



## **Pengembangan Instrumen Asesmen Literasi Matematika Berbasis PISA dengan Computer-Baset Test pada Siswa Jenjang Menengah Atas**

Hamidun Nafi' Syifauddin<sup>1</sup>, Ardhi Prabowo<sup>2\*</sup>

<sup>1,2</sup> Universitas Negeri Semarang

\* [ardhiprabowo@mail.unnes.ac.id](mailto:ardhiprabowo@mail.unnes.ac.id)

### **ABSTRAK**

Penelitian ini bertujuan mengembangkan instrumen asesmen literasi matematika berbasis *computer-based test* (CBT) yang mengadaptasi kerangka *Programme for International Student Assessment* (PISA) 2022. Kebutuhan ini berangkat dari fakta minimnya instrumen asesmen untuk jenjang menengah atas yang mana program PISA sendiri hanya sampai jenjang menengah pertama. Penelitian menggunakan model pengembangan ADDIE (*Analyze, Design, Develop, Implement, Evaluate*); instrumen berupa 9 butir soal diujicobakan pada 65 siswa SMA kelas XI pada tahap *Evaluate*. Data dianalisis menggunakan korelasi Pearson (validitas butir), reliabilitas *Cronbach's Alpha*, tingkat kesukaran, dan daya pembeda. Hasil menunjukkan seluruh 9 butir valid secara empiris dengan reliabilitas instrumen sebesar 0,641. Tujuh butir dinyatakan layak dipertahankan dan dua butir memerlukan revisi. Penelitian ini berkontribusi mengembangkan instrumen asesmen literasi matematika beracuan PISA yang siap diadaptasi untuk mengukur literasi matematika siswa jenjang menengah atas sekaligus membiasakan siswa melaksanakan tes dengan moda ujian berbasis komputer yang menyerupai kondisi asesmen internasional.

**Kata Kunci:** instrumen asesmen, PISA, literasi matematika, computer-based test, ADDIE

### **ABSTRACT**

*This study aims to develop a computer-based test (CBT) mathematical literacy assessment instrument that adapts the 2022 Programme for International Student Assessment (PISA) framework. This need stems from the scarcity of assessment instruments for the senior high school level, given that the PISA programme itself only reaches the junior high school level. The study employed the ADDIE (Analyze, Design, Develop, Implement, Evaluate) development model; the instrument, comprising nine items, was trialled on 65 eleventh-grade senior high school students at the Evaluate stage. The data were analysed using Pearson correlation (item validity), Cronbach's Alpha reliability, difficulty index, and discrimination index. The results show that all nine items are empirically valid, with an instrument reliability of 0.641. Seven items were deemed suitable to retain, and two items required revision. This study contributes to the development of a PISA-referenced mathematical literacy assessment instrument that is ready to be adapted for measuring the mathematical literacy of senior high school students while familiarising them with taking tests in a computer-based mode that resembles international assessment conditions.*

**Keywords:** assessment instrument, PISA, mathematical literacy, computer-based test, ADDIE

## 1. PENDAHULUAN

Literasi matematika merupakan kemampuan bernalar secara matematis serta merumuskan, menerapkan, dan menafsirkan matematika untuk menyelesaikan masalah dalam berbagai konteks kehidupan nyata (OECD, 2023a). Kemampuan ini menjadi salah satu fokus pengukuran PISA, sebuah studi internasional yang menilai literasi matematika, membaca, dan sains siswa berusia 15 tahun di lebih dari 80 negara. Hasil PISA 2022 menunjukkan skor rata-rata matematika siswa Indonesia berada jauh di bawah rata-rata negara OECD, dengan hanya sebagian kecil siswa yang mencapai kompetensi minimal level 2 (OECD, 2023b). Wijaya et al. (2024) mengonfirmasi temuan tersebut melalui perspektif guru matematika Indonesia, yang mengidentifikasi faktor kurikulum, keterbatasan sumber daya, dan dampak pandemi sebagai penyebab utama stagnasi capaian literasi matematika siswa. Maskar & Herman (2024) melaporkan kondisi senada pada siswa sekolah menengah pertama di Cianjur, dengan data PISA 2018 menunjukkan hanya 1 persen siswa Indonesia mencapai level kemahiran tinggi (level 5-6) sementara 72 persen berada pada level kemahiran rendah dalam literasi matematika. Kondisi ini menegaskan perlunya instrumen asesmen di tingkat sekolah yang dirancang selaras dengan kerangka dan karakteristik soal PISA. Instrumen semacam ini tidak dapat sepenuhnya mengandalkan butir-butir soal resmi PISA yang telah dipublikasikan, sebab PISA secara khusus dirancang untuk menilai siswa berusia 15 tahun menjelang akhir wajib belajar, dengan konten, bahasa, dan konteks soal yang sengaja disesuaikan agar tepat guna (*age-appropriate*) khusus untuk kelompok usia tersebut (OECD, 2023a), sehingga cakupannya tidak otomatis mewakili seluruh rentang usia siswa menengah atas. Kondisi inilah yang mendorong perlunya pengembangan instrumen bergaya PISA secara mandiri, agar guru dapat memetakan kemampuan literasi matematika siswa jenjang menengah atas secara lebih akurat.

