

Didaktika Dwija Indria

Jurnal Ilmiah Pendidikan

ISSN 2337-8786 (Print) | ISSN 2775-2917 (Online)

Analisis butir soal dengan AnBuSo pada penilaian sumatif tengah semester genap mata pelajaran IPAS kelas IV SD

Rafidah Uswatun Hasanah¹, Idam Ragil Widiyanto Atmojo², dan Roy Ardiansyah³
^{1,2,3} Program Studi Pendidikan Guru Sekolah Dasar, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sebelas Maret, Surakarta, Indonesia

Email penulis korespondensi: dwiyuniasihsaputri@staff.uns.ac.id

Dikirim: 1 Juni

DOI: <https://doi.org/10.20961/ddi.v14i4>

Direvisi: 1 Juli

Diterima: 1 Agustus

Kata Kunci:

Abstrak

Item Analysis; Elementary School; IPAS; AnBuSo. Accurate assessment of learning outcomes depends heavily on the quality of test instruments, but at the elementary school level, quantitative item analysis is often overlooked. This study aims to evaluate the empirical quality of Mid-Semester Summative Assessment (PSTS) items for Grade IV Science at Ta'mirul Islam Elementary School in Surakarta to ensure the instrument is capable of measuring student competency validly and reliably. The research method used is descriptive quantitative with a content analysis approach. Data were collected through a documentary study of the test scripts and answers from 36 students, which were then processed using AnBuSo and SPSS applications. The results showed that 56% of the test items were declared valid, but the instrument reliability level was relatively low (0.217–0.512) because they were dominated by easy category items (89%). In addition, 78% of the test items had poor discriminatory power, even though the distribution of cognitive levels included 35% of Higher Order Thinking Skills (HOTS). The implications of this study emphasize that qualitative validation alone is not sufficient to guarantee assessment quality; periodic psychometric analysis is needed to improve the effectiveness of distractors and the balance of item difficulty levels. These findings recommend the use of simple software such as AnBuSo for teachers to conduct more precise diagnostic evaluations, so that the assessment results can reflect students' scientific abilities objectively and support continuous learning improvement.



PENDAHULUAN

Penilaian hasil belajar merupakan instrumen krusial dalam siklus pendidikan yang berfungsi sebagai alat ukur akurat terhadap ketercapaian kompetensi peserta didik (Nasaruddin, 2026). Instrumen yang berkualitas harus memenuhi kriteria teknis seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh agar tidak menghasilkan gambaran kemampuan siswa yang bias (Hariati et al., 2026). Di era Kurikulum Merdeka, tuntutan terhadap kualitas asesmen semakin meningkat, terutama pada mata pelajaran IPAS yang menekankan kemampuan berpikir kritis dan logis (Fadila & Fitriyeni, 2024). Secara teoretis, siswa kelas IV SD berada pada tahap operasional konkret yang mulai berkembang menuju pemikiran sistematis (Magdalena et al., 2023), sehingga penyelarasan antara tingkat kognitif soal dengan kemampuan berpikir siswa menjadi sangat signifikan untuk memastikan efektivitas pembelajaran (constructive alignment) (Ait-hroch et al., 2025).

Meskipun penting, praktik di lapangan menunjukkan adanya kesenjangan yang nyata. Di SD Ta'mirul Islam Surakarta, proses validasi soal Penilaian Sumatif Tengah Semester (PSTS) masih bersifat kualitatif tanpa disertai analisis kuantitatif yang empiris. Masalah spesifik muncul ketika rata-rata nilai IPAS kelas IV (76,9) menjadi yang terendah dibandingkan mata pelajaran lainnya. Padahal, pembelajaran IPA diperlukan keterampilan proses dari siswa untuk menemukan sejumlah konsep-konsep yang dipelajari (Atmojo, 2015). Tanpa analisis butir soal, sulit dipastikan apakah rendahnya nilai tersebut disebabkan oleh rendahnya kompetensi siswa atau ketidakefektifan butir soal dalam mengukur kemampuan mereka. Selain itu, keterbatasan waktu dan penguasaan perangkat statistik yang kompleks menjadi hambatan bagi guru dalam melakukan evaluasi mandiri terhadap kualitas instrumen yang mereka susun.

