

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

STUDI MIXED METHODS CAPAIAN NUMERASI SISWA SEKOLAH MENENGAH BERDASARKAN DOMAIN, LEVEL, DAN JENIS SOAL

Jose Bonatua Hasibuan^{1*}, Deni Darmawan², Suhendra³, Deni Kurniawan⁴

^{1,2,3,4} Universitas Pendidikan Indonesia, Bandung, Indonesia

*Corresponding author. Jl. Flamboyan II Perum. Nuansa Griya Flamboyan G5, 28293, Pekanbaru, Indonesia.

E-mail: josehasibuan@upi.edu¹⁾

Receive 20 October 2025; Received in revise form 16 November 2025; Accepted 17 November 2025

Abstrak

Kemampuan numerasi siswa sekolah menengah di Indonesia menunjukkan kesenjangan yang jelas antara penguasaan soal rutin dan kemampuan menalar, menafsirkan data, serta memahami konteks. Kondisi ini menuntut pemetaan yang lebih komprehensif untuk memahami variasi capaian numerasi berdasarkan domain, level kognitif, dan karakteristik asesmen. Penelitian ini bertujuan memetakan kompetensi numerasi siswa SMA dan SMK di Provinsi Riau serta mengidentifikasi faktor pembelajaran, strategi guru, dan karakteristik soal yang memengaruhi hasil tersebut. Metode yang digunakan adalah *mixed methods* dengan desain sekuensial eksplanatori, melibatkan tes kuantitatif pada 288 siswa dan wawancara mendalam dengan guru serta siswa dari dua sekolah berprestasi tinggi. Instrumen tes mencakup 30 soal numerasi yang merepresentasikan tiga domain (geometri & pengukuran, aljabar, data & ketidakpastian), tiga level kognitif (*Knowing*, *Applying*, *Reasoning*), dan enam jenis soal. Hasil penelitian menunjukkan pola capaian hierarkis yang menurun dari *Knowing* ke *Reasoning* dengan perbedaan signifikan antar domain dan jenis soal. Capaian terendah ditemukan pada domain data dan ketidakpastian serta soal uraian yang menuntut penalaran tingkat tinggi dan interpretasi teks panjang. Temuan kualitatif menegaskan bahwa kesulitan utama siswa berkaitan dengan beban kognitif saat membaca konteks, keterbatasan hubungan konseptual–prosedural, dan kurangnya pembelajaran berbasis situasi nyata. Penelitian ini menyimpulkan bahwa peningkatan numerasi memerlukan pendekatan pembelajaran yang mengintegrasikan literasi membaca, penalaran matematis, dan strategi pedagogis reflektif secara berkelanjutan.

Kata kunci: Domain matematika; jenis soal; level kognitif; *mixed methods*; numerasi.

Abstract

Secondary school students' numeracy performance in Indonesia shows a clear gap between their mastery of routine tasks and their ability to reason, interpret data, and understand contextual information. This situation necessitates a more comprehensive mapping to examine variations in numeracy performance across mathematical domains, cognitive levels, and assessment characteristics. This study aims to map the numeracy competence of senior high school and vocational school students in Riau Province and to identify instructional factors, teacher strategies, and item characteristics that influence these outcomes. A mixed-methods approach with a sequential explanatory design was employed, involving quantitative testing of 288 students followed by in-depth interviews with teachers and students from two high-performing schools. The test instrument consisted of 30 numeracy items representing three domains (geometry & measurement, algebra, and data & uncertainty), three cognitive levels (*Knowing*, *Applying*, *Reasoning*), and six item formats. The findings reveal a hierarchical pattern of performance that declines from *Knowing* to *Reasoning*, with significant differences across domains and item types. The lowest performance occurred in the domain of data and uncertainty and in constructed-response items requiring higher-order *Reasoning* and interpretation of extended texts. Qualitative insights highlight students' challenges related to cognitive load when processing contextual information, limited conceptual–procedural integration, and insufficient use of real-world contexts in instruction. The study concludes that improving numeracy requires sustained instructional approaches that integrate reading literacy, mathematical Reasoning, and reflective pedagogical strategies.

Keywords: cognitive levels; item types; mathematical domains; *mixed methods*; numeracy.



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

PENDAHULUAN

Dalam pendidikan abad ke-21, literasi dan numerasi menjadi kompetensi fundamental yang menentukan kemampuan individu beradaptasi dalam era teknologi dan informasi. Literasi berperan dalam memahami serta mengomunikasikan informasi, sedangkan numerasi menuntut kemampuan menerapkan konsep matematika untuk menafsirkan data dan mengambil keputusan logis (OECD, 2023a). Kedua kompetensi ini berfungsi sebagai fondasi pembelajaran sepanjang hayat serta kecakapan warga digital. Berbagai kajian menegaskan bahwa numerasi tidak hanya berkaitan dengan prosedur berhitung, melainkan juga mencakup penalaran kuantitatif, literasi statistik, serta pemecahan masalah lintas konteks (Geiger, 2024). Oleh karena itu, pengembangan soal numerasi perlu mengintegrasikan konsep matematika dengan konteks sosial dan lingkungan nyata agar siswa dapat berpikir reflektif dan kritis (Sari et al., 2024).

Di Indonesia, capaian numerasi masih menghadapi tantangan serius. Milati et al. (2023) menekankan pentingnya kecerdasan logis-matematis dalam masalah kontekstual, sementara penelitian menunjukkan kelemahan siswa dalam penalaran dan interpretasi data meski cukup baik dalam soal rutin (Ekayanti, 2024). Kondisi ini memperlihatkan bahwa pembelajaran numerasi di sekolah menengah perlu mengintegrasikan literasi membaca, penalaran matematis, dan konteks nyata, bukan hanya latihan prosedural.

Penelitian sebelumnya telah mengembangkan berbagai inovasi pembelajaran numerasi, namun masih terbatas dari segi cakupan analisis. Listiawati dkk. (2023) menemukan bahwa pendekatan *realistic mathematics education (RME)* efektif

meningkatkan penalaran konseptual, tetapi penelitian tersebut hanya fokus pada domain tertentu tanpa menguji variasi capaian berdasarkan jenis soal. Kohar dkk. (2022) menunjukkan bahwa model pembelajaran berbasis PISA dapat meningkatkan keterlibatan dan pemahaman konsep, tetapi belum mengkaji perbedaan capaian numerasi pada level kognitif *Knowing*, *Applying*, dan *Reasoning*. Penelitian Sari dkk. (2024) menemukan bahwa siswa kesulitan memahami informasi dalam teks panjang, sehingga performa mereka menurun pada soal numerasi kontekstual; namun penelitian ini tidak membedakan capaian berdasarkan jenis soal seperti pilihan ganda, isian singkat, atau uraian. Putri dkk. (2025) mengembangkan bahan ajar numerasi berbasis budaya lokal dan melaporkan peningkatan literasi statistik, tetapi fokusnya terbatas pada domain data dan ketidakpastian.

Rangkaian temuan tersebut menunjukkan bahwa penelitian numerasi sebelumnya masih terfragmentasi dan lebih banyak menelaah satu aspek tunggal, baik domain matematika, level kognitif, ataupun jenis soal. Selain itu, penelitian yang menggunakan pendekatan *mixed methods* untuk menghubungkan capaian numerasi dengan faktor pembelajaran, strategi guru, dan konteks belajar masih sangat terbatas.

Secara global, asesmen internasional seperti PISA dan TIMSS menunjukkan bahwa tantangan numerasi tidak hanya terkait penguasaan prosedural, tetapi utamanya berkaitan dengan kemampuan memahami konteks, menganalisis data, dan menarik kesimpulan berbasis penalaran tingkat tinggi. Pola tersebut juga muncul di Indonesia melalui temuan AKM dan asesmen internasional lainnya, yang menegaskan bahwa kemampuan numerasi siswa

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

belum sepenuhnya memenuhi tuntutan kurikulum dan kebutuhan kompetensi abad ke-21.

