

Struktur Isi dan Tingkat Kognitif dalam Kisi-Kisi Penilaian Tertulis IPAS Tahun Ajaran 2024/2025

Bone Sugara Said^{*1}, Ruffi², Sabariah³

^{1,2,3}Universitas Adi Buana Surabaya, Surabaya, Indonesia

Email: ¹bonesaid26@gmail.com, ²ruffi@unipasby.ac.id, ³sabariah@unipasby.ac.id

Abstrak

Penilaian dalam pendidikan tidak lagi dipahami sekadar sebagai mekanisme pemberian nilai, melainkan sebagai proses terstruktur untuk merepresentasikan apa yang diharapkan dipelajari siswa dan bagaimana bukti pembelajaran dikumpulkan. Penelitian ini mengkaji kisi-kisi penilaian tertulis IPAS tahun ajaran 2024/2025 sebagai dokumen yang merepresentasikan struktur terencana dari evaluasi pembelajaran. Dengan menggunakan desain deskriptif kualitatif dan pendekatan analisis dokumen, penelitian ini menganalisis satu dokumen kisi-kisi yang memuat 25 butir soal pilihan ganda dengan alokasi waktu 90 menit. Analisis difokuskan pada pengorganisasian konten, distribusi tingkat kognitif, serta hubungan struktural antara capaian pembelajaran, cakupan materi, indikator, tingkat kognitif, nomor soal, dan format soal. Hasil penelitian menunjukkan bahwa dokumen tersusun secara sistematis dengan format yang konsisten pada seluruh 25 butir soal. Kisi-kisi mencakup beragam konten IPAS meliputi sistem organ manusia, energi, tata surya, komponen ekosistem, siklus air, geografi, sejarah, keragaman budaya, dan kegiatan ekonomi, serta mencerminkan tiga tingkat kognitif yaitu L1, L2, dan L3 yang mengindikasikan variasi tuntutan berpikir yang direncanakan. Namun, karena tidak tersedia data skor siswa, data respons, butir soal lengkap, maupun bukti psikometrik, penelitian ini tidak dapat membuat klaim empiris mengenai reliabilitas, tingkat kesukaran, daya pembeda, maupun kualitas pengecoh. Penelitian ini berkontribusi dengan menunjukkan bahwa kisi-kisi penilaian dapat dianalisis secara bermakna sebagai dokumen evaluasi sekaligus memperjelas batas inferensi dalam kondisi data yang terbatas. Guru dan pengembang instrumen disarankan memanfaatkan pendekatan analisis dokumen ini sebagai langkah awal dalam proses peningkatan kualitas penilaian sebelum data psikometrik tersedia.

Kata kunci: *analisis dokumen, evaluasi pendidikan, kisi-kisi penilaian, penilaian IPAS, tingkat kognitif,*

Assessment Blueprint Structure and Cognitive Levels in IPAS Written Assessment for Academic Year 2024/2025

Abstract

Assessment in contemporary education is no longer understood merely as a mechanism for assigning scores, but as a structured process for representing what students are expected to learn and how evidence of that learning is gathered. This study examines the IPAS written assessment blueprint for the 2024/2025 academic year as a document representing the planned structure of learning evaluation. Using a descriptive qualitative design and document analysis approach, the study analyzes a single blueprint document containing 25 multiple-choice items with a 90-minute testing allocation. The analysis focuses on content organization, cognitive-level distribution, and structural relationships among learning outcomes, content scope, indicators, cognitive levels, item numbers, and item format. Findings show the document is systematically organized with a consistent format across all 25 items. The blueprint covers a broad range of IPAS content including human organ systems, energy, the solar system, ecosystem components, the water cycle, geography, history, cultural diversity, and economic activities, and reflects three cognitive levels (L1, L2, L3) indicating planned variation in thinking demand. However, without student scores, response data, full test items, or psychometric evidence, no empirical claims can be made about reliability, difficulty, discrimination, or distractor quality. The study contributes by demonstrating that assessment blueprints can be meaningfully analyzed as evaluation documents while clarifying inferential limits under restricted data conditions. Teachers and instrument developers are encouraged to apply this document analysis approach as an initial step in improving assessment quality before psychometric data become available.

Keywords: *assessment blueprint, cognitive levels, document analysis, educational evaluation, IPAS assessment*

1. PENDAHULUAN

Penilaian dalam pendidikan kontemporer tidak lagi dipahami sekadar sebagai mekanisme pemberian nilai, melainkan sebagai proses terstruktur untuk merepresentasikan apa yang diharapkan dipelajari oleh siswa dan bagaimana bukti pembelajaran tersebut dikumpulkan. Dalam kerangka pemahaman yang lebih luas ini, kisi-kisi penilaian—yang kerap disebut sebagai *tabel spesifikasi*—memperoleh peran sentral karena secara eksplisit memetakan capaian pembelajaran, domain konten, dan tuntutan kognitif yang diharapkan ke dalam sebuah rencana pengujian yang terorganisir. Dalam literatur, kisi-kisi diperlakukan sebagai perangkat berorientasi validitas yang menjadi jangkar hubungan antara ekspektasi kurikulum dengan konstruksi butir soal, tidak hanya dalam tes bentuk tetap, tetapi juga dalam sistem penilaian berbasis program yang lebih luas [1]–[3].