PISA mengukur literasi matematika melalui tiga proses kognitif utama, yaitu *Formulate* (merumuskan situasi secara matematis), *Employ* (menerapkan konsep, fakta, dan prosedur matematika), dan *Interpret* (menafsirkan serta mengevaluasi hasil matematis dalam konteks nyata), yang secara kolektif dikenal sebagai indikator FEI (OECD, 2023a). Sejak siklus 2015, PISA juga beralih ke moda *computer-based test* (CBT) yang disebut *Computer-Based Assessment of Mathematics* (CBAM), dilengkapi alat hitung digital yang dapat digunakan siswa untuk menunjukkan proses perhitungan secara langsung pada layar ujian (OECD, 2023a). Karakteristik ganda ini, yaitu proses kognitif FEI dan moda CBT dengan alat hitung terintegrasi, menjadi ciri khas soal PISA generasi terbaru yang membedakannya dari soal ujian konvensional berbasis kertas.

Penelitian-penelitian sebelumnya telah menghasilkan sejumlah kontribusi penting pada dua kelompok yang berbeda. Pada kelompok pengembangan soal model PISA, Charmila et al. (2016) mengembangkan 14 butir soal matematika model PISA dengan konteks budaya Jambi yang valid, praktis, dan berpotensi meningkatkan literasi matematika siswa sekolah menengah pertama, menggunakan pendekatan design research. Tanjung et al. (2023) berada pada kelompok yang sama dengan mengembangkan soal model PISA berkonteks imunitas pada masa pandemi, mencakup satu unit stimulus dengan beberapa butir evaluasi pada level kesulitan PISA 3 hingga 5. Beberapa penelitian lain pada kelompok yang sama turut mengembangkan soal model PISA dengan konteks Indonesia yang beragam, antara lain konteks Asian Games (Jannah & Putri, 2019; Pratiwi & Putri, 2019), kesulitan siswa menyelesaikan soal PISA konten

shape and space pada masa pandemi (Sistiyawati & Sitisyarah, 2022), konteks pandemi COVID-19 (Nusantara et al., 2021; Zulkardi et al., 2020), aktivitas geometri berbantuan ICT (Meryansumayeka et al., 2022), dan konteks pemilihan umum (Mirandah & Putri, 2025). Penelitian-penelitian pada kelompok ini secara konsisten menerapkan design research dengan tahap preliminary dan formative evaluation sebagai metode baku, menghasilkan instrumen dalam moda cetak (paper based) yang divalidasi dari sisi isi, konstruk, bahasa, dan kepraktisan. Pada kelompok analisis kualitas butir, Desy Rizbudiani et al. (2021) menerapkan pendekatan item response theory model Rasch untuk menganalisis kualitas 20 butir pilihan ganda pada ujian akhir semester matematika materi sistem persamaan linear dua variabel pada 195 siswa, dengan hasil menunjukkan seluruh butir memiliki kecocokan yang baik terhadap model *Rasch* (INFIT MNSQ berkisar 0,89 sampai 1,17). Suwarno & Ardani (2022) mendiagnosis kemampuan literasi matematika siswa menggunakan tiga soal PISA level 4 dan menemukan variasi penguasaan kemampuan koneksi, reproduksi, dan refleksi antarsiswa, sementara Syifa' & Rohman (2023) serta Rahmaningtyas et al. (2021) menunjukkan bahwa karakteristik individual siswa, seperti gaya berpikir dan tipe kepribadian, turut memengaruhi capaian kemampuan matematis siswa pada instrumen yang diberikan.

Pada kelompok ketiga, penggunaan model ADDIE untuk pengembangan instrumen asesmen juga telah diterapkan pada berbagai konteks. Hari Sugiharto Setyaedhi (2025) mengembangkan 25 butir soal numerasi model PISA untuk siswa jenjang menengah atas melalui evaluasi formatif bertahap (*self-evaluation*, *expert review*, *one-to-one*, dan *small group*) dengan hasil validasi materi, konstruksi, dan bahasa berkategori sangat layak. Maxnun et al. (2024) mengembangkan instrumen asesmen kognitif berbasis higher order thinking skills menggunakan model ADDIE pada siswa sekolah menengah kejuruan dan menunjukkan instrumen tersebut dapat memetakan capaian HOTS siswa ke dalam kategori sangat baik, baik, dan cukup. Validasi ahli pada tahap Develop dalam kerangka ADDIE umumnya diukur menggunakan koefisien Aiken's V, sebagaimana diterapkan Retnawati (2016) yang membandingkan koefisien Aiken dengan indeks Gregory untuk membuktikan validitas isi skala *self-regulated learning*, dan Ahmad et al. (2024) yang memvalidasi instrumen learning management system pada institusi vokasi Malaysia melalui konsensus tujuh pakar. Penelitian-penelitian tersebut menegaskan model ADDIE sebagai kerangka pengembangan yang lazim digunakan untuk instrumen asesmen, dengan konteks penerapan yang beragam mulai dari numerasi model PISA berbasis kertas, asesmen kognitif HOTS, hingga validasi instrumen non-matematika.

Penelitian ini mengisi ruang di antara ketiga kelompok tersebut dengan mengembangkan instrumen asesmen literasi matematika yang sekaligus mengadaptasi bentuk dan cara penyajian soal PISA, indikator proses FEI, dan moda CBT lengkap dengan alat hitung digital di dalamnya, kemudian menguji kelayakan psikometriknya secara menyeluruh melalui uji validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Pengembangan dilakukan menggunakan model ADDIE pada siswa SMA kelas XI, jenjang yang relatif jarang menjadi subjek penelitian pengembangan soal model PISA dibandingkan jenjang sekolah menengah pertama pada penelitian-penelitian sebelumnya. Rumusan masalah penelitian ini adalah bagaimana proses pengembangan instrumen asesmen literasi matematika berbasis CBT beracuan PISA dengan indikator FEI, dan bagaimana kualitas psikometrik instrumen tersebut ditinjau dari validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal pada tahap *Evaluate*. Tujuan

penelitian ini adalah mendeskripsikan proses pengembangan instrumen tersebut dan menganalisis kelayakan psikometriknya berdasarkan data uji coba pada siswa SMA kelas XI.