Berdasarkan kondisi tersebut, problem statement penelitian ini dapat dirumuskan sebagai berikut: rendahnya capaian numerasi siswa pada tugas yang menuntut penalaran dan interpretasi data belum dipetakan secara komprehensif berdasarkan domain matematika, level kognitif, dan jenis soal. Selain itu, belum terdapat penelitian yang mengintegrasikan analisis kuantitatif dan kualitatif untuk memahami faktor pembelajaran dan strategi guru yang berpengaruh terhadap capaian numerasi tersebut.

Untuk menjawab permasalahan tersebut, penelitian ini menerapkan pendekatan *mixed methods* dengan desain sekuensial eksplanatori. Pendekatan ini digunakan untuk memperoleh gambaran kuantitatif mengenai variasi capaian numerasi siswa, yang kemudian diperdalam melalui wawancara kualitatif guna menelusuri faktor-faktor pembelajaran yang memengaruhinya. Integrasi kedua pendekatan ini memungkinkan pemahaman yang lebih holistik terhadap performa numerasi siswa.

Sejalan dengan itu, tujuan penelitian ini adalah: (1) memetakan capaian numerasi siswa SMA dan SMK berdasarkan domain matematika, level kognitif, dan jenis soal; (2) menganalisis variasi capaian numerasi menurut jenis dan karakteristik sekolah; dan (3) mengidentifikasi faktor pembelajaran, strategi guru, serta konteks belajar yang memengaruhi capaian numerasi siswa. Penelitian ini diharapkan memberikan kontribusi bagi pengembangan kebijakan, inovasi pembelajaran, dan penyempurnaan kurikulum numerasi yang berorientasi pada kemampuan berpikir tingkat tinggi dan literasi numerasi secara berkelanjutan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan campuran (*mixed methods*) dengan desain sekuensial eksplanatori (Creswell & Plano Clark, 2018). Desain ini dilaksanakan melalui dua tahap utama, yaitu tahap kuantitatif dan tahap kualitatif, yang disusun secara berurutan agar data kuantitatif menjadi dasar untuk menentukan fokus dan subjek pada tahap kualitatif. Pendekatan ini dipilih untuk memperoleh pemahaman yang lebih komprehensif mengenai capaian numerasi siswa dan faktor-faktor pembelajaran yang memengaruhinya.

Secara operasional, penelitian dilaksanakan melalui beberapa tahapan. Pada tahap awal, peneliti menyusun dan memvalidasi instrumen tes numerasi berdasarkan kerangka AKM. Setelah instrumen dinyatakan layak, penelitian memasuki tahap kuantitatif, yaitu pengadministrasian tes numerasi kepada sampel siswa SMA dan SMK. Hasil pada tahap ini dianalisis untuk memetakan capaian numerasi serta menentukan sekolah dan subjek yang menjadi sumber data kualitatif. Tahap berikutnya adalah pengumpulan data kualitatif melalui wawancara semi-terstruktur dengan siswa, guru, dan wakil kepala sekolah yang dipilih berdasarkan capaian numerasi tertinggi. Tahap akhir adalah integrasi temuan kuantitatif dan kualitatif untuk menghasilkan gambaran yang lebih mendalam tentang pembelajaran numerasi.

Subjek penelitian berjumlah 288 siswa SMA dan SMK di Provinsi Riau. Sampel dipilih menggunakan teknik *stratified random sampling* dengan mempertimbangkan heterogenitas berdasarkan jenis sekolah (SMA/SMK) dan lokasi (dalam/luar ibu kota provinsi). Pertimbangan ini digunakan agar sampel merepresentasikan variasi

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

karakteristik sekolah yang ada di wilayah penelitian. Pengumpulan data kuantitatif dilakukan pada semester genap tahun ajaran 2024/2025 di Kota Pekanbaru dan beberapa kabupaten/kota lainnya di Provinsi Riau. Tahap kualitatif dilakukan di dua sekolah dengan capaian numerasi tertinggi berdasarkan hasil analisis kuantitatif. Subjek kualitatif terdiri atas empat siswa, dua guru matematika, dan dua wakil kepala sekolah bidang kurikulum, yang dipilih secara *purposive* berdasarkan keterlibatan mereka dalam pembelajaran numerasi.

Materi penelitian ini adalah capaian numerasi siswa SMA dan SMK, yang diukur berdasarkan domain matematika, level kognitif, dan jenis soal sesuai kerangka AKM. Selain itu, penelitian juga menggali informasi kualitatif terkait pengalaman pembelajaran, kesulitan numerasi, strategi guru, asesmen, dan dukungan kurikulum.

Instrumen penelitian terdiri atas dua jenis. Instrumen pertama adalah tes numerasi yang memuat 30 butir soal berdasarkan kerangka AKM Kemdikbudristek, mencakup tiga domain (geometri & pengukuran, aljabar, serta data & ketidakpastian), tiga level kognitif (*Knowing*, *Applying*, *Reasoning*), dan enam jenis soal. Validasi isi instrumen dilakukan oleh tiga ahli, dan uji coba instrumen menunjukkan reliabilitas yang tinggi (Cronbach's Alpha = 0,82) (Tavakol & Dennick, 2011). Tes ini digunakan untuk memperoleh skor capaian numerasi siswa. Instrumen kedua adalah pedoman wawancara semi-terstruktur yang digunakan untuk menggali lebih jauh persepsi siswa dan guru mengenai kesulitan numerasi, strategi pembelajaran, asesmen, dan dukungan kurikulum. Wawancara dilakukan setelah analisis kuantitatif, sehingga

temuan awal dapat dijadikan dasar penyusunan pertanyaan yang lebih terarah.

Teknik pengumpulan data kuantitatif dilakukan melalui pelaksanaan tes numerasi kepada seluruh sampel pada waktu yang telah ditetapkan pihak sekolah. Tes diberikan secara tatap muka dan dikumpulkan oleh peneliti melalui koordinasi dengan guru di sekolah. Pengumpulan data kualitatif dilakukan melalui wawancara semi-terstruktur dengan subjek yang telah dipilih. Wawancara dilakukan secara langsung dan dicatat atau direkam sesuai persetujuan peserta, kemudian ditranskripsikan untuk dianalisis.

Analisis data kuantitatif dilakukan melalui analisis deskriptif dan inferensial. Uji normalitas dilakukan menggunakan Shapiro–Wilk, D'Agostino–Pearson, dan Anderson–Darling. Data yang berdistribusi normal dianalisis menggunakan uji parametrik, sedangkan data yang tidak normal dianalisis menggunakan uji nonparametrik. Perbandingan capaian antarjenis dan lokasi sekolah diuji dengan Mann–Whitney U, sedangkan perbedaan capaian antar domain, level kognitif, dan jenis soal dianalisis menggunakan uji Friedman dan Wilcoxon Signed-Rank dengan koreksi Bonferroni. Ukuran efek dihitung menggunakan Cohen's d dan eta kuadrat parsial (η^2_p).

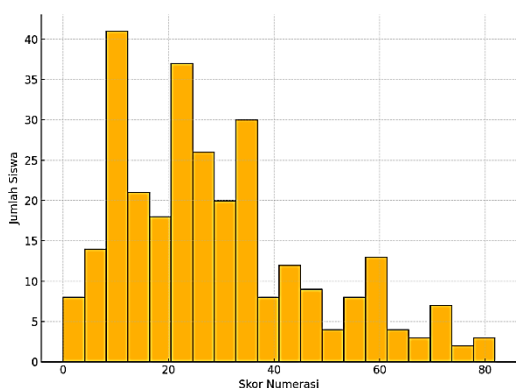
Analisis data kualitatif dilakukan terhadap transkrip wawancara untuk mengidentifikasi pola atau tema terkait kesulitan numerasi, strategi pembelajaran, asesmen, dan konteks kurikulum. Hasil analisis kualitatif digunakan untuk menjelaskan atau memperkaya interpretasi temuan kuantitatif. Integrasi hasil dilakukan pada tahap akhir dengan menghubungkan pola statistik dari tes numerasi dengan temuan naratif dari wawancara, sehingga diperoleh

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

gambaran yang lebih utuh mengenai faktor kognitif, pedagogis, dan struktural yang memengaruhi capaian numerasi siswa.

