta didik kelas 5^B yang belajar pada SDN 101774 Sampali. Teknik pendataan sampel yang dipakai ialah *random sampling*. Sesuai dengan penjelasan Taniredja serta Mustafidah (2012), "*Teknik random sampling* dikatakan juga random, serampangan, tidak pilih kasih, objektif, menyebabkan setiap komponen golongan memiliki peluang agar menjadi sampel observasi". Penggunaan teknik *random sampling* dipilih untuk memastikan bahwa semua subjek dalam populasi mempunyai kesempatan yang serupa guna digunakan sebagai sampel subjek. memilih teknik pengambilan *sampel random sampling* dengan alasan dalam jenjang SD, pengambilan siswa tidak berpedoman terhadap nilai khusus (Amelia, 2017b). Pengumpulan data tentang keterampilan beretorika dalam soal Bahasa Indonesia berstandar nasional pada siswa kelas V^B dilakukan dengan memakai tes opsi ganda yang terdiri dari empat opsi jawaban. Instrumen penelitian yang dipakai ialah 30 butir soal opsi ganda dalam bahasa Indonesia dengan standar nasional, mencakup berbagai jenis soal, termasuk soal UN serta USBN.

Metode analisis data yang diterapkan yaitu analisis deskriptif. Evaluasi kualitas data tes hasil belajar akan dilakukan memakai perangkat lunak TAP (Test Analysis Program) versi 14.7.4, yang tersedia secara gratis mempunyai hak cipta atas nama Gordon P. Brooks. Pemilihan TAP didasarkan pada kemudahan penggunaannya, di mana sekali input data bisa menghasilkan analisis validitas, reliabilitas, daya pembeda, tingkat kesulitan, serta distraktor dari soal tes (Amelia, 2017b).

Suatu tes dianggap mempunyai validitas apabila mampu secara tepat mengukur konstruk yang dimaksud. Validitas bisa dipahami dari dua perspektif: pertama, dalam proses penyusunan tes, pembuat berusaha untuk memilih item-item berdasarkan akal dapat menghitung konstruk yang dikehendaki, apakah berdasarkan perbandingan personal ataupun melalui diskusi dengan orang lain atau ahli dalam bidang terkait; kedua, ketika tes digunakan, validitasnya dapat dinilai dengan membandingkan hasilnya dengan tes lain yang telah terbukti validitasnya. Metode yang diterapkan pada observasi ini ialah korelasi biserial, yang diaplikasikan guna mengevaluasi validitas seluruh komponen tes. Rumus untuk menghitung korelasi biserial ialah antara lain (Amelia, 2017b):

$$r_{pbi} = \frac{Mp - Mt}{St} \sqrt{\frac{p}{q}}$$

Keterangan:

- rpbi = koefisien korelasi biserial
- Mp = rerata skor dari subjek yang menjawab benar untuk item yang dicari validitasnya.
- Mt = rerata skor jumlah
- St = standar deviasi dari skor jumlah
- p = proporsi peserta didik yang menjawab benar
- Q = proporsi peserta didik yang menjawab salah (q = 1-p)

$$(p = \frac{\text{banyaknya siswa yang menjawab benar}}{\text{jumlah seluruh siswa}})$$

Analisis validitas dalam observasi ini bisa diamati berdasarkan nilai *point biserial* yang diperoleh dari TAP. Nilai *point biserial* dipertimbangkan berdasarkan nilai rtabel pada tingkat signifikansi 5%. Jika nilai *point biserial* melebihi nilai rtabel, maka item tes tersebut dianggap valid. Nilai rtabel untuk sampel sebanyak 26 siswa adalah >0,388. Apabila nilai point biserial melebihi 0,388, maka *item* tes dianggap valid (Amelia, 2017a).

Reliabilitas merupakan faktor penting dalam mengevaluasi setiap item tes. Reliabilitas setiap item dalam sebuah tes mencerminkan seberapa stabil dan dapat diandalkannya setiap tes. Tes yang reliabel akan menunjukkan hasil yang konsisten jika diuji terhadap golongan yang serupa pada masa yang berbeda. Dalam observasi ini, peneliti mengadopsi metode belah 2 maupun split-half method. Prosedur dimaksud melibatkan pembagian setiap item tes menjadi dua bagian yang sebanding, yakni bagian ganjil dan genap berdasarkan nomor soal. Keandalan hasil kemudian dihitung menggunakan perangkat lunak TAP dan dianalisis dengan menggunakan kriteria Reliabilitas yang disusun oleh Masidjo (Amelia, 2017b).