## 2. METODE

### 2.1. Jenis Penelitian

Penelitian ini merupakan penelitian dan pengembangan (*Research and Development*) yang menggunakan model ADDIE, yaitu *Analyze*, *Design*, *Develop*, *Implement*, dan *Evaluate*, untuk mengembangkan instrumen asesmen literasi matematika berbasis CBT beracuan PISA. Artikel ini melaporkan proses pengembangan pada seluruh tahapan ADDIE dengan penekanan pada hasil analisis data kuantitatif tahap *Evaluate*.

### 2.2. Waktu dan Tempat Penelitian

Penelitian dilaksanakan di MAN 1 Jepara salah satu Sekolah Menengah Atas di Indonesia, pada siswa kelas XI. Penelitian dilaksanakan pada 9 September 2025. Sekolah ini dipilih dengan pertimbangan ketersediaan perangkat untuk pelaksanaan CBT.

### 2.3. Subjek Penelitian

Subjek uji coba tahap *Evaluate* berjumlah 65 siswa SMA kelas XI. Alasan dipilihnya Kelas XI sebagai subjek penelitian karena berada pada masa yang paling tepat dalam jenjang pendidikan SMA: siswa kelas X umumnya masih berada dalam tahap adaptasi terhadap lingkungan dan tuntutan akademik sekolah menengah atas, sedangkan siswa kelas XII telah memasuki masa asesmen atas kesiapan akademik yang dibangun sejak kelas X dan XI, sehingga kurang tepat dijadikan waktu uji coba instrumen yang masih bersifat pengembangan.

### 2.4. Prosedur Pengembangan

Tahap *Analyze* dilakukan untuk mengidentifikasi kebutuhan instrumen asesmen literasi matematika yang selaras dengan kerangka PISA 2022, mencakup analisis karakteristik siswa, kompetensi yang diukur, dan keterbatasan instrumen asesmen yang tersedia di sekolah. Tahap *Design* meliputi penyusunan kisi-kisi soal berdasarkan konten dan konteks PISA serta indikator proses FEI, perancangan bentuk penyajian soal bergaya unit PISA (satu stimulus dengan beberapa butir terkait), dan perancangan antarmuka CBT beserta alat hitung digital. Operasionalisasi indikator proses FEI pada tahap ini dilakukan dengan memetakan delapan kompetensi matematis Niss & Højgaard (2019) ke dalam tiga fase FEI PISA (OECD, 2023a), berdasarkan analisis pada fase mana setiap kompetensi paling dominan berperan sesuai deskripsi operasional masing-masing kompetensi. Hasil pemetaan ini menghasilkan indikator dan sub-indikator literasi matematika yang menjadi acuan penyusunan kisi-kisi soal, sebagaimana disajikan pada Tabel 1.