Peran kisi-kisi yang semakin meluas ini menjadi lebih signifikan seiring dengan penekanan penilaian pendidikan yang mengutamakan kesesuaian, keterlacakan, dan kegunaan instruksional. Hasil penelitian empiris dan konseptual menunjukkan bahwa kisi-kisi menyandikan struktur yang dimaksudkan dari penilaian dengan mengorganisasikan domain, tingkat, dan jenis tugas secara sistematis, sehingga membantu evaluasi berbasis kelas dan sekolah mencerminkan perkembangan kurikuler [4]–[6]. Pendekatan berbasis *learning progression* terkini semakin memperkuat gagasan bahwa penilaian harus dirancang dari akun eksplisit tentang seperti apa pembelajaran itu dan jenis kinerja apa yang harus dihitung sebagai bukti pembelajaran tersebut [5]–[7].

Meskipun demikian, perlakuan akademis terhadap kisi-kisi penilaian masih belum merata, khususnya dalam konteks di mana peneliti bekerja dengan bukti yang terbatas. Sebagian besar penelitian penilaian mengevaluasi instrumen melalui indikator berbasis respons seperti reliabilitas, tingkat kesukaran soal, daya pembeda, korelasi soal-total, dan fungsi pengecoh. Ukuran-ukuran ini sangat penting dalam menilai seberapa baik soal berfungsi secara empiris, namun bergantung pada ketersediaan data respons siswa, catatan penskoran, dan analisis statistik. Di banyak lingkungan sekolah, bukti yang tersedia tidak selalu mencakup lebih dari kisi-kisi itu sendiri. Ketika hanya dokumen penilaian yang tersedia, peneliti dihadapkan pada pertanyaan metodologis: sejauh mana instrumen penilaian dapat dikaji berdasarkan struktur dokumenternya saja?

Permasalahan ini menjadi lebih nyata dalam penilaian tertulis berbasis sekolah yang mencakup domain pengetahuan yang luas, karena desain kisi-kisi harus menyeimbangkan cakupan konten dan tuntutan kognitif secara bersamaan. Di berbagai disiplin ilmu, tabel spesifikasi digunakan untuk mendistribusikan area konten dan menentukan proses kognitif yang diharapkan—seperti mengingat, menerapkan, dan menganalisis—agar butir soal dapat mengambil sampel capaian yang dituju [8], [9]. Dalam penilaian sains terpadu, kebutuhan ini bahkan lebih terlihat karena bidang ini semakin membutuhkan representasi konsep yang saling terkait, penalaran lintas domain, dan kinerja tingkat tinggi yang tidak dapat ditangkap secara memadai hanya melalui pengingat faktual semata [4], [10].

Literatur juga menunjukkan bahwa keselarasan antara capaian pembelajaran, indikator, dan tugas merupakan hal yang sentral bagi validitas. Dalam pendekatan berbasis argumen, kekuatan penilaian bergantung pada apakah klaim tentang kompetensi siswa dapat didasarkan pada kinerja yang dapat diamati dan terhubung secara koheren dengan tujuan kurikuler [11], [12]. Ketika capaian pembelajaran, indikator, dan tugas dipetakan secara sistematis melalui kisi-kisi, hal ini mendukung penyelarasan konstruktif dan argumen validitas yang lebih kuat untuk menafsirkan pengetahuan dan keterampilan siswa [8], [11]. Sebaliknya, divergensi antar komponen dapat mengancam validitas dengan mendistorsi apa yang diukur dan memperlemah klaim interpretatif [8], [12], [13].

Kesenjangan yang ada dalam literatur terletak pada kurangnya kajian yang secara sengaja berhenti pada level kisi-kisi dan memeriksa apa yang dapat dipelajari dari struktur rencana penilaian itu sendiri. Penelitian ini mengisi kesenjangan tersebut dengan mengkaji kisi-kisi penilaian tertulis IPAS tahun ajaran 2024/2025 sebagai dokumen evaluasi yang direncanakan. Pemilihan mata pelajaran IPAS pada tahun ajaran 2024/2025 secara khusus dilatarbelakangi oleh sifatnya yang terpadu—menggabungkan sains alam dan konteks sosial dalam satu kerangka penilaian—sehingga menuntut representasi kognitif yang lebih kompleks dibandingkan mata pelajaran tunggal dan menjadikannya kasus yang relevan untuk dianalisis secara mendalam. Dokumen ini memuat 25 butir soal pilihan ganda dengan alokasi waktu 90 menit dan memetakan elemen-elemen utama seperti capaian pembelajaran, cakupan materi, kelas/semester, indikator soal, tingkat kognitif, nomor soal, dan format soal. Penelitian ini tidak bertujuan untuk mengestimasi reliabilitas, tingkat kesukaran, daya pembeda, korelasi soal-total, atau efektivitas pengecoh. Sebaliknya, penelitian ini bertujuan untuk mendeskripsikan bagaimana kisi-kisi terstruktur, bagaimana cakupan konten didistribusikan, bagaimana tingkat kognitif direpresentasikan, dan bagaimana hubungan antara capaian pembelajaran dan indikator soal terorganisasi dalam dokumen tersebut.

