

Pengembangan Asesmen Kompetensi Minimum Literasi Numerik Bentuk Uraian Dengan Menerapkan Generalized Partial Credit Model (GPCM) Untuk Siswa SD

Icha Maytasya Br Ginting¹, I Made Candiasa², Gusti Ayu Mahayukti³

¹Universitas Pendidikan Ganesha, ichamay23@gmail.com

²Universitas Pendidikan Ganesha, candiasaimade@undiksha.ac.id

³Universitas Pendidikan Ganesha, gustiayumahayukti@undiksha.ac.id

Abstrak

Penelitian ini bertujuan mengembangkan dan menguji kualitas instrumen Asesmen Kompetensi Minimum (AKM) literasi numerik bentuk uraian untuk siswa kelas V sekolah dasar dengan menerapkan Generalized Partial Credit Model (GPCM). Instrumen dikembangkan melalui model formative research dari Tessmer (1993) yang mencakup tahap preliminary, self-evaluation, dan formative evaluation. Tiga paket instrumen dikembangkan, masing-masing terdiri dari 10 butir uraian dengan penskoran politomus (skor 0–3) berdasarkan tiga level kognitif AKM: knowing, applying, dan reasoning. Uji coba dilakukan pada 100 siswa kelas V dari tiga sekolah dasar di Kabupaten Buleleng. Hasil penelitian menunjukkan: (1) seluruh butir soal valid secara isi dengan nilai CVR dan CVI sebesar 1,00; (2) konsistensi internal butir baik dengan korelasi product moment $> 0,1966$; (3) reliabilitas sangat tinggi dengan koefisien Alpha Cronbach sebesar 0,871, 0,888, dan 0,875; (4) daya beda butir berkisar antara 0,247–0,765; (5) taraf kesukaran bervariasi pada kategori sedang hingga sukar; (6) analisis ICC-GPCM menunjukkan peningkatan peluang skor tinggi seiring meningkatnya kemampuan siswa; dan (7) validitas konkuren menggunakan analisis jalur berbantuan SmartPLS menunjukkan hubungan positif dan signifikan ketiga paket terhadap hasil belajar matematika ($p < 0,05$). Instrumen yang dikembangkan sudah memenuhi persyaratan untuk mengukur kemampuan literasi numerik siswa SD secara akurat.

Kata Kunci: *Generalized Partial Credit Model, Asesmen Kompetensi Minimum, literasi numerik, instrumen uraian, Item Response Theory, Sekolah Dasar.*

Abstract

This study aims to develop and assess the quality of Minimum Competency Assessment (AKM) instruments for numeric literacy in essay form for fifth-grade elementary school students using the Generalized Partial Credit Model (GPCM). The instruments were developed through Tessmer's (1993) formative research model, covering the preliminary, self-evaluation, and formative evaluation stages. Three instrument packages were developed, each consisting of 10 essay items with polytomous scoring (scores 0–3) based on three AKM cognitive levels: knowing, applying, and reasoning. The field trial was conducted with 100 fifth-grade students from three elementary schools in Buleleng Regency. Results showed: (1) all items were content-valid with CVR and CVI values of 1.00; (2) adequate internal consistency, with item-total product moment correlations exceeding the critical value of 0.1966; (3) very high reliability with Cronbach's Alpha coefficients of 0.871, 0.888, and 0.875 for each package; (4) item discrimination ranging from 0.247 to 0.765; (5) item difficulty varied across moderate to difficult categories; (6) ICC-GPCM analysis showed that the probability of obtaining higher scores increased as student ability increased; and (7) concurrent validity via SmartPLS path analysis showed significant positive relationships between all three packages and mathematics learning outcomes ($p < 0.05$). The developed instruments meet the required psychometric criteria and are capable of measuring students' numeric literacy accurately.

Keywords: *Generalized Partial Credit Model, Minimum Competency Assessment, numeric literacy, essay instrument, Item Response Theory, Elementary School.*

Corresponding Author:

Ichha Maytasya

Universitas Pendidikan Ganesha, ichamay23@gmail.com

PENDAHULUAN

Perubahan peradaban yang ditandai akselerasi teknologi dan revolusi industri 4.0 menuntut reformulasi sistem pendidikan nasional. Pendidikan abad ke-21 mensyaratkan keterampilan berpikir kritis, komunikasi, kolaborasi, dan kreativitas (4C) sebagai kompetensi utama yang wajib dikuasai siswa. Dalam konteks ini, Indonesia melalui Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi menginisiasi Asesmen Nasional (AN) sebagai transformasi besar dalam mengevaluasi mutu pendidikan. Asesmen Kompetensi Minimum (AKM) menjadi salah satu pilar utama AN yang dirancang untuk mengukur literasi membaca dan literasi numerik secara kontekstual dan lintas disiplin (Kemendikbud, 2021).

Literasi numerik mencakup kemampuan siswa dalam memahami, menafsirkan, dan menggunakan informasi matematika secara logis untuk menyelesaikan masalah kehidupan nyata. OECD (2019) mendefinisikannya sebagai kemampuan mengakses, menggunakan, menafsirkan, dan mengkomunikasikan informasi matematika dalam berbagai bentuk. Namun, berdasarkan laporan PISA 2022, capaian literasi numerik siswa Indonesia masih memprihatinkan dengan skor rata-rata 366 poin tertinggal 106 poin dari rata-rata global, dan sekitar 82% siswa berada di bawah level dua (OECD, 2023). Di Kabupaten Buleleng, kemampuan numerik siswa SD hanya mencapai 39,5% pada pelaksanaan Asesmen Nasional 2022, dengan skor literasi numerik hanya 1,55 dari rentang 1–3 (Suardipa, 2025).

