

ANALISIS KUALITAS BUTIR SOAL BERBASIS LITERASI DAN NUMERASI DALAM MENDUKUNG IMPLEMENTASI KURIKULUM MERDEKA DI SEKOLAH DASAR

Suryadi Ishak^{1*}, Syahrul², Irfawandi Samad³, Naharuddin⁴

Program Studi Penelitian dan Evaluasi Pendidikan^{1,2,4}, Pendidikan Matematika³

Program Pascasarjana^{1,2,4}, Fakultas Keguruan dan Ilmu Pendidikan³, Universitas

Negeri Makassar^{1,2,4} Universitas Al Asyariah Mandar³

suryadi.ishak@unm.ac.id¹, syahrul@unm.ac.id², ippang05@gmail.com³

250012301006Student@unm.ac.id⁴

Abstrak

Asesmen berbasis literasi dan numerasi menempati posisi sentral dalam Kurikulum Merdeka, namun mutu butir soal yang disusun guru di sekolah dasar belum banyak diuji secara sistematis, padahal ketepatan pemetaan kemampuan peserta didik sepenuhnya bergantung pada kualitas instrumen tersebut. Penelitian ini bertujuan menganalisis kualitas butir soal berbasis literasi dan numerasi ditinjau dari validitas isi, tingkat kesukaran, daya pembeda, dan keberfungsian pengecoh, serta memetakan profil penguasaan peserta didik pada setiap aspek konten numerasi. Penelitian menggunakan pendekatan kuantitatif dengan metode analisis butir. Populasi penelitian adalah peserta didik kelas V sekolah dasar di Polewali Mandar, dengan sampel sebanyak 65 peserta didik yang ditentukan melalui teknik purposive sampling. Instrumen pengumpulan data berupa tes objektif sebanyak 20 butir soal pilihan ganda dengan empat alternatif jawaban, disertai lembar validasi ahli untuk menilai kesesuaian isi butir. Data validitas isi dianalisis menggunakan formula V Aiken, sedangkan data hasil tes dianalisis melalui analisis butir klasik (CTT) untuk memperoleh indeks tingkat kesukaran, daya pembeda, dan distribusi pemilihan pengecoh. Hasil penelitian menunjukkan seluruh butir memenuhi validitas isi pada kategori sedang hingga tinggi dengan rerata V sebesar 0,76, memiliki daya pembeda sangat baik pada rentang 0,792–0,944, dan seluruh pengecoh berfungsi karena dipilih di atas 5% peserta didik. Tingkat kesukaran berkisar 0,193–0,463 dengan rerata 0,318, terdiri atas 45% butir sulit dan 55% butir sedang tanpa satu pun butir mudah. Penguasaan peserta didik pada keempat aspek konten seluruhnya berada di bawah 50%, dengan capaian terendah pada aspek Kuantitas sebesar 32%. Disimpulkan bahwa instrumen layak digunakan namun perlu penambahan butir pada tingkat kesukaran rendah. Temuan ini mengimplikasikan perlunya penguatan kapasitas guru dalam merumuskan stimulus soal serta pembelajaran numerasi berbasis konteks nyata sebagai dasar pembelajaran terdiferensiasi.

Kata Kunci: Kualitas Butir Soal, Literasi Numerasi, Kurikulum Merdeka, Sekolah Dasar

A. Pendahuluan

Kurikulum Merdeka merupakan kebijakan strategis pemerintah dalam rangka meningkatkan mutu pendidikan nasional melalui penguatan kompetensi esensial peserta didik, khususnya literasi dan numerasi (Kemendikbudristek, 2022). Literasi dan numerasi tidak lagi dipahami sebatas kemampuan membaca teks dan melakukan operasi hitung, melainkan kemampuan memahami informasi, menalar secara logis, serta memecahkan masalah dalam berbagai konteks kehidupan nyata (Abidin, 2018). Orientasi ini sejalan dengan kerangka global pengukuran kompetensi yang dikembangkan OECD melalui *Programme for International Student Assessment* (PISA), yang menekankan penerapan pengetahuan dalam situasi autentik (OECD, 2019).

Penguatan literasi dan numerasi juga diwujudkan melalui kebijakan Asesmen Nasional, khususnya Asesmen Kompetensi Minimum (AKM), yang dirancang untuk mengukur kemampuan mendasar peserta didik tanpa berfokus pada penguasaan konten semata (Kemendikbudristek, 2021). Dengan demikian, sistem evaluasi pembelajaran di sekolah dasar dituntut mampu mengakomodasi karakteristik asesmen yang menilai kemampuan berpikir kritis, analitis, dan reflektif. Namun demikian, keberhasilan implementasi Kurikulum Merdeka sangat dipengaruhi oleh kualitas instrumen asesmen yang dikembangkan dan digunakan guru. Instrumen yang baik harus memenuhi prinsip validitas, reliabilitas, objektivitas, serta memiliki tingkat kesukaran dan daya pembeda yang memadai (Arikunto, 2018). Selain itu, efektivitas pengecoh (distraktor) dalam soal pilihan ganda juga menjadi indikator penting dalam menentukan kualitas butir soal (Sudijono, 2015).

Berbagai penelitian menunjukkan bahwa masih banyak butir soal di sekolah dasar yang berada pada level kognitif rendah dan belum sepenuhnya mengukur kemampuan berpikir tingkat tinggi atau pemecahan masalah kontekstual (Widana, 2016). Kondisi ini mengindikasikan adanya kesenjangan antara tuntutan kebijakan nasional dengan praktik evaluasi di lapangan. Analisis butir soal berbasis teori tes klasik (Classical Test Theory) menjadi pendekatan yang relevan untuk mengevaluasi kualitas instrumen secara empiris (Retnawati, 2016). Melalui analisis ini, dapat diidentifikasi butir soal yang layak digunakan, direvisi, atau dibuang

sehingga kualitas instrumen penilaian dapat ditingkatkan dan hasil pengukuran kemampuan peserta didik menjadi lebih akurat serta dapat dipertanggungjawabkan.

