



## Instrumen Diagnostik SPLDV Berbasis *Cognitive Diagnostic Model*

Nugro Krismanto, Wardani Rahayu, Achmad Ridwan, & Maria Sumunaringtyas

<sup>1</sup> Program Studi Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta, Jakarta, Indonesia

<sup>2</sup> Program Studi Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta, Jakarta, Indonesia

<sup>3</sup> Program Studi Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta, Jakarta, Indonesia

<sup>4</sup> BPK PENABUR Jakarta, Jakarta, Indonesia

\*Correspondence: E-mail: [nugro.krismanto@bpkpenaburjakarta.or.id](mailto:nugro.krismanto@bpkpenaburjakarta.or.id)

### ABSTRACT

The low mathematics achievement of Indonesian students in PISA 2022 highlights the need for assessment instruments that not only measure learning outcomes but also diagnose students' learning difficulties in a detailed manner. This study aimed to develop a diagnostic instrument for Systems of Linear Equations in Two Variables (SPLDV) using a complex multiple-choice format to support data-driven instructional decisions. Content validity was evaluated by five experts using Aiken's V. The instrument was tested through a pilot study involving 200 students and a field test with 960 eighth-grade students from Jakarta, West Java, and Banten. Data analysis employed Classical Test Theory (CTT), Item Response Theory using the Generalized Partial Credit Model (GPCM), and the Cognitive Diagnostic Model based on GDINA. The results indicated that the final instrument consisting of 13 items demonstrated strong content validity (Aiken's  $V = 0.80-1.00$ ) and good internal consistency ( $\alpha = 0.829$ ). GPCM analysis showed that the instrument provided optimal measurement information at low to moderate ability levels, while GDINA analysis confirmed accurate attribute specification, with all items achieving  $PVAF \geq 0.95$ . The novelty of this study lies in the integration of a complex multiple-choice format with polytomous IRT and GDINA-based diagnostic analysis to generate detailed cognitive attribute profiles that are directly applicable for instructional diagnosis. Overall, the instrument is empirically valid, practically informative, and effective for supporting diagnostic assessment and differentiated instruction in SPLDV learning.

### ARTICLE INFO

#### Article History:

Submitted/Received 16 Des 2025

First Revised 2 Jan 2026

Accepted 27 Jan 2026

First Available online 01 Feb 2026

Publication Date 01 Feb 2026

#### Keyword:

*cognitive diagnostic model, GDINA, diagnostic assessment, mathematics literacy, SPLDV*

**ABSTRAK**

Rendahnya capaian literasi matematika siswa Indonesia pada PISA 2022 menunjukkan perlunya instrumen asesmen yang tidak hanya mengukur hasil belajar, tetapi juga mampu memberikan informasi diagnosis kesulitan belajar siswa secara rinci. Penelitian ini bertujuan mengembangkan instrumen diagnostik literasi matematika pada materi Sistem Persamaan Linear Dua Variabel (SPLDV) menggunakan format pilihan ganda kompleks dan Cognitive Diagnostic Model (CDM) berbasis Generalized Deterministic Input, Noisy "And" gate (GDINA). Validitas isi instrumen dievaluasi oleh lima pakar menggunakan koefisien Aiken's V. Uji coba awal melibatkan 200 siswa kelas VIII, sedangkan uji lapangan melibatkan 960 siswa kelas VIII SMP di Jakarta, Jawa Barat, dan Banten. Analisis data dilakukan menggunakan Classical Test Theory (CTT), Item Response Theory dengan Generalized Partial Credit Model (GPCM), serta analisis diagnostik berbasis GDINA. Hasil penelitian menunjukkan bahwa instrumen final yang terdiri atas 13 butir memiliki validitas isi yang kuat (Aiken's V = 0,80–1,00) dan reliabilitas yang baik ( $\alpha = 0,829$ ). Analisis GPCM menunjukkan bahwa instrumen memberikan informasi pengukuran optimal pada tingkat kemampuan rendah hingga sedang, sedangkan analisis GDINA mengonfirmasi ketepatan spesifikasi atribut kognitif dengan seluruh butir mencapai nilai PVAF  $\geq 0,95$ . Kebaruan penelitian ini terletak pada integrasi format pilihan ganda kompleks dengan analisis IRT politomus dan Cognitive Diagnostic Model berbasis GDINA untuk menghasilkan profil penguasaan atribut kognitif siswa yang rinci, bermakna secara pedagogis, dan aplikatif bagi pengambilan keputusan pembelajaran berbasis data. Secara keseluruhan, instrumen yang dikembangkan valid secara empiris, informatif secara diagnostik, dan efektif untuk mendukung asesmen diagnostik serta diferensiasi pembelajaran SPLDV.