Salah satu faktor yang berkontribusi terhadap rendahnya capaian literasi numerik adalah keterbatasan instrumen asesmen yang digunakan, yang umumnya masih bersifat prosedural, berbentuk pilihan ganda konvensional, dan tidak mampu mengungkap proses berpikir tingkat tinggi siswa (Wilson, 2016). Untuk mengatasi hal ini, diperlukan instrumen asesmen berbentuk uraian yang mampu merekam kualitas jawaban siswa secara bertahap. Soal uraian memungkinkan pengujian kemampuan menganalisis, mengevaluasi, dan menyintesis informasi serta menyusun argumen secara logis, sejalan dengan tujuan AKM yang menekankan kemampuan bernalar (Kemendikbud, 2021).

Salah satu pendekatan yang sangat sesuai untuk menganalisis soal uraian dengan penskoran bertingkat adalah Generalized Partial Credit Model (GPCM). GPCM merupakan model dalam kerangka Item Response Theory (IRT) yang dikembangkan Muraki sebagai penyempurnaan Partial Credit Model (PCM). GPCM menggunakan dua parameter—tingkat kesukaran dan daya beda—sehingga lebih fleksibel dan mampu membedakan kemampuan siswa secara lebih rinci berdasarkan kualitas jawaban bertingkat (Kurnia, 2019). Dibandingkan model politomus lainnya seperti Graded Response Model (GRM), GPCM lebih adaptif terhadap variasi skor dalam soal-soal uraian (Dai dkk., 2021).

Beberapa penelitian sebelumnya telah membahas pengembangan instrumen AKM maupun penerapan IRT dalam asesmen pendidikan. Noviantini dkk. (2023) mengembangkan instrumen AKM numerasi untuk siswa SMP kelas VII dan membuktikan instrumen yang disusun sistematis dapat memenuhi kriteria validitas yang baik, namun belum menerapkan analisis IRT pada butir soalnya. Mertasari & Candiasa

(2023) menunjukkan bahwa model politomus IRT efektif digunakan dalam pengembangan dan validasi instrumen tes, namun penelitian tersebut menggunakan Graded Response Model (GRM) pada konteks tes metakognitif, bukan instrumen AKM literasi numerik untuk siswa SD. Damayanti dkk. (2025) mengidentifikasi bahwa kemampuan numerasi siswa SD pada konten bilangan masih didominasi kategori sedang, mengindikasikan perlunya instrumen diagnostik yang lebih akurat. Namun demikian, belum ditemukan penelitian yang secara khusus mengembangkan instrumen AKM literasi numerik bentuk uraian untuk siswa SD kelas V dengan mengintegrasikan penskoran politomus berbasis GPCM sekaligus menguji validitas konkurensinya terhadap hasil belajar matematika. Kesenjangan inilah yang menjadi dasar dan kebaruan penelitian ini.

Berdasarkan latar belakang dan kesenjangan penelitian yang telah diuraikan, penelitian ini bertujuan untuk: (1) mengembangkan instrumen AKM literasi numerik bentuk uraian berbasis GPCM untuk siswa kelas V SD di Kabupaten Buleleng; (2) mendeskripsikan validitas isi, konsistensi internal, reliabilitas, daya beda, dan tingkat kesukaran instrumen yang dikembangkan; (3) mendeskripsikan estimasi parameter butir menggunakan analisis GPCM; dan (4) mengetahui validitas konkuren instrumen terhadap hasil belajar matematika siswa.

METODE

Penelitian ini merupakan penelitian pengembangan (research and development) yang mengacu pada model formative research Tessmer (1993), mencakup tiga tahap utama: (1) tahap pendahuluan (preliminary), (2) evaluasi diri (self-evaluation), dan (3) evaluasi formatif (formative evaluation).

Pada tahap pendahuluan, dilakukan kajian pustaka mengenai AKM, literasi numerik, dan karakteristik soal uraian, serta observasi dan wawancara dengan guru untuk memperoleh informasi kontekstual mengenai kondisi belajar siswa, tingkat kemampuan literasi numerik, serta tantangan guru dalam merancang asesmen numerik sesuai prinsip AKM.

Tahap evaluasi diri mencakup analisis kurikulum, analisis siswa, analisis materi, dan perancangan kisi-kisi instrumen. Instrumen disusun dalam tiga paket soal, masing-masing 10 butir uraian dengan penskoran politomus (skor 0–3) berdasarkan level kognitif AKM (knowing, applying, reasoning) dan mencakup empat domain konten: bilangan, pengukuran dan geometri, aljabar, serta data dan ketidakpastian.

Tahap evaluasi formatif meliputi uji ahli (expert review) oleh 7 validator (3 dosen dan 4 guru), uji one-to-one dengan 9 siswa, uji kelompok kecil (small group) dengan 30 siswa, dan uji lapangan (field test) dengan 100 siswa kelas V dari tiga sekolah dasar di Kabupaten Buleleng, yaitu SD Negeri 1 Astina, SD Negeri 1 Paket Agung, dan SD Negeri 1 Banjar Bali.