Penelitian terkait analisis butir soal telah banyak dilakukan, terutama dalam konteks evaluasi pembelajaran berbasis HOTS maupun kurikulum sebelumnya (Arikunto, 2018; Retnawati, 2016). Beberapa studi menunjukkan bahwa kualitas soal guru masih didominasi level kognitif rendah dan belum optimal dalam mengukur kemampuan analitis siswa (Nitko & Brookhart, 2014). Namun, penelitian yang secara spesifik menganalisis kualitas butir soal berbasis literasi dan numerasi dalam konteks implementasi Kurikulum Merdeka di Sekolah Dasar masih terbatas. Sebagian besar penelitian sebelumnya berfokus pada jenjang SMP atau SMA serta belum secara eksplisit mengaitkan hasil analisis dengan kebijakan AKM dan penguatan Profil Pelajar Pancasila (OECD, 2019; Brookhart, 2010). Sedangkan Kebaruan penelitian ini terletak pada, Fokus pada asesmen berbasis literasi dan numerasi dalam konteks Kurikulum Merdeka. Analisis komprehensif kualitas butir soal di jenjang Sekolah Dasar. Penyusunan rekomendasi perbaikan instrumen asesmen yang selaras dengan kebijakan pendidikan nasional terkini.

Penelitian ini bertujuan untuk mengembangkan instrumen literasi dan numerasi. Instrumen yang memiliki kualitas baik dengan pendekatan analisis butir seperti tingkat kesukaran, daya beda, dan distraktor, validitas isi, dan reliabilitas instrumen. Selanjutnya, melihat profil kemampuan siswa terhadap literasi dan numerasi pada seperti kontan.

B. Metode Penelitian

Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis butir soal (item analysis) untuk mencapai tujuan menganalisis kualitas instrumen asesmen berbasis literasi dan numerasi dalam mendukung implementasi Kurikulum Merdeka. Pendekatan kuantitatif dipilih karena penelitian ini bertujuan memperoleh data empiris berupa skor hasil tes siswa yang kemudian dianalisis menggunakan parameter statistik untuk menentukan kualitas setiap butir soal. Desain penelitian bersifat evaluatif, yaitu mengevaluasi instrumen yang telah digunakan guru dalam pembelajaran. Analisis dilakukan berdasarkan teori tes

klasik (Classical Test Theory/CTT), yang menekankan karakteristik butir soal melalui indikator validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor (Fraenkel & Hyun, 2012).

Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh siswa Sekolah Dasar di Kabupaten Polewali Mandar yang telah menerapkan Kurikulum Merdeka. Sampel penelitian ditentukan secara purposive sampling dengan lokasi penelitian di SDN 066 Pekkabata, Kabupaten Polewali Mandar. Sampel terdiri dari siswa kelas V (atau kelas yang relevan dengan implementasi literasi dan numerasi), dengan pertimbangan bahwa pada jenjang ini kemampuan literasi dan numerasi telah berkembang secara memadai untuk dianalisis melalui instrumen tes tertulis. Pemilihan sekolah didasarkan pada kesiapan sekolah dalam menerapkan Kurikulum Merdeka serta kesediaan menjadi mitra penelitian

Teknik Pengumpulan Data

Data dalam penelitian ini dikumpulkan melalui teknik dokumentasi dan tes. Dokumentasi dilakukan untuk memperoleh perangkat soal yang digunakan guru dalam pembelajaran. Teknik tes digunakan untuk memperoleh data kuantitatif berupa skor jawaban siswa terhadap butir soal literasi dan numerasi. Selain itu, dilakukan wawancara terbatas dengan guru untuk memperoleh informasi pendukung mengenai proses penyusunan butir soal yang digunakan dalam penelitian.

Prosedur Penelitian

Prosedur penelitian dilaksanakan melalui beberapa tahapan sistematis sebagai berikut; Tahap pertama adalah studi pendahuluan dan identifikasi instrumen yang telah digunakan guru dalam pembelajaran literasi dan numerasi. Pada tahap ini dilakukan telaah dokumen soal dan koordinasi dengan pihak sekolah. Tahap kedua adalah pelaksanaan tes kepada siswa menggunakan instrumen yang telah dikumpulkan. Lembar jawaban siswa dikumpulkan sebagai data utama penelitian. Tahap ketiga adalah analisis butir soal yang meliputi perhitungan validitas item, reliabilitas tes, indeks tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Analisis dilakukan secara statistik menggunakan perangkat lunak pengolahan data. Tahap keempat adalah interpretasi hasil analisis untuk menentukan kategori butir

soal (baik, direvisi, atau dibuang) serta penyusunan rekomendasi perbaikan instrumen. Tahap kelima adalah penyusunan laporan akhir dan diseminasi hasil penelitian kepada pihak sekolah sebagai bentuk umpan balik akademik.

Teknik Analisis Data

Data dianalisis menggunakan teknik analisis butir soal berdasarkan teori tes klasik (Gronlund & Linn, 2000). Validitas item dihitung menggunakan korelasi product moment antara skor item dan skor total. Reliabilitas tes dihitung menggunakan koefisien reliabilitas (misalnya Alpha Cronbach). Tingkat kesukaran dihitung berdasarkan proporsi siswa yang menjawab benar. Daya pembeda dihitung dengan membandingkan kelompok atas dan bawah. Efektivitas distraktor dianalisis berdasarkan distribusi pemilihan jawaban oleh siswa. Hasil analisis dikategorikan berdasarkan kriteria baku dalam evaluasi pendidikan (Azwar, 2018).