© 2025 Teknologi Pendidikan UPI

## 1. PENDAHULUAN

Hasil *Programme for International Student Assessment (PISA) 2022* menunjukkan bahwa capaian literasi matematika siswa Indonesia masih berada di bawah rata-rata *Organisation for Economic Co-operation and Development (OECD)*, yang mencerminkan tantangan berkelanjutan dalam kemampuan siswa menerapkan pengetahuan matematika pada konteks pemecahan masalah dunia nyata (OECD, 2019; OECD, 2023). Literasi matematika tidak hanya menekankan penguasaan prosedural, tetapi juga kemampuan untuk memformulasikan, menerapkan, dan menafsirkan konsep matematika secara bermakna. Temuan ini menegaskan pentingnya praktik asesmen yang mampu menangkap proses kognitif dan pemahaman konseptual siswa, bukan semata-mata mengandalkan pengukuran hasil belajar berbasis skor akhir.

Salah satu peran utama asesmen dalam pendidikan matematika adalah menyediakan informasi diagnostik yang dapat digunakan sebagai dasar pengambilan keputusan pembelajaran. Asesmen konvensional umumnya mengelompokkan siswa berdasarkan skor total, sehingga hanya memberikan gambaran terbatas mengenai kekuatan dan kelemahan kognitif siswa secara spesifik. Akibatnya, guru sering mengalami kesulitan dalam mengidentifikasi hambatan belajar siswa pada tingkat atribut kognitif, khususnya pada materi yang menuntut integrasi konseptual yang kuat seperti Sistem Persamaan Linear Dua Variabel (SPLDV). SPLDV merupakan materi fundamental pada jenjang pendidikan menengah pertama yang menuntut penguasaan berbagai atribut kognitif, termasuk representasi masalah, manipulasi aljabar, dan penalaran logis. Pembelajaran berbasis diagnosis dan diferensiasi sejalan dengan pendekatan *deep learning* dan penguatan asesmen abad ke-21 yang menekankan pemahaman konseptual dan regulasi diri siswa (Fullan et al., 2018; Zimmerman & Schunk, 2021; OECD, 2021).

*Cognitive Diagnostic Models (CDM)* semakin diakui sebagai pendekatan yang efektif dalam asesmen diagnostik karena mampu mengidentifikasi profil penguasaan siswa terhadap berbagai atribut kognitif secara rinci (Rupp et al., 2010; de la Torre, 2011). Di antara berbagai model CDM, *Generalized Deterministic Input, Noisy "And" gate (GDINA)* menawarkan fleksibilitas tinggi dalam memodelkan hubungan yang kompleks antara atribut kognitif dan respons butir. Model GDINA memungkinkan analisis interaksi atribut yang bersifat kompensatoris maupun non-kompensatoris, sehingga sangat sesuai untuk mendiagnosis kompetensi matematika yang kompleks.

Selain pemilihan model, format butir soal juga berperan penting dalam meningkatkan ketepatan diagnosis. Butir pilihan ganda kompleks mampu menyediakan informasi respons yang lebih kaya dibandingkan format dikotomis tradisional karena dapat menangkap pengetahuan parsial dan tingkat pemahaman yang beragam. Ketika dikombinasikan dengan *Item Response Theory (IRT)* politomus, seperti *Generalized Partial Credit Model (GPCM)*, format pilihan ganda kompleks dapat meningkatkan ketelitian pengukuran pada berbagai tingkat kemampuan sekaligus mendukung analisis diagnostik kognitif yang lebih mendalam.

Penelitian mengenai CDM terus berkembang, studi empiris yang mengintegrasikan format pilihan ganda kompleks, analisis IRT politomus, dan asesmen diagnostik berbasis GDINA masih terbatas, khususnya dalam konteks asesmen literasi matematika pada jenjang pendidikan menengah pertama. Sebagian besar penelitian yang ada masih berfokus pada format butir dikotomis atau menghasilkan diagnosis yang bersifat deskriptif, sehingga belum sepenuhnya menggambarkan penguasaan parsial atribut kognitif siswa. Kesenjangan ini menunjukkan perlunya pengembangan instrumen diagnostik yang mampu menghasilkan profil kognitif siswa yang rinci, bermakna secara pedagogis, dan mendukung pembelajaran berdiferensiasi berbasis data.

Tujuan penelitian ini adalah untuk mengembangkan dan memvalidasi instrumen diagnostik literasi matematika pada materi Sistem Persamaan Linear Dua Variabel (SPLDV)

menggunakan format pilihan ganda kompleks dan *Cognitive Diagnostic Model* berbasis GDINA.

## 2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain pengembangan instrumen diagnostik literasi matematika pada materi Sistem Persamaan Linear Dua Variabel (SPLDV). Proses pengembangan dan validasi instrumen mengintegrasikan analisis validitas isi, *Classical Test Theory (CTT)*, *Item Response Theory (IRT)*, dan *Cognitive Diagnostic Model (CDM)*, yang secara luas digunakan dalam pengembangan instrumen diagnostik pembelajaran (Rupp et al., 2010; de la Torre, 2011). Uji coba awal melibatkan 200 siswa kelas VIII SMP, sedangkan uji lapangan melibatkan 760 siswa kelas VIII SMP. Seluruh analisis psikometrik utama dilaporkan berdasarkan data uji lapangan.