Sejumlah penelitian sebelumnya telah mengkaji kualitas soal menggunakan berbagai perangkat analisis, misalnya model Rasch, SPSS, Itean, Anates, hingga Excel (Anwar et al., 2024)(Basri et al., 2021). Hasil penelitian tersebut menegaskan pentingnya analisis butir soal dalam menjamin mutu instrumen penilaian. Salah satu perangkat yang sederhana dan praktis adalah AnBuSo (Analisis Butir Soal) dikembangkan berbasis Microsoft Excel, sehingga lebih mudah dioperasikan oleh guru dibandingkan program statistik seperti SPSS atau Itean (Muhson, 2017). Mata Pelajaran IPAS berperan penting dalam melatih kemampuan berpikir kritis, logis, dan analitis siswa (Sutinah et al., 2025). Kualitas soal yang rendah dapat menghambat tercapainya tujuan pembelajaran abad ke-21, termasuk pengembangan keterampilan *problem solving* dan *critical thinking* (Luna, 2015).

Kondisi terkini dalam bidang evaluasi pendidikan menekankan pentingnya analisis butir soal yang menyeluruh (Fitrianawati, 2015). Beberapa penelitian sebelumnya telah mengevaluasi pengelolaan kelas dan instrumen penilaian, namun sering kali terfokus pada sistem daring selama pandemi atau pada objek kelas tinggi. Penelitian oleh Muhson dkk. (Muhson, 2017) memperkenalkan perangkat lunak AnBuSo sebagai solusi praktis berbasis Excel untuk mengatasi hambatan teknis guru dalam menganalisis soal. Tinjauan literatur menunjukkan bahwa integrasi antara analisis statistik butir soal dengan distribusi tingkat kognitif Bloom sangat jarang

dilakukan secara spesifik pada mata pelajaran IPAS di jenjang awal kelas atas sekolah dasar.

Kebaruan (novelty) penelitian ini terletak pada penggunaan aplikasi AnBuSo untuk membedah kualitas soal PSTS IPAS secara integratif, yang mencakup aspek kuantitatif sekaligus distribusi tingkat kognitif pasca-pandemi di sekolah dasar swasta. Kesenjangan penelitian yang diisi adalah minimnya data empiris mengenai validitas dan reliabilitas soal IPAS yang selama ini hanya divalidasi secara bahasa di lokasi penelitian tersebut. Tujuan utama karya ini adalah untuk menganalisis validitas, reliabilitas, daya beda, tingkat kesukaran, efektivitas pengecoh, dan tingkat kognitif soal PSTS Genap IPAS Kelas IV. Kesimpulan utama yang diharapkan adalah tersedianya profil kualitas instrumen asesmen yang objektif sebagai dasar perbaikan penyusunan soal dan perencanaan program remedial yang lebih tepat sasaran bagi guru.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain deskriptif non-eksperimen melalui analisis isi kuantitatif. Desain ini dipilih karena penelitian tidak memberikan perlakuan terhadap variabel, melainkan bertujuan mendeskripsikan kualitas soal sebagaimana adanya (Arikunto, 2013; Ary et al., 2010; Sugiyono, 2020).

Lokasi dan waktu penelitian ditetapkan di SD Ta'mirul Islam Surakarta sebagai sumber utama soal Penilaian Sumatif Tengah Semester (PSTS) mata pelajaran IPAS, serta SD Muhammadiyah 5 Surakarta sebagai lokasi uji coba instrumen. Penelitian dilaksanakan pada semester genap tahun ajaran 2024/2025. Subjek penelitian terdiri dari (1) seluruh butir soal PSTS Genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta, dan (2) peserta didik kelas IV SD Muhammadiyah 5 Surakarta yang berjumlah 36 siswa (18 siswa kelas IVA dan 18 siswa kelas IVB).

Teknik sampling yang digunakan adalah total sampling untuk butir soal, dan purposive sampling untuk peserta didik. Pemilihan peserta didik mempertimbangkan kesesuaian kurikulum, jumlah siswa yang mencukupi, dan kemudahan administratif. Teknik pengumpulan data dilakukan melalui studi dokumentasi dengan mengumpulkan kisi-kisi, soal, kunci jawaban, serta lembar jawaban siswa.