C. Hasil dan Pembahasan

Hasil

Penyajian hasil analisis data yang didasarkan pada empirik seperti uji validitas isi menggunakan V Aikens dan pembuktian reliabilitas dengan pendekatan cronbach alpha. Pengujian kualitas butir menggunakan pendekatan CTT yakni memeriksa tingkat kesukaran soal, daya beda soal, dan keberfungsian distraktor. Selain itu, pada bagian ini juga menyajikan profil kemampuan siswa dengan analisis deskriptif untuk melihat sebaran data dan profil penguasaan siswa dengan persentase untuk setiap aspek yang telah dikuasai oleh siswa. Oleh karena itu, hasil analisis disajikan dalam tabel untuk memudahkan pembacaan data, kemudian dibahas pada bagian berikutnya.

Uji Validitas Isi

Uji validitas isi melibatkan 5 orang validator, empat orang diantaranya dari unsur dosen yang ahli pada bidang literasi dan Numerasi serta asesmen, dan melibatkan satu orang guru yang ahli pada literasi dan numerasi. Tabel 1 berikut merupakan hasil analisis validitas isi dengan pendekatan V Aikens.

Tabel 1 Validitas isi instrumen dengan V Aikens

No Item	V Aiken	No Item	V Aiken
1	0,75	11	0,85
2	0,7	12	0,7
3	0,85	13	0,95
4	0,75	14	0,65
5	0,65	15	0,75
6	0,8	16	0,6
7	0,65	17	0,75
8	0,75	18	0,85
9	0,75	19	0,8
10	0,9	20	0,8

Hasil analisis validitas isi menggunakan koefisien Aiken's V menunjukkan bahwa 20 butir soal yang dikembangkan memiliki nilai koefisien antara 0,60 sampai 0,95, dengan nilai rata-rata sebesar 0,7625. Sebanyak 16 butir (80%) memperoleh nilai Aiken's $V \geq 0,70$, sedangkan empat butir (20%), yaitu butir 5, 7, 14, dan 16, memperoleh nilai di bawah 0,70. Selain itu, delapan butir (40%) menunjukkan nilai Aiken's $V \geq 0,80$, dengan nilai tertinggi terdapat pada butir 13 ($V = 0,95$), diikuti butir 10 ($V = 0,90$), serta butir 3, 11, dan 18 ($V = 0,85$). Temuan ini menunjukkan bahwa secara umum terdapat tingkat kesepakatan yang cukup baik di antara para ahli mengenai kesesuaian isi butir soal dengan aspek yang dinilai. Nilai Aiken's V sebesar $\geq 0,70$ umumnya digunakan sebagai batas minimal penerimaan validitas isi, sedangkan nilai yang semakin mendekati 1 menunjukkan semakin kuatnya kesepakatan ahli terhadap relevansi dan kesesuaian suatu butir.

Reliabilitas

Pembuktian reliabilitas ditujukan untuk melihat konsistensi suatu instrumen. Berdasarkan data empiris reliabilitas instrumen literasi dan Numerasi yang berisi 20 butir dengan 4 pilihan jawaban dan didasarkan pada aspek konten dari konstruk Literasi Numerasi. Pembuktian reliabilitas menggunakan pendekatan Cronbach Alpha diperoleh sebesar 0.893, hasil ini membuktikan bahwa instrumen memiliki konsistensi yang cukup tinggi, dimana besaran ini telah lebih tinggi dari nilai minimal yakni 0,4 (Baharuddin., et al. 2025).

Kualitas Butir

Pengujian kualitas butir untuk mengetahui

Tingkat Kesukaran Butir

Tingkat kesukaran menunjukkan proporsi peserta didik yang mampu menjawab suatu butir dengan benar, dan menjadi salah satu penanda utama mutu perangkat tes. Butir yang terlalu mudah tidak memberi informasi mengenai perbedaan kemampuan peserta didik, sedangkan butir yang terlalu sulit berisiko mendorong jawaban spekulatif sehingga skor tidak lagi mencerminkan penguasaan yang sebenarnya. Hasil analisis tingkat kesukaran terhadap 20 butir soal disajikan pada Tabel 2.

Table 2 Tingkat Kesukaran Butir Literasi dan Numerasi

Butir	Kesukaran		Butir	Kesukaran	
1	0.243	Sulit	11	0.263	Sulit
2	0.380	Sedang	12	0.203	Sulit
3	0.277	Sulit	13	0.350	Sedang
4	0.387	Sedang	14	0.237	Sulit
5	0.377	Sedang	15	0.317	Sedang
6	0.193	Sulit	16	0.390	Sedang
7	0.423	Sedang	17	0.373	Sedang
8	0.360	Sedang	18	0.310	Sedang
9	0.293	Sulit	19	0.243	Sulit
10	0.270	Sulit	20	0.463	Sedang

Analisis tingkat kesukaran terhadap 20 butir soal pilihan ganda dengan empat alternatif jawaban menunjukkan bahwa indeks kesukaran berada pada rentang 0,193–0,463, dengan rata-rata sebesar 0,318. Berdasarkan kriteria yang digunakan dalam penelitian ini, sebanyak 11 butir (55%) berada pada kategori sedang, sedangkan 9 butir (45%) berada pada kategori sulit. Tidak terdapat butir yang termasuk dalam kategori mudah. Komposisi tersebut menunjukkan bahwa instrumen secara keseluruhan memiliki tingkat kesukaran yang relatif tinggi, dengan proporsi butir sedang sedikit lebih dominan dibandingkan butir sulit. Butir dengan tingkat kesukaran terendah adalah butir 6 (0,193), sedangkan tingkat kesukaran tertinggi terdapat pada butir 20 (0,463).

Daya Pembeda

Daya pembeda menunjukkan kemampuan suatu butir dalam membedakan peserta didik berkemampuan tinggi dari peserta didik berkemampuan rendah. Butir dengan daya pembeda memadai akan lebih banyak dijawab benar oleh kelompok atas dibandingkan kelompok bawah, sehingga skor yang dihasilkan konsisten dengan kemampuan yang sesungguhnya. Sebaliknya, butir dengan daya pembeda rendah tidak mampu memisahkan kedua kelompok tersebut dan justru

mengaburkan hasil pengukuran, baik karena rumusannya ambigu maupun karena jawaban benar diperoleh secara kebetulan. Hasil analisis daya pembeda terhadap 20 butir soal disajikan pada Tabel 3.