2. TINJAUAN PUSTAKA

2.1. Evaluasi Instrumen sebagai Upaya Konseptual dan Empiris

Evaluasi instrumen dalam pengukuran pendidikan secara umum dipahami sebagai pengumpulan bukti secara sistematis untuk mendukung klaim tentang kualitas alat penilaian. Kualitas ini bertumpu pada apakah instrumen mengukur apa yang dimaksudkan dan apakah instrumen tersebut melakukannya secara konsisten lintas konteks. Literatur membedakan dua jalur yang saling melengkapi dalam evaluasi instrumen: pertama, evaluasi berbasis dokumen yang menekankan tinjauan desain, validitas isi, keselarasan kurikuler, dan penilaian ahli; kedua, analisis butir soal empiris yang mengandalkan data respons yang diamati, uji coba, dan pemodelan psikometrik [10], [14], [15]. Penelitian terkini berargumen bahwa kedua jalur ini tidak boleh diperlakukan sebagai pendekatan yang saling bersaing, melainkan sebagai dimensi yang sekuensial dan saling melengkapi dari argumen validitas yang koheren.

Evaluasi berbasis dokumen sangat penting pada tahap awal pengembangan penilaian. Pada tahap ini, perhatian bukan pada bagaimana siswa merespons butir soal, melainkan pada apakah instrumen secara konseptual terbentuk dengan baik, selaras dengan standar yang dimaksudkan, dan mampu merepresentasikan konstruk target. Para ahli menggambarkan fase ini sebagai saat di mana makna yang dimaksudkan dari penilaian dirakit melalui desain kisi-kisi, pemetaan konten, penyelarasan standar, dan validasi ahli [10], [14]. Dari perspektif ini, dokumen penilaian itu sendiri menjadi objek evaluasi karena ia mewujudkan asumsi tentang pengetahuan apa yang layak diukur, bagaimana konten harus disampling, dan apa yang merupakan tingkat tuntutan kognitif yang tepat.

2.2. Komponen Inti Kualitas Tes

Dalam berbagai kerangka pengukuran pendidikan, beberapa komponen kualitas tes muncul secara konsisten: reliabilitas, tingkat kesukaran butir soal, daya pembeda, korelasi soal-total, dan efektivitas pengecoh [1], [2], [9]. Reliabilitas mengacu pada konsistensi skor di berbagai kondisi pengukuran yang berulang atau relasi butir soal internal. Tingkat kesukaran butir soal menunjukkan sejauh mana suatu soal mudah atau sulit bagi kelompok peserta ujian tertentu. Daya pembeda mengacu pada kemampuan soal untuk membedakan peserta ujian dengan tingkat kemampuan yang berbeda. Korelasi soal-total menunjukkan seberapa baik suatu soal individual selaras dengan skor keseluruhan tes. Untuk tes pilihan ganda, efektivitas pengecoh menambah lapisan evaluasi tersendiri karena pilihan jawaban yang salah harus berfungsi dengan plausibel untuk menarik peserta ujian yang kurang siap.

Meskipun demikian, komponen-komponen kualitas ini mensyaratkan tersedianya data respons empiris. Ketika data tersebut tidak ada, komponen-komponen ini tetap menjadi titik referensi konseptual yang penting. Relevansinya bagi penelitian ini terletak pada klarifikasi tentang apa yang kisi-kisi dapat dan tidak dapat ungkapkan. Dokumen tersebut dapat menunjukkan bagaimana tes dimaksudkan untuk mendistribusikan materi dan tingkat kognitif, tetapi tidak dapat dengan sendirinya mengungkapkan apakah butir soal yang dihasilkan reliabel, diskriminatif, atau didukung oleh pengecoh yang berfungsi.

2.3. Hubungan antara Kisi-Kisi, Indikator, dan Kualitas Tes

Literatur memperlakukan kisi-kisi penilaian—yang sering dioperasionalkan sebagai tabel spesifikasi—sebagai pemetaan eksplisit dari capaian pembelajaran ke domain konten dan indikator kognitif. Dalam peran ini, kisi-kisi mendefinisikan bagaimana konten penilaian harus disampling, bagaimana tingkat kesukaran butir soal dapat didistribusikan, dan bagaimana keselarasan dapat didokumentasikan untuk argumen validitas selanjutnya [9], [12], [13]. Kisi-kisi oleh karena itu berfungsi sebagai infrastruktur konseptual dari sebuah tes.

Indikator butir soal memainkan peran penting dalam menerjemahkan kisi-kisi ke dalam bentuk operasional. Jika kisi-kisi mengekspresikan apa yang harus dicakup dan pada tingkat apa, maka indikator mengartikulasikan kinerja yang dapat diamati yang diharapkan dari siswa sebagai respons terhadap tugas tertentu. Literatur menunjukkan bahwa indikator yang diturunkan secara cermat dari kisi-kisi memungkinkan bukti tingkat butir soal terakumulasi secara koheren [13], [16]. Sebaliknya, divergensi antara kisi-kisi dan indikator dapat mengancam validitas karena melemahkan rantai penghubung antara niat kurikulum, desain tugas, dan pola respons yang diamati.