Tabel 1. Kriteria Reliabilitas

Koefisien Korelasi	Kualifikasi
0,91-1,00	Sangat Tinggi
0,71-0,90	Tinggi
0,41-0,70	Cukup
0,21-0,40	Rendah
Negatif-0,20	Sangat Rendah

Daya pembeda sebuah item tes merujuk pada kemampuan item tersebut untuk memisahkan terhadap pelajar yang sudah memahami pembelajaran yang diujikan serta pelajar yang kurang menguasainya. Tujuan dari daya pembeda pada suatu tes ialah agar mengetahui pelajar yang mempunyai keahlian lebih ketimbang pelajar yang mempunyai keahlian minim. Pada observasi ini, peneliti mengadopsi indeks daya pembeda yang dijelaskan, secara penjumlahan sesuai terhadap metode yang telah ditetapkan sebagai berikut (Amelia, 2017a):

$$DP = \frac{BA - BB}{\frac{1}{2}N}$$

Keterangan:

- D = indeks daya pembeda soal (Indeks Diskriminasi)
 BA = total jawaban benar dalam kelompok atas
 BB = total jawaban benar dalam kelompok bawah
 N = total anggota tes

Kategori sebagai daya pembeda maupun indeks diskriminatif yang disajikan oleh Cracker serta Algina pada Kusaeri yang diterapkan dalam analisis daya pembeda pada observasi ini ialah antara lain (Amelia, 2017c):

Tabel 2. Kriteria Daya Pembeda

Range Daya Pembeda	Kategori	Keputusan
0,40 - 1,00	Sangat memuaskan	Diterima
0,30 - 0,39	Memuaskan	Diterima
0,20 - 0,29	Tidak memuaskan	Ditolak/direvisi
0,00 - 0,19	Sangat tidak memuaskan	Direvisi menyeluruh

Soal yang memiliki kualitas bagus ialah soal yang sedang-sedang saja. Indeks kesulitan dinyatakan dalam rentang 0,00 hingga 1,0, yang mengindikasikan tingkat kesulitan suatu soal. Nilai indeks kesukaran mencerminkan tingkat kesulitan dari soal tersebut. Sebuah soal dengan indeks kesulitan 0,0 menandakan soal tersebut sangat sulit, sementara nilai 1,0 menggambarkan soal tersebut cukup ringan (Amelia, 2017b).

Tabel 3. Indeks Kesukaran

Indeks kesukaran	Kategori
0,00 - 0,30	Sulit
0,31 - 0,70	Menengah
0,71 - 1,00	Mudah

Dalam terminologi evaluasi, indeks kesulitan tersebut disimbolkan sebagai P, yang merupakan kependekan terhadap "proporsi". Oleh karena itu, sebuah soal dengan P = 0,70 dianggap cenderung ringan daripada soal dengan P = 0,20. Adapun soal dengan P = 0,30 dianggap cenderung sulit ketimbang soal P = 0,80 (Amelia, 2017b).

$$P = \frac{B}{JS}$$

Keterangan:

- P = indeks kesulitan
 B = banyaknya peserta yang menjawab soal itu dengan benar
 JS = total keseluruhan peserta didik anggota tes

Indeks kesukaran bisa ditemukan dalam Tabel 3. Distribusi tingkat kesulitan, mengikuti pandangan Widoyoko (2014: 165), adalah: 25% mudah, 50% menengah, serta 25% sulit (Amelia, 2017b).

Kunci jawaban dan *distraktor*, juga dikenal sebagai pengecoh, merupakan dua opsi yang tersedia dalam soal opsi ganda. Analisis *distraktor* bermaksud agar mengevaluasi persentase peserta didik yang mengambil *distraktor* serta yang menjawab dengan baik dari kunci jawaban (Elviana, n.d.). Adapun *distraktor* digunakan untuk mengetahui peserta didik dengan keahlian lebih juga kurang (Fiska et al., 2021). Keberhasilan *distraktor* diukur dari anggota tes yang berasal pada golongan terendah atau peserta didik dengan keahlian kurang yang

menetapkannya. *Distraktor* dikatakan kurang tepat apabila ditetapkan terhadap peserta didik dengan keahlian yang tergolong Lebih (Elviana, 2020).