HASIL DAN PEMBAHASAN

Analisis skor numerasi menunjukkan kesenjangan besar antar siswa sekolah menengah di Riau. Skor terendah 0,00 menunjukkan ada siswa yang tidak menjawab benar, sedangkan skor tertinggi 81,82 mencerminkan pemahaman yang belum optimal. Dengan rata-rata 29,18 dan median 27,27, kemampuan numerasi tergolong rendah dan terkonsentrasi pada level dasar. Hanya 25% siswa yang mencapai skor di atas 39,39, menandakan lemahnya penguasaan numerasi siswa.



Gambar 1. Histogram distribusi skor numerasi siswa

Distribusi skor numerasi seperti terlihat pada Gambar 1 menunjukkan ketimpangan yang jelas, dengan sebagian besar siswa berada pada rentang skor rendah (10–40) dan hanya sedikit yang mencapai skor tinggi. Pola miring ke kanan (*skewness* positif) ini menandakan dominasi siswa berkemampuan dasar dan ketidak-merataan capaian numerasi di kalangan siswa sekolah menengah di Riau. Temuan ini konsisten dengan laporan PISA 2022 yang menunjukkan hanya sebagian kecil siswa Indonesia mencapai level profisiensi minimal dalam matematika (OECD, 2023b) serta hasil AKM 2023 yang menempatkan mayoritas siswa pada kategori dasar literasi numerasi (Kemdikbudristek, 2023). Kondisi ini memperkuat pandangan Beatty et al. (2021) bahwa capaian belajar siswa Indonesia masih stagnan baik dari sisi kualitas maupun pemerataan.

Hasil uji normalitas dengan Shapiro–Wilk, D’Agostino–Pearson, dan Anderson–Darling (Tabel 1) menunjukkan bahwa data tidak memenuhi asumsi normalitas. Kondisi ini sejalan dengan pola histogram yang terdistribusi tidak simetris dan cenderung ke kanan. Karena data tidak berdistribusi normal, analisis selanjutnya menggunakan uji nonparametrik yang sesuai dengan karakteristik data.

Tabel 1. Hasil uji normalitas skor numerasi siswa.

Jenis Uji	Statistik Uji	p-value	Kesimpulan
Shapiro-Wilk	0,939	<0,001	Tidak normal
D’Agostino & Pearson	25,566	<0,001	Tidak normal
Anderson-Darling	5,216	–	Tidak normal

Perbandingan Berdasarkan Jenis Sekolah

Analisis perbandingan seperti yang tersaji pada Tabel 2 menunjukkan bahwa capaian numerasi siswa SMA sedikit lebih tinggi daripada SMK,

dengan rata-rata 30,91 dan 27,06, meski median keduanya sama (27,27). Skor minimum SMK 0,00 menandakan masih ada siswa yang tidak menjawab benar, sementara skor maksimum SMA (81,82) lebih tinggi dari SMK (69,70).

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

Standar deviasi yang lebih besar pada SMA menunjukkan variasi capaian lebih luas. Sebaliknya, hasil uji Mann–Whitney U ($U = 10147$; $p = 0,568$) menunjukkan perbedaan tersebut tidak menyatakan signifikansi statistik. Artinya, meski skor SMA sedikit lebih tinggi, capaian numerasi di kedua jenis sekolah relatif serupa, dengan mayoritas siswa masih berada pada kategori rendah, menunjukkan tantangan numerasi yang bersifat umum.

Temuan ini menunjukkan bahwa perbedaan orientasi kurikulum antara SMA dan SMK tidak secara langsung memengaruhi kemampuan numerasi siswa. Hal ini sejalan dengan laporan OECD (2023c) yang menegaskan bahwa capaian numerasi lebih ditentukan oleh kualitas pembelajaran, dukungan guru, serta peluang siswa memecahkan masalah kontekstual daripada jenis sekolah.

Tabel 2. Perbandingan skor numerasi berdasarkan jenis sekolah

Jenis Sekolah	N	Rata-rata	Median	Minimum	Maksimum	Standar Deviasi
SMA	179	30,91	27,27	3,03	81,82	18,43
SMK	109	27,06	27,27	0,00	69,70	16,93

Tabel 3. Perbandingan skor numerasi berdasarkan lokasi sekolah

Lokasi Sekolah	N	Rata-rata	Median	Minimum	Maksimum	Standar Deviasi
Pekanbaru	137	39,38	36,36	3,03	81,82	18,01
Luar Pekanbaru	151	19,65	18,18	0,00	51,52	11,58

Perbandingan Berdasarkan Lokasi Sekolah

Analisis menunjukkan perbedaan signifikan capaian numerasi antara siswa di Pekanbaru dan luar Pekanbaru seperti terlihat pada Tabel 3. Rata-rata skor Pekanbaru mencapai 39,38, hampir dua kali lipat dari luar Pekanbaru (19,65), dengan skor maksimum 81,82 dibanding 51,52. Uji Mann–Whitney U ($U = 16644$, $p < 0,001$) mengonfirmasi perbedaan ini signifikan secara statistik. Meski sebagian besar siswa di kedua wilayah masih berkemampuan rendah, siswa Pekanbaru secara konsisten menunjukkan capaian lebih tinggi. Temuan ini menegaskan adanya kesenjangan numerasi antarwilayah yang perlu menjadi perhatian utama dalam peningkatan mutu proses belajar mengajar matematika pada jenjang pendidikan menengah.

Sejalan dengan Beatty dkk. (2021) yang mengaitkan perbedaan capaian numerasi antar wilayah dengan ketimpangan kualitas pendidikan. Baharuddin dan Burhan (2025) turut menunjukkan bahwa kesenjangan hasil belajar mencerminkan belum meratanya pelaksanaan kebijakan pendidikan. Sejalan dengan itu, Jannah dkk. (2024) menegaskan bahwa disparitas numerasi antarwilayah masih menjadi tantangan yang memerlukan strategi pembelajaran kontekstual dan adaptif. Hasil penelitian ini memperkuat bahwa perbedaan capaian antara Pekanbaru dan luar Pekanbaru merupakan cerminan ketimpangan pendidikan di Indonesia.

Analisis Berdasarkan Domain Kompetensi

Setelah menganalisis perbedaan capaian numerasi berdasarkan wilayah sekolah, pembahasan dilanjutkan

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

dengan perbandingan capaian siswa pada tiga domain kompetensi numerasi. Ketiga domain ini mencerminkan aspek kemampuan matematis yang berbeda, dan seluruh siswa memiliki skor pada masing-masing domain sehingga data

bersifat berpasangan (*within-subjects design*). Untuk menggambarkan karakteristik distribusi skor, dilakukan analisis deskriptif, sebagaimana dirangkum pada Tabel 4.