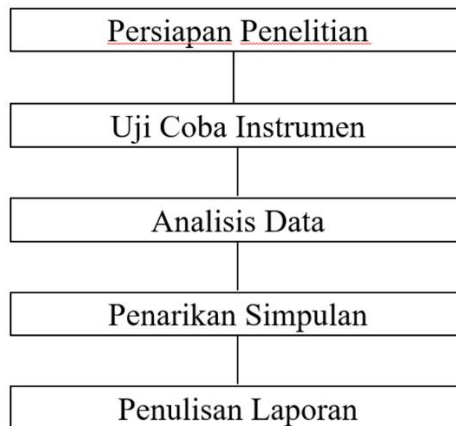
Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan metode analisis isi (content analysis). Desain ini dipilih untuk mendeskripsikan kualitas empiris butir soal tanpa memberikan perlakuan (non-experimental) terhadap subjek penelitian (Sugiyono, 2022). Fokus utama penelitian adalah mengevaluasi karakteristik teknis instrumen Penilaian Sumatif Tengah Semester (PSTS) Genap mata pelajaran IPAS melalui parameter statistik dan distribusi tingkat kognitif.

Data dalam penelitian ini terdiri dari data kualitatif berupa dokumen kisi-kisi dan butir soal, serta data kuantitatif berupa skor jawaban peserta didik. Sumber data penelitian berasal dari dua lokasi:

1. SD Ta'mirul Islam Surakarta: Sumber dokumen soal PSTS IPAS Kelas IV yang disusun secara mandiri oleh guru.

2. SD Muhammadiyah 5 Surakarta: Sumber data empiris hasil uji coba soal yang melibatkan 36 peserta didik kelas IV (Kelas IVA dan IVB). Sekolah ini dipilih melalui teknik purposive sampling berdasarkan kesamaan kurikulum dan aksesibilitas untuk mendapatkan sampel yang representatif (Sugiyono, 2017).

Seluruh instrumen yang dianalisis mencakup total populasi butir soal (total sampling) pada perangkat PSTS tersebut. Skema prosedur penelitian digambarkan pada Gambar 1.



Gambar 1. Skema Prosedur Penelitian

Teknik pengumpulan data menggunakan studi dokumentasi. Peneliti menghimpun dokumen fisik dan digital berupa Kisi-kisi soal dan naskah soal PSTS Genap IPAS Kelas IV; Kunci jawaban; dan Lembar jawab hasil pekerjaan peserta didik dari subjek uji coba.

Analisis data dilakukan melalui dua tahap integratif untuk mengekstraksi informasi kegunaan butir soal:

1. Analisis Kualitatif (Level Kognitif): Memetakan Kata Kerja Operasional (KKO) pada butir soal ke dalam taksonomi Bloom revisi (C1–C6) untuk melihat distribusi Low Order Thinking Skills (LOTS) dan High Order Thinking Skills (HOTS) (Anderson & Krathwohl, 2001).
2. Analisis Kuantitatif (Statistik Deskriptif): Mengolah data numerik hasil uji coba menggunakan perangkat lunak AnBuSo (Analisis Butir Soal) berbasis Microsoft Excel dan SPSS. Parameter analisis mengikuti kriteria yang ditetapkan oleh Budiyo (2023) dan Arikunto (2018) sebagaimana dirangkum dalam Tabel 1.

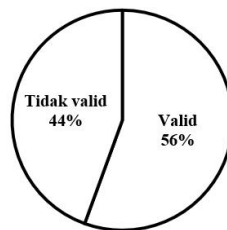
Tabel 1. Kriteria Analisis Butir Soal

Aspek Analisis	Kriteria Keberhasilan/kategori	Sumber
Validitas	Valid jika $r_{hitung} > r_{tabel}$ atau taraf signifikansi 5%,	Arikunto (2018)
Reliabilitas	Reliabel jika koefisiennya	Budiyono (2023)
Tingkat Kesukaran	Baik jika $0,30 \leq P \leq 0,70$	Budiyono (2023)
Daya Pembeda	Baik jika $D \geq 0,3$	Budiyono (2023)
Efektivitas Pengecoh	Berfungsi jika dipilih minimal 5%	Budiyono (2023)

Data hasil analisis ini kemudian digunakan untuk menentukan kebutuhan remedial per butir soal, sehingga memberikan umpan balik diagnostik yang spesifik bagi perbaikan pembelajaran (Nursobah, 2017).

HASIL

Tingkat validitas pada penelitian ini diuji dan dianalisis menggunakan SPSS dengan simpulan dikatakan valid apabila $r_{hitung} > r_{tabel}$ ($r_{tabel} = 0,329$ karena N atau total peserta didik adalah 36 orang). Hasil pengujian tersebut sebagian besar menghasilkan tingkat validitas pada rentang 0,336 – 1 yang bermakna bahwa soal Penilaian Sumatif Tengah Semester genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta tahun ajaran 2024/2025 sebagian besar dinyatakan valid.