Kemampuan pengecoh dalam menyesatkan peserta didik yang kurang cakap dalam memilih alternatif jawaban adalah yang dimaksud dengan efektivitas pengecoh. Fungsinya adalah guna mengecoh anggota tes yang tidak menjawab dengan tepat (Putu et al., 2013). Penilaian efektivitas pengecoh (*Distractor*) terletak pada kemiripan antara jawaban pengecoh dan jawaban yang benar, sehingga seringkali membingungkan siswa dalam menjawab soal tes yang diberikan (Setiawati, 2019). Pengecoh dianggap efektif jika ditetapkan paling kurang 5% dari jumlah anggota tes (Hamzah B. Uno, 2018) (Amelia, 2017b).

Opsi atau pilihan alternatif dalam setiap butir tes pilihan ganda dapat terdiri dari tiga hingga lima opsi, di mana salah satunya adalah jawaban yang benar (kunci jawaban), sementara yang lainnya ialah jawaban yang tidak benar. Jawaban-jawaban yang tidak benar tersebut yang dikatakan pengecoh. Maka dari itu, pengecoh adalah opsi alternatif yang mampu menyesatkan jawaban yang sebenarnya (Amelia, 2017b).

HASIL DAN PEMBAHASAN

Analisis soal tes hasil belajar siswa berstandar nasional bahasa Indonesia di kelas 5 SDN 101774Sampali menunjukkan bahwa evaluasi telah dilakukan terhadap bobot tes, termasuk validitas, reliabilitas, daya pembeda, tingkat kesulitan, dan pengecoh.

Analisis Validitas

Validitas sebagai karakteristik pertama tes yang baik diartikan sejauh mana tes tersebut mengukur apa yang seharusnya diukur. (Brown, 2004). Dalam penelitian (Solichin, 2017) validitas soal dikatakan valid apabila memiliki dukungan besar terhadap skor total. Skor butir soal menyebabkan skor total menjadi tinggi atau rendah. Artinya dapat dikatakan butir soal memiliki validitas tinggi jika skor pada butir soal memiliki kesejajaran dengan skor total Soal dikatakan reliabel apabila soal diujikan beberapa kali menunjukkan hasil yang sama. Analisis validitas memakai test analysis program berguna agar menentukan apakah suatu soal valid atau tidak. Evaluasi hasil analisis dilakukan melalui TAP secara memperhatikan nilai point biserial. Nilai point biserial kemudian dibandingkan menggunakan r_{tabel} dengan tingkat signifikansi 5%. Koefisien validitas pada r_{tabel} dari tingkat signifikansi 5% bagi 26 peserta didik adalah sekitar 0.388. apabila nilai point biserial melebihi 0.388, dikatakan soal itu dianggap valid. Informasi mengenai hasil analisis validitas soal bisa ditemukan pada Tabel 4.

Tabel 4. Hasil Analisis Validitas

No Item	r hitung	r tabel	Keterangan
Item 01	0.209	0.388	Tidak Valid
Item 02	0.23	0.388	Tidak Valid
Item 03	0.427	0.388	Valid
Item 04	0.109	0.388	Tidak Valid
Item 05	0.609	0.388	Valid
Item 06	0.439	0.388	Valid
Item 07	0.457	0.388	Valid
Item 08	0.025	0.388	Tidak Valid
Item 09	0.318	0.388	Tidak Valid
Item 10	-0.032	0.388	Tidak Valid
Item 11	0.574	0.388	Valid
Item 12	-0.139	0.388	Tidak Valid
Item 13	0.188	0.388	Tidak Valid
Item 14	0.706	0.388	Valid

No Item	r hitung	r tabel	Keterangan
Item 15	0.557	0.388	Valid
Item 16	0.502	0.388	Valid
Item 17	0.478	0.388	Valid
Item 18	0.495	0.388	Valid
Item 19	0.5	0.388	Valid
Item 20	0.356	0.388	Tidak Valid
Item 21	0.382	0.388	Tidak Valid
Item 22	0.73	0.388	Valid
Item 23	0.424	0.388	Valid
Item 24	0.621	0.388	Valid
Item 25	0.48	0.388	Valid
Item 26	0.365	0.388	Tidak Valid
Item 27	0.279	0.388	Tidak Valid
Item 28	0.349	0.388	Tidak Valid
Item 29	0.035	0.388	Tidak Valid
Item 30	0.603	0.388	Valid

Dari Tabel 4, terlihat bahwa beberapa butir soal dianggap valid, sementara yang lainnya tidak valid karena perbedaan koefisien point biserial yang diperoleh. Beberapa butir soal melebihi nilai t tabel, sementara lainnya kurang dari nilai t tabel sekitar 0.388. Oleh karena itu, sebagian butir soal dianggap telah mengukur kemampuan peserta didik terkait soal tes hasil belajar siswa berstandar nasional Bahasa Indonesia.