Tabel 1. Fase dan Indikator Literasi Matematika

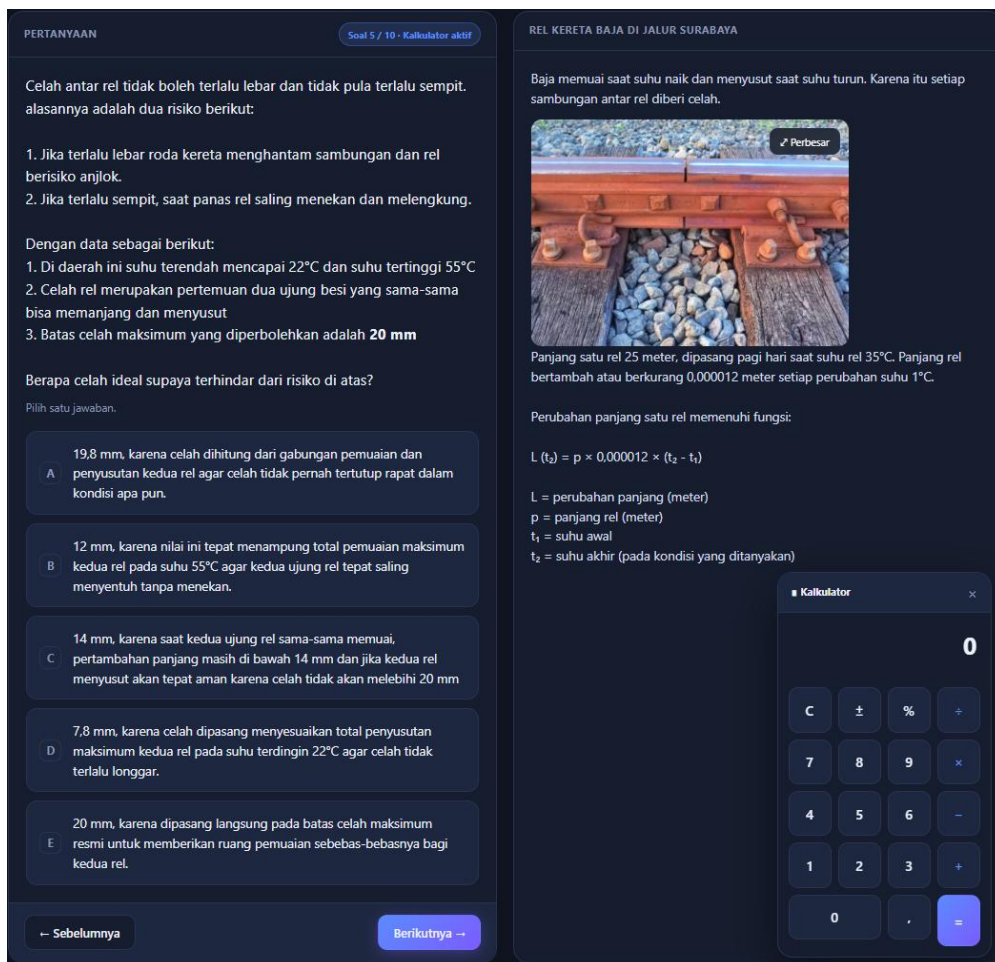
| Fase         | Kode | Indikator                                                                                              |
|--------------|------|--------------------------------------------------------------------------------------------------------|
| Formulating  | F1   | Menyaring dan mengidentifikasi informasi penting dari situasi nyata                                    |
|              | F2   | Mengenali struktur, keteraturan, dan pola dalam situasi                                                |
|              | F3   | Menyederhanakan atau mengurai situasi nyata agar dapat dianalisis secara matematis                     |
|              | F4   | Mengidentifikasi konsep matematis yang relevan dari situasi yang diberikan                             |
|              | F5   | Merepresentasikan situasi menggunakan variabel, simbol, diagram, dan model standar                     |
|              | F6   | Menerjemahkan masalah ke dalam bahasa atau representasi matematis                                      |
|              | F7   | Menyusun model atau struktur matematika untuk menyederhanakan masalah                                  |
|              | F8   | Memilih dan menggunakan alat komputasi yang paling efektif                                             |
| Employing    | E1   | Merancang dan mengimplementasikan strategi untuk menemukan solusi                                      |
|              | E2   | Menerapkan fakta, aturan, algoritma, dan struktur matematis                                            |
|              | E3   | Memanipulasi bilangan, ekspresi aljabar, data statistik, dan representasi geometris                    |
|              | E4   | Menggunakan dan beralih di antara berbagai representasi                                                |
|              | E5   | Membuat generalisasi dan konjektur berdasarkan hasil prosedur                                          |
|              | E6   | Menggunakan simbol, definisi, dan fakta matematika untuk memanipulasi ekspresi matematis               |
|              | E7   | Menerapkan algoritma atau aturan kalkulasi secara presisi untuk menemukan solusi numerik atau simbolik |
|              | E8   | Memanfaatkan alat bantu secara efektif untuk mendukung penyelesaian                                    |
| Interpreting | I1   | Menginterpretasikan hasil matematis kembali ke konteks masalah dunia nyata semula                      |
|              | I2   | Mengevaluasi ketepatan dan kelayakan solusi dalam konteks masalah                                      |
|              | I3   | Menjelaskan mengapa suatu hasil masuk akal atau tidak                                                  |
|              | I4   | Mengkritisi dan mengidentifikasi batas-batas model yang digunakan                                      |
|              | I5   | Menggunakan pemikiran matematis dan komputasional untuk membuat prediksi                               |
|              | I6   | Merumuskan kesimpulan atau keputusan akhir yang logis berdasarkan hasil perhitungan                    |
|              | I7   | Memvalidasi apakah solusi yang ditemukan memenuhi kriteria permasalahan awal                           |

Tahap *Develop* meliputi penulisan butir soal, penyusunan kunci jawaban dengan pedoman penskoran biner (skor 1 untuk jawaban benar dan skor 0 untuk jawaban salah), serta validasi konstruk oleh dua orang ahli menggunakan lembar validasi berskala 1 sampai 4 yang dianalisis dengan koefisien Aiken's V (Aiken, 1985), mengikuti kelaziman validasi ahli pada penelitian pengembangan instrumen berbasis ADDIE (Ahmad et al., 2024; Retnawati, 2016). Hasil validasi konstruk disajikan pada bagian 3.1. Tahap *Implement* meliputi uji coba terbatas instrumen menggunakan moda CBT pada subjek penelitian. Tahap *Evaluate*, yang menjadi

fokus analisis pada artikel ini, meliputi pengujian validitas, reliabilitas, tingkat kesukaran, dan daya pembeda instrumen berdasarkan data hasil uji coba.

## 2.5. Instrumen Penelitian

Instrumen penelitian terdiri atas 9 butir soal literasi matematika dalam bentuk isian singkat dan pilihan ganda, mengadaptasi bentuk penyajian unit PISA serta konten dan konteks kehidupan nyata sesuai kategori PISA 2022 (OECD, 2023a). Setiap butir diskor secara biner, yaitu skor 1 untuk jawaban yang sesuai kunci jawaban dan skor 0 untuk jawaban yang tidak sesuai, tanpa pemberian skor parsial. Instrumen disajikan melalui platform CBT yang dikembangkan sendiri oleh peneliti (tanpa nama komersial khusus), dilengkapi alat hitung digital berupa kalkulator dasar (operasi tambah, kurang, kali, bagi) mengikuti karakteristik editor tool pada CBAM PISA (OECD, 2023a), sehingga siswa dapat melakukan dan menunjukkan proses perhitungan langsung pada layar ujian. Contoh tampilan antarmuka CBT disajikan pada Gambar 1. Naskah lengkap instrumen beserta kunci jawaban disajikan sebagai Lampiran.