Instrumen dikembangkan dalam bentuk 15 butir soal *pilihan ganda kompleks* yang dirancang untuk menangkap pemahaman parsial siswa. Butir soal mengukur empat atribut kognitif SPLDV, yaitu (1) mengidentifikasi SPLDV, (2) memodelkan masalah, (3) menyelesaikan SPLDV, serta (4) mengevaluasi dan menafsirkan solusi. Setiap butir dipetakan ke dalam *Q-matrix* sebagai dasar analisis CDM, sebagaimana direkomendasikan dalam pengembangan instrumen diagnostik berbasis atribut (Rupp et al., 2010). Penskoran menggunakan skala politomus (0–1–2) untuk merepresentasikan tingkat penguasaan siswa.

Validitas isi dievaluasi oleh lima pakar menggunakan koefisien *Aiken's V* untuk menilai kesepakatan ahli terhadap relevansi butir instrumen (Aiken, 1985). Analisis empiris dilakukan secara bertahap, dimulai dengan analisis *CTT* untuk menguji reliabilitas instrumen, kemudian dilanjutkan dengan analisis *IRT* politomus menggunakan *Generalized Partial Credit Model (GPCM)* yang sesuai untuk data respons bertingkat (Muraki, 1992). Analisis diagnostik selanjutnya dilakukan menggunakan *Cognitive Diagnostic Model* berbasis *Generalized Deterministic Input, Noisy "And" gate (GDINA)*. Model *GDINA* dipilih karena memiliki fleksibilitas yang lebih tinggi dibandingkan model *DINA* dan *DINO* dalam memodelkan interaksi antaratribut tanpa asumsi konjungtif atau disjungtif yang ketat, sehingga lebih akurat untuk instrumen dengan struktur atribut yang kompleks (de la Torre, 2011).

## 3. HASIL DAN PEMBAHASAN

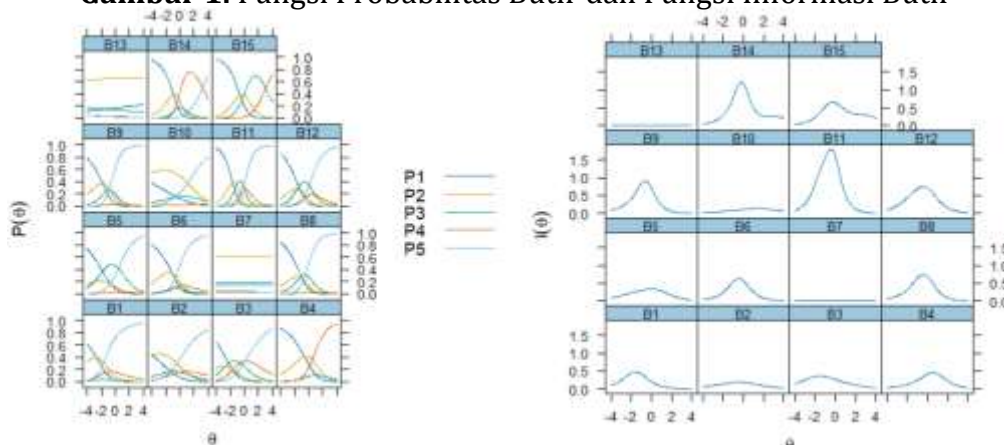
Analisis reliabilitas berdasarkan data uji coba awal pada 200 siswa menunjukkan nilai *Alpha Cronbach* sebesar 0,829, yang mengindikasikan konsistensi internal instrumen yang tinggi. Meskipun instrumen bersifat multidimensi, nilai reliabilitas ini digunakan sebagai indikator awal kelayakan instrumen sebelum dilakukan analisis psikometrik lanjutan berbasis *Item Response Theory (IRT)* dan *Cognitive Diagnostic Model (CDM)*. Temuan ini menunjukkan bahwa butir-butir instrumen secara konsisten mengukur konstruk literasi matematika SPLDV.

Analisis kualitas butir menggunakan *Classical Test Theory (CTT)* menunjukkan bahwa sebagian besar butir berada pada tingkat kesukaran yang optimal dan memiliki daya beda yang memadai. Dua butir, yaitu B7 dan B13, dieliminasi karena memiliki daya beda rendah atau negatif, yang berpotensi menurunkan kestabilan estimasi parameter pada analisis *IRT* dan *CDM*. Seleksi butir berbasis *CTT* menghasilkan 13 butir final yang memenuhi prasyarat metodologis untuk analisis lanjutan secara stabil dan andal.