Tabel 4. Deskriptif per domain

Domain	Median	Q1	Q2	IQR	Min	Max
Geometri & Pengukuran	30	20	40	20	0	90
Aljabar	27	15	38	23	0	92
Data & Ketidakpastian	20	10	40	30	0	90

Tabel 4 menunjukkan bahwa median tertinggi terdapat pada domain Geometri & Pengukuran (30), disusul Aljabar (27), dan terendah Data & Ketidakpastian (20). Perbedaan median ini mengindikasikan bahwa capaian siswa relatif lebih baik pada domain Geometri & Pengukuran dibandingkan dua domain lainnya. Dari sisi sebaran, domain Data & Ketidakpastian memiliki IQR terbesar (30), yang menunjukkan variasi capaian antar siswa lebih luas, sedangkan Geometri & Pengukuran memiliki IQR terkecil (20), menandakan distribusi skor yang lebih terkonsentrasi di sekitar median. Rentang skor keseluruhan juga cukup lebar, dengan nilai minimum 0 dan maksimum berkisar antara 90 hingga 92, mencerminkan adanya perbedaan kemampuan yang nyata di antara siswa pada ketiga domain numerasi.

Selanjutnya dilakukan analisis inferensial non-parametrik untuk menguji hipotesis apakah terdapat perbedaan distribusi skor antar domain menggunakan uji Friedman, yang

merupakan padanan non-parametrik dari *repeated measures* ANOVA.

Berdasarkan hasil perhitungan uji Friedman diperoleh Chi-Square = 26.971, dengan $df = 2$, dan $p = 0.000001$. Nilai p yang lebih kecil dari 0,05 menegaskan bahwa terdapat perbedaan signifikan antara ketiga domain. Dengan kata lain, uji omnibus ini memberikan bukti bahwa distribusi skor siswa pada domain Geometri & Pengukuran, Aljabar, dan Data & Ketidakpastian tidak identik. Namun, hasil Friedman tidak menunjukkan pasangan domain mana yang berbeda, sehingga diperlukan uji lanjut (*post hoc*).

Untuk mengidentifikasi pasangan domain mana yang berbeda signifikan, dilakukan uji Wilcoxon Signed-Rank untuk setiap pasangan domain. Karena terdapat tiga perbandingan, digunakan koreksi Bonferroni untuk menjaga tingkat kesalahan keluarga (*family-wise error rate*). Dengan α awal 0,05, maka α terkoreksi menjadi 0,0167. Hasil uji Wilcoxon Berpasangan dengan Bonferroni sebagaimana terlihat pada Tabel 5.

Tabel 5. Hasil uji Wilcoxon Berpasangan dengan Bonferroni

Perbandingan	N efektif	W	Z (apros.)	p mentah	p Bonferroni	α Bonf.	r (efek)
Geometri & Pengukuran vs Aljabar	284	16949.0	-2.372	0.017654	0.052961	0.0167	0.141

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

Perbandingan	N efektif	W	Z (aproks.)	p mentah	p Bonferroni	α Bonf.	r (efek)
Geometri & Pengukuran vs Data & Ketidakpastian	229	8179.5	-4.970	0.000000	0.000001	0.0167	0.328
Geometri & Pengukuran vs Aljabar	277	13714.5	-4.149	0.000033	0.000099	0.0167	0.249

Hasil analisis menunjukkan adanya perbedaan signifikan pada sebagian domain numerasi. Uji Wilcoxon Signed-Rank memperlihatkan bahwa perbedaan antara Geometri & Pengukuran dan Aljabar tidak signifikan setelah koreksi Bonferroni ($W = 16949.0$, $Z = -2.372$, $p = 0.052961 > 0.0167$). Namun, perbandingan Geometri & Pengukuran dengan Data & Ketidakpastian ($W = 8179.5$, $Z = -4.970$, $p < 0.001$) serta Aljabar dengan Data & Ketidakpastian ($W = 13714.5$, $Z = -4.149$, $p < 0.001$) menunjukkan hasil signifikan. Dengan demikian, perbedaan capaian terutama muncul pada domain Data & Ketidakpastian, sedangkan Geometri & Pengukuran dan Aljabar menunjukkan capaian yang relatif sebanding.

Meskipun domain Geometri & Pengukuran menunjukkan capaian tertinggi, hasil keseluruhan tetap rendah dibandingkan standar PISA (OECD, 2023a) temuan penelitian mengenai rendahnya literasi matematika siswa Indonesia dalam menyelesaikan soal berbasis PISA (Rivai dkk., 2024). Perbedaan signifikan hanya terlihat pada domain Data & Ketidakpastian yang lebih rendah dari dua domain lainnya, menandakan kelemahan utama

siswa dalam memahami dan menafsirkan data. Kondisi ini berlawanan dengan negara-negara berprestasi tinggi dalam PISA yang menunjukkan keseimbangan antar domain. Menurut Jablonski dan Ludwig (2023), pencapaian tersebut didukung oleh pembelajaran integrasi berbagai domain, menggunakan konteks nyata, serta menekankan penalaran dan pemecahan masalah matematika.

Analisis Berdasarkan Level Kognitif

Bagian ini menganalisis capaian numerasi siswa berdasarkan tiga level kognitif: *Knowing*, *Applying*, dan *Reasoning*. *Knowing* mencerminkan pemahaman konsep dasar, *Applying* menunjukkan penerapan pengetahuan dalam masalah rutin, sedangkan *Reasoning* menggambarkan kemampuan bernalar untuk menyelesaikan situasi non-rutin. Instrumen disusun untuk mengukur ketiganya sehingga data bersifat berpasangan dan memungkinkan perbandingan yang akurat antarlevel. Analisis deskriptif meliputi rata-rata, median, simpangan baku, kuartil, rentang interkuartil, serta skor minimum dan maksimum disajikan pada Tabel 6 untuk memperlihatkan variasi capaian tiap level kognitif.

Tabel 6. Statistik deskriptif hasil tes numerasi per level kognitif

Level	N	Mean	SD	Med	Q1	Q3	IQR	Min	Max
<i>Knowing</i>	288	44.34	28.42	40	20	60	40	0	100
<i>Applying</i>	288	27.74	24.17	20	10	40	30	0	100
<i>Reasoning</i>	288	18.59	15.02	15	8	23	15	0	69

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

Level *Knowing* menunjukkan capaian tertinggi dengan variasi skor yang lebar, berkisar antara 0 hingga 100. Nilai rata-rata 44,34, median 40, dan simpangan baku 28,42 menunjukkan adanya kelompok siswa dengan capaian sangat tinggi, sementara sebagian lainnya masih berada pada tingkat rendah. Sekitar 43,4% siswa mencapai skor minimal 50, dan 12,5% di antaranya mencapai ≥ 85 , menandakan sebagian siswa telah menguasai konsep dasar dengan baik. Namun, sebaran skor yang luas mencerminkan adanya kesenjangan pemahaman konseptual yang besar.

Pada level *Applying*, capaian siswa menurun tajam dengan skor rata-rata 27,74 dan median 20. Sebagian besar siswa (56,6%) berada di bawah skor 30, sedangkan hanya 19,4% melampaui 50. IQR yang lebih sempit menunjukkan variasi yang lebih kecil, menandakan bahwa mayoritas siswa mengalami kesulitan dalam menerapkan konsep dasar ke situasi baru atau

kontekstual. Hal ini mengindikasikan keterbatasan kemampuan transfer pengetahuan.

Level *Reasoning* menunjukkan capaian terendah dengan rata-rata 18,59 dan median 15, di mana lebih dari separuh siswa (58,3%) hanya mencapai skor 0–19 dan tidak ada yang melampaui 70. Pola ini menunjukkan lemahnya kemampuan penalaran matematis tingkat tinggi, terutama dalam berpikir abstrak dan menyelesaikan masalah non-rutin. Uji asumsi statistik menunjukkan data tidak memenuhi syarat analisis parametrik, dengan hasil uji Shapiro–Wilk dan Levene menunjukkan distribusi tidak normal dan varians berbeda signifikan ($p < 0,05$). Untuk itu, uji Friedman diterapkan dalam analisis data yang menghasilkan $\chi^2(2) = 208,265$; $p = 5,97 \times 10^{-46}$ dengan ukuran efek Kendall's $W = 0,362$, menandakan adanya perbedaan signifikan dan konsisten antar ketiga level kognitif.