Validitas isi dianalisis menggunakan Content Validity Ratio (CVR) dan Content Validity Index (CVI) metode Lawshe (1975). Konsistensi internal butir diuji dengan korelasi product moment. Reliabilitas dihitung menggunakan rumus Alpha Cronbach. Daya beda dan tingkat kesukaran dihitung menggunakan formula butir politomus. Analisis parameter butir dilakukan menggunakan pendekatan Generalized Partial Credit Model (GPCM) yang dikembangkan oleh Muraki (1992) dengan bantuan bahasa pemrograman Python. GPCM dipilih karena mampu mengestimasi dua parameter sekaligus pada setiap butir, yaitu parameter daya beda (discrimination, a) dan parameter kesukaran kategori (threshold, b), sehingga lebih fleksibel dalam menggambarkan karakteristik butir soal uraian dengan penskoran politomus.

Hasil analisis GPCM divisualisasikan melalui kurva karakteristik butir (Item Characteristic Curve/ICC) yang menggambarkan hubungan antara kemampuan peserta (θ) dengan probabilitas memperoleh skor tertentu, serta kurva fungsi informasi butir (Item Information Function/IIF) yang menggambarkan seberapa besar kontribusi setiap butir dalam mengukur kemampuan literasi numerik pada rentang θ tertentu. Estimasi kemampuan peserta tes dilakukan menggunakan metode Maximum Likelihood Estimation (MLE) yang menghasilkan nilai θ secara langsung berdasarkan pola respons siswa tanpa memerlukan asumsi distribusi prior, sehingga menghasilkan estimasi yang konsisten dan stabil pada model politomus (Luo, 2018). Nilai θ yang diperoleh selanjutnya dikategorikan menjadi tiga kelompok, yaitu rendah ($\theta \leq -1$), sedang ($-1 < \theta \leq 1$), dan tinggi ($\theta > 1$).

Validitas konkuren diuji menggunakan analisis jalur berbantuan SmartPLS dengan teknik bootstrapping, karena tidak mensyaratkan distribusi normal dan sesuai untuk ukuran sampel sedang (Hair dkk., 2019; Subhaktiyasa, 2024).

HASIL DAN PEMBAHASAN

Instrumen AKM literasi numerik yang dikembangkan berupa soal uraian berbasis konteks kehidupan sehari-hari dalam tiga paket, masing-masing terdiri dari 10 butir soal. Setiap paket disusun berdasarkan indikator literasi numerik yang mencakup domain bilangan, pengukuran dan geometri, aljabar, serta data dan ketidakpastian, dengan sebaran level kognitif knowing, applying, dan reasoning.

Validitas Isi

Uji validitas isi dilakukan menggunakan metode Lawshe (1975) dengan melibatkan tujuh validator. Penilaian awal menunjukkan beberapa butir perlu direvisi. Setelah dilakukan revisi, hasil perhitungan CVR pada seluruh butir di ketiga paket memperoleh nilai 1,00 dan CVI keseluruhan juga sebesar 1,00, melampaui nilai minimum $CVR \geq 0,99$ untuk tujuh validator. Hal ini menunjukkan seluruh instrumen berada pada kategori sangat valid dari segi isi. Hasil rekapitulasi validitas isi disajikan pada Tabel 1.

Tabel 1. Rekapitulasi Validitas Isi

Paket	Jumlah Butir	CVR	CVI	Keterangan
Paket 1	10	1,00	1,00	Valid
Paket 2	10	1,00	1,00	Valid
Paket 3	10	1,00	1,00	Valid

Konsistensi Internal Butir

Hasil uji korelasi product moment menunjukkan seluruh butir soal memiliki nilai r_{hitung} yang lebih besar dari r_{tabel} (0,1966) pada taraf signifikansi 5%. Nilai korelasi pada Paket 1 berkisar antara 0,384–0,859; Paket 2 antara 0,469–0,877; dan Paket 3 antara 0,454–0,850. Seluruh butir dinyatakan konsisten dan mampu mengukur konstruk literasi numerik yang sama. Hasil uji konsistensi internal butir disajikan pada Tabel 2.

Tabel 2. Hasil Uji Konsistensi Internal Butir

No Butir	Paket 1 (r_{xy})	Paket 2 (r_{xy})	Paket 3 (r_{xy})	r_{tabel}	Keterangan
1	0,644	0,649	0,634	0,1966	Konsisten
2	0,812	0,829	0,790	0,1966	Konsisten
3	0,859	0,877	0,850	0,1966	Konsisten
4	0,734	0,753	0,738	0,1966	Konsisten
5	0,811	0,834	0,827	0,1966	Konsisten
6	0,636	0,642	0,639	0,1966	Konsisten
7	0,740	0,763	0,753	0,1966	Konsisten
8	0,618	0,649	0,609	0,1966	Konsisten

9	0,384	0,469	0,454	0,1966	Konsisten
10	0,566	0,594	0,556	0,1966	Konsisten

Reliabilitas

Analisis reliabilitas menggunakan Alpha Cronbach berbantuan Python menghasilkan koefisien reliabilitas Paket 1 sebesar 0,871, Paket 2 sebesar 0,888, dan Paket 3 sebesar 0,875. Ketiga paket tergolong dalam kategori sangat tinggi, menandakan instrumen memiliki konsistensi pengukuran yang sangat baik. Hasil uji reliabilitas disajikan pada Tabel 3.