Gambar 1. Distribusi Tingkat Validitas Soal PSTS Genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta

Gambar 1 menunjukkan bahwa berdasarkan hasil analisis uji validitas yang dilakukan dengan bantuan SPSS menghasilkan koefisien korelasi tiap butir soal pada taraf signifikansi 5% atau r_{tabel} sebesar 0,329 dengan persebaran sebanyak 44% soal tidak valid dan 56% soal valid. Berdasarkan hasil analisis tersebut dapat disimpulkan bahwa soal Penilaian Sumatif Tengah Semester genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta tahun ajaran 2024/2025 sebagian besar dinyatakan valid dan mampu menjalankan fungsinya sebagai alat ukur dengan baik. Faktor yang berpengaruh terhadap validitas hasil tes, yaitu faktor internal instrumen tes, administrasi dan penskoran, serta jawaban peserta didik (Sukardi, 2017). Masih adanya soal tidak valid pada penelitian ini disebabkan oleh faktor internal instrumen tes yaitu konstruksi item tes yang kurang baik, penggunaan diksi

yang terlalu mudah, jawaban item tes dapat diprediksi oleh peserta didik, dan mudahnya tingkat kesulitan soal.

Tabel 1. Uji Reliabilitas soal Pilihan Ganda

<i>Reliability Statistics</i>	
Cronbach's Alpha	N of Items
.512	20

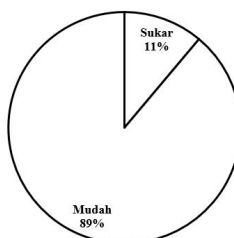
Tabel 2. Uji Reliabilitas soal Jawaban Singkat

<i>Reliability Statistics</i>	
Cronbach's Alpha	N of Items
.217	4

Tabel 3. Uji Reliabilitas soal Uraian

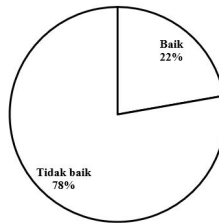
<i>Reliability Statistics</i>	
Cronbach's Alpha	N of Items
.463	3

Tabel 1, 2, dan 3 menunjukkan bahwa berdasarkan hasil analisis uji reliabilitas Alpha Cronbach menggunakan SPSS diperoleh nilai alpha berturut-turut sebesar 0,512, 0,217, dan 0,463 yang berarti reliabilitas pada butir soal pilihan ganda, jawaban singkat, dan uraian pada Penilaian Sumatif Tengah Semester genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta tahun ajaran 2024/2025 ini tergolong rendah dan belum dapat diterima. Faktor yang berdampak terhadap tingkat reliabilitas instrumen soal, yaitu panjang tes, persebaran skor, kesukaran tes, dan objektivitas (Karomah et al., 2022). Adapun salah satu penyebab rendahnya reliabilitas pada penelitian ini yaitu karena faktor kesukaran tes yang mayoritas mudah bagi peserta didik.



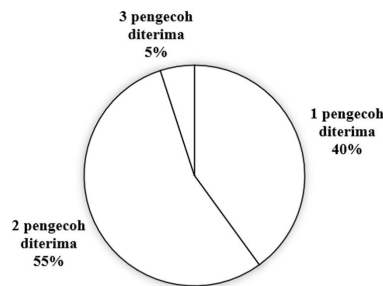
Gambar 2. Distribusi Tingkat Kesukaran Soal PSTS Genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta

Gambar 2 menunjukkan bahwa berdasarkan analisis uji tingkat kesukaran dengan bantuan program AnBuSo didapatkan hasil sebanyak 89% soal dengan tingkat kesukaran mudah dan 11% soal dengan tingkat sukar. Pada penelitian ini diketahui bahwa mayoritas butir soal memiliki tingkat kesukaran mudah. Hal tersebut disebabkan oleh tidak seimbangnya perbandingan peserta didik dengan jawaban benar dan salah terhadap butir soal yang diujikan. Hal tersebut selaras dengan pendapat yang mengungkapkan bahwasanya tingkat kesukaran butir soal adalah perbandingan peserta didik menjawab benar terhadap butir soal yang diujikan pada tes yang diikutinya (Rohma, 2023). Semakin besar rasio peserta tes dengan jawaban benar, maka tingkat kesukaran butir soal akan semakin rendah. Selain itu, banyaknya butir soal yang mudah menyebabkan soal kurang mampu melatih kemampuan berpikir peserta didik. Meskipun begitu, soal yang mudah dan sulit bukan berarti tidak boleh digunakan, akan tetapi persebarannya dalam sebuah instrumen tes haruslah berimbang.