Analisis Reliabilitas

Analisis reliabilitas soal dengan TAP menunjukkan nilai reliabilitas Split-Half (odd/even) sebesar 0,858. Dalam penelitian (Ambarwati & Ismiyati, 2022a) ada beberapa faktor yang memengaruhi reliabilitas tinggi dari suatu soal antara lain jumlah butir soal, lama tes, rentang kesulitan butir soal, dan objektivitas penskoran, hasil uji reliabilitas soal termasuk pada kriteria "tinggi." Dengan demikian, lembaran soal tersebut menggambarkan konsistensi yang lebih pada keterampilan siswa dalam menjawab soal Tes Hasil Belajar Siswa Soal Berstandar Nasional Bahasa Indonesia.

Analisis Daya Pembeda

Daya beda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah. Angka yang menunjukkan besarnya daya beda disebut indeks diskriminasi (D). Indeks diskriminasi butir soal uraian dapat dihitung dengan rumus perbandingan antara selisih mean kelompok atas dan mean kelompok bawah dengan skor maksimal tiap butir soal. (Boopathiraj & Chellamani, 2013) Hasil negatif memiliki daya pembeda yang buruk, sedangkan nilai antara 0.71 dan 1.00 dianggap sangat baik (Monica et al., 2019). Dianggap bisa diakui apabila berada dalam rentang 0,307 hingga 0,769. Namun, soal akan ditolak atau direvisi jika indeks daya pembeda berada dalam rentang - 0,076 hingga 0,230, dan akan ditolak jika bernilai -0,00. Informasi tentang hasil analisis daya pembeda bisa ditemukan pada Tabel 5.

Tabel 5. Hasil Analisis Daya Pembeda

No Item	Dicrimination Index	Kriteria
Item 01	0,076	Direvisi total
Item 02	-0,076	Direvisi total
Item 03	0,384	Diterima
Item 04	0	Direvisi total

No Item	Dicrimination Index	Kriteria
Item 05	0,153	Direvisi total
Item 06	0,538	Diterima
Item 07	0,153	Direvisi total
Item 08	0,307	Diterima
Item 09	0	Direvisi total
Item 10	0,076	Direvisi total
Item 11	0,076	Direvisi total
Item 12	-0,384	Direvisi total
Item 13	0,076	Direvisi total
Item 14	0,076	Direvisi total
Item 15	0,769	Diterima
Item 16	0,153	Direvisi total
Item 17	0,076	Direvisi total
Item 18	0,307	Diterima
Item 19	0,230	Direvisi
Item 20	0,307	Diterima
Item 21	0,230	Direvisi
Item 22	0,230	Direvisi
Item 23	0,230	Direvisi
Item 24	-0,230	Direvisi total
Item 25	0	Direvisi total
Item 26	0	Direvisi total
Item 27	-0,153	Direvisi total
Item 28	0	Direvisi total
Item 29	-0,307	Direvisi total
Item 30	0	Direvisi total

Dari Tabel 5, berdasarkan 30 poin soal tes, ditemukan bahwa 24 poin soal memerlukan revisi. Soal yang perlu direvisi adalah nomor 1, 2, 4, 5, 7, 9, 10, 11, 12, 13, 14, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, dan 30. Sementara itu, 6 soal lainnya dapat diterima tanpa revisi. Butir soal yang memerlukan revisi disebabkan oleh indeks daya pembeda yang masih rendah, menyebabkan belum mampu membandingkan siswa dengan keterampilan tinggi dan rendah. Sebaliknya, soal-soal yang telah bisa diakui terdapat indeks daya pembeda yang bagus, yang mana mampu membandingkan siswa dengan keterampilan lebih dari yang berkemampuan rendah.

Analisis Tingkat Kesulitan

Tingkat kesulitan soal adalah proporsi peserta didik menjawab dengan benar terhadap suatu butir soal (Riani et al., n.d.) Tingkat kesulitan butir soal artinya mengkaji soal-soal tes dari segi kesulitannya dengan tujuan mendapatkan informasi mengenai soal-soal mana saja yang tergolong mudah, sedang dan sukar (Valen & Satria, 2021). Ini berarti bahwa semakin banyak peserta tes yang menjawab dengan benar, semakin besar indeks tingkat kesulitannya, yang berarti semakin mudah soal tersebut. sebaliknya, semakin sedikit individu yang dapat menjawab pertanyaan dengan benar, maka soal tersebut makin sukar.