Tabel 3. Hasil Uji Reliabilitas

Paket	Koefisien Alpha Cronbach	Kategori
Paket 1	0,871	Sangat Tinggi
Paket 2	0,888	Sangat Tinggi
Paket 3	0,875	Sangat Tinggi

Tingkat Kesukaran

Sebagian besar butir soal berada pada kategori sedang (IK 0,31–0,70). Pada ketiga paket, butir soal nomor 1 hingga 8 dan nomor 10 termasuk kategori sedang, sedangkan butir nomor 9 termasuk kategori sukar (IK < 0,30). Keberadaan butir sukar penting untuk melatih siswa menghadapi soal-soal bernalar tingkat tinggi (Marhami dkk., 2024). Hasil rekapitulasi tingkat kesukaran disajikan pada Tabel 4.

Tabel 4. Rekapitulasi Tingkat Kesukaran Butir Soal

No Butir	Paket 1	Kategori	Paket 2	Kategori	Paket 3	Kategori
1	0,481	Sedang	0,543	Sedang	0,549	Sedang
2	0,531	Sedang	0,525	Sedang	0,543	Sedang
3	0,481	Sedang	0,549	Sedang	0,543	Sedang
4	0,469	Sedang	0,475	Sedang	0,488	Sedang
5	0,321	Sedang	0,377	Sedang	0,370	Sedang
6	0,340	Sedang	0,315	Sedang	0,309	Sedang
7	0,432	Sedang	0,383	Sedang	0,389	Sedang
8	0,636	Sedang	0,506	Sedang	0,537	Sedang
9	0,185	Sukar	0,241	Sukar	0,235	Sukar
10	0,364	Sedang	0,383	Sedang	0,389	Sedang

Daya Beda

Nilai daya beda butir pada Paket 1 berkisar antara 0,247–0,716; Paket 2 antara 0,333–0,765; dan Paket 3 antara 0,272–0,728. Sebagian besar butir berada pada kategori baik hingga sangat baik, kecuali butir nomor 9 yang berada pada kategori cukup baik di ketiga paket. Hasil rekapitulasi daya beda disajikan pada Tabel 5.

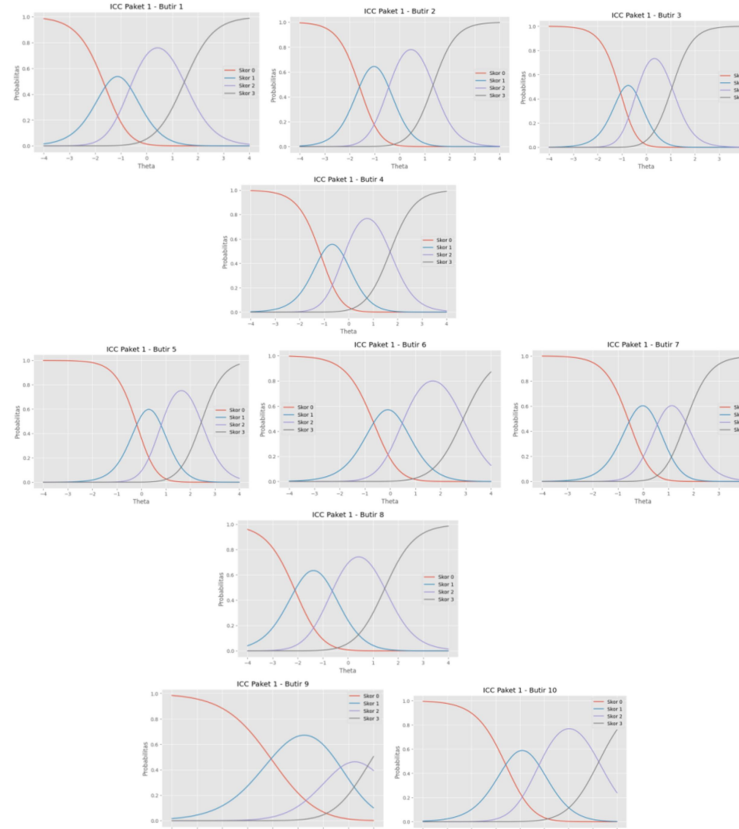
Tabel 5. Rekapitulasi Daya Beda

No Butir	Paket 1	Kategori	Paket 2	Kategori	Paket 3	Kategori
1	0,494	Baik	0,469	Baik	0,506	Baik
2	0,568	Baik	0,679	Baik	0,667	Baik
3	0,716	Sangat Baik	0,704	Sangat Baik	0,716	Sangat Baik
4	0,543	Baik	0,704	Sangat Baik	0,679	Baik
5	0,617	Baik	0,728	Sangat Baik	0,691	Baik
6	0,506	Baik	0,556	Baik	0,568	Baik
7	0,667	Baik	0,765	Sangat Baik	0,728	Sangat Baik
8	0,432	Baik	0,617	Baik	0,556	Baik
9	0,247	Cukup Baik	0,333	Cukup Baik	0,272	Cukup Baik

10	0,481	Baik	0,494	Baik	0,454	Baik
----	-------	------	-------	------	-------	------