Perbandingan model *IRT* menunjukkan bahwa *Generalized Partial Credit Model (GPCM)* merupakan model yang paling sesuai untuk karakteristik data respons instrumen. Meskipun *Graded Response Model (GRM)* menghasilkan nilai informasi yang sedikit lebih tinggi, asumsi respons ordinal kumulatif pada model tersebut kurang sejalan dengan karakter penskoran parsial pada butir *pilihan ganda kompleks*. Selain itu, hasil uji *independensi lokal* menunjukkan bahwa hampir seluruh pasangan butir memenuhi asumsi independensi, dengan hanya satu pasangan (0,95%) yang memiliki nilai residual  $Q_3$  di atas 0,30, jauh di bawah ambang toleransi umum sebesar 5%. Temuan ini mengindikasikan bahwa estimasi parameter *IRT* bersifat stabil dan layak digunakan untuk analisis diagnostik lanjutan (Cohen et al., 2018).

Analisis fungsi probabilitas butir dan fungsi informasi butir (Gambar 1) menunjukkan bahwa sebagian besar butir memenuhi asumsi *monotonisitas*, yang ditandai dengan meningkatnya probabilitas kategori skor yang lebih tinggi seiring dengan peningkatan kemampuan laten peserta didik. Pola kurva yang teratur mengindikasikan bahwa butir-butir instrumen mampu merepresentasikan struktur atribut secara konsisten serta memberikan informasi pengukuran yang memadai pada rentang kemampuan yang relevan. Sebaliknya, butir B7 dan B13 menunjukkan pola respons yang tidak stabil dan tidak monotonik, sehingga dieliminasi dari instrumen final karena tidak layak secara diagnostik.

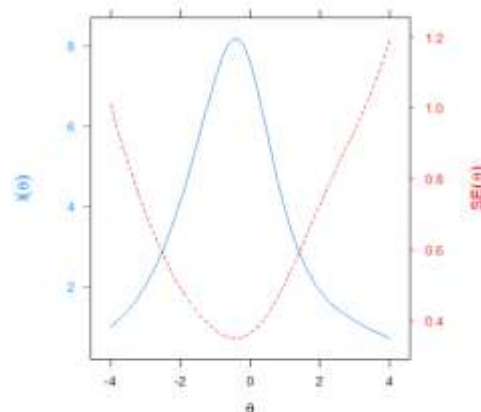
**Gambar 1.** Fungsi Probabilitas Butir dan Fungsi Informasi Butir



Hasil analisis fungsi informasi tes (Gambar 2) menunjukkan bahwa instrumen memberikan informasi pengukuran maksimum pada rentang kemampuan menengah, dengan puncak fungsi informasi berada pada nilai kemampuan laten sekitar  $\theta \approx 0$ . Rentang efektivitas pengukuran instrumen berada antara  $\theta = -1,932$  hingga  $\theta = 1,874$ , yang menunjukkan bahwa instrumen mampu mengukur kemampuan peserta didik dari tingkat rendah hingga tinggi secara memadai. Kesesuaian antara tingkat kesukaran instrumen dan distribusi kemampuan responden mengindikasikan bahwa instrumen memiliki karakteristik pengukuran yang stabil dan presisi pada populasi sasaran.

**Gambar 2.** Fungsi Informasi Tes

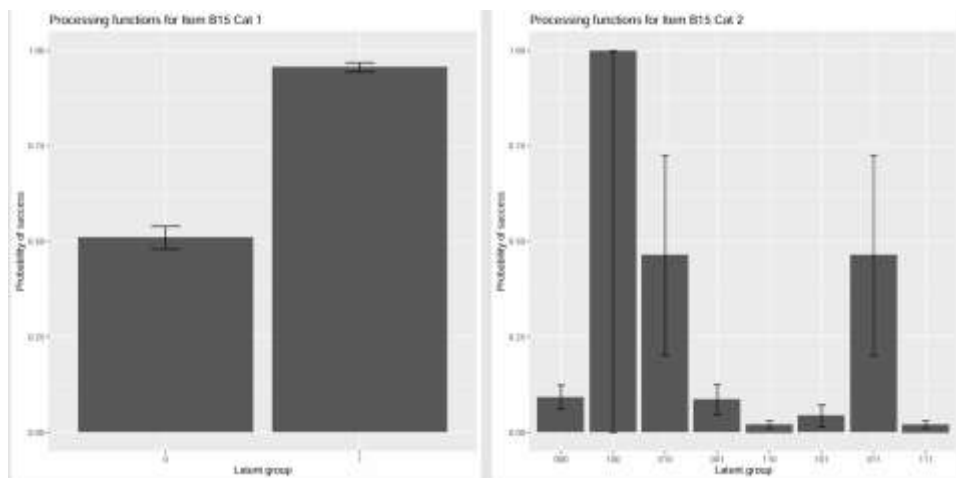
Kemampuan ( $\theta$ ) max	1,874
kemampuan ( $\theta$ ) min	-1,932
Rata-rata kemampuan ( $\theta$ )	0,0365



Rentang efektivitas pengukuran instrumen berada antara  $\theta$  minimum =  $-1,932$  hingga  $\theta$  maksimum =  $1,874$ , yang menunjukkan bahwa instrumen mampu mengukur kemampuan peserta didik dari tingkat rendah hingga tinggi secara memadai. Pada rentang tersebut, fungsi informasi tes relatif tinggi dan galat baku pengukuran relatif rendah, sehingga estimasi kemampuan dapat dianggap stabil dan presisi. Di luar rentang tersebut, informasi tes menurun dan galat baku meningkat, yang menandakan berkurangnya ketepatan pengukuran pada kemampuan yang sangat ekstrem.