Tingkat kesulitan umumnya dikelompokkan menjadi tiga kriteria: sulit, menengah, serta mudah. Kriteria sulit terdapat dalam kisaran nilai 0,00-0,460, kriteria menengah terdapat dalam kisaran nilai 0,461-0,653, serta kriteria mudah terdapat pada kisaran 0,692-0,923. Distribusi tingkat kesulitan berdasarkan 30 soal yang dipakai ialah sebagai berikut: 27% mudah (8 soal), 63% sedang (20 soal), serta 0,6% sulit (2 soal). Hasil analisis tingkat kesulitan bisa dicermati di Tabel 6.

Tabel 6. Analisis Tingkat Kesukaran

No Item	Item Difficulty	Kriteria
Item 01	0,653	Menengah
Item 02	0,423	Sulit
Item 03	0,423	Sulit
Item 04	0,846	Mudah
Item 05	0,923	Mudah
Item 06	0,576	Menengah
Item 07	0,461	Menengah
Item 08	0,538	Menengah
Item 09	0,769	Mudah
Item 10	0,576	Menengah
Item 11	0,884	Mudah
Item 12	0,576	Menengah
Item 13	0,5	Menengah
Item 14	0,807	Mudah
Item 15	0,615	Menengah
Item 16	0,461	Menengah
Item 17	0,653	Menengah
Item 18	0,615	Menengah
Item 19	0,576	Menengah
Item 20	0,615	Menengah
Item 21	0,807	Mudah
Item 22	0,653	Menengah
Item 23	0,576	Menengah
Item 24	0,576	Menengah
Item 25	0,615	Menengah
Item 26	0,692	Mudah
Item 27	0,615	Menengah
Item 28	0,692	Mudah
Item 29	0,461	Menengah
Item 30	0,615	Menengah

Dari Tabel 6, ditemukan bahwa 8 soal (27%) terdapat tingkat kesulitan kriteria mudah, 20 soal (63%) mempunyai tingkat kesulitan kriteri menengah, serta 2 soal (0,6%) mempunyai tingkat kesulitan kriteri a sulit. Aspek tersebut menggambarkan bahwa distribusi soal tidak mencukupi syarat untuk soal yang efektif. agar mencapai distribusi soal yang optimal, diperlukan penyesuaian dengan menambahkan 2 soal pada kriteria mudah, mengurangi 8 soal pada kriteria menengah, serta menambahkan 8 soal pada kriteri sulit.

Analisis Pengecoh

Berdasarkan (Vina Nur Indah Sari et al., 2022), pengecoh merupakan opsi dalam tes yang ditujukan untuk menarik perhatian peserta tanpa menjadi jawaban yang benar. Jika lebih banyak peserta memilih pengecoh tersebut, itu menunjukkan efektivitasnya. Pengecoh butir soal dikatakan berfungsi dengan baik apabila ketiga pengecohnya berfungsi atau dipilih oleh peserta didik. Namun, jika salah satu atau lebih pengecoh tidak berfungsi atau tidak dipilih, maka pengecoh butir soal tersebut tidak berfungsi dengan baik (Riswanda Himawan, 2022). Sedangkan (Ambarwati & Ismiyati, 2022b) menyatakan bahwa fungsi pengecoh tidak memiliki daya tarik bagi peserta didik yang kurang menguasai konsep atau materi. Oleh karena itu, pengecoh

yang berfungsi dengan baik sebaiknya disimpan dan pengecoh yang tidak berfungsi sebaiknya diperbaiki agar soal dapat berfungsi, sehingga soal tersebut dapat digunakan kembali untuk tes berikutnya. Efektivitas distraktor diukur dengan minimal 5% peserta memilihnya. Analisis pengecoh dapat ditemukandalam Tabel 7 di bawah ini.