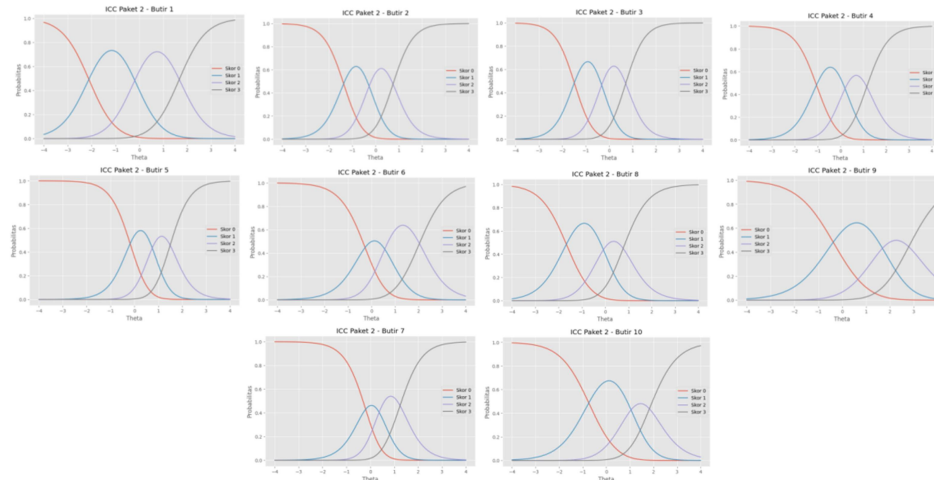
Analisis GPCM

Analisis Generalized Partial Credit Model (GPCM) dilakukan dengan bantuan Python untuk menghasilkan kurva karakteristik butir (Item Characteristic Curve/ICC) yang menggambarkan hubungan antara kemampuan peserta (θ) dengan probabilitas memperoleh kategori skor tertentu pada setiap butir soal. Setiap butir memiliki empat kurva yang merepresentasikan probabilitas memperoleh skor 0 (merah), skor 1 (biru), skor 2 (ungu), dan skor 3 (abu-abu) pada rentang kemampuan $\theta = -4$ hingga $\theta = +4$.



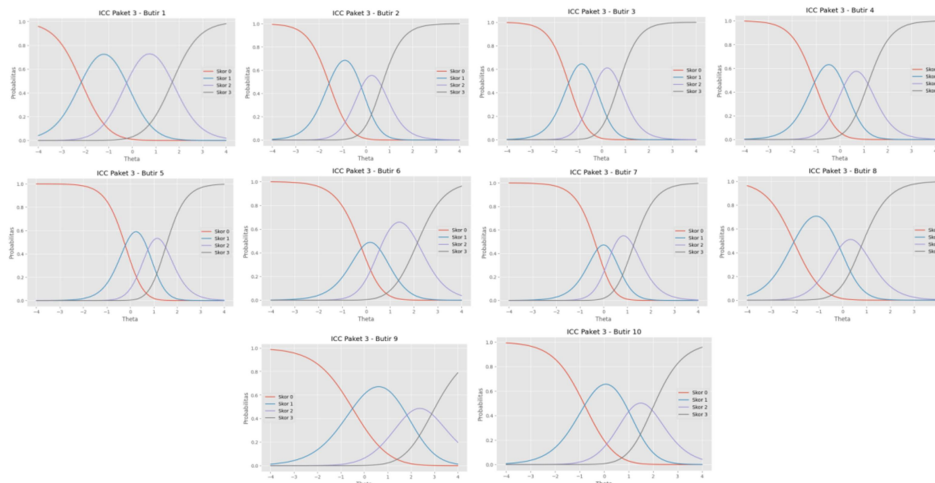
Gambar 1. Kurva ICC Paket 1

Hasil analisis ICC Paket 1 menunjukkan bahwa seluruh butir memiliki pola perpindahan probabilitas yang konsisten sesuai peningkatan kemampuan peserta. Siswa berkemampuan rendah ($\theta \leq -2$) didominasi probabilitas skor 0, kemampuan sedang ($\theta = 0$ hingga 1) menunjukkan peningkatan probabilitas skor 1 dan skor 2, sedangkan kemampuan tinggi ($\theta \geq 2$) didominasi probabilitas skor 3. Butir nomor 1 pada domain bilangan menunjukkan kemampuan diskriminasi yang baik dengan nilai daya beda $a = 1,031$. Butir nomor 3 yang merupakan soal domain pengukuran dan geometri pada level reasoning menunjukkan transisi antarkategori paling tajam dengan nilai daya beda tertinggi $a = 1,374$, mencerminkan kemampuan butir tersebut dalam membedakan siswa secara presisi. Butir nomor 5 pada domain pengukuran dan geometri juga memberikan diskriminasi sangat baik dengan nilai $a = 1,298$. Sementara itu, butir nomor 9 yang merupakan soal domain bilangan pada level reasoning memerlukan kemampuan lebih tinggi untuk mencapai skor maksimum, sebagaimana tercermin dari nilai daya beda terendah pada Paket 1 yaitu $a = 0,615$ dengan kategori cukup baik. Pola yang sistematis pada seluruh butir membuktikan instrumen Paket 1 memenuhi karakteristik model GPCM.