2.4. Kerangka Teoritis untuk Menganalisis Tingkat Kognitif dalam Penilaian

Analisis tuntutan kognitif dalam penilaian pendidikan telah lama bergantung pada kerangka taksonomi. Taksonomi Bloom dan revisinya tetap menjadi salah satu model yang paling banyak digunakan karena memberikan bahasa hierarkis untuk menggambarkan tuntutan butir soal, mulai dari mengingat dan memahami hingga menerapkan, menganalisis, mengevaluasi, dan mencipta. Banyak penelitian menggunakan pengkodean berbasis Bloom untuk mengklasifikasikan butir soal atau indikator penilaian, terutama ketika meneliti apakah penilaian kelas atau sekolah terlalu menekankan kognisi tingkat rendah dengan mengorbankan pemikiran tingkat tinggi [17], [18].

Literatur juga menunjukkan bahwa label tingkat kognitif tidak boleh diterima begitu saja. Sebaliknya, label tersebut harus dibaca dalam kaitannya dengan domain konten, rumusan indikator, dan teori pembelajaran yang terkandung dalam desain penilaian [17], [19]. Bagi penelitian ini, tubuh literatur ini memberikan lensa interpretatif yang penting. Kisi-kisi menyertakan tiga tingkat kognitif yang dinyatakan, yaitu L1, L2, dan L3. Meskipun tingkat-tingkat ini lebih sederhana dari rangkaian penuh model taksonomi yang digunakan dalam penelitian, ketiganya tetap menyiratkan pandangan bertingkat tentang tuntutan kognitif.

3. METODE PENELITIAN

3.1. Desain Penelitian

Penelitian ini menggunakan desain kualitatif deskriptif dengan pendekatan analisis dokumen. Pemilihan desain ini ditentukan oleh sifat bukti yang tersedia, yaitu sebuah dokumen kisi-kisi penilaian tunggal untuk ujian tertulis IPAS tahun ajaran 2024/2025. Karena material yang tersedia tidak mencakup respons siswa, skor, naskah ujian lengkap, kunci jawaban, maupun laporan psikometrik, penelitian ini tidak dirancang sebagai evaluasi uji empiris. Penelitian ini diposisikan dalam ranah evaluasi instrumen berbasis dokumen yang membedakan evaluasi berbasis desain dari evaluasi empiris, serta secara sengaja menempati sisi berbasis desain dari argumen validitas [10], [14], [15].

3.2. Sumber Data dan Unit Analisis

Satu-satunya sumber data adalah dokumen kisi-kisi penilaian untuk ujian tertulis IPAS tahun ajaran 2024/2025. Kisi-kisi memuat 25 butir soal pilihan ganda dan menetapkan total waktu ujian 90 menit. Setiap entri dalam dokumen mencakup elemen-elemen berikut: capaian pembelajaran, cakupan materi, kelas/semester, indikator soal, tingkat kognitif, nomor soal, dan format soal. Unit analisis adalah entri kisi-kisi pada tingkat butir soal. Setiap entri soal diperlakukan sebagai unit analitis karena menghubungkan satu atau lebih komponen pendidikan ke dalam satu garis terstruktur perencanaan penilaian.

Tabel 1. Ringkasan Dataset Berbasis Dokumen

Komponen	Keterangan
Jumlah dokumen	1 kisi-kisi penilaian
Mata pelajaran	IPAS
Tahun ajaran	2024/2025
Format tes	Tes tertulis pilihan ganda
Jumlah butir soal	25
Alokasi waktu	90 menit
Tingkat kognitif yang dinyatakan	L1, L2, L3
Elemen data per entri	Capaian pembelajaran, cakupan materi, kelas/semester, indikator soal, tingkat kognitif, nomor soal, format soal
Data tidak tersedia	Skor siswa, kunci jawaban, respons soal, koefisien reliabilitas, indeks kesukaran, daya pembeda, korelasi soal-total, statistik pengecoh

Tabel 1 merangkum dataset berbasis dokumen yang digunakan dalam penelitian ini, menunjukkan kekayaan struktural sekaligus keterbatasan empirisnya.

3.3. Prosedur Pengumpulan dan Pengorganisasian Data

Pengumpulan data tidak melibatkan kerja lapangan, observasi, wawancara, atau administrasi survei. Langkah pertama adalah menetapkan dokumen kisi-kisi sebagai korpus analisis. Langkah kedua adalah mengidentifikasi semua elemen struktural yang berulang dalam entri kisi-kisi. Setelah identifikasi, data diorganisasikan pada tingkat entri soal sesuai tiga dimensi utama: cakupan konten, formulasi indikator, dan klasifikasi tingkat kognitif. Langkah prosedural tambahan melibatkan penetapan batas analisis: penelitian ini secara eksplisit memisahkan fitur dokumenter yang dapat diamati dari klaim yang memerlukan verifikasi empiris. Misalnya, kehadiran indikator dan tingkat kognitif dalam kisi-kisi dapat dideskripsikan sebagai bukti struktur penilaian yang direncanakan, tetapi

bukan sebagai bukti bahwa butir soal yang dihasilkan benar-benar berfungsi pada tingkat tersebut dalam kinerja siswa yang sesungguhnya [14], [15].