Table 3 Daya beda butir instrumen Literasi dan Numerasi

Butir	Daya beda	Butir	Daya beda
V1	0.883 Sangat baik	V11	0.903 Sangat baik
V2	0.906 Sangat baik	V12	0.792 Sangat baik
V3	0.899 Sangat baik	V13	0.876 Sangat baik
V4	0.906 Sangat baik	V14	0.871 Sangat baik
V5	0.902 Sangat baik	V15	0.944 Sangat baik
V6	0.805 Sangat baik	V16	0.915 Sangat baik
V7	0.879 Sangat baik	V17	0.891 Sangat baik
V8	0.930 Sangat baik	V18	0.939 Sangat baik
V9	0.92 Sangat baik	V19	0.886 Sangat baik
V10	0.913 Sangat baik	V20	0.821 Sangat baik

Analisis daya pembeda menunjukkan seluruh butir berada pada kategori sangat baik dengan indeks berkisar antara 0,792 hingga 0,944 dan tidak satu pun butir tergolong lemah. Butir nomor 15 memiliki daya pembeda tertinggi (0,944), diikuti butir nomor 18 (0,939) dan nomor 8 (0,930), sedangkan nilai terendah dimiliki butir nomor 12 (0,792) yang tetap berada jauh di atas ambang minimum 0,40. Temuan ini menunjukkan bahwa seluruh butir mampu membedakan peserta didik berkemampuan tinggi dari peserta didik berkemampuan rendah secara konsisten. Yang lebih penting, hasil ini mengoreksi dugaan yang muncul dari analisis tingkat kesukaran. Pada lima butir yang indeks kesukarannya berada di sekitar peluang tebakan acak sebesar 0,25 yaitu butir nomor 1, 6, 12, 14, dan 19 sempat muncul kemungkinan bahwa jawaban benar berasal dari tebakan. Kemungkinan tersebut tidak didukung oleh data, sebab kelima butir justru memiliki daya pembeda yang tinggi (0,792–0,903). Apabila jawaban benar dihasilkan melalui tebakan, sebarannya akan merata pada semua kelompok kemampuan dan daya pembeda akan mendekati nol. Dengan demikian, kesulitan butir-butir tersebut bersifat substantif: butir memang menuntut kompetensi yang belum dikuasai sebagian besar peserta didik, bukan sekadar membingungkan.

Distraktor

Pengecoh atau distraktor merupakan alternatif jawaban selain kunci yang disediakan pada soal pilihan ganda. Sebuah pengecoh dinyatakan berfungsi apabila dipilih sekurang-kurangnya 5% peserta didik, yang menandakan alternatif tersebut

cukup meyakinkan untuk dipertimbangkan sebagai jawaban. Pengecoh yang tidak berfungsi menyebabkan peluang menjawab benar secara kebetulan meningkat karena peserta didik dapat mengeliminasinya tanpa penalaran, sehingga butir menjadi lebih mudah dari yang seharusnya. Distribusi pilihan jawaban peserta didik pada 20 butir soal disajikan pada Tabel 4

Table 4 Distraktor Literasi dan Numerasi

Butir	Pilihan				keberfungsian >5%
	A	B	C	D	
1	29,23%	32,31%	23,08%	15,38%	Berfungsi
2	46,15%	23,08%	15,38%	15,38%	Berfungsi
3	33,85%	23,08%	32,31%	10,77%	Berfungsi
4	32,31%	46,15%	12,31%	9,23%	Berfungsi
5	16,92%	10,77%	23,08%	49,23%	Berfungsi
6	26,15%	24,62%	21,54%	27,69%	Berfungsi
7	26,15%	15,38%	9,23%	49,23%	Berfungsi
8	46,15%	18,46%	10,77%	24,62%	Berfungsi
9	23,08%	35,38%	27,69%	13,85%	Berfungsi
10	20,00%	23,08%	32,31%	24,62%	Berfungsi
11	33,85%	12,31%	13,85%	40,00%	Berfungsi
12	36,92%	29,23%	12,31%	21,54%	Berfungsi
13	16,92%	30,77%	38,46%	13,85%	Berfungsi
14	38,46%	12,31%	23,08%	26,15%	Berfungsi
15	30,77%	38,46%	23,08%	7,69%	Berfungsi
16	12,31%	47,69%	27,69%	12,31%	Berfungsi
17	41,54%	16,92%	27,69%	13,85%	Berfungsi
18	10,77%	15,38%	35,38%	38,46%	Berfungsi
19	29,23%	27,69%	30,77%	12,31%	Berfungsi
20	12,31%	12,31%	23,08%	52,31%	Berfungsi

Seluruh persentase merupakan kelipatan 1,538%, yang mengonfirmasi jumlah responden 65 orang ($7,69\% = 5/65$; $49,23\% = 32/65$). Nilai terendah di seluruh tabel adalah 7,69%, sehingga tidak ada satu pun opsi di bawah ambang 5% berapa pun kunci jawabannya, seluruh pengecoh berfungsi.

Evaluasi Butir

Penilaian kelayakan butir tidak dapat didasarkan pada satu kriteria tunggal, melainkan menuntut pemaduan seluruh indikator yang telah dianalisis. Butir dengan tingkat kesukaran rendah, misalnya, masih dapat dipertahankan apabila daya pembedanya tinggi dan pengecohnya berfungsi, sebaliknya butir dengan tingkat kesukaran ideal tetap perlu direvisi bila validitas isinya belum memadai. Atas dasar itu, hasil analisis validitas isi, tingkat kesukaran, daya pembeda, dan keberfungsian pengecoh disandingkan untuk menghasilkan keputusan akhir

terhadap setiap butir, yang dikelompokkan menjadi diterima, diterima dengan revisi, atau ditolak. Rekapitulasi keputusan tersebut disajikan pada Tabel 5.