Gambar 2. Kurva ICC Paket 2

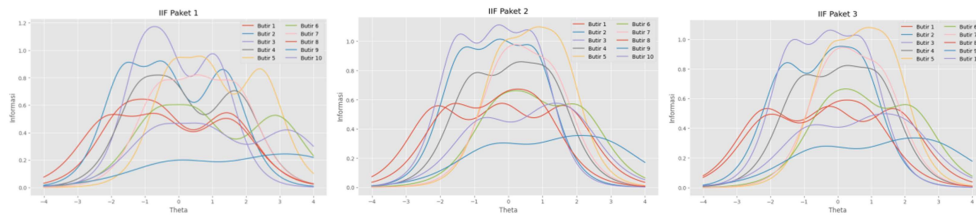
Hasil analisis ICC Paket 2 menunjukkan pola yang serupa dengan Paket 1. Seluruh butir memperlihatkan perpindahan probabilitas yang bertahap dari skor rendah menuju skor tinggi seiring meningkatnya nilai θ . Butir nomor 1 pada domain bilangan menunjukkan kemampuan diskriminasi yang sangat baik dengan nilai $a = 1,039$. Butir nomor 3 yang merupakan soal domain pengukuran dan geometri kembali menjadi butir dengan kemampuan diskriminasi tertinggi ($a = 1,404$), diikuti butir nomor 5 ($a = 1,334$) pada domain pengukuran dan geometri, serta butir nomor 7 ($a = 1,253$) pada domain data dan ketidakpastian, yang ketiganya menunjukkan transisi antarkategori yang tegas. Butir nomor 9 pada Paket 2 menunjukkan nilai daya beda yang lebih tinggi dibanding Paket 1 ($a = 0,750$) sehingga mampu membedakan kemampuan peserta lebih baik, meskipun tingkat kesukarannya tetap relatif lebih tinggi dibanding butir lain. Keseluruhan pola ICC Paket 2 membuktikan instrumen memenuhi karakteristik model GPCM.



Gambar 3. Kurva ICC Paket 3

Hasil analisis ICC Paket 3 menunjukkan pola perpindahan probabilitas yang paling konsisten dan merata di antara ketiga paket. Butir nomor 3 ($a = 1,360$) pada domain aljabar dan butir nomor 5 ($a = 1,323$) pada domain pengukuran dan geometri menunjukkan transisi antarkategori yang sangat tegas, sedangkan butir nomor 7 ($a = 1,204$) pada domain data dan ketidakpastian juga memberikan kemampuan diskriminasi yang sangat baik. Butir nomor 9 ($a = 0,726$) tetap menunjukkan pola positif meskipun skor 3 baru dominan pada kemampuan sangat tinggi, selaras dengan karakteristiknya

sebagai soal reasoning domain bilangan dengan tingkat kesukaran tertinggi. Pola yang konsisten pada seluruh butir Paket 3 membuktikan instrumen memenuhi karakteristik model GPCM dan layak digunakan untuk mengukur kemampuan literasi numerik siswa secara bertahap dan diagnostik.



Gambar 4. Kurva IIF Paket 1,2 dan 3

Analisis fungsi informasi butir (Item Information Function/IIF) dilakukan untuk mengetahui besarnya kontribusi masing-masing butir dalam mengukur kemampuan laten literasi numerik pada rentang θ tertentu. Hasil analisis menunjukkan bahwa butir nomor 3 dan butir nomor 5 secara konsisten menjadi butir paling informatif pada ketiga paket dengan puncak nilai informasi di atas 1,00. Pada Paket 1, butir nomor 3 mencapai puncak informasi mendekati 1,20 pada θ sekitar -1 ($a = 1,374$), pada Paket 2 mencapai nilai di atas 1,00 pada θ -1 hingga 0 ($a = 1,404$), dan pada Paket 3 tetap menjadi yang tertinggi bersama butir nomor 5 ($a = 1,360$ dan $a = 1,323$) pada rentang θ sedang hingga tinggi. Sebaliknya, butir nomor 9 secara konsisten memiliki nilai informasi paling rendah pada ketiga paket, yaitu sekitar $0,20$ – $0,25$ pada Paket 1, $0,30$ – $0,35$ pada Paket 2, dan $0,25$ – $0,30$ pada Paket 3, selaras dengan nilai daya bedanya yang lebih kecil dibanding butir lain. Secara keseluruhan, ketiga paket instrumen memberikan kontribusi informasi pengukuran yang optimal pada rentang kemampuan sedang hingga tinggi, yang merupakan rentang kemampuan dominan siswa berdasarkan hasil estimasi θ . Kurva fungsi informasi butir ketiga paket disajikan pada Gambar 4.

Estimasi Kemampuan Peserta Tes

Estimasi kemampuan menggunakan Maximum Likelihood Estimation (MLE) dalam kerangka GPCM berhasil dilakukan pada seluruh 100 siswa. Distribusi kemampuan pada ketiga paket disajikan pada Tabel 6.

Tabel 6. Distribusi Frekuensi Estimasi Kemampuan (θ) Peserta Tes

Kategori	Keterangan	Paket 1	Paket 2	Paket 3
$\theta \leq -1$	Rendah	21%	22%	25%
$-1 < \theta \leq 1$	Sedang	63%	59%	57%
$\theta > 1$	Tinggi	16%	19%	18%

Sebagian besar siswa berada pada kategori kemampuan sedang. Korelasi antara nilai θ dan skor total mencapai $0,99$, menunjukkan estimasi MLE sangat akurat. Temuan ini konsisten dengan Supriyanto dkk. (2024) yang melaporkan mayoritas siswa SD berada pada kategori kemampuan numerik sedang.

Validitas Konkuren

Analisis jalur menggunakan SmartPLS menunjukkan bahwa ketiga paket memiliki hubungan positif dan signifikan terhadap hasil belajar matematika. Hasil analisis disajikan pada Tabel 7.