3.4. Prosedur Analitis

Penelitian ini menggunakan analisis konten kualitatif sebagai prosedur analitis utama, diterapkan secara deskriptif. Terdapat empat fokus analitis: (1) distribusi konten—mengkaji area materi pelajaran yang muncul dalam kisi-kisi dan cakupan kurikulum IPAS yang direpresentasikan; (2) representasi tingkat kognitif—menganalisis bagaimana label L1, L2, dan L3 muncul dalam dokumen dan bagaimana label tersebut dikaitkan dengan indikator; (3) kesesuaian antara capaian pembelajaran, cakupan materi, dan indikator—membaca setiap entri sebagai struktur pemetaan untuk mengkaji apakah elemen-elemen yang tercantum tampak terhubung secara sistematis; dan (4) konsistensi format—menilai tingkat kestabilan pola dokumentasi di seluruh 25 entri. Untuk memastikan reliabilitas pengkodean tingkat kognitif (L1, L2, L3), setiap label yang tercantum dalam kisi-kisi dibaca secara berpasangan dengan redaksi indikator soal yang menyertainya, sehingga proses pengkodean dapat diverifikasi secara internal berdasarkan konsistensi antara label dan pernyataan kinerja yang dituntut. Tidak ada analisis statistik yang dilakukan karena data respons tidak tersedia.

4. HASIL DAN PEMBAHASAN

4.1. Konsistensi Format Dokumenter

Hasil yang paling menonjol dari analisis dokumen adalah tingginya konsistensi struktur formal kisi-kisi. Seluruh 25 entri mengikuti pola organisasi yang sama, masing-masing memuat capaian pembelajaran, cakupan materi, kelas/semester, indikator soal, tingkat kognitif, nomor soal, dan format soal. Tidak ditemukan entri yang menghilangkan kolom-kolom ini, maupun terdapat deviasi yang terlihat dalam tata letak dokumenter. Konsistensi ini mengindikasikan bahwa kisi-kisi dirancang menggunakan templat yang terstandarisasi, yang memungkinkan elemen-elemen terkait kurikulum dapat ditelusuri dari satu entri ke entri lainnya [1], [4]. Komponen struktural yang berulang disajikan pada Tabel 2.

Tabel 2. Komponen Struktural yang Berulang pada Seluruh Entri Kisi-Kisi

Komponen Struktural	Kehadiran pada 25 Entri
Capaian pembelajaran	Hadir pada semua entri
Cakupan materi	Hadir pada semua entri
Kelas/semester	Hadir pada semua entri
Indikator soal	Hadir pada semua entri
Tingkat kognitif	Hadir pada semua entri
Nomor soal	Hadir pada semua entri
Format soal	Hadir pada semua entri

Tabel 2 menunjukkan bahwa seluruh entri kisi-kisi mengikuti templat struktural yang sama, mengindikasikan format dokumenter yang seragam.

4.2. Cakupan Konten

Kisi-kisi menunjukkan cakupan konten yang luas dalam mata pelajaran IPAS. Domain konten yang dapat diidentifikasi secara eksplisit meliputi sistem organ manusia, energi, tata surya, komponen biotik dan abiotik, siklus air, kondisi geografis, perjuangan bersejarah, keragaman budaya, dan kegiatan ekonomi. Luasnya cakupan ini menunjukkan bahwa kisi-kisi berupaya merepresentasikan IPAS sebagai bidang terpadu, bukan sebagai topik yang tersegmentasi secara sempit. Temuan ini mendukung observasi dalam literatur bahwa kisi-kisi sering digunakan untuk mengorganisasikan cakupan konten secara non-acak dan memastikan beberapa domain kurikulum disampling secara disengaja [4], [5]. Domain konten yang teridentifikasi disajikan pada Tabel 3.

No.	Domain Konten yang Diidentifikasi
1	Sistem organ manusia
2	Energi
3	Tata surya
4	Komponen biotik dan abiotik
5	Siklus air
6	Kondisi geografis
7	Perjuangan bersejarah
8	Keragaman budaya
9	Kegiatan ekonomi

Kisi-kisi menunjukkan cakupan konten yang luas dalam mata pelajaran IPAS. Domain konten yang dapat diidentifikasi secara eksplisit meliputi sistem organ manusia, energi, tata surya, komponen biotik dan abiotik, siklus air, kondisi geografis, perjuangan bersejarah, keragaman budaya, dan kegiatan ekonomi. Luasnya cakupan ini menunjukkan bahwa kisi-kisi berupaya merepresentasikan IPAS sebagai bidang terpadu, bukan sebagai topik yang tersegmentasi secara sempit. Temuan ini mendukung observasi dalam literatur bahwa kisi-kisi sering digunakan untuk mengorganisasikan cakupan konten secara non-acak dan memastikan beberapa domain kurikulum disampling secara disengaja [4], [5]. Domain konten yang teridentifikasi disajikan pada Tabel 3.

Tabel 3. Domain Konten yang Direpresentasikan dalam Kisi-Kisi Penilaian IPAS

No.	Domain Konten yang Diidentifikasi
1	Sistem organ manusia
2	Energi
3	Tata surya
4	Komponen biotik dan abiotik
5	Siklus air
6	Kondisi geografis
7	Perjuangan bersejarah
8	Keragaman budaya
9	Kegiatan ekonomi

Tabel 3 memperlihatkan bahwa kisi-kisi mencakup tema sains alam maupun konteks sosial, konsisten dengan karakter terpadu mata pelajaran IPAS.