Tabel 7. Hasil Analisis Pengecoh

No Item	Pilihan Jawaban				Keterangan
	A	B	C	D	
Item 1	Kunci jawaban	0,12	0,00	0,23	Pengecoh C Tidak Bekerja
Item 2	Kunci jawaban	0,42	0,08	0,08	-
Item 3	0,46	0,04	0,08	Kunci jawaban	-
Item 4	0,08	Kunci jawaban	0,04	0,04	-
Item 5	0,08	0,00	Kunci jawaban	0,00	Pengecoh B dan D Tidak Bekerja
Item 6	Kunci jawaban	0,31	0,08	0,04	
Item 7	0,035	Kunci jawaban	0,12	0,08	
Item 8	0,27	0,19	Kunci jawaban	0,00	Pengecoh D Tidak Bekerja
Item 9	Kunci jawaban	0,12	0,08	0,04	
Item 10	0,08	0,19	Kunci jawaban	0,15	
Item 11	0,08	0,04	Kuci jawaban	0,00	Pengecoh D Tidak Bekerja
Item 12	0,12	Kunci jawaban	0,58	0,19	-
Item 13	Kunci jawaban	0,31	0,12	0,08	-
Item 14	Kunci jawaban	0,04	0,04	0,08	-
Item 15	Kunci jawaban	0,15	0,15	0,08	-
Item 16	0,23	0,04	0,27	Kunci jawaban-	
Item 17	0,23	Kunci jawaban	0,00	0,12	Pengecoh C Tidak Bekerja
Item 18	0,31	0,15	Kunci jawaban	0,00	Pengecoh D Tidak Bekerja
Item 19	0,04	Kunci jawaban	0,08	0,23	-
Item 20	0,04	Kunci jawaban	0,04	0,27	-
Item 21	0,08	Kunci jawaban	0,00	0,00	Pengecoh C dan D Tidak Bekerja
Item 22	0,23	0,04	0,08	Kunci jawaban-	
Item 23	0,27	0,08	0,08	Kunci jawaban-	
Item 24	Kunci jawaban	0,12	0,8	0,19	-
Item 25	0,04	0,019	Kunci jawaban	0,08	-
Item 26	Kunci jawaban	0,12	0,08	0,08	-
Item 27	Kunci jawaban	0,19	0,08	0,12	-
Item 28	0,15	Kunci jawaban	0,15	0,00	Pengecoh D Tidak Bekerja
Item 29	0,15	0,38	Kunci jawaban	0,00	Pengecoh D Tidak Bekerja
Item 30	0,12	0,19	0,19	Kunci jawaban-	

Berdasarkan data pada tabel di atas, terdapat total 9 tipuan yang tidak bekerja secara keseluruhan. tipuan dikatakan tidak berguna apabila jumlahnya sedikit dari 5% dari total anggota tes. Oleh karena itu, penyesatan yang tidak berguna mesti diperbaiki lagi.

Hasil analisis Validitas, Reliabilitas, Daya Pembeda, Tingkat Kesulitan, dan Pengecoh butir soal kemudian dianalisis secara bersama-sama untuk mengetahui kualitas butir soal. Kriteria butir soal dibagi ke dalam 3 (tiga) kategori yaitu: sulit, menengah dan mudah.

Dari hasil analisis, sebanyak 2 butir soal atau 0,6% termasuk ke dalam kriteria sulit, 20 butir soal atau 63% termasuk dalam kriteria menengah, dan 8 butir soal atau 27% termasuk dalam kriteria mudah. Butir soal yang kriterianya sulit bisa dimasukkan ke dalam bank soal dengan tetap menjaga kerahasiaan soal tersebut sehingga dapat digunakan lagi untuk ujian yang akan datang. Butir soal yang kriterianya menengah belum bisa dimasukkan ke dalam bank soal karena harus dilakukan revisi terlebih dahulu sesuai dengan indikator

kegagalannya. Butir soal yang kriterianya mudah lebih baik dibuang karena butir tersebut membutuhkan revisi yang signifikan.

Hasil analisis menunjukkan sebanyak 24 butir soal memerlukan revisi dan 6 butir soal yang tidak direvisi dan dapat diterima. Hal tersebut menunjukkan bahwa butir soal belum bisa menjalankan fungsinya dengan baik. Penyebab kegagalan butir soal tersebut dapat ditelusuri dari aspek validitas, daya pembeda, tingkat kesulitan, dan pengecoh tiap butir soal.

Penyebab kegagalan butir soal pertama yaitu pada validitas soal. Hal tersebut menunjukkan bahwa validitas setiap butir soal belum memiliki dukungan besar terhadap skor total tes.

Penyebab kegagalan butir soal kedua yaitu pada daya pembeda. Hal tersebut menunjukkan bahwa butir soal yang memerlukan revisi disebabkan oleh indeks daya pembeda yang masih rendah, menyebabkan belum mampu membandingkan siswa dengan keterampilan tinggi dan rendah.

Penyebab kegagalan butir soal ketiga yaitu pada tingkat kesulitan. Dari segi tingkat kesulitan, soal Soal Tes Berstandar Nasional Bahasa Indonesia yang diujikan kepada siswa kelas 5 ditemukan bahwa 8 soal (27%) terdapat tingkat kesulitan kriteria mudah, 20 soal (63%) mempunyai tingkat kesulitan kriteria menengah, serta 2 soal (0,6%) mempunyai tingkat kesulitan kriteria sulit. Aspek tersebut menggambarkan bahwa distribusi soal tidak mencukupi syarat untuk soal yang efektif.