Gambar 1. Tampilan Antarmuka CBT dengan Alat Hitung Digital Terintegrasi

## 2.6. Teknik Analisis Data

Data hasil uji coba dianalisis secara kuantitatif. Validitas butir dihitung menggunakan korelasi Pearson product moment antara skor tiap butir dan skor total, dengan kriteria butir

dinyatakan valid apabila nilai signifikansi lebih kecil dari 0,01 (2-tailed). Reliabilitas instrumen dihitung menggunakan koefisien Cronbach's Alpha. Tingkat kesukaran dihitung berdasarkan proporsi jawaban benar terhadap seluruh subjek, dengan kategori sangat sukar, sukar, sedang, dan mudah. Daya pembeda dihitung berdasarkan selisih proporsi jawaban benar antara kelompok atas dan kelompok bawah (masing-masing 18 siswa), dengan kategori buruk, cukup, baik, dan sangat baik. Seluruh analisis kuantitatif dilakukan menggunakan perangkat lunak SPSS. Sebagai data pendukung, dilakukan pula wawancara lisan singkat dengan sebagian subjek pada butir-butir yang tergolong sukar untuk membantu menginterpretasikan kemungkinan penyebab kesulitan tersebut.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Pengembangan Instrumen

Hasil tahap Analyze hingga Implement menghasilkan instrumen asesmen literasi matematika berbasis CBT yang terdiri atas 9 butir, mengadaptasi bentuk penyajian soal PISA, indikator proses FEI, dan moda CBT dengan alat hitung digital yang terintegrasi pada antarmuka ujian (OECD, 2023a). Naskah lengkap instrumen beserta kunci jawaban disajikan sebagai Lampiran.

Tabel 2. Hasil Validasi Konstruk oleh Ahli (Koefisien Aiken's V per Aspek)

| No        | Aspek Validitas Konstruk                                                     | V           | Kategori      |
|-----------|------------------------------------------------------------------------------|-------------|---------------|
| 1         | Kesesuaian butir dengan indikator penerapan konsep fungsi pada situasi nyata | 1,00        | Sangat tinggi |
| 2         | Kesesuaian jenis dan tingkat operasi kognitif dengan tujuan pengukuran       | 0,83        | Tinggi        |
| 3         | Kejelasan rumusan soal sehingga konteks tidak menimbulkan tafsir ganda       | 0,83        | Tinggi        |
| 4         | Kecukupan dan relevansi data pada stimulus terhadap pertanyaan               | 0,83        | Tinggi        |
| 5         | Kesesuaian tingkat kesulitan dengan jenjang peserta didik                    | 0,67        | Cukup         |
| 6         | Homogenitas dan kelogisan pengecoh pada soal pilihan ganda (butir 2)         | 0,83        | Tinggi        |
| 7         | Ketepatan dan kejelasan format jawaban yang diminta                          | 0,83        | Tinggi        |
| 8         | Konsistensi antar-butir dalam mengukur konstruk fungsi secara berjenjang     | 0,83        | Tinggi        |
| 9         | Kebebasan soal dari bias budaya dan ambiguitas bahasa                        | 0,83        | Tinggi        |
| Rata-rata |                                                                              | <b>0,83</b> | <b>Tinggi</b> |

Validasi konstruk pada tahap Develop dilakukan oleh dua orang ahli menggunakan lembar validasi yang mencakup sembilan aspek validitas konstruk serta penilaian relevansi tiap butir terhadap indikator yang diukur, keduanya berskala 1 sampai 4. Validator pertama memberi skor 4 pada aspek kesesuaian butir dengan indikator penerapan konsep fungsi pada situasi nyata, dan skor 3 (cukup sesuai, perlu revisi kecil) pada delapan aspek lainnya. Validator kedua

memberi skor 4 pada delapan aspek dan skor 3 pada satu aspek (kesesuaian tingkat kesulitan dengan jenjang peserta didik). Koefisien Aiken's V (Aiken, 1985) dihitung dari kedua skor tersebut menggunakan rumus  $V = S/[n(c-1)]$ , dengan  $n = 2$  validator dan  $c = 4$  kategori skala; hasil selengkapnya disajikan pada Tabel 2.

Penilaian relevansi tiap butir terhadap indikator yang diukur menghasilkan koefisien Aiken's V butir dengan rata-rata 0,93 (kategori sangat tinggi). Kedua validator sepakat memberi skor 4 (sangat relevan,  $V = 1,00$ ) pada tujuh dari sembilan butir, sedangkan butir 5 dan butir 7 masing-masing memperoleh  $V = 0,67$  karena validator pertama memberi skor 2 (perlu revisi besar) disertai catatan perbaikan tertulis. Kedua validator menyatakan kesimpulan akhir yang sama, yaitu instrumen layak digunakan dengan revisi sesuai saran. Catatan revisi pada butir 5 konsisten dengan hasil uji empiris pada bagian 3.4 yang juga menunjukkan butir tersebut tergolong sukar dengan daya pembeda hanya cukup, sedangkan catatan pada butir 7 tidak sejalan dengan hasil empiris karena butir tersebut justru menunjukkan daya pembeda sangat baik. Rincian catatan revisi per butir dari kedua validator disajikan bersama naskah instrumen pada Lampiran.

### 3.2. Hasil Uji Validitas Butir

Hasil uji validitas menggunakan korelasi Pearson product moment antara skor tiap butir dan skor total pada 65 subjek disajikan pada Tabel 3. Seluruh 9 butir soal dinyatakan valid secara empiris karena nilai signifikansi seluruh butir berada di bawah 0,01.