Table 5 Evaluasi Butir

Butir	Kesukaran	daya beda	Distraktor	Evaluasi awal
1	Sulit	Sangat baik	Berfungsi	Dipertahankan
2	Sedang	Sangat baik	Berfungsi	Dipertahankan
3	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
4	Sedang	Sangat baik	Berfungsi	Dipertahankan
5	Sedang	Sangat baik	Berfungsi	Dipertahankan
6	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
7	Sedang	Sangat baik	Berfungsi	Dipertahankan
8	Sedang	Sangat baik	Berfungsi	Dipertahankan
9	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
10	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
11	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
12	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
13	Sedang	Sangat baik	Berfungsi	Dipertahankan
14	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
15	Sedang	Sangat baik	Berfungsi	Dipertahankan
16	Sedang	Sangat baik	Berfungsi	Dipertahankan
17	Sedang	Sangat baik	Berfungsi	Dipertahankan
18	Sedang	Sangat baik	Berfungsi	Dipertahankan
19	Sulit	Sangat baik	Berfungsi	Dipertahankan/review
20	Sedang	Sangat baik	Berfungsi	Dipertahankan

Profil Kemampuan siswa

Analisis Deskriptif

Statistik deskriptif menyajikan kecenderungan pemusatan dan penyebaran skor yang menjadi dasar penafsiran analisis selanjutnya, sebab rerata capaian yang sama dapat berasal dari sebaran yang sangat berbeda dan menuntut tindak lanjut pembelajaran yang tidak serupa pula. Rangkuman statistik deskriptif hasil pengukuran kemampuan literasi dan numerasi peserta didik disajikan pada Tabel 6.

Table 6 Analisis deskriptif

Deskriptif	Nilai
Skor maksimum	20
Skor minimum	0
Rata-rata skor	7,78
Median	4
Rata-rata persentase jawaban benar	38.89%

Kemampuan Siswa Terhadap Penguasaan

Skor perolehan menjadi dasar bagi seluruh penafsiran berikutnya, baik dalam menentukan posisi capaian peserta didik terhadap kompetensi yang diharapkan

maupun dalam memetakan keragaman kemampuan di dalam kelas. Skor perolehan peserta didik pada 20 butir soal yang diujikan disajikan pada Tabel 7

Table 7 Skor Jawaban perolehan siswa

Rentang skor	Interpretasi deskriptif	Jumlah
0–5	Sangat rendah	34
6–10	Rendah	7
11–15	Sedang	6
16–20	Tinggi	18
Total		65

Profil Penguasaan Siswa terhadap Aspek Konten

Analisis kualitas butir memberikan gambaran mengenai mutu instrumen, namun belum menjelaskan kondisi kemampuan peserta didik yang diukurnya. Karena itu, hasil pengukuran ditelaah lebih lanjut untuk memperoleh profil penguasaan peserta didik pada setiap aspek konten numerasi, yaitu Perubahan dan Hubungan, Ruang dan Bentuk, Kuantitas, serta Ketidakpastian dan Data. Penelaahan per aspek diperlukan karena capaian yang setara pada skor total dapat menyembunyikan kesenjangan penguasaan antardomain sementara penentuan prioritas tindak lanjut pembelajaran justru bergantung pada letak kesenjangan tersebut. Dalam kerangka Kurikulum Merdeka, informasi semacam ini menjadi dasar bagi guru dalam merancang pembelajaran terdiferensiasi sesuai kebutuhan aktual peserta didik. Rata-rata penguasaan peserta didik pada keempat aspek konten disajikan pada Tabel 8.

Table 8 Profil Penguasaan Siswa

Aspek Konten	Rata-rata Penguasaan
Perubahan dan Hubungan	41%
Ruang dan Bentuk	37%
Kuantitas	32%
Ketidakpastian dan Data	41%

Secara umum, kemampuan literasi numerasi peserta masih relatif rendah dan menunjukkan heterogenitas yang tinggi. Hal tersebut ditunjukkan oleh rata-rata capaian sebesar 38,89%, dengan hampir separuh peserta (47,22%) memperoleh skor 0. Pada tingkat butir, proporsi jawaban benar berkisar antara 19,4% hingga 52,8%, menunjukkan bahwa sebagian besar butir relatif sulit bagi peserta. Meskipun demikian, terdapat kelompok peserta dengan kemampuan sangat tinggi, ditunjukkan oleh 16,67% peserta yang memperoleh skor sempurna. Kondisi ini

menunjukkan adanya kesenjangan kemampuan literasi numerasi yang cukup besar antarpeserta.

Pembahasan

Hasil penelitian menunjukkan bahwa kualitas butir soal berbasis literasi dan numerasi perlu dipahami tidak hanya berdasarkan karakteristik statistik setiap butir, tetapi juga berdasarkan sejauh mana butir tersebut mampu merepresentasikan tuntutan kompetensi yang dikembangkan dalam pembelajaran pada konteks Kurikulum Merdeka. Analisis terhadap karakteristik butir memberikan gambaran mengenai kemampuan instrumen dalam mengukur kompetensi peserta didik secara tepat, baik dari aspek tingkat kesukaran, daya pembeda, maupun karakteristik lainnya yang relevan. Dalam konteks literasi dan numerasi, kualitas tersebut menjadi semakin penting karena soal diharapkan tidak sekadar mengukur kemampuan mengingat atau menerapkan prosedur rutin, tetapi mendorong peserta didik untuk memahami informasi, menggunakan representasi, melakukan penalaran, menginterpretasikan data, serta mengambil keputusan berdasarkan permasalahan yang kontekstual. Oleh karena itu, pembahasan hasil penelitian ini diarahkan untuk menginterpretasikan kualitas butir secara menyeluruh dan mengaitkannya dengan karakteristik asesmen literasi dan numerasi serta kebutuhan implementasi Kurikulum Merdeka di sekolah dasar.