Gambar 3. Distribusi Daya Pembeda Soal PSTS Genap mata pelajaran IPAS kelas IV SD Ta'mirul Islam Surakarta

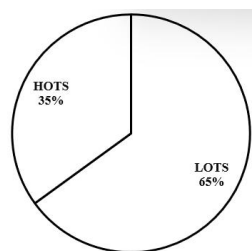
Gambar 3 menunjukkan bahwa berdasarkan analisis uji daya pembeda dengan bantuan program AnBuSo diperoleh hasil sebanyak 78% soal berdaya pembeda tidak baik dan 22% baik. Berdasarkan hasil analisis tersebut, daya pembeda pada penelitian ini kurang mampu dalam membedakan kemampuan peserta didik, sebab sebagian besar memiliki daya pembeda tidak baik. Hal ini karena analisis daya pembeda bermaksud untuk mengetahui seberapa jauh kekuatan butir soal dalam memisahkan penguasaan materi peserta didik berdasarkan kriteria tertentu (Arifin, 2017). Selain itu, faktor yang menyebabkan butir soal penelitian ini berdaya pembeda kurang baik, yaitu tingkat kesukaran soal yang terlampaui mudah sehingga tidak mampu membedakan kemampuan peserta didik dan soal memiliki instruksi yang terlalu jelas sehingga peserta didik dapat dengan mudah mengetahui jawaban tanpa benar-benar memahami materi yang diujikan.



Gambar 4. Diagram Kualitas Pengecoh

Gambar 4 menunjukkan bahwa berdasarkan analisis uji kualitas pengecoh pada soal pilihan ganda dan diperoleh hasil sebanyak 1 pengecoh diterima ditemukan di 40% butir soal, 2 pengecoh yang diterima ditemukan di 55% butir soal, dan 3 pengecoh diterima ditemukan di 5% butir soal. Distraktor dengan kriteria ditolak berarti pengecoh butir soal tersebut belum efektif sehingga membutuhkan perbaikan atau diganti, sedangkan distraktor dengan kriteria diterima berarti pengecoh butir soal tersebut sudah efektif atau sudah berfungsi (Rahayu & Djazari, 2016).

Analisis remedial menggunakan AnBuSo menunjukkan bahwa seluruh peserta didik lulus pada soal pilihan ganda dan uraian, tetapi masih terdapat sebagian kecil yang belum tuntas pada isian singkat, khususnya terkait konsep dasar gaya. Temuan ini menunjukkan bahwa pembelajaran IPAS telah berlangsung efektif, namun tetap diperlukan dukungan belajar tambahan bagi siswa yang mengalami kesulitan. Mekanisme remedial AnBuSo berbasis batas minimal 55% ketepatan individu dan ketuntasan klasikal kurang dari 15% menjamin proses penilaian yang objektif.



Gambar 5. Diagram Level Kognitif

Gambar 5 menunjukkan bahwa berdasarkan analisis level kognitif yang dilakukan dengan mengidentifikasi KKO pada kisi-kisi dan menyesuaikannya dengan taksonomi bloom revisi diperoleh hasil sebanyak 35% atau 7 butir soal memiliki level kognitif HOTS dan 65% soal atau 13 butir soal memiliki level kognitif LOTS. Hal tersebut menunjukkan bahwasanya sebagian besar butir soal yang disusun mengandalkan keterampilan mengingat, mengacu pada pengertian dan konsep saja (Nafiati, 2021)(Subandi et al., 2024), serta butir soal disampaikan secara gamblang sehingga jawaban mudah diidentifikasi benar dan salahnya. Walau demikian masih terdapat butir soal HOTS sehingga terdapat soal yang menuntut

berpikir tingkat tinggi.