Tabel 7. Hasil uji Wilcoxon Signed-Rank antarlevel kognitif

Pasangan	W	Z (approx)	p_Bonferroni	Median Δ	Mean Δ	r (efek)	Signifikansi
<i>Knowing</i> vs <i>Applying</i>	3113.0	10.507	$2,41 \times 10^{-25}$	10.00	16.60	0.619	Signifikan
<i>Knowing</i> vs <i>Reasoning</i>	2468.5	12.792	$5,44 \times 10^{-37}$	22.00	25.75	0.754	Signifikan
<i>Applying</i> vs <i>Reasoning</i>	9923.5	6.927	$1,29 \times 10^{-11}$	7.00	9.16	0.408	Signifikan

Uji Wilcoxon Signed-Rank dengan koreksi Bonferroni ($\alpha^* = 0,0167$) menunjukkan bahwa seluruh pasangan level berbeda signifikan (Tabel 7). Perbedaan terbesar terdapat antara *Knowing* dan *Reasoning* dengan efek sangat besar ($r = 0,754$), diikuti oleh *Knowing–Applying* dengan efek besar ($r = 0,619$), dan *Applying–Reasoning* dengan efek sedang ($r = 0,408$). Hasil ini menegaskan bahwa

meskipun ketiga level berbeda secara signifikan, kesenjangan paling tajam terjadi antara penguasaan konsep dasar dan kemampuan penalaran tingkat tinggi

Hasil penelitian menunjukkan pola hierarkis capaian numerasi yang menurun seiring meningkatnya tuntutan kognitif (*Knowing* > *Applying* > *Reasoning*). Siswa umumnya mampu memahami konsep dasar, tetapi

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

kesulitan saat harus menerapkan atau menalar secara abstrak, menandakan kompleksitas kognitif belum sejalan dengan kesiapan mereka.

Rendahnya capaian *Applying* mengindikasikan pembelajaran masih berorientasi prosedural tanpa konteks bermakna. Zana dkk. (2022) menegaskan bahwa transisi dari *LOTS* ke *HOTS* menuntut pembelajaran berbasis eksplorasi dan pemecahan masalah. Sejalan dengan itu, Aydın dan Birgili (2023) menemukan bahwa evaluasi berbasis hafalan menghambat transfer pengetahuan dan pengembangan berpikir tingkat tinggi.

Kelemahan utama tampak pada *Reasoning*, di mana sebagian besar siswa gagal menunjukkan penalaran sistematis. Jablonski dan Ludwig (2023) menyoroti peran strategi pengajaran yang memberi ruang pada argumentasi dan refleksi. Dengan demikian, keterbatasan penalaran tidak hanya bersifat kognitif, tetapi juga pedagogis.

Transformasi pembelajaran melalui *problem-based learning* dan asesmen autentik menjadi kunci untuk menumbuhkan berpikir kritis dan reflektif. Darling-Hammond dan Bransford (2017) menegaskan bahwa pembelajaran berbasis *HOTS* menuntut guru yang mampu memfasilitasi eksplorasi dan kolaborasi. Karena itu, peningkatan kompetensi profesional pendidik perlu diprioritaskan agar pengembangan kemampuan berpikir tingkat tinggi efektif di SMA dan SMK.

Analisis Berdasarkan Jenis Soal

Analisis hasil numerasi menunjukkan adanya variasi capaian siswa berdasarkan jenis soal, meskipun mengukur konsep yang sama. Jenis soal memengaruhi tingkat kesulitan dan strategi berpikir yang digunakan. Marcq dkk. (2024) menemukan bahwa jenis soal menjelaskan sekitar 12% varians kesulitan dalam asesmen PISA 2018, menunjukkan bahwa struktur item turut menentukan tantangan kognitif siswa. Temuan serupa melalui studi e-TIMSS menunjukkan soal pilihan ganda cenderung lebih mudah dibanding uraian di berbagai negara (Sangwin & Jones, 2017). Dengan demikian, jenis soal berpengaruh tidak hanya terhadap capaian numerasi, tetapi juga terhadap kedalaman proses berpikir siswa.

Untuk mendeskripsikan pencapaian siswa pada berbagai jenis soal, dilakukan analisis statistik deskriptif yang mencakup rata-rata, nilai tengah, simpangan baku, batas nilai terendah dan tertinggi serta koefisien variasi. Hasil pada Tabel 8 memperlihatkan perbedaan mencolok antarjenis soal. Soal pilihan ganda memiliki rata-rata tertinggi (54,69) dan median 55, menunjukkan performa terbaik siswa pada jenis soal ini. Sebaliknya, soal uraian mencatat rata-rata terendah (19,73) dan median 20, menunjukkan kesulitan pada bentuk terbuka. Dari sisi variasi, isian singkat memiliki standar deviasi tertinggi (37,02), menunjukkan perbedaan kemampuan yang lebar, sedangkan uraian memiliki deviasi terendah (17,74), menandakan capaian rendah namun relatif seragam.

Analisis korelasi Pearson dilakukan untuk mengetahui hubungan antarjenis soal (Tabel 9).

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

Tabel 8. Statistik deskriptif per jenis soal

Jenis Soal	Mean	Median	Std	Min	Max	CV (Std/Mean)
PG	54.69	55.00	27.83	0.0	100	0.51
Checklist	22.37	20.00	24.36	0.0	100	1.09
BS	22.14	20.00	21.57	0.0	100	0.97
Menjodohkan	27.31	20.00	30.52	0.0	100	1.12
Isian	33.56	30.00	37.02	0.0	100	1.10
Uraian	19.73	20.00	17.74	0.0	100	0.90

Tabel 9. Korelasi Pearson antar jenis soal

Jenis Soal	PG	Checklist	BS	Menjodohkan	Isian	Uraian
PG	1.00	0.45	0.41	0.52	0.58	0.39
Checklist	0.45	1.00	0.71	0.54	0.49	0.44
BS	0.41	0.71	1.00	0.56	0.46	0.40
Menjodohkan	0.52	0.54	0.56	1.00	0.54	0.42
Isian	0.58	0.49	0.46	0.54	1.00	0.47
Uraian	0.39	0.44	0.40	0.42	0.47	1.00

Tabel 9 menunjukkan bahwa korelasi tertinggi terdapat antara *checklist* dan Benar-Salah (0,71). Hal ini menunjukkan bahwa siswa yang dapat mengerjakan soal *checklist* dengan baik cenderung juga mampu

menjawab soal Benar-Salah dengan baik. Sebaliknya, soal Uraian menunjukkan korelasi yang rendah dengan semua jenis soal lain. Ini menandakan bahwa capaian siswa pada soal uraian berdiri lebih independen.

Tabel 10. Perbandingan berpasangan (Selisih rata-rata dan *effect size*)

Perbandingan (A–B)	Mean Diff	95% CI		Effect Size
	(A–B)	Low	High	dz
PG – Checklist	32.32	29.10	35.41	1.23
PG – BS	32.55	29.56	35.80	1.29
PG – Menjodohkan	27.38	24.11	30.41	1.00
PG – Isian	21.13	17.61	24.67	0.66
PG – Uraian	34.96	31.97	37.78	1.15
Checklist – BS	0.23	-1.87	2.21	0.01
Checklist – Menjodohkan	-4.94	-7.78	-1.87	-0.19
Checklist – Isian	-11.19	-14.34	-7.82	-0.39
Checklist – Uraian	-2.64	-4.98	-0.31	-0.12
BS – Menjodohkan	-5.17	-7.97	-2.12	-0.21
BS – Isian	-11.42	-14.65	-8.01	-0.41
BS – Uraian	-2.87	-5.22	-0.47	-0.13
Menjodohkan – Isian	-6.25	-9.53	-2.89	-0.22
Menjodohkan – Uraian	2.30	0.11	4.67	0.10
Isian – Uraian	8.55	5.87	11.12	0.27

Analisis perbandingan (Tabel 10) menunjukkan perbedaan capaian mencolok antarjenis soal. Pilihan Ganda unggul jauh atas Uraian (selisih 34,96; efek besar = 1,15). Skor tertinggi diperoleh pada pilihan ganda (rata-rata

54,69) dan terendah pada uraian (19,73) Temuan ini menegaskan pengaruh signifikan jenis soal terhadap capaian numerasi, sejalan dengan temuan bahwa soal pilihan ganda memungkinkan jawaban benar karena menebak, sedangkan

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

soal uraian menuntut proses penalaran yang mendalam dan konstruksi jawaban (Sangwin & Jones, 2017), serta komunikasi matematis yang lebih eksplisit (Hubbard dkk., 2017). Hasil AKM juga menunjukkan siswa masih kesulitan pada uraian (Sangwin & Jones, 2017).

