



## **Penerapan Teori Respon Butir dalam Pengembangan Instrumen Literasi Matematis Elemen Bilangan Kelas X**

**Endah Susilowati<sup>1</sup>, Sintha Sih Dewanti<sup>2</sup>**

<sup>1,2</sup>Pendidikan Matematika, UIN Sunan Kalijaga. Jl. Laksda Adisucipto, Papringan, Caturtunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta  
e-mail: endasusilo24@gmail.com<sup>1</sup>, sintha.dewanti@uin-suka.ac.id<sup>2</sup>

### **ABSTRAK**

Kebutuhan alat ukur literasi matematis elemen bilangan sangat penting guna mendesain pembelajaran yang tepat, namun instrumen yang tersedia belum terstandar. Penelitian ini bertujuan mengembangkan instrumen tes literasi matematis elemen bilangan kelas X yang dianalisis menggunakan teori respon butir model PCM. Penelitian ini menggunakan metode *Research and Development (R&D)* dengan prosedur pengembangan tes: *overall plan, content definition, test specifications, item development, tes design and assembly, test production, test administration, scoring test responses, cut scores, test score report*. Hasil penelitian menunjukkan delapan butir tes terbukti valid dan reliabel. Seluruh butir valid secara isi dengan nilai Aiken  $> 0,80$ . Delapan butir tes valid secara konstruk dengan nilai  $SLF \geq 0,45$ , reliabel secara konstruk dengan  $CR = 0,75$ . Tingkat kesulitan butir berada pada kategori baik dengan nilai  $b$  berada pada interval  $-2 < b < 2$ . Terdapat 2 butir dengan plot ICC ideal dan 6 butir dengan plot ICC tidak ideal. Fungsi informasi menunjukkan bahwa instrumen tes memberikan informasi paling baik dan akurat untuk responden dengan tingkat kemampuan sedikit di atas rata-rata. Analisis *item fit* menunjukkan terdapat tujuh butir yang *fit* dan satu butir yang tidak *fit*. Analisis *person fit* menunjukkan 82 dari 195 responden terdeteksi tidak cocok atau menyimpang dari pola respon jawaban.

**Kata Kunci:** bilangan, *Item Response Theory (IRT)*, literasi matematis.

### **ABSTRACT**

*The urgency of mathematical literacy measurement tools for number elements is crucial for designing appropriate learning, but the available instruments are not yet standardized. This study aims to develop a mathematical literacy test instrument for number elements for 10th grade students, analyzed using the PCM item response theory model. The research used the Research and Development (R&D) method with test development procedures: overall plan, content definition, test specifications, item development, test design and assembly, test production, test administration, scoring test responses, cut scores, and test score reports. All items were content-valid with Aiken value  $>0.80$ . The eight items were construct-valid with  $SLF \geq 0.45$  and construct-reliable with  $CR$  of 0.75. The items difficulty is in the good category with  $b$  value in the interval  $-2 < b < 2$ . There were 2 items with ideal ICC plots and six items with non-ideal ICC plots. The information function shows that the test instrument provides the best and most accurate information for respondents with abilities slightly above average. Item fit analysis shows that there are seven fit items and one unfit item. Person fit analysis shows that 82 of the 195 respondents were detected as not fitting or deviating from the response pattern.*

**Keywords:** quantity, *Item Response Theory (IRT)*, mathematical literacy.

## PENDAHULUAN

Literasi matematis menurut *Organization for Economic Cooperation and Development* (OECD) ialah kemampuan seseorang untuk merumuskan, mengaplikasikan, serta menginterpretasikan matematika dalam berbagai konteks (OECD, 2023a). Menurut Stacey (2015), literasi matematis digunakan dalam pemecahan masalah dalam kehidupan sehari-hari sehingga lebih siap menghadapi tantangan kehidupan. Kemampuan literasi matematis sangat bermanfaat bagi peserta didik, di antaranya dapat mencerna masalah dengan benar, mengembangkan ide dan konsep matematis, dan memecahkan masalah secara sistematis (Kurniawan & Djidu, 2021). Oleh karena itu, penting bagi peserta didik untuk memiliki kemampuan literasi matematis.

Literasi matematis merupakan salah satu kemampuan yang diukur pada skala internasional melalui *Programme for International Student Assessment* (PISA). Program ini diselenggarakan tiga tahun sekali oleh OECD untuk peserta didik usia 15 tahun. Penilaian literasi matematis PISA bertujuan mengukur sejauh mana suatu negara mempersiapkan peserta didiknya untuk memanfaatkan matematika dalam berbagai aspek kehidupan (OECD, 2023a). Dalam kehidupan yang semakin berkembang, yang terpenting bukan hanya pengetahuan matematika tapi bagaimana cara memanfaatkan matematika dalam kehidupan.