PEMBAHASAN

Kualitas instrumen Penilaian Sumatif Tengah Semester (PSTS) Genap mata pelajaran IPAS Kelas IV di SD Ta'mirul Islam Surakarta menunjukkan karakteristik yang beragam namun cenderung memerlukan penguatan pada aspek reliabilitas dan daya pembeda. Hasil analisis validitas menunjukkan bahwa secara umum instrumen ini memiliki tingkat validitas sedang, dengan 56% butir soal dinyatakan valid ($r_{hitung} > 0,329$). Temuan ini mengindikasikan bahwa sebagian besar soal telah selaras dengan indikator pembelajaran yang ditetapkan. Namun, adanya 44% soal tidak valid mencerminkan adanya hambatan internal instrumen, seperti konstruksi item yang ambigu dan penggunaan diksi yang terlalu mudah ditebak. Hal ini sejalan dengan temuan Kurniawati dan Rahmawati (2022) bahwa ketidaksesuaian antara indikator materi dengan perumusan kalimat sering kali menjadi penyebab utama rendahnya validitas pada mata pelajaran sains di tingkat sekolah dasar.

Kesenjangan yang paling menonjol ditemukan pada aspek reliabilitas, di mana nilai koefisien Cronbach's Alpha untuk seluruh jenis soal (pilihan ganda 0,512; isian singkat 0,217; uraian 0,463) berada di bawah standar minimum 0,70. Rendahnya reliabilitas ini berkorelasi kuat dengan temuan pada tingkat kesukaran soal yang didominasi oleh kategori "mudah" (89%). Secara psikometris, tes yang terlalu mudah tidak memberikan persebaran skor yang cukup luas untuk menghasilkan konsistensi pengukuran yang stabil (Budiyono, 2023). Kondisi ini diperparah oleh daya pembeda soal yang 78% di antaranya masuk kategori tidak baik. Hal ini membuktikan bahwa instrumen tersebut belum mampu memisahkan kelompok siswa berkemampuan tinggi dan rendah secara efektif, karena instruksi soal yang terlalu gamblang memungkinkan siswa menjawab benar tanpa benar-benar melakukan proses kognitif yang mendalam.

Distribusi level kognitif dalam instrumen ini memperlihatkan dominasi Lower Order Thinking Skills (LOTS) sebesar 65%, sementara Higher Order Thinking Skills (HOTS) menempati porsi 35%. Meskipun didominasi LOTS, proporsi HOTS tersebut sebenarnya sudah cukup representatif untuk siswa kelas IV yang berada pada masa transisi menuju pemikiran sistematis (Putri & Fatmawati, 2024). Namun, implikasi praktis dari dominasi LOTS yang dikombinasikan dengan pengecoh yang kurang efektif (di mana mayoritas soal hanya memiliki 1-2 pengecoh berfungsi) menyebabkan instrumen ini lebih cenderung mengukur aspek hafalan daripada pemahaman konseptual. Sebagaimana ditegaskan oleh Biggs dan Tang (2011), tanpa constructive alignment yang kuat antara tujuan pembelajaran berpikir kritis dengan kualitas distractor soal, asesmen berisiko menjadi formalitas administratif belaka.

Menariknya, meskipun secara statistik kualitas butir soal perlu banyak perbaikan, analisis melalui aplikasi AnBuSo memberikan kontribusi signifikan dalam memetakan kebutuhan remedial secara presisi. Temuan bahwa tidak ada siswa yang remedi pada pilihan ganda namun terdapat kesulitan spesifik pada konsep gaya di isian singkat menunjukkan bahwa evaluasi ini berhasil menjalankan fungsi diagnostiknya. Arah penelitian masa depan perlu difokuskan pada

pengembangan bank soal yang terstandarisasi secara empiris di tingkat sekolah dasar, serta pelatihan guru dalam menyusun pengecoh (distraktor) yang logis dan homogen. Hal ini penting agar instrumen penilaian tidak hanya menjadi alat ukur hasil akhir, tetapi juga menjadi instrumen pendorong kualitas literasi sains siswa secara berkelanjutan.