Hasil analisis validitas isi terhadap 20 butir soal menunjukkan nilai V Aiken berkisar antara 0,60 hingga 0,95 dengan rerata 0,76. Merujuk kriteria Retnawati (2016), seluruh butir berada pada kategori sedang hingga tinggi dan tidak satu pun tergolong rendah, dengan rincian 5 butir (25%) berkategori tinggi ($V > 0,80$) yaitu butir nomor 3, 10, 11, 13, dan 18, serta 15 butir (75%) berkategori sedang. Butir nomor 13 memperoleh nilai tertinggi ($V = 0,95$), sedangkan butir nomor 16 terendah ($V = 0,60$), disusul butir nomor 5, 7, dan 14 yang masing-masing bernilai 0,65. Berdasarkan catatan penilai, kelemahan butir-butir tersebut terletak pada aspek bahasa, sehingga menjadi dasar dilakukannya revisi. Kecenderungan ini merupakan temuan yang lazim pada pengembangan instrumen literasi dan numerasi jenjang sekolah dasar: soal berbasis literasi menuntut stimulus berupa teks, tabel, atau grafik, sementara keterbacaannya sering belum disesuaikan dengan tingkat perkembangan bahasa peserta didik, sehingga kesulitan yang muncul bersumber

dari beban membaca dan bukan dari kompleksitas kompetensi yang diukur. Perbedaan kesimpulan antara kedua kriteria tersebut menegaskan bahwa keputusan mempertahankan atau merevisi butir tidak dapat bertumpu pada angka semata, melainkan harus dipadukan dengan telaah kualitatif atas masukan validator serta analisis butir secara empiris.

Analisis tingkat kesukaran terhadap 20 butir soal pilihan ganda dengan empat alternatif jawaban menunjukkan indeks kesukaran berkisar antara 0,193 hingga 0,463 dengan rerata 0,318. Sebanyak 9 butir (45%) tergolong sulit dan 11 butir (55%) tergolong sedang, sementara tidak satu pun butir berada pada kategori mudah. Butir nomor 6 merupakan butir tersulit ($p = 0,193$), diikuti butir nomor 12 ($p = 0,203$) dan nomor 14 ($p = 0,237$), sedangkan butir nomor 20 menjadi butir termudah ($p = 0,463$). Sebaran tersebut menyimpang dari proporsi ideal 25% mudah, 50% sedang, dan 25% sulit yang lazim disarankan dalam penyusunan perangkat tes. Yang lebih perlu dicermati, dengan empat alternatif jawaban peluang menjawab benar secara acak adalah 0,25, sementara lima butir nomor 1, 6, 12, 14, dan 19 memiliki indeks kesukaran pada atau di bawah ambang tersebut. Kondisi ini mengindikasikan bahwa proporsi jawaban benar pada butir-butir tersebut belum dapat dibedakan secara meyakinkan dari hasil tebakan. Apabila diterapkan koreksi terhadap peluang tebakan, rerata perangkat diperkirakan turun menjadi sekitar 0,09, yang menegaskan bahwa perangkat soal ini tergolong terlalu berat bagi populasi sasaran. Ketiadaan butir mudah turut memperlemah kemampuan tes memetakan peserta didik pada kelompok kemampuan rendah, karena tidak tersedia butir yang berfungsi sebagai pijakan awal padahal fungsi pemetaan inilah yang justru dibutuhkan dalam asesmen diagnostik,

Tingginya proporsi butir sulit dapat dipahami sebagai konsekuensi karakteristik soal berbasis literasi dan numerasi yang menuntut peserta didik memahami stimulus terlebih dahulu sebelum menyelesaikan persoalan. Beban kognitif ganda tersebut membaca dan menafsirkan stimulus, lalu menerapkan penalaran menjadikan butir terasa lebih berat dibandingkan soal konvensional yang bersifat langsung, dan sejalan dengan pola capaian literasi serta numerasi peserta didik sekolah dasar yang secara nasional masih tertinggal [sitasi]. Namun demikian, kesulitan butir tidak sepenuhnya sejajar dengan hasil validitas isi, sehingga penyebabnya perlu ditelusuri

secara berbeda-beda. Butir nomor 14 memperoleh nilai V Aiken yang relatif rendah (0,65) sekaligus indeks kesukaran yang rendah (0,237), sehingga kesulitannya patut diduga bersumber dari kelemahan rumusan butir dan menuntut revisi redaksional. Sebaliknya, butir nomor 6 dan 12 memiliki validitas isi yang memadai ($V = 0,80$ dan $0,70$) namun tetap sangat sulit, yang mengindikasikan kesulitan bersifat substantif dan berkaitan dengan tuntutan kompetensi itu sendiri. Pola sebaliknya tampak pada butir nomor 16 yang justru memiliki tingkat kesukaran sedang ($p = 0,390$) meskipun nilai validitas isinya terendah ($V = 0,60$), sehingga kelemahannya terletak pada representasi konstruk, bukan pada keterjangkauan butir bagi peserta didik. Perbedaan ketiga pola ini penting karena menentukan arah tindak lanjut yang berbeda: perbaikan redaksi, pemertahanan butir sebagai penantang sepanjang daya pembedanya memadai, atau penyelarasan ulang butir terhadap indikator kompetensi.