Variasi jenis soal diperlukan agar penilaian numerasi lebih representatif. Ketergantungan pada pilihan ganda dapat menimbulkan ilusi kemampuan (Hubbard dkk., 2017); karena itu, Uraian perlu diperkuat untuk melatih penalaran matematis (Kurniawan dkk., 2022). Instrumen yang baik harus menampilkan variasi kemampuan sesuai prinsip asesmen autentik yang mendorong berpikir reflektif (Klingbeil dkk., 2024). Secara keseluruhan, capaian numerasi terbukti bergantung pada jenis soal Kim dkk. (2018) dan membuka ruang penelitian lanjutan mengenai strategi pedagogis berbasis masalah dan penerapan *scaffolding* penalaran, yang berkontribusi signifikan terhadap pengembangan kemampuan berpikir tingkat lanjut dan keterampilan *problem solving* siswa (Kim dkk., 2018).

Hasil Wawancara Tahap Kualitatif

Tahap kualitatif bertujuan memperdalam hasil kuantitatif dan memahami makna pengalaman belajar numerasi dari perspektif siswa dan guru. Wawancara dengan empat siswa dari dua sekolah berprestasi tinggi mengungkap perbedaan strategi dan persepsi dalam menyelesaikan soal numerasi. Variasi latar belakang dan pendekatan pembelajaran menunjukkan bahwa capaian numerasi dipengaruhi bukan hanya oleh kemampuan berhitung, tetapi juga oleh pemahaman konteks dan bimbingan guru dalam membangun proses berpikir matematis.

Hasil wawancara memperlihatkan bahwa kesulitan utama siswa bukan

terletak pada perhitungan, melainkan pada pemahaman terhadap teks panjang dan soal berbasis konteks. Banyak siswa harus membaca ulang untuk menemukan inti pertanyaan atau menafsirkan instruksi dengan benar. Hal ini menunjukkan bahwa numerasi erat kaitannya dengan kemampuan literasi matematis (*disciplinary literacy*). Berdasarkan *Cognitive Load Theory*, teks yang kompleks dapat meningkatkan *extraneous cognitive load* sehingga fokus siswa bergeser dari penalaran matematis ke pemahaman bacaan (Sweller, 1988). Rendahnya performa siswa Indonesia dalam PISA juga dikaitkan dengan lemahnya kemampuan menafsirkan konteks bacaan dalam soal dunia nyata (OECD, 2023b).

Persepsi siswa terhadap tingkat kesulitan soal sering berbeda dengan hasil capaian aktual. Domain Data dan Ketidakpastian yang dianggap mudah karena hanya melibatkan substitusi rumus, justru menghasilkan skor terendah. Siswa cenderung mampu menghitung, tetapi kesulitan menafsirkan makna data dan menarik kesimpulan dari grafik atau tabel, menunjukkan perlunya penguatan *data literacy* berbasis proyek sosial dan lingkungan (OECD, 2023b). Pada domain lain, geometri dianggap sulit karena menuntut integrasi visual dan simbolik, sedangkan aljabar menuntut keseimbangan pemahaman konsep dan keterampilan prosedur (Rittle-Johnson dkk., 2015). Selain itu, jenis soal juga memengaruhi capaian; soal benar-salah cenderung hanya menguji pengenalan klaim dan rentan terhadap jawaban menebak, sedangkan uraian menuntut kemampuan menjelaskan langkah berpikir dan memberikan alasan secara eksplisit, meskipun memiliki risiko subjektivitas dalam penskoran (Breuer dkk., 2023).

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

Variasi pengalaman belajar dan ketersediaan fasilitas turut berperan besar dalam membentuk kesiapan numerasi. Siswa di sekolah berprestasi tinggi terbiasa menghadapi asesmen berbasis stimulus dan teknologi, sedangkan siswa di sekolah lain menghadapi keterbatasan sumber daya dan latihan soal nonrutin. Ketimpangan akses teknologi dan perbedaan kesempatan belajar ini berdampak pada capaian numerasi dan kualitas penalaran matematis siswa, sejalan dengan temuan mengenai pengaruh lingkungan digital terhadap pengajaran, pembelajaran, dan asesmen matematika (Weigand dkk., 2024). Sekolah A yang menerapkan kurikulum ganda (nasional dan Singapura) unggul dalam literasi data dan *problem solving*, sejalan dengan temuan bahwa kurikulum matematika Singapura secara eksplisit mengintegrasikan literasi matematika termasuk analisis data dan pemecahan masalah dalam kerangka kurikulumnya (Sumliyah dkk., 2025), sementara Sekolah B dengan kurikulum nasional menonjol melalui refleksi pedagogis dan adaptasi strategi pembelajaran (Hidayah dkk., 2025). Sekolah A berupaya menerapkan prinsip-prinsip perencanaan berbasis bukti dan instruksional penelitian dalam pembelajaran matematika (Fernández & Ortiz Galarza, 2023), sedangkan Sekolah B memperkuat *teacher agency* dan menerapkan *differentiated instruction* (Baharuddin & Burhan, 2025).

Kedua sekolah telah menyesuaikan asesmen dengan format AKM berbasis stimulus, meskipun menghadapi tantangan berbeda. Sekolah A masih terkendala literasi dalam memahami teks, grafik, dan data, sedangkan Sekolah B terkendala waktu dalam pengembangan soal berbasis konteks. Penerapan *formative assessment*

terbukti meningkatkan capaian siswa ketika digunakan secara konsisten (Maskos dkk., 2025). Guru di Sekolah A unggul dalam keterampilan prosedural, sedangkan Sekolah B lebih kuat dalam penalaran konseptual, mendukung pentingnya *Mathematical Knowledge for Teaching* (Fernández & Ortiz Galarza, 2023).

Inisiatif peningkatan numerasi diimplementasikan melalui berbagai inovasi pembelajaran. Sekolah A mengoptimalkan program formal seperti klub matematika, kemitraan universitas, dan pembinaan olimpiade, sedangkan Sekolah B mengutamakan pendekatan kontekstual melalui bimbingan dan latihan numerasi berbasis kehidupan sehari-hari. Pendekatan berbasis konteks ini juga sejalan dengan temuan (Mahmuti dkk., 2025), yang menunjukkan bahwa pembelajaran yang menghubungkan konsep dengan pengalaman nyata membantu siswa memahami situasi numerasi secara lebih bermakna. Temuan penelitian ini mengindikasikan bahwa capaian numerasi tidak hanya ditentukan oleh keterampilan berhitung, tetapi juga oleh kemampuan siswa memahami konteks, literasi yang terintegrasi, asesmen yang digunakan, serta strategi pembelajaran yang menumbuhkan penalaran matematis dalam kehidupan sehari-hari.