Secara keseluruhan, temuan ini menunjukkan bahwa instrumen memiliki karakteristik pengukuran yang optimal pada rentang kemampuan inti populasi, serta layak digunakan untuk tujuan pengukuran dan diagnosis kemampuan peserta didik secara andal.

**Gambar 3.** Fungsi Pemrosesan Respons Butir dalam *Cognitive Diagnostic Model*

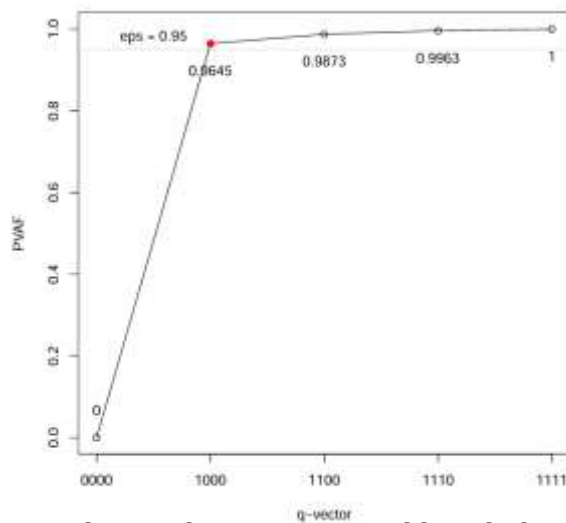


Secara keseluruhan, kumpulan *processing functions* menunjukkan bahwa probabilitas respons peserta didik sangat ditentukan oleh kombinasi penguasaan *cognitive attributes*, bukan semata-mata oleh kemampuan umum. Pada sebagian besar butir, termasuk butir B15, peserta didik dengan *attribute profiles* yang relevan memperlihatkan probabilitas keberhasilan yang lebih tinggi dan stabil pada berbagai kategori skor, sementara profil dengan penguasaan atribut yang parsial atau tidak sesuai menunjukkan probabilitas yang lebih rendah. Pola ini mengindikasikan bahwa butir-butir instrumen memiliki *diagnostic discriminating power* yang baik dalam membedakan struktur penguasaan atribut peserta didik.

Variasi probabilitas antarprofil atribut pada kategori skor yang berbeda juga merefleksikan adanya interaksi atribut yang kompleks, termasuk pola *non-compensatory interaction*, di mana penguasaan satu atribut tidak selalu dapat menggantikan kekurangan atribut lainnya. Hal ini tampak dari beberapa profil dengan penguasaan atribut parsial yang menunjukkan probabilitas respons lebih rendah dibandingkan profil dengan penguasaan atribut penuh, meskipun jumlah atribut yang dikuasai relatif sama. Temuan ini menegaskan bahwa struktur butir tidak bersifat linier dan menuntut pendekatan pemodelan yang mampu menangkap interaksi atribut secara fleksibel.

Secara agregat, pola *processing functions* ini mendukung kesesuaian penggunaan *Cognitive Diagnostic Model* berbasis *GDINA*, karena model tersebut mampu merepresentasikan hubungan nonlinier dan interaksi antaratribut secara akurat. Dengan demikian, instrumen yang dikembangkan tidak hanya andal dalam pengukuran, tetapi juga efektif untuk tujuan diagnosis kognitif dalam konteks pembelajaran berbasis teknologi, khususnya untuk mengidentifikasi kekuatan dan kelemahan atribut peserta didik secara spesifik sebagai dasar perancangan intervensi dan *adaptive learning*.