Tabel 7. Koefisien Jalur dan Signifikansi (SmartPLS)

Kategori	Koefisien Jalur	R ²	T-Statistik	P-Value
Paket 1	0,356	0,127	3,754	0,000
Paket 2	0,252	0,063	2,707	0,007
Paket 3	0,252	0,064	2,669	0,008

Ketiga paket memenuhi kriteria signifikansi ($p < 0,05$ dan $t > 1,96$), menunjukkan validitas konkuren yang baik. Paket 1 memiliki kemampuan prediktif tertinggi ($R^2 = 12,7\%$), sedangkan Paket 3 dinilai paling representatif secara psikometrik karena distribusi kemampuan siswa paling merata .

Pembahasan

Instrumen AKM literasi numerik bentuk uraian yang dikembangkan telah memenuhi seluruh kriteria kualitas instrumen yang baik. Validitas isi sempurna ($CVR = CVI = 1,00$) menunjukkan kesesuaian penuh antara butir soal dengan indikator AKM, sejalan dengan temuan Noviantini dkk. (2023) yang menyatakan instrumen AKM yang disusun sistematis dapat memenuhi kriteria validitas yang baik. Konsistensi internal seluruh butir yang terpenuhi ($r_{xy} > 0,1966$) mengindikasikan bahwa setiap butir soal mengukur dimensi yang sama, yaitu kemampuan literasi numerik. Tingkat kesukaran yang didominasi kategori sedang menunjukkan instrumen tidak terlalu mudah maupun terlalu sukar bagi siswa kelas V SD. Satu-satunya butir berkategori sukar pada ketiga paket, yaitu butir nomor 9, dirancang khusus pada level reasoning yang memang menuntut kemampuan penalaran tingkat tinggi. Hal ini disengaja karena soal-soal non-rutin dengan tingkat kesulitan tinggi diperlukan agar siswa terlatih menghadapi tantangan kognitif, sebagaimana ditegaskan oleh Marhami dkk. (2024). Reliabilitas sangat tinggi ($\alpha \geq 0,871$) menandakan bahwa penskoran bertingkat berbasis GPCM menghasilkan data yang konsisten. Hal ini mendukung keunggulan GPCM dalam menangani data politomus yang lebih adaptif dibanding model klasik (Harsana dkk., 2023). Temuan daya beda yang baik hingga sangat baik pada sebagian besar butir membuktikan instrumen mampu membedakan siswa berkemampuan tinggi dan rendah secara efektif.

Analisis ICC-GPCM mengungkap bahwa setiap butir bekerja sesuai prinsip model: probabilitas respons meningkat secara bertahap dari skor rendah ke skor tinggi mengikuti peningkatan kemampuan (θ). Hal ini selaras dengan pernyataan Kurnia (2019) bahwa GPCM mampu menangkap perbedaan kualitas jawaban siswa secara bertingkat dalam asesmen berbasis uraian, karena setiap butir memiliki parameter diskriminasi tersendiri yang menggambarkan ketajaman pembeda kemampuan. Temuan ini juga sejalan dengan Harsana dkk. (2023) yang menyatakan bahwa model IRT politomus seperti GPCM lebih unggul dibanding model klasik dalam mengidentifikasi karakteristik butir soal literasi secara akurat. Butir nomor 3 dan 5 yang secara konsisten paling informatif pada ketiga paket dapat dijadikan acuan prioritas dalam pengembangan instrumen serupa ke depannya.

Distribusi kemampuan yang mayoritas berada pada kategori sedang (57%–63%) mengindikasikan bahwa sebagian besar siswa kelas V SD di Kabupaten Buleleng belum sepenuhnya menguasai kompetensi literasi numerik, khususnya pada level reasoning yang menuntut penalaran tingkat tinggi. Temuan ini konsisten dengan Supriyanto dkk. (2024) yang melaporkan mayoritas siswa SD berada pada kategori kemampuan numerik sedang, dan menunjukkan perlunya intervensi pembelajaran yang lebih intensif berbasis pemecahan masalah kontekstual. Validitas konkuren yang signifikan pada ketiga paket ($p < 0,05$) mengkonfirmasi bahwa kemampuan literasi numerik dan hasil belajar matematika memiliki keterkaitan konseptual yang kuat, karena keduanya sama-sama menuntut penalaran logis dan pemrosesan informasi kuantitatif (Mutik dkk., 2025). Sejalan dengan temuan Nursyifa & Masyithoh (2023) bahwa literasi numerasi memiliki hubungan positif terhadap hasil belajar siswa. Perbedaan nilai R^2 antara Paket 1 (12,7%) dan Paket 2–3 (6,3%–6,4%) menunjukkan bahwa Paket 1 memiliki kemampuan prediksi terhadap hasil belajar yang lebih kuat, sehingga lebih sesuai digunakan dalam konteks evaluasi formatif yang menghendaki prediksi capaian belajar. Sebaliknya, Paket 3 dengan distribusi

kemampuan siswa yang paling merata dan konsistensi pola ICC yang paling stabil menjadikannya lebih unggul untuk keperluan diagnostik kemampuan literasi numerik secara komprehensif.