Perbandingan dengan Praktik Internasional dan Arah Pengembangan Numerasi

Hasil penelitian menunjukkan bahwa tantangan numerasi tidak hanya disebabkan oleh kelemahan konseptual dan prosedural, tetapi juga oleh faktor lain seperti kurikulum, strategi pembelajaran, asesmen, dan kemampuan literasi dalam memahami teks serta data. Kesulitan membaca stimulus panjang dan kecenderungan pada soal

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

rutin menjadi pola umum, sementara variasi implementasi kurikulum menegaskan bahwa keberhasilan numerasi lebih ditentukan oleh kualitas pedagogi daripada jenis kurikulum.

Perbandingan dengan praktik internasional menunjukkan berbagai pendekatan efektif. Di Singapura, model Concrete–Pictorial–Abstract (CPA) terbukti meningkatkan motivasi belajar dan capaian matematis melalui tahapan representasi bertingkat yang membantu menurunkan beban kognitif siswa (Iyamuremye & Burns, 2025). Jepang menerapkan *lesson study* Fauziah dan Pandra (2024) untuk memperkuat refleksi guru dan penalaran siswa, sedangkan Belanda melalui Realistic Mathematics Education (RME) menekankan konteks nyata. Finlandia mengintegrasikan multiliterasi melalui pembelajaran berbasis fenomena yang menerapkan literasi multimodal, dialog inklusif, dan representasi visual-simbolik di ruang kelas matematika (Kangas & Rasi, 2021), Australia menonjol dengan penerapan *formative assessment* berbasis rubrik penalaran matematis (Herbert dkk., 2022), dan Inggris melalui pendekatan *Teaching for Mastery* yang menerjemahkan temuan riset ke dalam kebijakan dan praktik kelas menekankan pemahaman bertahap melalui latihan bervariasi (Coles & Helme, 2022).

Dari perbandingan tersebut, penguatan numerasi di Indonesia perlu difokuskan pada penerapan pendekatan kontekstual seperti CPA dan RME, peningkatan profesionalisme guru melalui *lesson study*, asesmen formatif reflektif, dan integrasi literasi dalam pembelajaran matematika. Prinsip-prinsip ini sejalan dengan *Cognitive Load Theory* dan teori relasi konseptual–prosedural (Rittle-Johnson dkk., 2015), menuju pembelajaran

matematika yang reflektif, kontekstual, dan adaptif terhadap tantangan abad ke-21.

Hasil penelitian ini memiliki beberapa implikasi penting bagi praktik pembelajaran, asesmen, dan kebijakan pendidikan. Temuan bahwa capaian *Reasoning* dan domain Data & Ketidakpastian merupakan area terlemah menunjukkan perlunya penguatan literasi teks matematika dan pemahaman konteks autentik dalam proses belajar. Implikasi bagi guru adalah perlunya merancang pembelajaran yang tidak hanya berfokus pada prosedur, tetapi juga pada penalaran, interpretasi informasi, dan keterampilan membaca matematis. Dari sisi asesmen, hasil penelitian menegaskan pentingnya variasi jenis soal yang mampu menangkap kemampuan berpikir tingkat tinggi, bukan hanya kemampuan menyelesaikan soal rutin. Sementara itu, bagi pengambil kebijakan, temuan ini memberikan kontribusi berupa dasar empirik bagi pengembangan modul numerasi kontekstual, peningkatan kapasitas guru dalam pedagogi literasi-numerasi, serta penataan asesmen yang lebih selaras dengan tuntutan numerasi abad ke-21. Penelitian ini juga memperkuat pemahaman bahwa numerasi merupakan keterampilan multidimensi yang memerlukan sinergi antara literasi, penalaran, dan konteks pembelajaran.

KESIMPULAN DAN SARAN

Integrasi data kuantitatif dan kualitatif menghasilkan tiga temuan utama. Pertama, capaian numerasi menunjukkan pola menurun sesuai level kognitif: *Knowing* tertinggi, *Applying* menengah, dan *Reasoning* terendah, dengan domain Data dan Ketidakpastian paling lemah. Kedua, kesulitan utama siswa terletak pada memahami teks dan konteks soal akibat rendahnya literasi

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

dan terbatasnya pengalaman dengan variasi soal. Ketiga, numerasi terbukti sebagai keterampilan multidimensi yang menghubungkan literasi, penalaran, dan penerapan konsep matematika dalam konteks nyata.

Sintesis kedua pendekatan menunjukkan bahwa rendahnya capaian tidak hanya disebabkan lemahnya penguasaan konsep, tetapi juga kesulitan memahami konteks bacaan dan jenis soal. Berdasarkan *Cognitive Load Theory*, guru disarankan mengurangi beban kognitif melalui pelatihan istilah dan visualisasi pendukung, serta memperkuat hubungan konseptual-prosedural. Pendekatan *lesson study* dan asesmen hibrida seperti *multiple-choice with Reasoning* efektif menilai penalaran, sedangkan integrasi *multiliteracy* dan *National Numeracy Learning Progression* memperluas penerapan numerasi lintas disiplin. Dari sisi kebijakan, pemerataan sumber belajar, adaptasi modul kontekstual berbasis CPA, serta strategi *small steps* dan *intelligent practice* diperlukan untuk menyinergikan literasi, pedagogi reflektif, dan asesmen kontekstual dalam memperkuat penalaran berkelanjutan.

Penelitian selanjutnya disarankan memperluas cakupan sampel agar pemetaan numerasi lebih representatif dan memungkinkan perbandingan antarwilayah. Kajian lanjutan juga perlu menelaah hubungan antara capaian numerasi dan strategi pembelajaran maupun asesmen guru untuk mengidentifikasi faktor-faktor yang berpengaruh secara lebih spesifik. Selain itu, penelitian dapat menguji efektivitas strategi pembelajaran dalam meningkatkan kompetensi numerasi peserta didik.

DAFTAR PUSTAKA

- Aydın, U., & Birgili, B. (2023). Assessing mathematical higher-order thinking skills: An analysis of Turkish university entrance examinations. *Educational Assessment*. <https://eprints.gla.ac.uk/296378>
- Baharuddin, B., & Burhan, B. (2025). Urban and rural teacher perspectives on Indonesian educational reform: Challenges and policy implications. *Cogent Education*, 12(1), 2497142. <https://doi.org/10.1080/2331186X.2025.2497142>
- Beatty, A., Berkhout, E., Bima, L., Pradhan, M., & Suryadarma, D. (2021). Schooling progress, learning reversal: Indonesia's learning profiles between 2000 and 2014. *International Journal of Educational Development*, 85, 102436. <https://doi.org/10.1016/j.ijedudev.2021.102436>
- Breuer, S., Scherndl, T., & Ortner, T. M. (2023). Effects of response format on achievement and aptitude assessment results: Multi-level random effects meta-analyses. *Royal Society Open Science*, 10(5), 220456. <https://doi.org/10.1098/rsos.220456>
- Coles, A. T., & Helme, R. (2022). Teaching for mastery in primary mathematics: A study of translating research into policy and practice. *Review of Education*, 10(1), e3326. <https://doi.org/10.1002/rev3.3326>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Sage.
- Darling-Hammond, L., & Bransford, J. (Eds.). (2017). *Preparing teachers for a changing world: What teachers should learn and be able to do*. Jossey-Bass.