**Gambar 4.** *Mesa Plot* Butir 5

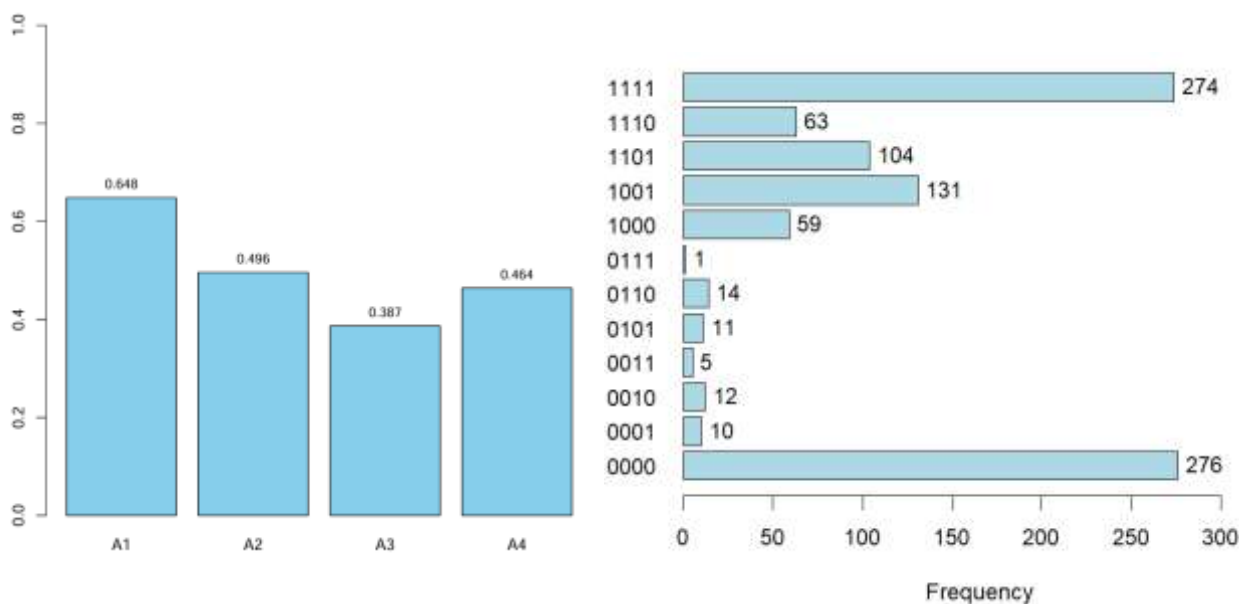


Hasil mesa plot pada Gambar 4 menunjukkan hubungan yang konsisten antara probabilitas respons dan kombinasi penguasaan *cognitive attributes*. Sebagian besar butir memperlihatkan pola nonlinier dan bertingkat, yang mencerminkan adanya *non-compensatory interaction*, di mana penguasaan satu atribut tidak sepenuhnya dapat menggantikan kekurangan atribut lainnya. Variasi pola antarbutir mengindikasikan bahwa setiap butir memiliki karakteristik diagnostik yang unik, sehingga memperkuat kemampuan instrumen dalam mengidentifikasi perbedaan struktur penguasaan atribut siswa secara rinci.

Sebagian besar *mesa plots* memperlihatkan bentuk kurva yang tidak linier dan cenderung bertingkat, yang menandakan adanya interaksi atribut yang kompleks. Pada beberapa butir, profil dengan penguasaan atribut parsial menunjukkan probabilitas respons yang berbeda secara signifikan dibandingkan profil dengan penguasaan atribut penuh, meskipun jumlah atribut yang dikuasai relatif sama. Temuan ini mencerminkan adanya *non-compensatory interaction*, di mana kekurangan pada satu atribut tidak dapat sepenuhnya dikompensasi oleh penguasaan atribut lainnya. Dengan demikian, keberhasilan respons lebih ditentukan oleh kombinasi atribut tertentu daripada akumulasi atribut secara sederhana.

Variasi pola antarbutir pada *mesa plots* juga menunjukkan bahwa setiap butir memiliki karakteristik diagnostik yang unik. Beberapa butir menunjukkan transisi probabilitas yang sangat tajam antarprofil atribut, mengindikasikan *high diagnostic sensitivity*, sementara butir lainnya memperlihatkan transisi yang lebih gradual, yang mencerminkan tuntutan kognitif yang lebih berlapis. Keberagaman pola ini memperkuat fungsi instrumen dalam mengidentifikasi perbedaan struktur penguasaan atribut peserta didik secara rinci.

Secara agregat, keseluruhan pola *mesa plots* mendukung kesesuaian penggunaan *Cognitive Diagnostic Model* berbasis *GDINA*, karena model ini mampu merepresentasikan hubungan nonlinier dan interaksi antaratribut secara fleksibel. Dengan demikian, instrumen yang dikembangkan tidak hanya valid untuk tujuan pengukuran, tetapi juga efektif dalam konteks *educational technology* untuk mendukung *diagnostic assessment*, *personalized learning*, dan perancangan intervensi pembelajaran berbasis profil kognitif peserta didik.