KESIMPULAN

Hasil penelitian menunjukkan bahwa kualitas instrumen Penilaian Sumatif Tengah Semester (PSTS) Genap mata pelajaran IPAS Kelas IV di SD Ta'mirul Islam Surakarta secara umum belum optimal dan memerlukan perbaikan signifikan pada aspek teknis. Analisis empiris mengungkap bahwa hampir separuh butir soal (45% pilihan ganda) dinyatakan tidak valid dengan tingkat reliabilitas yang rendah, yaitu berkisar antara 0,217 hingga 0,512. Kesenjangan ini berdampak pada rendahnya daya pembeda soal, di mana 80% soal pilihan ganda dan seluruh soal isian singkat belum mampu membedakan kemampuan siswa secara efektif. Dari sisi tingkat kesukaran, instrumen ini didominasi oleh soal kategori mudah (95% pada pilihan ganda), yang menjelaskan mengapa rata-rata nilai siswa tampak cukup tinggi namun kurang mampu menjadi alat seleksi kompetensi yang akurat. Hal ini membuktikan bahwa validasi kualitatif yang selama ini dilakukan oleh tim kurikulum sekolah belum cukup untuk menjamin kualitas instrumen tanpa disertai analisis statistik yang objektif.

Meskipun demikian, penelitian ini menemukan sisi positif pada distribusi level kognitif yang telah mencakup aspek HOTS sebesar 35%, yang menunjukkan upaya guru dalam mendorong kemampuan berpikir kritis siswa pada fase operasional konkret. Temuan mengenai efektivitas pengecoh yang masih lemah memberikan dasar praktis bagi guru untuk merevisi alternatif jawaban agar lebih berfungsi secara diagnostik. Signifikansi penelitian ini terletak pada penggunaan aplikasi AnBuSo yang berhasil memetakan kebutuhan remedial secara personal per butir soal, bukan sekadar berdasarkan nilai akhir. Hal ini mewujudkan prinsip *constructive alignment* di mana hasil evaluasi langsung menjadi kompas untuk perbaikan instruksional.

Sebagai implikasi, sekolah perlu mengintegrasikan analisis butir soal berbasis teknologi sederhana sebagai prosedur standar pasca-ujian untuk menjamin akuntabilitas penilaian. Guru diharapkan mulai menyeimbangkan proporsi tingkat kesukaran dan memperkuat konsistensi antarbutir soal sebelum ujian dilaksanakan. Peneliti selanjutnya disarankan untuk memperluas cakupan subjek penelitian dan mengombinasikan metode kualitatif melalui wawancara mendalam dengan pengembang soal guna memahami kendala konstruksi soal dari sisi pedagogis. Dengan demikian, sistem evaluasi di sekolah dasar dapat berkembang menjadi instrumen yang tidak hanya mengukur, tetapi juga meningkatkan mutu pendidikan secara berkelanjutan.

DAFTAR PUSTAKA

- Ait-hroch, A., Gazi, S. El, & Ibrahim, A. (2025). The constructive alignment of objectives , tasks , and assessment in hybrid language training. *World Journal of Advanced Engineering Technology and Sciences*.
<https://doi.org/https://10.30574/wjaets>
- Anwar, Z., Masrifatin, Y., & Rahmania, N. (2024). *Analisis Butir Soal Penilaian Akhir Semester MI Menggunakan Rasch Model*. 2(2), 151–161.
- Arifin, Z. (2017). *Evaluasi Pembelajaran*. Direktorat Jenderal Pendidikan Islam Kementerian Agama RI.
- Arikunto, S. (2013). *Prosedur penelitian: Suatu pendekatan praktik (Edisi revisi)*. Rineka Cita.
- Ary, D., Jacobs, L. C., Sorensen, C., & Razavieh, A. (2010). *Introduction to research in education*.
- Atmojo, I. R. W. (2015). *Pengaruh Penggunaan Metode Discovery Berbasis Media Realita Terhadap Hasil Belajar Mata Kuliah Konsep Dasar*. 2(2), 130–139.
<https://doi.org/10.17509/mimbar-sd.v2i2.1324>
- Basri, K., Baidowi, B., Junaidi, J., & Turmuzi, M. (2021). Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP. *Griya Journal of Mathematics Education and Application*, 1(4), 682–694.
<https://doi.org/10.29303/griya.v1i4.107>
- Biggs, J., & Biggs, J. (2013). John Biggs. *Medical Education*, 33(8), 624–625.
<https://doi.org/10.1046/j.1365-2923.1999.00431.x>
- Budiyono. (2023). *Pengantar Metodologi Penelitian Pendidikan*. Surakarta: UNSPRESS.
- Bujuri, D. A. (2018). Analisis Perkembangan Kognitif Anak Usia Dasar dan Implikasinya dalam Kegiatan Belajar Mengajar. *LITERASI (Jurnal Ilmu Pendidikan)*, 9(1), 37-50. <https://doi:10.21927>
- Fadila, F., & Fitriyeni. (2024). Implementasi Kurikulum Merdeka pada Pembelajaran IPAS di Sekolah Dasar. *Jurnal Kependidikan*, 13(4), 4357–4366.
<https://doi.org/10.31004/basicedu.v7i5.5766>
- Fitrianawati, M. (2015). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, 5(3), 282–295.
- Hariati, B. S., Hakim, L., Pratiwi, V., & Ayu, R. C. (2026). Analisis Kualitas Butir Soal Pilihan Ganda Elemen Dokumen Berbasis Digital Menggunakan ANATES. *Jurnal Jendela Pendidikan*, 6(01), 46–58. <https://doi.org/10.57008/jjp.v6i01.1956>
- Karomah, A. K., Istiyati, S., & Surya, A. (2022). Analisis tingkat kognitif soal pada