SIMPULAN

Kesimpulan penelitian ini adalah sebagai berikut. Pertama, instrumen AKM literasi numerik bentuk uraian berhasil dikembangkan dalam tiga paket soal yang mencakup materi bilangan, pengukuran dan geometri, aljabar, serta data dan ketidakpastian dengan level kognitif knowing, applying, dan reasoning. Kedua, kualitas instrumen sangat baik ditinjau dari validitas isi ($CVR = CVI = 1,00$), konsistensi internal, reliabilitas sangat tinggi ($\alpha \geq 0,871$), tingkat kesukaran yang proporsional, serta daya beda yang baik hingga sangat baik. Ketiga, analisis GPCM menunjukkan seluruh butir mampu mengukur kemampuan literasi numerik secara bertahap dan konsisten. Keempat, instrumen memiliki validitas konkuren yang baik terhadap hasil belajar matematika pada ketiga paket ($p < 0,05$), dengan Paket 1 menunjukkan koefisien jalur tertinggi sebesar 0,356 ($R^2 = 12,7\%$), menjadikannya paling tepat untuk konteks prediksi capaian belajar. Paket 3 direkomendasikan untuk keperluan diagnostik karena memiliki distribusi kemampuan siswa yang paling merata. Konsistensi pola ICC Paket 3 yang paling stabil di antara ketiga paket juga menjadikannya pilihan utama untuk pengukuran kemampuan literasi numerik secara komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Prof. Dr. I Made Candiasa, M.I.Kom. dan Prof. Dr. Gst. Ayu Mahayukti, M.Si. selaku dosen pembimbing, serta kepada seluruh guru dan siswa di SD Negeri 1 Astina, SD Negeri 1 Paket Agung, dan SD Negeri 1 Banjar Bali atas partisipasi dan kerja samanya dalam penelitian ini.

DAFTAR RUJUKAN

- Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of polytomous IRT models with rating scale data: An investigation over sample size, instrument length, and missing data. *Frontiers in Education*, 6, 721963.
- Damayanti, N. L. P. R., Mahayukti, G. A., & Sudiarta, I. G. P. (2025). Profil kemampuan numerasi siswa dalam menyelesaikan soal Asesmen Kompetensi Minimum (AKM) pada konten bilangan. *Jurnal IKA*, 23(1), 23–30.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24.
- Harsana, F. N., Retnawati, H., Dewanti, S. R., Lumenyela, R. A., Adzima, M. F., & Septiana, A. R. (2023). Comparison of item characteristic analysis models of reading literacy test with polytomous Item Response Theory. *REID (Research and Evaluation in Education)*, 10(2), 7.
- Kemendikbud. (2021). Asesmen Nasional, lembar tanya jawab. Kementerian Pendidikan dan Kebudayaan.
- Kılıç, A. F., & Doğan, N. (2021). Comparison of confirmatory factor analysis estimation methods on mixed-format data. *International Journal of Assessment Tools in Education*, 8(1), 21–37.
- Kurnia, A. (2019). Analisis tes kemampuan berpikir kritis matematika siswa dengan menggunakan Generalized Partial Credit Model (GPCM). *Pediamatika*, 1(02).
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.

- Luo, Y. (2018). Parameter recovery with marginal maximum likelihood and Markov chain Monte Carlo estimation for the generalized partial credit model. arXiv preprint arXiv:1809.07359.
- Marhami, M., Juandi, D., & Dasari, D. (2024). Indonesian students' numeracy skills based on PISA mathematical problems in secondary school: A meta-synthesis. *Jurnal Pendidikan MIPA*, 25(1), 14–26.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, 1992(1), i–30.
- Mertasari, N. M. S., & Candiasa, I. M. (2023). Development and validation of metacognitive test in programming using graded response model. *IConVET 2022 Proceedings*, 70. European Alliance for Innovation.
- Mutik, R. I., Zaenuri, Z., Walid, W., Sugiman, S., & Agoetanto, A. (2025). Analisis kemampuan literasi numerasi ditinjau dari kecemasan matematika pada pembelajaran ELPSA kelas XI SMA. *PRISMA: Prosiding Seminar Nasional Matematika*, 318–325.
- Noviantini, N. M. H., Sariyasa, S., & Ardana, I. M. (2023). Development of minimum competency assessment (AKM) instruments to measure the numeration ability of seventh grade of junior high school students. *International Journal of Trends in Mathematics Education Research*, 6(1), 65–71.
- Nursyifa, A., & Masyithoh, S. (2023). Analisis hubungan literasi numerasi dan hasil belajar siswa. *Jurnal Pendidikan Dasar dan Keguruan*, 8(1), 22–29.
- OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing, Paris.
- OECD. (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing.
- Suardipa, I. P. (2025). Pengembangan Learning Management System berorientasi kearifan lokal pada pembelajaran matematika untuk meningkatkan literasi numerasi dan literasi digital pada siswa SD (Disertasi, Universitas Pendidikan Ganesha).
- Subhaktiyasa, P. G. (2024). PLS-SEM for multivariate analysis: A practical guide to educational research using SmartPLS. *EduLine: Journal of Education and Learning Innovation*, 4(3), 353–365.
- Supriyanto, A., Mariana, N., Wiryanto, W., & Hendratno, H. (2024). Profil kemampuan numerasi siswa sekolah dasar dalam pembelajaran abad 21. *Kognitif: Jurnal Riset HOTS Pendidikan Matematika*, 4(4), 1531–1544.
- Tessmer, M. (1993). *Planning and conducting formative evaluation*. Routledge Taylor and Francis Group.
- Wilson, L. O. (2016). Bloom's taxonomy revised – Understanding the new version of Bloom's taxonomy. *A Taxonomy for Learning, Teaching, and Assessing*, 1(1), 1–8.