Tabel 3. Hasil Uji Validitas Butir Soal

| Butir | r hitung | Signifikansi | Keterangan |
|-------|----------|--------------|------------|
| 1     | 0,355    | 0,004        | Valid      |
| 2     | 0,396    | 0,001        | Valid      |
| 3     | 0,736    | 0,000        | Valid      |
| 4     | 0,656    | 0,000        | Valid      |
| 5     | 0,355    | 0,004        | Valid      |
| 6     | 0,592    | 0,000        | Valid      |
| 7     | 0,491    | 0,000        | Valid      |
| 8     | 0,352    | 0,004        | Valid      |
| 9     | 0,556    | 0,000        | Valid      |

Butir 3 dan 4 menunjukkan koefisien validitas tertinggi (0,736 dan 0,656), sedangkan butir 8 menunjukkan koefisien validitas terendah meskipun masih berada pada kategori valid (0,352).

### 3.3. Hasil Uji Reliabilitas

Hasil uji reliabilitas menggunakan Cronbach's Alpha terhadap 9 butir soal pada 65 subjek menghasilkan koefisien sebesar 0,641. Nilai ini menunjukkan instrumen memiliki reliabilitas pada kategori dapat diterima untuk instrumen penelitian pendidikan berskala kecil dengan jumlah butir terbatas, meskipun berada di bawah ambang 0,70 yang umum dijadikan rujukan

untuk instrumen baku. Analisis item-total statistics menunjukkan bahwa nilai Alpha akan meningkat apabila butir 1 dihapus (dari 0,641 menjadi 0,649), mengindikasikan butir tersebut memberikan kontribusi konsistensi internal yang relatif rendah dibandingkan butir lain, sejalan dengan koefisien validitasnya yang juga tergolong lebih rendah.

### 3.4. Hasil Analisis Tingkat Kesukaran dan Daya Pembeda

Hasil analisis tingkat kesukaran dan daya pembeda beserta keputusan akhir terhadap setiap butir disajikan pada Tabel 4.

Tabel 4. Tingkat Kesukaran, Daya Pembeda, dan Keputusan Butir Soal

| Butir | Tingkat Kesukaran | Kategori     | Daya Pembeda | Kategori    | Keputusan     |
|-------|-------------------|--------------|--------------|-------------|---------------|
| 1     | 0,80              | Mudah        | 0,39         | Baik        | Dipertahankan |
| 2     | 0,78              | Mudah        | 0,39         | Baik        | Dipertahankan |
| 3     | 0,46              | Sedang       | 0,94         | Sangat Baik | Dipertahankan |
| 4     | 0,38              | Sedang       | 0,83         | Sangat Baik | Dipertahankan |
| 5     | 0,11              | Sukar        | 0,22         | Cukup       | Direvisi      |
| 6     | 0,18              | Sukar        | 0,50         | Sangat Baik | Dipertahankan |
| 7     | 0,45              | Sedang       | 0,72         | Sangat Baik | Dipertahankan |
| 8     | 0,06              | Sangat Sukar | 0,17         | Buruk       | Direvisi      |
| 9     | 0,37              | Sedang       | 0,61         | Sangat Baik | Dipertahankan |

Tujuh dari sembilan butir (butir 1, 2, 3, 4, 6, 7, dan 9) memenuhi kriteria tingkat kesukaran dan daya pembeda yang memadai sehingga dinyatakan layak dipertahankan tanpa perbaikan. Butir 5 memiliki tingkat kesukaran sukar (0,11) dengan daya pembeda pada kategori cukup (0,22), sehingga dinyatakan perlu direvisi. Butir 8 memiliki tingkat kesukaran sangat sukar (0,06) dengan daya pembeda buruk (0,17), sehingga dinyatakan perlu direvisi total.

### 3.5. Pembahasan

Hasil uji validitas dan reliabilitas pada bagian 3.2 dan 3.3 menunjukkan bahwa instrumen yang dikembangkan memiliki kualitas psikometrik yang memadai untuk mengukur literasi matematika siswa. Verifikasi kualitas butir secara empiris semacam ini sejalan dengan semangat penelitian Desy Rizbudiani et al. (2021), yang juga menempatkan pengujian empiris atas kualitas butir instrumen matematika sebagai syarat kelayakan instrumen, meskipun penelitian tersebut menempuh pendekatan berbeda, yaitu item response theory model Rasch, dibandingkan pendekatan teori tes klasik yang digunakan pada penelitian ini. Koefisien reliabilitas 0,641 pada instrumen ini perlu dibaca dalam konteks jumlah butir yang relatif sedikit (9 butir) dan format butir yang sebagian besar berupa constructed response, format yang secara umum menghasilkan variasi jawaban siswa lebih besar dibandingkan format pilihan ganda sehingga cenderung menekan nilai konsistensi internal. Nilai ini sebanding dengan temuan Priyani & Sugiharto (2024) pada analisis butir soal ujian tengah semester biologi, yang melaporkan koefisien Cronbach's Alpha sebesar 0,619 pada pendekatan teori tes klasik

(kategori cukup) berdampingan dengan reliabilitas model Rasch sebesar 0,85 (kategori baik) untuk data yang sama, menunjukkan bahwa reliabilitas pada kategori cukup dari pendekatan teori tes klasik adalah hal yang lazim dijumpai pada instrumen dengan jumlah butir terbatas.