**Gambar 5.** Rata-Rata Profil Penguasaan Atribut dan Distribusi Pola Atribut

Berdasarkan Gambar 5, rata-rata probabilitas penguasaan atribut menunjukkan pola  $A1 > A2 > A3$ , dengan A3 (0,387) sebagai titik kesulitan utama pada tahap penyelesaian prosedural SPLDV. Nilai A4 (0,464) yang lebih tinggi daripada A3 mengindikasikan bahwa sebagian siswa mampu mengevaluasi dan menafsirkan solusi melalui penalaran kontekstual meskipun penguasaan prosedural belum optimal. Pola ini sejalan dengan karakteristik literasi matematika pada tahap *interpret* dalam kerangka PISA serta sifat estimasi *GDINA* yang tidak memaksakan hierarki atribut secara ketat (OECD, 2019; de la Torre, 2011).

Distribusi pola atribut menunjukkan polarisasi kompetensi yang kuat, dengan proporsi terbesar siswa berada pada pola 0000 dan 1111, yang menandakan kesenjangan kemampuan yang jelas. Dominannya pola parsial 1001 dan 1101 mengindikasikan bahwa sebagian siswa mampu mengidentifikasi masalah dan mengevaluasi solusi meskipun penguasaan prosedural SPLDV belum optimal. Pola ini konsisten dengan karakteristik literasi matematika PISA, di mana kemampuan *interpret* dapat muncul tanpa penguasaan penuh pada tahap *employ*, serta mencerminkan estimasi atribut non-hierarkis dalam Cognitive Diagnostic Model berbasis *GDINA* (OECD, 2019; de la Torre, 2011).

Analisis *Cognitive Diagnostic Model* berbasis *GDINA* mampu mengungkap variasi dan polarisasi penguasaan atribut kognitif siswa, termasuk dominasi pola ekstrem dan kemunculan pola parsial yang mencerminkan strategi kognitif yang beragam dalam penyelesaian SPLDV (de la Torre, 2011; Maradani & Wilson, 2021). Keakuratan spesifikasi Q-matrix yang dikonfirmasi melalui mesa plot ( $PVAF \geq 0,95$ ) menegaskan tidak adanya misspecification baik pada butir satu atribut maupun multiatribut, sehingga memperkuat validitas diagnostik instrumen (de la Torre, 2011; Rizki & Retnawati, 2020). Dengan demikian, instrumen ini tidak hanya berfungsi sebagai alat ukur capaian belajar, tetapi juga sebagai alat diagnosis pembelajaran yang efektif untuk memetakan kesulitan dan strategi kognitif siswa, serta mendukung perencanaan intervensi dan diferensiasi pembelajaran SPLDV yang terarah, adaptif, dan berbasis data.

#### 4. SIMPULAN

Berdasarkan hasil analisis validitas dan kualitas pengukuran, instrumen diagnostik SPLDV disusun dalam bentuk final yang terdiri atas 13 butir terverifikasi. Seluruh butir memenuhi kriteria validitas isi (*Aiken's V*  $\geq 0,80$ ) dan menunjukkan konsistensi internal yang baik dengan koefisien reliabilitas  $\alpha = 0,829$  (Nunnally & Bernstein, 1994; Yusrizal et al., 2020). Analisis *Classical Test Theory (CTT)* menunjukkan bahwa sebagian besar butir memiliki tingkat kesukaran moderat dan daya beda  $\geq 0,30$ , sehingga layak digunakan untuk mengukur variasi kemampuan siswa (Allen & Yen, 2002). Dua butir (B7 dan B13) dieliminasi karena memiliki daya beda rendah, pola respons yang tidak konsisten, serta informasi butir yang sangat rendah pada analisis *Item Response Theory (IRT)*, yang berpotensi mengganggu kestabilan estimasi parameter (Cohen et al., 2018).

Analisis *IRT* menegaskan bahwa *Generalized Partial Credit Model (GPCM)* merupakan model yang paling sesuai untuk karakteristik butir politomus pada instrumen ini, dengan kecocokan model yang lebih baik dibandingkan *Partial Credit Model (PCM)* dan *Graded Response Model (GRM)* (Muraki, 1992). Ketiga belas butir final menunjukkan kontribusi informasi yang stabil pada rentang kemampuan rendah hingga sedang, sehingga efektif untuk tujuan diagnosis pembelajaran dan identifikasi kesulitan belajar siswa (Cohen et al., 2018).

Analisis *Cognitive Diagnostic Model (CDM)* berbasis *Generalized Deterministic Input, Noisy "And" gate (GDINA)* menghasilkan distribusi pola atribut yang terpolarisasi, dengan dominasi siswa pada pola 0000 dan 1111, serta kemunculan pola parsial seperti 1001 dan 1101, yang mengindikasikan adanya variasi strategi kognitif siswa dalam menyelesaikan SPLDV. Pola ini mencerminkan estimasi atribut yang bersifat non-hierarkis dan diagnostik, sesuai dengan karakteristik model *GDINA* (de la Torre, 2011; Maradani & Wilson, 2021).