Penyebab kegagalan butir soal yang keempat yaitu pada pengecoh. Dari segi pengecoh, Soal Tes Berstandar Nasional Bahasa Indonesia memiliki pengecoh yang belum berfungsi dengan baik, terdapat total 9 tipuan yang tidak bekerja secara keseluruhan, Oleh karena itu, penyesatan yang tidak berguna mesti diperbaiki lagi.

Dari uraian di atas, dapat disimpulkan jika Soal Tes Berstandar Nasional Bahasa Indonesia merupakan soal yang belum berkualitas baik. Hal tersebut dikarenakan hanya ada 6 butir soal yang berkriteria baik dan dapat langsung disimpan di bank soal sedangkan untuk 24 butir soal memerlukan revisi baik revisi minor maupun sedang sesuai dengan indikator keagalannya.

Dalam proses melakukan penelitian ini, terdapat keterbatasan yang mungkin dapat memengaruhi hasil penelitian, yaitu Adanya keterbatasan waktu penelitian, tenaga, dan kemampuan peneliti. Serta Adanya kemampuan responden yang kurang dalam memahami pernyataan pada soal dan juga kejujuran dalam mengisi Soal sehingga ada kemungkinan hasilnya kurang akurat.

Adapun Implikasi dari penelitian ini terhadap perkembangan keilmuan meliputi: Penyusunan kembali soal yang tidak memenuhi standar. Penelitian ini juga merekomendasikan langkah-langkah untuk menganalisis dan memperbaiki soal yang tidak memenuhi standar. Serta memperbaiki dan menganalisis ulang peneliti merekomendasikan agar peneliti selanjutnya melakukan perbaikan dan analisis ulang terhadap soal yang belum memadai. Dengan demikian, implikasi dari penelitian ini adalah bahwa menata ulang soal yang tidak memenuhi standar, perbaikan dan analisis ulang, serta penekanan pada pembuatan soal yang memiliki pengecoh yang dapat bekerja sehingga dapat memengaruhi perkembangan ilmu pendidikan.

SIMPULAN

Tujuan penelitian ini adalah menganalisis lembar soal bahasa Indonesia yang berstandar nasional, yaitu UN dan USBN Bahasa Indonesia guna mengambil langkah terkait seluruh penilaian. Analisis validitas menunjukkan tingkat kevalidan yang lebih tinggi, analisis reliabilitas menunjukkan konsistensi yang tinggi dalam keterampilan siswa menjawab soal, sementara analisis daya pembeda dan analisis tingkat kesulitan mengungkapkan kebutuhan akan revisi dan penyesuaian pada beberapa soal agar dapat secara efektif membedakan antara siswa yang berbeda tingkat kemampuan. Analisis pengecoh menyatakan bahwa lebih banyak pengecoh yang bekerja secara efektif untuk mengukur pemahaman siswa secara akurat. Dari hasil penelitian, disimpulkan bahwa soal tes berstandar nasional Bahasa Indonesia masih belum mencapai standar

yang diinginkan, namun penelitian ini memberikan arah dan langkah konkret untuk memperbaiki dan meningkatkan kualitasnya.

UCAPAN TERIMAKASIH

Kami ingin mengucapkan terima kasih kepada dosen, guru, dan para siswa yang terlibat langsung dalam penelitian ini. Tanpa kontribusi dan dorongan dari kalian, pencapaian ini tidak akan terwujud. Kami sungguh menghargai kontribusi, saran, dan bimbingan yang telah diberikan. diharapkan jurnal ini memberikan kegunaan bagi kita semua, terimakasih.