4.3. Representasi Tingkat Kognitif

Kisi-kisi secara eksplisit mengidentifikasi tiga tingkat kognitif: L1, L2, dan L3. Penggunaannya mengindikasikan bahwa dokumen tidak memperlakukan semua soal sebagai setara secara kognitif; setiap soal ditetapkan ke salah satu dari tiga tingkat tuntutan yang diharapkan. L1 merepresentasikan tingkat kognitif yang lebih rendah, L2 tingkat menengah, dan L3 tingkat yang lebih tinggi. Hal ini sejalan dengan rekomendasi literatur bahwa kisi-kisi penilaian perlu memasukkan variasi tuntutan kognitif untuk menghindari konsentrasi berlebihan pada keterampilan berpikir tingkat rendah [17], [18].

Namun, dokumen tidak menyediakan tabel frekuensi yang menunjukkan berapa banyak soal yang masuk ke setiap tingkat. Sebagai akibatnya, analisis ini dapat mengonfirmasi keberadaan tiga kategori kognitif yang berbeda tetapi tidak dapat melaporkan distribusi numerik yang tepat dari soal-soal di L1, L2, dan L3. Untuk memperkaya pemahaman tentang perbedaan antar tingkat, redaksi indikator dalam kisi-kisi memberi petunjuk penting: indikator L1 cenderung menggunakan kata kerja seperti "menyebutkan" atau "mengidentifikasi" yang menuntut pengingatan faktual; indikator L2 menggunakan kata kerja seperti "menjelaskan" atau "menghubungkan" yang menuntut pemahaman konseptual; sedangkan indikator L3 menggunakan kata kerja seperti "menganalisis" atau

"mengevaluasi" yang menuntut penalaran tingkat tinggi. Kehadiran ketiga tingkat tersebut tetap merupakan temuan struktural yang penting karena kisi-kisi membuat klasifikasi kognitif menjadi terlihat dalam perencanaan penilaian. Apa yang belum tersedia adalah bukti empiris yang menunjukkan apakah butir soal yang sesungguhnya berfungsi sesuai dengan tingkat yang dinyatakan tersebut [17], [19].

4.4. Kesesuaian antara Capaian Pembelajaran, Cakupan Materi, dan Indikator

Hasil utama lainnya adalah struktur kesesuaian yang tertanam dalam dokumen. Setiap entri soal menghubungkan capaian pembelajaran, cakupan materi, indikator, dan tingkat kognitif. Ini berarti kisi-kisi tidak hanya terorganisir sebagai daftar topik atau nomor soal; setiap entri dibingkai sebagai pemetaan antara komponen terkait kurikulum dan penilaian. Capaian pembelajaran tampak beroperasi sebagai jangkar kurikuler, cakupan materi sebagai referensi domain, indikator sebagai pernyataan operasional dari kinerja yang diharapkan, dan tingkat kognitif sebagai penanda formal dari tuntutan yang dimaksudkan. Hubungan yang didokumentasikan ini disajikan pada Tabel 4.

Tabel 4. Struktur Kesesuaian yang Didokumentasikan dalam Entri Kisi-Kisi

Elemen	Peran Dokumenter dalam Kisi-Kisi
Capaian pembelajaran	Jangkar kurikuler untuk apa yang akan dinilai
Cakupan materi	Referensi domain untuk topik atau materi
Indikator soal	Pernyataan operasional dari kinerja yang diharapkan
Tingkat kognitif	Penanda formal dari tuntutan yang dimaksudkan
Nomor soal	Referensi administratif untuk pengurutan
Format soal	Spesifikasi formal dari jenis respons

Tabel 4 menunjukkan bahwa kisi-kisi memuat komponen struktural minimum yang diperlukan untuk mendokumentasikan hubungan yang direncanakan antara capaian, konten, dan tuntutan tugas.

Kesesuaian yang terdokumentasi ini tidak membuktikan kualitas tes secara empiris, tetapi memberikan bukti bahwa penilaian dikonseptualisasikan melalui logika evaluatif yang dapat dikenali. Dalam pengertian ini, penelitian ini mengonfirmasi argumen sentral dalam literatur: kualitas penilaian dimulai bukan dari statistik respons, melainkan dari arsitektur desain yang menentukan apa yang dimaksudkan oleh penilaian untuk direpresentasikan [1], [8], [11]. Lebih jauh, kesesuaian antara capaian pembelajaran dan indikator yang teridentifikasi dalam kisi-kisi ini memiliki dampak instruksional yang nyata: guru dapat menggunakan peta kesesuaian ini sebagai panduan untuk menyelaraskan aktivitas belajar dengan tuntutan kognitif yang direncanakan, sehingga proses pembelajaran dan penilaian dapat saling mendukung secara konstruktif.

4.5. Keterbatasan Bukti dan Implikasi Metodologis

Pada saat yang sama, hasil analisis juga mengungkap ketidakhadiran bukti yang substansial. Dokumen tidak memuat data kinerja siswa maupun indikator psikometrik apa pun. Oleh karena itu, penelitian tidak dapat melaporkan reliabilitas, kesukaran empiris, daya pembeda, korelasi soal-total, efektivitas pengecoh, atau statistik keadilan. Keterbatasan ini tidak membatalkan nilai penelitian, tetapi mendefinisikan batasan argumennya dengan jelas [14], [15]. Analisis ini mendukung klaim tentang arsitektur perencanaan penilaian, bukan tentang fungsi tes yang telah selesai. Implikasi metodologis dari temuan ini adalah bahwa penelitian berbasis dokumen dapat membuat kontribusi akademis yang sah ketika klaim-klaimnya tetap proporsional dengan bukti yang tersedia.