Temuan pada bagian 3.4 menunjukkan pola yang selaras dengan rancangan tingkat kesulitan butir berjenjang pada tahap Design (bagian 2.4). Butir 5, yang dirancang menuntut penggabungan beberapa langkah perhitungan sebelumnya menjadi satu simpulan, tergolong sukar dengan daya pembeda hanya cukup, sejalan dengan kompleksitas kognitif berjenjang yang memang menjadi ciri khas unit soal bergaya PISA dan konsisten dengan karakteristik unit soal PISA yang dirancang oleh Tanjung et al. (2023), yang juga menyusun satu unit stimulus dengan beberapa butir evaluasi berjenjang pada level kesulitan PISA 3 hingga 5. Adapun butir 8 menunjukkan tingkat kesukaran sangat tinggi disertai daya pembeda buruk, kombinasi yang mengindikasikan bahwa kesulitan pada butir ini kemungkinan besar juga dipengaruhi oleh kelemahan rumusan soal atau instruksi, bukan semata-mata oleh kompleksitas kognitif yang diharapkan, karena daya pembeda yang buruk berarti siswa berkemampuan tinggi maupun rendah sama-sama cenderung gagal menjawab benar. Dugaan ini turut dikuatkan oleh wawancara lisan terhadap seorang siswa berkemampuan tinggi pada salah satu butir sukar, yang semula menyatakan tidak memahami maksud soal namun mampu menjawab dengan benar setelah kalimat soal disederhanakan peneliti, mengindikasikan bahwa sebagian kesulitan bersumber dari kompleksitas kalimat, bukan semata-mata dari kompleksitas matematis yang dituju. Permasalahan ini memang kerap terjadi pada soal cerita, sebagaimana diungkapkan Thevenot dan Barrouillet (2014) dan dikuatkan oleh kajian yang lebih baru bahwa kompleksitas linguistik memengaruhi pemahaman dan performa penyelesaian soal matematika (Cruz Neri & Retelsdorf, 2022; Verschaffel et al., 2020). Temuan ini menunjukkan bahwa validasi ahli pada tahap Develop belum cukup menangkap kesulitan yang baru terlihat secara empiris pada tahap Evaluate, sehingga revisi lanjutan pascauji coba tetap diperlukan.

Instrumen pada penelitian ini menambahkan dimensi moda CBT dengan alat hitung digital terintegrasi sebagaimana karakteristik CBAM PISA. Penggunaan konteks yang dekat dengan keseharian siswa pada instrumen ini juga sejalan dengan temuan Edwar et al. (2023) pada pengembangan soal berpikir tingkat tinggi berkonteks lokal bagi guru matematika, yang menunjukkan bahwa konteks yang familiar bagi siswa membantu proses interpretasi dan penalaran matematis. Dimensi ini relevan mengingat capaian literasi matematika siswa Indonesia yang masih rendah pada PISA 2022 turut dipengaruhi oleh keterbatasan keakraban siswa dengan moda ujian berbasis komputer di sekolah (Wijaya et al., 2024). Dengan demikian, instrumen hasil pengembangan ini berpotensi digunakan tidak hanya untuk mengukur literasi matematika siswa, tetapi juga untuk membiasakan siswa dengan format dan moda ujian yang menyerupai kondisi asesmen internasional.

#### 4. SIMPULAN

Penelitian ini menghasilkan instrumen asesmen literasi matematika berbasis CBT yang mengadaptasi kerangka PISA 2022, indikator FEI, dan alat hitung digital terintegrasi, dikembangkan melalui model ADDIE dan diujicobakan pada 65 siswa SMA kelas XI. Instrumen yang dihasilkan terbukti valid secara empiris pada seluruh 9 butir soal ( $r$  hitung 0,352 sampai 0,736;  $p < 0,01$ ) dengan reliabilitas sebesar 0,641. Berdasarkan analisis tingkat

kesukaran dan daya pembeda, tujuh butir layak dipertahankan, sedangkan dua butir (butir 5 dan butir 8) perlu direvisi, masing-masing pada aspek tingkat kesukaran dan daya pembeda. Temuan ini menunjukkan bahwa instrumen asesmen literasi matematika berbasis CBT beracuan PISA dengan indikator FEI dapat dikembangkan melalui model ADDIE dan menghasilkan kualitas psikometrik yang secara umum memadai, dengan catatan perbaikan pada butir-butir yang menuntut penggabungan hasil perhitungan bertingkat.

## 5. DAFTAR PUSTAKA

- Ahmad, N. A., Mayouf, A. A., Elias, N. F., & Mohamed, H. (2024). Learning management system instrument development based on Aiken's V technique. *International Journal of Evaluation and Research in Education*, 13(5), 3211–3219. <https://doi.org/10.11591/ijere.v13i5.28925>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1). <https://doi.org/10.1177/0013164485451012>
- Charmila, N., Zulkardi, Z., & Darmawijoyo, D. (2016). Pengembangan soal matematika model PISA menggunakan Konteks Jambi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 198–207. <https://doi.org/10.21831/pep.v20i2.7444>
- Cruz Neri, N., & Retelsdorf, J. (2022). The role of linguistic features in science and math comprehension and performance: A systematic review and desiderata for future research. *Educational Research Review*, 36, 100460. <https://doi.org/10.1016/j.edurev.2022.100460>
- Desy Rizbudiani, A., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables(SLETV). In *Jurnal Pendidikan Matematika* (Vol. 12, Number 2). <http://ejournal.radenintan.ac.id/index.php/al-jabar/index>
- Edwar, Putri, R. ilma I., Zulkardi, & Darmawijoyo. (2023). Developing a workshop model for high school mathematics teachers constructing HOTS questions through the Pendidikan Matematika Realistik Indonesia approach. *Journal on Mathematics Education*, 14(4), 603–626. <https://doi.org/10.22342/jme.v14i4.pp603-626>
- Hari Sugiharto Setyaedhi. (2025). PISA Model Numeracy Questions to Improve Student Competence. *JPI (Jurnal Pendidikan Indonesia)*, 14(1), 200–211. <https://doi.org/10.23887/jpiundiksha.v14i1.84401>
- Jannah, R. D., & Putri, R. I. I. (2019). Soft tennis and volleyball contexts in asian games for pisa-like mathematics problems. *Journal on Mathematics Education*, 10(1), 157–169. <https://doi.org/10.22342/jme.10.1.5248.157-170>
- Maskar, S., & Herman, T. (2024). The relation between teacher and students' mathematical mindsets to the student's comprehension of mathematics concepts. In *Journal on Mathematics Education* (Vol. 15, Number 1, pp. 27–54). Sriwijaya University. <https://doi.org/10.22342/jme.v15i1.pp27-54>
- Maxnun, L., Kristiani, K., & Sulistyaningrum, C. D. (2024). Development of hots-based cognitive assessment instruments: ADDIE model. *Journal of Education and Learning*, 18(2), 489–498. <https://doi.org/10.11591/edulearn.v18i2.21079>