DAFTAR PUSTAKA

- Ambarwati, Y. F., & Ismiyati, I. (2022a). Analisis Butir Soal Pilihan Ganda Ulangan Akhir Semester Genap Mata Pelajaran Kearsipan. *Measurement In Educational Research (Meter)*, 1(2), 64. <https://doi.org/10.33292/meter.v1i2.144>
- Ambarwati, Y. F., & Ismiyati, I. (2022b). Analisis Butir Soal Pilihan Ganda Ulangan Akhir Semester Genap Mata Pelajaran Kearsipan. *Measurement In Educational Research (Meter)*, 1(2), 64. <https://doi.org/10.33292/meter.v1i2.144>
- Amelia, M. A. (2017). *Analisis Soal Tes Hasil Belajar High Order Thinking Skills (HOTS) Matematika Materi Pecahan untuk Kelas 5 Sekolah Dasar*.
- Aulia Ulfa Dewi. (2016). Analisis Butir Soal Ujian Nasional Mata Pelajaran IPA. *Analisis Butir Soal Ujian Nasional Mata Pelajaran IPA Sekolah Dasarmadrasah Ibtidaiyah Tahun Ajaran 2014/2015, DIY*.
- Boopathiraj, C., & Chellamani, D. K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index In The Test For Research In Education. In *International Journal of Social Science & Interdisciplinary Research ISSN* (Vol. 2, Issue 2). February.
- Brown, H. D. (2004). *Language Assessment : Principles and Classroom Practices*. Pearson/Longman.
- Riswanda Himawan. (2022). *dipa, +12. +Riswanda*.
- Ebel, R. L., & Frisbie, D. A. (2016). *Essentials of Educational Measurement*.
- Eko Putra Widoyoko. (2009). *Evaluasi Program Pembelajaran*.
- Elviana. (n.d.). *Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates* (Vol. 10, Issue 2).
- Elviana. (2020). *Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates* (Vol. 10, Issue 2).
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian IPA Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Jurnal Natural Science Educational Research*, 4(1), 2021.
- Khasanah, I., Fuady, A., & Islam Malang, U. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. In *Mathema Journal E-ISSN* (Vol. 5, Issue 2).
- Latifah, I., Ardiyaningrum, M., Putri, M. A., & Arifin, A. S. (2024). Analisis Hasil Ujian Sekolah Berstandar Nasional di Sekolah Dasar Kecamatan Bulu pada Muatan Matematika. *Indonesian Journal of Elementary Education and Teaching Innovation*, 3(1), 52. [https://doi.org/10.21927/ijeeti.2024.3\(1\).52-63](https://doi.org/10.21927/ijeeti.2024.3(1).52-63)
- Lesta Ariany, R., Al, A., Kedua, G., Program,), Matematika, S., Tarbiyah, F., Uin, K., Gunung, S., & Bandung, D. (2018). *Penggunaan Software Anates Untuk Validasi Instrumen Tes*.
- Monica, S., Sendratasik, J., & Sudarman, Y. (2019). Analisis Butir Soal Ujian Tengah Semester Ganjil Seni Budaya Kelas VII di SMPN 29 Sijunjung. In *3 Seri A* (Vol. 7).

- 1277 *Analisis Soal Tes Hasil Belajar Siswa Soal Berstandar Nasional Bahasa Indonesia Kelas 5 Sekolah Dasar*, Syahria
DOI: <https://doi.org/10.31004/basicedu.v8i2.7354>
- Putu, I. A., Putri, G., Koyan, W., Program, C., Penelitian, S., & Pendidikan, E. (2013). *Program Pascasarjana Universitas Pendidikan Ganesha Program Studi Penelitian dan Evaluasi Pendidikan* (Vol. 3, Issue 1).
- Ratni, Pembeda, D., Pengecoh, E., Ulangan, S., Indonesia, B., Sekolah, D., & Ratni, D. (2023). *Prosa Jurnal Penelitian Pendidikan Guru Sekolah Dasar*. <https://doi.org/10.35326/prosa.v8i4.4234>
- Riani, D., Almujab, S., Dina, A., & Budiarto, R. (n.d.). *Analisis Butir Soal Dan Kemampuan Siswa dalam Menjawab Soal Ujian Nasional pada Mata Pelajaran Ekonomi*.
- Setiawati, S. (2019). Analisis Higher Order Thinking Skills (HOTS) Siswa Sekolah Dasar dalam Menyelesaikan Soal Bahasa Indonesia. *Prosiding Seminar Nasional Pendidikan Kaluni*, 2. <https://doi.org/10.30998/prokaluni.v2i0.143>
- Shermis, M. D., & di Vesta, F. J. (2011). *Classroom Assessment in Action*. Rowman & Littlefield Publishers.
- Solichin, M. (2017). *Analisis Daya Beda Soal*. www.depdiknas.go.id/evaluasi-proses-
- 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>
- Vina Nur Indah Sari, Nur Indah Sari, V., & Purwo Yudi Utomo, A. (2022). *JPBSI 11 (2) (2022) Kualitas Soal Bahasa Indonesia di SMP Muhammadiyah 1 Pontianak: Analisis Butir Soal*. <https://journal.unnes.ac.id/sju/index.php/jpbs>