Dari perspektif praktis, kisi-kisi ini dapat menjadi titik awal bagi proses peningkatan kualitas bertahap. Pertama, keselarasan yang ada antara capaian pembelajaran, indikator, dan tingkat dapat diaudit lebih detail. Kedua, spesifikasi butir soal dapat disempurnakan untuk memastikan bahwa stimulus tes yang sesungguhnya benar-benar mencerminkan tuntutan kognitif yang dinyatakan. Ketiga, setelah butir soal lengkap tersedia, uji coba dapat dilakukan untuk mengumpulkan data empiris minimal. Keempat, indikator empiris seperti tingkat kesukaran, daya pembeda, fungsi pengecoh, dan diagnostik keadilan dapat secara bertahap diintegrasikan ke dalam proses evaluasi [1], [9], [20].

5. KESIMPULAN

Penelitian ini menunjukkan bahwa kisi-kisi penilaian tertulis IPAS tahun ajaran 2024/2025 dapat dibaca sebagai dokumen evaluasi yang terstruktur, bukan sekadar catatan administratif. Dokumen tersebut

mendemonstrasikan format organisasi yang konsisten pada 25 butir soal pilihan ganda dan mencerminkan pemetaan sistematis antara capaian pembelajaran, cakupan materi, indikator soal, tingkat kognitif, nomor soal, dan format soal. Kisi-kisi juga mengungkapkan cakupan materi yang luas di sembilan domain IPAS dan penggunaan tiga tingkat kognitif—L1, L2, dan L3—mengindikasikan bahwa penilaian yang direncanakan berupaya mengakomodasi variasi dalam hal materi pelajaran maupun tuntutan berpikir.

Kontribusi penelitian ini terletak bukan pada penilaian kualitas empiris tes, melainkan pada klarifikasi batas antara analisis dokumen struktural dan evaluasi instrumen penuh. Penelitian ini menambahkan perspektif metodologis yang sederhana namun berguna pada penelitian penilaian pendidikan dengan mendemonstrasikan bahwa studi dengan data terbatas masih dapat memberikan kontribusi akademis yang valid ketika klaim-klaimnya tetap proporsional dengan bukti yang tersedia. Penelitian di masa depan sebaiknya mengintegrasikan analisis kisi-kisi dengan tinjauan butir soal dan data respons siswa untuk menghasilkan evaluasi yang lebih komprehensif terhadap kualitas penilaian IPAS. Ke depannya, pendekatan analisis dokumen serupa berpotensi diadaptasi untuk mengevaluasi kisi-kisi asesmen pada tingkat sekolah atau nasional, sehingga dapat mendukung perbaikan sistemik dalam desain penilaian secara lebih luas.

DAFTAR PUSTAKA

- [1] J. Schneider et al., "The Assessment for Learning Exchange: A Framework for Expanding the Roles of Assessment in Learning," *Stud. Educ. Eval.*, vol. 68, p. 100975, 2021. <https://doi.org/10.1016/j.stueduc.2021.100975>
- [2] C. Bengs, K. Hamberg, dan G. Westman, "Student and Teacher Perspectives on Assessment of Knowledge: A Tool for Learning or Surveillance?" *Adv. Health Sci. Educ.*, vol. 25, no. 5, hal. 1179–1196, 2020. <https://doi.org/10.1007/s10459-020-09961-4>
- [3] C. Udeozor, F. R. Abegão, dan J. Glassey, "Measuring Learning in Digital Games: Applying a Game-based Assessment Framework," *Br. J. Educ. Technol.*, vol. 55, no. 3, hal. 957–991, 2024. <https://doi.org/10.1111/bjet.13407>
- [4] S. R. Student dan B. Gong, "Designing for Coherence Across Assessment Claims: Lessons Learned from a Multi-grade Science Assessment System," *Educ. Meas. Issues Pract.*, vol. 41, no. 4, hal. 30–44, 2022. <https://doi.org/10.1111/emip.12505>
- [5] L. R. Harris, L. Adie, dan C. Wyatt-Smith, "Learning Progression–Based Assessments: A Systematic Review of Student and Teacher Uses," *Rev. Educ. Res.*, vol. 92, no. 6, hal. 996–1040, 2022. <https://doi.org/10.3102/00346543221081552>
- [6] A. C. Alonzo, M. Wooten, dan J. Christensen, "Learning Progressions as a Simplified Model: Examining Teachers' Reported Uses to Inform Classroom Assessment Practices," *Sci. Educ.*, vol. 106, no. 4, hal. 852–889, 2022. <https://doi.org/10.1002/sce.21713>
- [7] M. E. Walker et al., "Culturally Responsive Assessment: Provisional Principles," *ETS Res. Rep. Ser.*, vol. 2023, no. 1, hal. 1–24, 2023. <https://doi.org/10.1002/ets2.12374>
- [8] N. Teig dan T. Nilsen, "Profiles of Instructional Quality in Primary and Secondary Education: Patterns, Predictors, and Relations to Student Achievement and Motivation in Science," *Stud. Educ. Eval.*, vol. 74, p. 101170, 2022. <https://doi.org/10.1016/j.stueduc.2022.101170>
- [9] L. N. Pitot dan M. M. Balgopal, "Science Education Reform Conundrum: An Analysis of Teacher Developed Common Assessments," *Sch. Sci. Math.*, vol. 121, no. 5, hal. 299–309, 2021. <https://doi.org/10.1111/ssm.12472>
- [10] D. W. Dai et al., "Expanding Kane's Argument-based Validity Framework: What Can Validation Practices in Language Assessment Offer Health Professions Education?" *Med. Educ.*, vol. 58, no. 12, hal. 1462–1468, 2024. <https://doi.org/10.1111/medu.15452>
- [11] K. Kanaki, M. Kalogiannakis, E. Poulakis, dan P. Politis, "Investigating the Association Between Algorithmic Thinking and Performance in Environmental Study," *Sustainability*, vol. 14, no. 17, p. 10672, 2022. <https://doi.org/10.3390/su141710672>
- [12] M. A. Arrafii, "Teachers' Conceptions of Assessment in the Context of Assessment Reform in Indonesia," *Curric. J.*, vol. 33, no. 1, hal. 103–125, 2021. <https://doi.org/10.1002/curj.130>
- [13] V. Yaneva, B. E. Clauser, A. Morales, dan M. Paniagua, "Using Eye-Tracking Data as Part of the Validity Argument for Multiple-Choice Questions: A Demonstration," *J. Educ. Meas.*, vol. 58, no. 4, hal. 515–537, 2021. <https://doi.org/10.1111/jedm.12304>
- [14] C. F. Collares, "Cognitive Diagnostic Modelling in Healthcare Professions Education: An Eye-Opener," *Adv. Health Sci. Educ.*, vol. 27, no. 2, hal. 427–440, 2022. <https://doi.org/10.1007/s10459-022-10093-y>