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

- Ekayanti, I. G. A. P. W. (2024). Indonesian Realistic Mathematics Education: Alternative Mathematics Learning Approach to Improve Fifth Grade Numeracy Literacy Skills. *Journal for Lesson and Learning Studies*, 7(1), 54–60. <https://doi.org/10.23887/jlls.v7i1.68943>
- Fauziah, A., & Pandra, V. (2024). Developing PISA-like Mathematics Tasks in Musi Rawas Regency Contexts using Lesson Study. *Mathematics Education Journal*, 17(3), 279–296.
- Fernández, L. M., & Ortiz Galarza, M. (2023). Contextualizing the Mathematical Knowledge for Teaching Framework for Teachers of Emergent Bilinguals. *Frontiers in Education*, 8, 1146797. <https://doi.org/10.3389/educ.2023.1146797>
- Geiger, V. (2024). A critical turn in numeracy education and practice. *Frontiers in Education*, 9, 1363566. <https://doi.org/10.3389/educ.2024.1363566>
- Herbert, S., Vale, C., White, P., & Bragg, L. A. (2022). Engagement with a formative assessment rubric: A case of mathematical reasoning. *International Journal of Educational Research*, 111, 101899. <https://doi.org/10.1016/j.ijer.2021.101899>
- Hidayah, I. R., Sa'dijah, C., Anwar, L., Yerizon, Y., & Arnawa, I. M. (2025). Empowering students' numeracy skills: Mathematics teachers' perceptions regarding the effectiveness and challenges of Indonesian national curriculum and the programs – A mixed method study. *Infinity Journal*, 14(1), 163–188. <https://doi.org/10.22460/infinity.v14i1.p163-188>
- Hubbard, J. K., Potts, M. A., & Couch, B. A. (2017). How question types reveal student thinking: Multiple-true-false item formats capture critical differences in student reasoning. *CBE—Life Sciences Education*, 16(3), ar41. <https://doi.org/10.1187/cbe.16-12-0339>
- Iyamuremye, E., & Burns, D. (2025). Concrete–Pictorial–Abstract instruction: Enhancing students' learning motivation and achievement in mathematics. *Cogent Education*, 12(1), 2558303. <https://doi.org/10.1080/2331186X.2025.2558303>
- Jablonski, S., & Ludwig, M. (2023). Teaching and learning of geometry—A literature review on current developments in theory and practice. *Education Sciences*, 13(7), 682. <https://doi.org/10.3390/educsci13070682>
- Jannah, M., Pesik, A., Kumesan, S., Ermita, & Rahim, N. (2024). Analysis of student errors in solving equations and inequalities problems in algebra and trigonometry based on Newman procedure. *Journal of Education and Teaching Learning*, 6(1), 51–63. <https://doi.org/10.51178/jetl.v6i1.1616>
- Kangas, M., & Rasi, P. (2021). Phenomenon-based learning of multiliteracy in a Finnish upper secondary school. *Media Practice and Education*, 22(4), 342–359. <https://doi.org/10.1080/25741136.2021.1977769>
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2023). *Pengelompokan satuan pendidikan berdasarkan hasil Asesmen Nasional tahun 2021*. Pusat Asesmen Pendidikan (Pusmendik).

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

- <https://pusmendik.kemdikbud.go.id/pdf/file-136>
- Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review*, 30, 397–429. <https://doi.org/10.1007/s10648-017-9419-1>
- Klingbeil, K., Rösken, F., Barzel, B., Schacht, F., Stacey, K., Steinle, V., & Thurm, D. (2024). Validity of multiple-choice digital formative assessment for detecting misconceptions in mathematics. *ZDM – Mathematics Education*. <https://doi.org/10.1007/s11858-024-01556-0>
- Kohar, A. W., Rahaju, E. B., & Rohim, A. (2022). Prospective teachers' design of numeracy tasks using a physical distancing context. *Journal on Mathematics Education*, 13(2), 191–210. <https://doi.org/10.22342/jme.v13i2.pp191-210>
- Kurniawan, A. P., Budiarto, M. T., & Ekawati, R. (2022). Pengembangan soal numerasi berbasis konteks nilai budaya primbon Jawa. *JRPM (Jurnal Review Pembelajaran Matematika)*, 7(1), 20–34. <https://doi.org/10.15642/jrpm.2022.7.1.20-34>
- Listiawati, N., Sabon, S. S., Wibowo, S. W. A., Zulkardi, & Riyanto, B. S. (2023). Analysis of implementing Realistic Mathematics Education principles to enhance mathematics competence of slow learner students. *Journal on Mathematics Education*, 14(4), 683–700. <https://doi.org/10.22342/jme.v14i4.pp683-700>
- Mahmuti, A., Hamzić, D. K., & Thaqi, X. (2025). The impact of contextual teaching and learning on improving student achievement in economic mathematics. *International Electronic Journal of Mathematics Education*, 20(3), em0833. <https://doi.org/10.29333/iejme/16233>
- Marcq, K., Chalén Donayre, E. J., & Braeken, J. (2024). The role of item format in the PISA 2018 mathematics literacy assessment: A cross-country study. *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2024.101401>
- Maskos, K., Schulz, A., Oeksuez, S. S., & Rakoczy, K. (2025). Formative assessment in mathematics education: A systematic review. *ZDM – Mathematics Education*, 57(2), 679–693. <https://doi.org/10.1007/s11858-025-01696-x>
- Milati, A., Baiduri, & Khusna, A. H. (2023). Kemampuan numerasi siswa dalam menyelesaikan masalah kontekstual berdasarkan kecerdasan logis-matematis. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(3), 2592–2604. <https://doi.org/10.24127/ajpm.v12i3.7163>
- OECD. (2023a). *PISA 2022 results: Learning in the digital world, vol. I*. OECD Publishing. <https://doi.org/10.1787/7f6d1f92-en>
- OECD. (2023b). *PISA 2022 results: Learning in the digital world, vol. I*. OECD Publishing. <https://doi.org/10.1787/7f6d1f92-en>
- OECD. (2023c). *PISA 2022 results (Vol. I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Putri, R. I. I., Zulkardi, Sari, N., Sagita, L., Siligar, E. I. P., & Sukma, Y. (2025). Learning numeracy using new Pempek mathematics. *Journal*

DOI: <https://doi.org/10.24127/ajpm.v14i4.14490>

- on *Mathematics Education*, 16(1), 1–22.
<https://doi.org/10.22342/jme.v16i1.pp1-22>
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a One-Way Street: Bidirectional Relations Between Procedural and Conceptual Knowledge of Mathematics. *Educational Psychology Review*, 27(4), 587–597.
<https://doi.org/10.1007/s10648-015-9302-x>
- Rivai, A., Lestari, A., Munir, N. P., & Anas, A. (2024). Students' Mathematical Literacy in Solving PISA Problems Observed by Learning Styles. *Mathematics Education Journal*, 17(1), 121–134.
- Sangwin, C. J., & Jones, I. (2017). Asymmetry in student achievement on multiple-choice and constructed-response items in reversible mathematics processes. *Educational Studies in Mathematics*, 94(2), 205–222.
<https://doi.org/10.1007/s10649-016-9725-4>
- Sari, Y. M., Kohar, A. W., El Milla, Y. I., Fiangga, S., & Rahayu, D. S. (2024). Aligning numeracy task design with SDG goals: Nutrition facts as a context for prospective mathematics teachers' problem posing. *Journal on Mathematics Education*, 15(1), 191–206.
<https://doi.org/10.22342/jme.v15i1.pp191-206>
- Sumliyah, I., Junaedi, I., & Mulyono. (2025). Mathematical literacy: A comparative analysis of school mathematics curricula in Indonesia, Singapore, and China. *Inovasi Matematika (Inomatika)*, 7(1), 105–129.
<https://doi.org/10.35438/inomatika.v7i1.488>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
https://doi.org/10.1207/s15516709cog1202_4
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
<https://doi.org/10.5116/ijme.4dfb.8dfd>
- Weigand, H.-G., Trgalová, J., & Tabach, M. (2024). Mathematics teaching, learning, and assessment in the digital age. *ZDM – Mathematics Education*, 56, 525–541.
<https://doi.org/10.1007/s11858-024-01612-9>
- Zana, F. M., Sa'dijah, C., & Susiswo, S. (2022). LOTS to HOTS: How do mathematics teachers improve students' higher-order thinking skills in the class? *International Journal of Trends in Mathematics Education Research*, 5(3), 251–260.
<https://doi.org/10.33122/ijtmer.v5i3.143>