- buku matematika kelas IV kurikulum merdeka. *Didaktika Dwija Indria*, 201–205. <https://doi.org/https://doi.org/10.20961/jpiuns.v9i4.82137>
- Luna, C. (2015). *Education Research And Foresight What Kind Of Learning*. 1–14.
- Magdalena, I., Nurchayati, A., Suhirman, D. P., Fathya, N. N., & Tangerang, U. M. (2023). Implementasi Teori Pengembangan Kognitif Jean Piaget Dalam Pembelajaran Ipa Di Sekolah Dasar. *ANWARUL Jurnal Pendidikan Dan Dakwah*, 3, 960–969. <https://doi.org/10.58578/anwarul.v3i5.1431>
- Muhson, A. (2017). Panduan Penggunaan AnBuso 2017. *Universitas Negeri Yogyakarta*.
- Nafiati, D. A. (2021). Revisi taksonomi Bloom: Kognitif, afektif, dan psikomotorik. *Humanika*, 21(2), 151–172. <https://doi.org/10.21831/hum.v21i2.29252>
- Nasaruddin, W. R. (2026). Analisis Penilaian Proses Dan Hasil Belajar Dalam Meningkatkan Kualitas Pembelajaran Abad Ke-21. *Kreatif*, 24, 169–180. <https://doi.org/https://10.21831/pep.v18i2.2481>
- Nouraey, P., Al-Badi, A., Riasati, M. J., & Maata, R. L. (2020). Educational Program and Curriculum Evaluation Models: A Mini Systematic Review of the Recent Trends. *Universal Journal of Educational Research*, 8(9), 4048–4055. <https://doi.org/10.13189/ujer.2020.080930>
- Nursobah, A. (2017). Perencanaan Pembelajaran MI/SD.Pamekasan : Duta Media Publishing.
- Putri, E., & Fatmawati, L. (2024). Analysis of High Order Thinking Skills (HOTS) in Objective Test Questions for Science Learning Material in Fourth Grade Elementary School. *International Journal of Learning Reformation in Elementary Education*, 3(03), 137–144. <https://doi.org/10.56741/ijlree.v3i03.624>
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akutansi Indonesia*, 0–2.
- Rohma, S. N. (2023). Analisis butir dan level kognitif pada soal penilaian tengah semester gasal mata pelajaran bahasa indonesia kelas iv sekolah dasar. *Didaktika Dwija Indria*, 449. <https://doi.org/https://doi.org/10.20961/ddi.v11i2.76111>
- Subandi, I. W., Atmojo, I. R. W., & Ardiansyah, R. (2024). Pengaruh case based learning terhadap hasil belajar kognitif mahasiswa ditinjau dari computational thinking. *Didaktika Dwija Indria*, 447–452. <https://doi.org/https://doi.org/10.20961/ddi.v13i3.93070>
- Sugiyono. (2020). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sukardi. (2017). *Evaluasi Pendidikan (Prinsip dan Operasionalnya)* (pp. 38–39). Bumi

Aksara.

Sutinah, C., Mahardhika, R. L., & Kartini, D. (2025). Mata Pelajaran Ilmu Pengetahuan Alam dan Sosial (IPAS) Ilmu Pengetahuan Alam dan Sosial (IPAS) Fase B dan C. *Badan Standar, Kurikulum, Dan Asesmen Pendidikan Kementerian Pendidikan Dasar Dan Menengah Republik Indonesia*.

Pernyataan Konflik Kepentingan: Penulis menyatakan bahwa penelitian dilakukan tanpa adanya hubungan komersial atau finansial yang dapat ditafsirkan sebagai potensi konflik kepentingan.