-
- [15] P. Tscholl, F. Stampfer, dan T. Hell, "Does a Centralized Written Final Examination in Mathematics Indeed Improve Pupils' Subject-Related Study Ability?" *Eur. J. Sci. Math. Educ.*, vol. 12, no. 1, hal. 38–59, 2024. <https://doi.org/10.30935/scimath/13829>
- [16] A. N. Garguilo, "Differences in Question Reading and Responding Behaviour Across Reading Achievement Skills: An Eye-Tracking Study," *J. Res. Read.*, vol. 49, no. 1, 2025. <https://doi.org/10.1111/1467-9817.70018>
- [17] K. Marcq dan J. Braeken, "From Framework to Functionality: A Cross-Country Analysis of PISA 2018 Reading Assessment Framework's Item Features as Determinants of Item Difficulty," *Large-Scale Assess. Educ.*, vol. 13, no. 1, 2025. <https://doi.org/10.1186/s40536-025-00261-y>
- [18] Ö. F. TAVŞANLI, Y. Soysal, dan A. Kaldırım, "A Chronological Analysis of Intellectual Demands in the Turkish Writing Curricula," *SAGE Open*, vol. 13, no. 3, 2023. <https://doi.org/10.1177/21582440231190245>
- [19] S. Adofo, S. Kärkkäinen, dan T. Keinonen, "Examining Ghanaian Lower Secondary School Science Test Through Revised Bloom's Taxonomy," *Sci. Educ. Int.*, vol. 35, no. 4, hal. 382–393, 2024. <https://doi.org/10.33828/sei.v35.i4.9>
- [20] R. S. Baker, U. Boser, dan E. L. Snow, "Learning Engineering: A View on Where the Field Is At, Where It's Going, and the Research Needed," *Technol. Mind Behav.*, vol. 3, no. 1, hal. 1–23, 2022. <https://doi.org/10.1037/tmb0000058>
- [21] I. Rafi et al., "What Might Be Frequently Overlooked Is Actually Still Beneficial: Learning From Post National-Standardized School Examination," *Pedagog. Res.*, vol. 8, no. 1, 2023. <https://doi.org/10.29333/pr/12657>
- [22] W. Wang et al., "Applying Evidence-Centered Design in the Development of a Multidimensional Adaptive Reading Motivation Measure," *Educ. Meas. Issues Pract.*, vol. 40, no. 4, hal. 91–100, 2021. <https://doi.org/10.1111/emip.12468>
- [23] W. Ye, R. Strietholt, dan S. Blömeke, "Academic Resilience: Underlying Norms and Validity of Definitions," *Educ. Assess. Eval. Account.*, vol. 33, no. 1, hal. 169–202, 2021. <https://doi.org/10.1007/s11092-020-09351-7>
- [24] J. Bartolomé dan P. Garaizar, "Design and Validation of a Novel Tool to Assess Citizens' Netiquette and Information and Data Literacy Using Interactive Simulations," *Sustainability*, vol. 14, no. 6, p. 3392, 2022. <https://doi.org/10.3390/su14063392>
- [25] A. Akbari dan M. Shahrokhi, "Unveiling Fairness in Scoring: A Thorough Method for Precise Cutoff Score Calculation in Education Assessment," *Qual. Assur. Educ.*, vol. 32, no. 3, hal. 493–509, 2024. <https://doi.org/10.1108/qae-12-2023-0208>