Konsistensi daya pembeda yang tinggi pada seluruh butir mengindikasikan bahwa perangkat soal ini mengukur konstruk yang relatif homogen, yakni kemampuan literasi dan numerasi sebagai satu kesatuan kemampuan bernalar atas stimulus. Peserta didik yang mampu menafsirkan stimulus pada satu butir cenderung mampu pula melakukannya pada butir lain, sehingga pola respons terhadap seluruh butir saling selaras. Namun demikian, keseragaman nilai yang tinggi ini perlu ditafsirkan secara berhati-hati. Nilai daya pembeda yang seluruhnya berada di atas 0,79 tanpa variasi berarti merupakan pola yang tidak lazim ditemukan pada uji coba instrumen di lapangan, dan berpotensi dipengaruhi oleh jumlah butir dan jumlah siswa yang terbatas. Secara keseluruhan, pemaduan ketiga kriteria menghasilkan kesimpulan bahwa instrumen ini layak digunakan dengan catatan: seluruh butir memiliki daya pembeda memadai, sebagian besar memenuhi validitas isi, namun sebarannya terlalu terkonsentrasi pada tingkat kesukaran sulit dan sedang sehingga perlu ditambahkan butir dengan tingkat kesukaran rendah agar perangkat mampu memetakan peserta didik pada seluruh rentang kemampuan.

Analisis keberfungsian pengecoh terhadap 20 butir soal pilihan ganda menunjukkan hasil yang konsisten. Merujuk kriteria bahwa sebuah pengecoh dinyatakan berfungsi apabila dipilih sekurang-kurangnya 5% peserta didik, seluruh alternatif jawaban pada semua butir memenuhi kriteria tersebut. Proporsi pemilihan

terendah tercatat 7,69% pada butir nomor 15, disusul 9,23% pada butir nomor 4 dan nomor 7, serta 10,77% pada butir nomor 3, 5, 8, dan 18. Dengan demikian tidak ditemukan pengecoh mati, yakni alternatif yang tidak dipilih atau nyaris tidak dipilih karena kekeliruannya terlalu kentara. Temuan ini menunjukkan bahwa alternatif jawaban tidak disusun sebagai pelengkap semata, melainkan merepresentasikan kemungkinan jawaban yang secara nalar masuk akal bagi peserta didik sekolah dasar. Hasil ini sekaligus menjelaskan temuan pada analisis sebelumnya, yakni mengapa butir dengan tingkat kesukaran tinggi tetap memiliki daya pembeda yang sangat baik. Pengecoh yang efektif menuntut peserta didik membedakan jawaban yang benar dari jawaban yang tampak benar, sehingga hanya peserta didik dengan pemahaman memadai yang mampu memilih kunci kondisi yang justru mempertajam kemampuan butir dalam membedakan tingkat kemampuan.

Meskipun demikian, pola sebaran pilihan pada beberapa butir menuntut perhatian tersendiri. Pada butir nomor 6, keempat alternatif dipilih dalam proporsi yang hampir merata, yaitu 26,15%, 24,62%, 21,54%, dan 27,69%; pola serupa tampak pada butir nomor 1 dan nomor 19 dengan selisih antar opsi tertinggi dan terendah yang relatif sempit. Sebaran mendekati rata semacam ini berarti kunci jawaban tidak lebih menarik dibandingkan pengecohnya, dan lazimnya menandakan bahwa peserta didik belum memiliki dasar pengetahuan yang cukup untuk mengeliminasi alternatif. Ketiga butir tersebut termasuk butir dengan indeks kesukaran terendah, sehingga pola ini konsisten dengan temuan bahwa kompetensi yang diukur memang belum dikuasai sebagian besar peserta didik. Sebaliknya, butir nomor 20 memperlihatkan pola yang paling terkonsentrasi dengan satu alternatif dipilih 52,31% peserta didik, yang menunjukkan bahwa sebagian besar peserta didik telah mampu mengarahkan pilihan secara terarah pada butir tersebut. Informasi inilah yang seharusnya dimanfaatkan guru dalam merancang pembelajaran terdiferensiasi sebagaimana diamanatkan Kurikulum Merdeka, sebab analisis pengecoh tidak berhenti pada penilaian mutu butir melainkan menjadi sumber data mengenai cara berpikir peserta didik

Profil penguasaan peserta didik pada keempat aspek konten numerasi menunjukkan capaian yang seluruhnya berada di bawah 50%, dengan rentang yang

sempit yaitu 32% hingga 41%. Aspek Perubahan dan Hubungan serta aspek Ketidakpastian dan Data memperoleh capaian tertinggi dan setara (41%), disusul aspek Ruang dan Bentuk (37%), sedangkan aspek Kuantitas menempati posisi terendah (32%). Sempitnya rentang antar aspek merupakan temuan yang perlu dimaknai secara khusus. Perbedaan capaian yang hanya berkisar sembilan poin mengindikasikan bahwa kelemahan peserta didik tidak bersifat spesifik pada domain konten tertentu, melainkan bersifat umum dan menyeluruh. Dengan kata lain, hambatan yang dihadapi peserta didik lebih berkaitan dengan kemampuan bernalar atas stimulus memahami konteks, mengekstraksi informasi, dan menerjemahkannya ke dalam operasi matematis daripada dengan penguasaan materi pada domain tertentu.

Temuan yang paling menarik justru terletak pada rendahnya capaian aspek Kuantitas (32%), padahal aspek ini umumnya dipandang paling fundamental dan memperoleh porsi waktu terbesar dalam pembelajaran matematika di sekolah dasar. Kondisi ini mengindikasikan adanya kesenjangan antara kemampuan prosedural dan kemampuan kontekstual. Butir numerasi pada aspek Kuantitas tidak berhenti pada operasi hitung, melainkan menuntut penalaran proporsional, estimasi, dan pemilihan operasi yang tepat berdasarkan konteks yang disajikan. Peserta didik yang terlatih menyelesaikan soal hitung dalam bentuk baku dapat mengalami kesulitan ketika operasi yang sama disajikan dalam situasi nyata [sitasi]. Secara keseluruhan, profil ini menegaskan bahwa penguatan numerasi di sekolah dasar perlu diarahkan pada pembelajaran berbasis konteks nyata, bukan pada penambahan latihan berhitung, sekaligus menjadi dasar bagi guru dalam merancang pembelajaran terdiferensiasi sebagaimana diamanatkan Kurikulum Merdeka.