- Meryansumayeka, Zulkardi, Putri, R. I. I., & Hiltrimartin, C. (2022). Designing geometrical learning activities assisted with ICT media for supporting students' higher order thinking skills. *Journal on Mathematics Education*, 13(1), 135–148. <https://doi.org/10.22342/jme.v13i1.pp135-148>
- Mirandah, S., & Putri, R. I. I. (2025). Development of PISA-Type Questions in Election Context to Strengthen Students' Mathematical Reasoning Skills. *Mathematics Education Journal*, 19(4), 785–806. <https://doi.org/10.22342/mej.v19i4.pp785-806>
- Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102(1), 9–28. <https://doi.org/10.1007/s10649-019-09903-9>
- Nusantara, D. S., Zulkardi, Z., & Putri, R. I. I. (2021). DESIGNING PISA-LIKE MATHEMATICS TASK USING A COVID-19 CONTEXT (PISACOMAT). *Journal on Mathematics Education*, 12(2), 349–364. <https://doi.org/10.22342/jme.12.2.13181.349-364>
- OECD. (2023a). *PISA 2022 Assessment and Analytical Framework*. OECD. <https://doi.org/10.1787/dfe0bf9c-en>
- OECD. (2023b). PISA 2022 Results The State of Learning and Equity in Education. In *Journal Pendidikan: I. PISA*, OECD Publishing, Paris. <https://doi.org/10.1787/53f23881-en>
- Pratiwi, I., & Putri, R. I. I. (2019). Long jump in asian games: Context of pisa-like mathematics problems. *Journal on Mathematics Education*, 10(1), 81–92. <https://doi.org/10.22342/jme.10.1.5250.81-92>
- Priyani, T., & Sugiharto, B. (2024). Analysis of biology midterm exam items using a comparison of the classical theory test and the Rasch model. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 10(3), 939–958. <https://doi.org/10.22219/jpbi.v10i3.34345>
- Rahmaningtyas, F., Purwati, H., & Nugroho, A. A. (2021). Analisis Kemampuan Koneksi Matematis Ditinjau dari Tipe Kepribadian Sanguinis dan Melancholis Siswa SMP. *Square : Journal of Mathematics and Mathematics Education*, 3(2), 107–116. <https://doi.org/10.21580/square.2021.3.2.8527>
- Retnawati, H. (2016). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). *REID (Research and Evaluation in Education)*, 2(2), 155–164. <https://doi.org/10.21831/reid.v2i2.11029>
- Sistyawati, R. I., & Sitisyarah, K. (2022). The Development of PISA Questions and Activities Using Social Distancing Context During Pandemic. *Mathematics Education Journal*, 16(3), 239–256. <https://doi.org/10.22342/jpm.16.3.17358.239-256>
- Suwarno, M., & Ardani, R. A. (2022). Kemampuan Literasi Matematika Siswa Berdasarkan PISA Level 4. *Square : Journal of Mathematics and Mathematics Education*, 4(2), 107–115. <https://doi.org/10.21580/square.2022.4.2.12401>
- Syifa', A. M., & Rohman, A. A. (2023). Analisis Kemampuan Berpikir Kreatif Guilford Berdasarkan Gaya Berpikir Siswa. *Square : Journal of Mathematics and Mathematics Education*, 5(2), 67–81. <https://doi.org/10.21580/square.2023.5.2.18387>
- Tanjung, A. L., Zulkardi, Z., & Putri, R. I. I. (2023). The development of PISA-like problems using immunity context during pandemic. *Jurnal Elemen*, 9(1), 31–48. <https://doi.org/10.29408/jel.v9i1.6152>

- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM*, 52(1), 1–16. <https://doi.org/10.1007/s11858-020-01130-4>
- Wijaya, T. T., Hidayat, W., Hermita, N., Alim, J. A., & Talib, C. A. (2024). Exploring contributing factors to PISA 2022 mathematics achievement: Insights from Indonesian teachers. *Infinity Journal*, 13(1), 139–156. <https://doi.org/10.22460/infinity.v13i1.p139-156>
- Zulkardi, Z., Meryansumayeka, M., Putri, R. I. I., Alwi, Z., Nusantara, D. S., Ambarita, S. M., Maharani, Y., & Puspitasari, L. (2020). HOW STUDENTS WORK WITH PISA-LIKE MATHEMATICAL TASKS USING COVID-19 CONTEXT. *Journal on Mathematics Education*, 11(3), 405–416. <https://doi.org/10.22342/jme.11.3.12915.405-416>