Hasil *mesa plot* pada seluruh butir menunjukkan bahwa nilai *Proportion of Variance Accounted For (PVAF)* melampaui ambang  $\varepsilon = 0,95$  pada kombinasi atribut yang ditetapkan dalam *Q-matrix*, dengan pola kenaikan tajam yang diikuti oleh *plateau*. Temuan ini mengonfirmasi bahwa spesifikasi *Q-matrix* telah akurat, baik untuk butir satu atribut maupun multiatribut, serta tidak ditemukan indikasi *misspecification*, yang merupakan prasyarat penting dalam analisis diagnostik berbasis *CDM* (de la Torre, 2011; Rizki & Retnawati, 2020).

Secara keseluruhan, hasil penelitian ini menegaskan bahwa instrumen diagnostik SPLDV yang dikembangkan valid secara teoretis, kuat secara empiris, dan efektif dalam memetakan profil penguasaan atribut kognitif siswa. Dengan demikian, instrumen ini dapat dimanfaatkan untuk mendukung intervensi pembelajaran dan diferensiasi pembelajaran yang terarah, adaptif, dan berbasis data.

#### 5. PERNYATAAN PENULIS

Penulis menyatakan bahwa tidak terdapat konflik kepentingan terkait penerbitan artikel ini. Penulis menegaskan bahwa naskah artikel bebas dari plagiarisme.

#### 6. REFERENSI

- Amelia, A. (2016). Pengembangan instrumen literasi numerasi berbasis *higher order thinking skills* pada materi aritmatika sosial. *Jurnal Riset Pendidikan Matematika*, 3(2), 112–125. <https://journal.unirow.ac.id>
- Anjarrani, A., & Kurniasih, R. (2023). Analisis kualitas soal numerasi dalam meningkatkan *higher order thinking skills* siswa. *Jurnal Pendidikan Matematika*, 10(2), 123–135.
- Anwar, F., & Kartika, R. (2023). Kemampuan literasi matematis calon guru dalam pengajuan masalah berorientasi data. *Jurnal Matematika dan Pendidikan Matematika*, 11(1), 65–78. <https://ojs.fkip.ummetro.ac.id/index.php/matematika/article/view/6463>

- Aulia, M., & Umi, M. (2019). Pengembangan soal matematika model *PISA* level 5 untuk program pengayaan SMP. *Jurnal Pendidikan Matematika*, 6(2), 123–135. <https://www.researchgate.net/publication/335496256>
- Auliya, L. (2022). *Pengembangan instrumen tes literasi matematis berbasis budaya Betawi untuk meningkatkan pemahaman siswa dalam konteks budaya lokal* [Skripsi, Universitas Islam Negeri Syarif Hidayatullah Jakarta]. <https://repository.uinjkt.ac.id>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- de la Torre, J. (2011). The generalized *DINA* model framework. *Psychometrika*, 76(2), 179–199. <https://doi.org/10.1007/s11336-011-9207-7>
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.
- Fatimah, S., & Handayani, D. (2020). Pengembangan soal matematika model *PISA* menggunakan konteks Kalimantan Timur. *Jurnal Pendidikan Matematika*, 7(1), 68–77. <https://www.researchgate.net/publication/339938499>
- Fullan, M., Quinn, J., & McEachen, J. (2018). *Deep learning: Engage the world, change the world*. Corwin Press.
- Kemendikbudristek. (2022). *Panduan pelaksanaan Asesmen Nasional*. Pusat Asesmen Pendidikan.
- Maradani, R., & Wilson, M. (2021). Development of diagnostic instruments in education: Applying the construct modeling approach. *Journal of Educational Measurement and Evaluation*, 15(2), 55–70.
- McCrinkle, M., & Fell, J. (2020). *Understanding Generation Alpha: Educating and engaging the youngest generation*. McCrinkle Research.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176. <https://doi.org/10.1177/014662169201600206>
- OECD. (2019a). *PISA 2018 assessment and analytical framework: Mathematics, reading, science and global competence*. OECD Publishing.
- OECD. (2019b). *PISA 2018 results: What students know and can do*. OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- OECD. (2021). *The future of education and skills: Assessment in the 21st century*. OECD Publishing.
- OECD. (2023). *Mathematics literacy and PISA performance: Trends and challenges in digital learning*. OECD Publishing.
- Retnawati, H. (2016). *Validitas dan reliabilitas instrumen penelitian*. Parama Publishing.
- Retnawati, H., Hadi, S., & Nugraha, A. (2016). Difficulties in the implementation of *higher order thinking skills* in mathematics learning curriculum. *Journal of Physics: Conference Series*, 755, 012047. <https://doi.org/10.1088/1742-6596/755/1/012047>
- Rizki, S., & Retnawati, H. (2020). Developing diagnostic test instruments in mathematics education: A systematic approach. *Jurnal Riset Pendidikan Matematika*, 7(2), 145–160.
- Yusrizal, B., Susanti, T., & Astuti, R. (2020). Validasi instrumen dengan koefisien *Aiken's V*. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, 9(3), 221–230.
- Zimmerman, B. J., & Schunk, D. H. (2021). *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.