D. Kesimpulan

Analisis terhadap 20 butir soal pilihan ganda berbasis literasi dan numerasi menunjukkan bahwa instrumen ini layak digunakan dengan sejumlah catatan perbaikan. Seluruh butir memenuhi validitas isi pada kategori sedang hingga tinggi (rerata V Aiken 0,76), memiliki daya pembeda sangat baik (0,792–0,944), dan seluruh pengecohnya berfungsi karena dipilih di atas 5% peserta didik. Kelemahan utama terletak pada sebaran tingkat kesukaran yang terkonsentrasi pada kategori

sulit dan sedang dengan rerata 0,318 tanpa satu pun butir mudah, sehingga instrumen belum mampu memetakan peserta didik pada kelompok kemampuan rendah. Profil penguasaan pada keempat aspek konten seluruhnya berada di bawah 50% dengan rentang yang sempit (32%–41%) dan capaian terendah justru pada aspek Kuantitas. Dengan demikian, penguatan asesmen literasi dan numerasi dalam Kurikulum Merdeka menuntut perbaikan pada dua sisi sekaligus: penyempurnaan mutu butir yang disusun guru, khususnya pada aspek perumusan stimulus dan keseimbangan tingkat kesukaran, serta pembelajaran berbasis konteks nyata yang melatih peserta didik menafsirkan informasi sebelum menerapkan operasi matematis. Meski demikian masih terdapat keterbatasan dalam penelitian ini yakni jumlah siswa yang menjadi responden masih relatif sedikit sehingga disarankan penelitian selanjutnya melibatkan jumlah siswa yang relatif lebih besar dan memeriksa validitas konstruk dengan analisis CFA.

Temuan penelitian ini memiliki implikasi yaitu instrumen yang dikembangkan terbukti memenuhi validitas isi pada kategori sedang hingga tinggi, memiliki daya pembeda yang sangat baik pada seluruh butir, serta pengecoh yang berfungsi tanpa kecuali, sehingga layak dimanfaatkan guru sebagai perangkat asesmen literasi dan numerasi. Profil penguasaan yang seluruhnya berada di bawah 50% dengan rentang antar aspek yang sempit mengindikasikan bahwa hambatan peserta didik bersifat menyeluruh dan berkaitan dengan kemampuan bernalar atas stimulus, bukan dengan penguasaan domain konten tertentu; karena itu intervensi yang disarankan bukan penambahan latihan berhitung, melainkan pembelajaran berbasis konteks nyata yang melatih peserta didik menafsirkan informasi sebelum menerapkan operasi matematis.

Daftar Pustaka

- Abidin, Y. 2018. Pembelajaran Literasi. Jakarta: Bumi Aksara.
- Allen, M.J., & Yen, W.M. 2002. Introduction to Measurement Theory. Long Grove: Waveland Press.
- Anastasi, A., & Urbina, S. 1997. Psychological Testing. Upper Saddle River: Prentice Hall.
- Arikunto, S. 2018. Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara.

- Azwar, S. 2015. Reliabilitas dan Validitas. Yogyakarta: Pustaka Pelajar
- Baharuddin, H., Riskawati., Fitrawardana., 2025. Identifikasi Kemampuan Prosedural dan Konseptual Siswa dalam Memecahkan Masalah Matematika dengan Model DINA. Jurnal Studi Guru dan Pembelajaran (JSGP), Vol. 8, No. 3.
- Brookhart, S.M. 2010. How to Assess Higher-Order Thinking Skills in Your Classroom. Alexandria: ASCD.
- Brookhart, S.M. 2013. How to Create and Use Rubrics for Formative Assessment and Grading. Alexandria: ASCD.
- Creswell, J.W. 2014. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks: Sage Publications.
- Ebel, R.L., & Frisbie, D.A. 1991. Essentials of Educational Measurement. Englewood Cliffs: Prentice Hall.
- Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. 2012. How to Design and Evaluate Research in Education. New York: McGraw-Hill.
- Gronlund, N.E., & Linn, R.L. 2000. Measurement and Evaluation in Teaching. New York: Macmillan.
- Ishak, S. (2018). Penerapan Metode Eksplorasi Dalam Meningkatkan Hasil Belajar Matematika Pada Peserta Didik Kelas IX SMP Negeri 5 Campalagian. *Pepatudzu: Media Pendidikan dan Sosial Kemasyarakatan*, 13(1), 75-87.
- Ishak, S., Dermawan, D., & Karmila, K. (2025, May). Efektivitas Penerapan Pembelajaran Berdiferensiasi Pada Kurikulum Merdeka Terhadap Hasil Belajar Matematika Siswa. In *Journal Peqguruang: Conference Series* (Vol. 6, No. 2, pp. 566-570).
- Kemendikbudristek. 2022. Panduan Implementasi Kurikulum Merdeka. Jakarta: Kemendikbudristek.
- Kemendikbudristek. 2021. Asesmen Nasional: AKM dan Implikasinya terhadap Pembelajaran. Jakarta: Kemendikbudristek.
- Nitko, A.J., & Brookhart, S.M. 2014. Educational Assessment of Students. Boston: Pearson.
- OECD. 2019. PISA 2018 Assessment and Analytical Framework. Paris: OECD Publishing.
- Retnawati, H. 2016. Analisis Kuantitatif Instrumen Penelitian. Yogyakarta: Parama Publishing.

- Sudijono, A. 2015. Pengantar Evaluasi Pendidikan. Jakarta: Raja Grafindo Persada.
- Tasni, N., Ishak, S., Azis, S., Syukriani, A., & Wajdi, F. (2024). Literasi dan Numerasi Pada Pendidikan Anak Usia Dini. *Journal Pengabdian Kepada Masyarakat Sawyerigading*, 3(2), 12-22.
- Widana, I.W. 2017. Higher Order Thinking Skills Assessment. Jakarta: Gramedia.