PISA menggunakan 3 domain utama untuk mengukur literasi matematis yaitu domain proses, konten, dan konteks (OECD, 2023a). Domain proses literasi matematis meliputi *formulate*, *employ*, *interpret*, dan *reasoning*. Domain konteks meliputi konteks pribadi, konteks pekerjaan, konteks sosial, dan konteks ilmiah. Domain konten meliputi *space and shape* (ruang dan bentuk), *change and relationship* (perubahan dan hubungan), *quantity* (bilangan), dan *uncertainty and data* (probabilitas/ ketidakpastian data).

*Quantity* atau bilangan merupakan salah satu konten dalam literasi matematis. Di Indonesia, konten *quantity* berkaitan dengan elemen bilangan dalam kurikulum merdeka. Konten bilangan merupakan bagian yang paling luas dan sangat penting dalam matematika. Hal ini dikarenakan bilangan merupakan dasar perhitungan matematika yang berguna untuk menyelesaikan masalah matematis (Nurjanah & Hakim, 2019). Konten bilangan berkaitan dengan kemampuan memahami tentang pengukuran, satuan, pola bilangan, dan segala sesuatu yang berkaitan dengan bilangan dalam kehidupan (Putra & Vebrian, 2019). Oleh karena itu penting untuk menguasai bilangan sebelum melanjutkan materi yang lebih kompleks.

Pentingnya literasi matematis, khususnya pada konten bilangan ternyata belum diikuti dengan prestasi peserta didik Indonesia. Berdasarkan hasil PISA 2022 skor literasi matematis Indonesia konten bilangan sebesar 363, yang mana masih jauh dari rata-rata OECD sebesar 472 (OECD, 2023b). Hal ini sejalan dengan hasil penelitian yang dilakukan Purwanto dan Jailani (2024), juga menunjukkan bahwa literasi matematis pada konten bilangan peserta didik masih rendah, di mana tidak ada separuh peserta didik yang mampu menjawab dengan sempurna.

Penelitian [Takaria et al. \(2022\)](#) menyebutkan penyebab rendahnya literasi matematis dikarenakan peserta didik belum terbiasa mengerjakan tes literasi matematis. Pendidik kesulitan menyusun instrumen yang sesuai dengan indikator literasi matematis [Wibowo et al. \(2020\)](#). Instrumen yang tersedia kurang mengaitkan dengan permasalahan sehari-hari sehingga kurang mendukung kemampuan literasi matematis peserta didik ([Farida et al., 2021](#)). Selain itu, terdapat serangkaian prosedur untuk dapat menghasilkan suatu instrumen yang baik, akurat, dan sistematis.

Salah satu prosedur pengembangan tes yaitu analisis butir tes. Secara umum, analisis karakteristik tes selama ini masih menggunakan *Classical Test Theory* (CTT) yang memiliki beberapa kekurangan. Menurut Gulliksen, CTT memiliki beberapa kekurangan, yaitu karakteristik butir tergantung pada karakteristik peserta didik atau bersifat group dependent ([Azwar, 2019](#)). Oleh karena itu, IRT dikembangkan untuk mengatasi kelemahan CTT. Teori respons butir atau *Item Response Theory* (IRT) secara spesifik memberikan respons butir individu berdasarkan pada karakteristik individu dan karakteristik butir ([Dewanti, 2021](#)). Dengan demikian, hasil IRT akan lebih akurat dibandingkan dengan hasil CTT yang masih bergantung pada sampel.

IRT adalah kerangka pengukuran yang digunakan dalam analisis penilaian pendidikan dan psikologis seperti ujian prestasi, skala penilaian, atau instrumen lain yang mengukur ciri-ciri sifat laten yang diamati ([Ogunsakin & Shogbesan, 2018](#)). [Sudaryono \(2012\)](#) menyebutkan bahwa IRT terdiri atas 3 model logistik yaitu model 1 Paramater Logistik (1PL), model 2 Paramater Logistik (2PL), dan model 3 Paramater Logistik (3PL). Model 1PL hanya mengukur tingkat kesulitan ( $b$ ). Model 2PL mengukur 2 parameter yaitu tingkat kesulitan ( $b$ ) dan daya beda ( $a$ ). Model 3PL mengukur tingkat kesulitan ( $b$ ), daya beda ( $a$ ), dan *pseudo-guessing* ( $c$ ).

IRT memiliki tiga asumsi utama yang perlu dipenuhi agar analisis valid yaitu uji unidimensi, independensi lokal, dan invariansi parameter ([Hambleton et al., 1991](#)). Asumsi unidimensi terpenuhi apabila tes hanya mengukur satu kemampuan utama ([Sudaryono, 2012](#)). Independensi lokal mengacu pada independensi respon terhadap butir. Asumsi ini menunjukkan bahwa jawaban responden pada satu butir tes dengan butir tes yang lain tidak saling bergantung. Asumsi terakhir, invariansi parameter diartikan bahwa karakteristik butir tes independen terhadap parameter kemampuan peserta tes dan parameter kemampuan peserta tes independen terhadap butir tes ([Retnawati, 2014](#)).

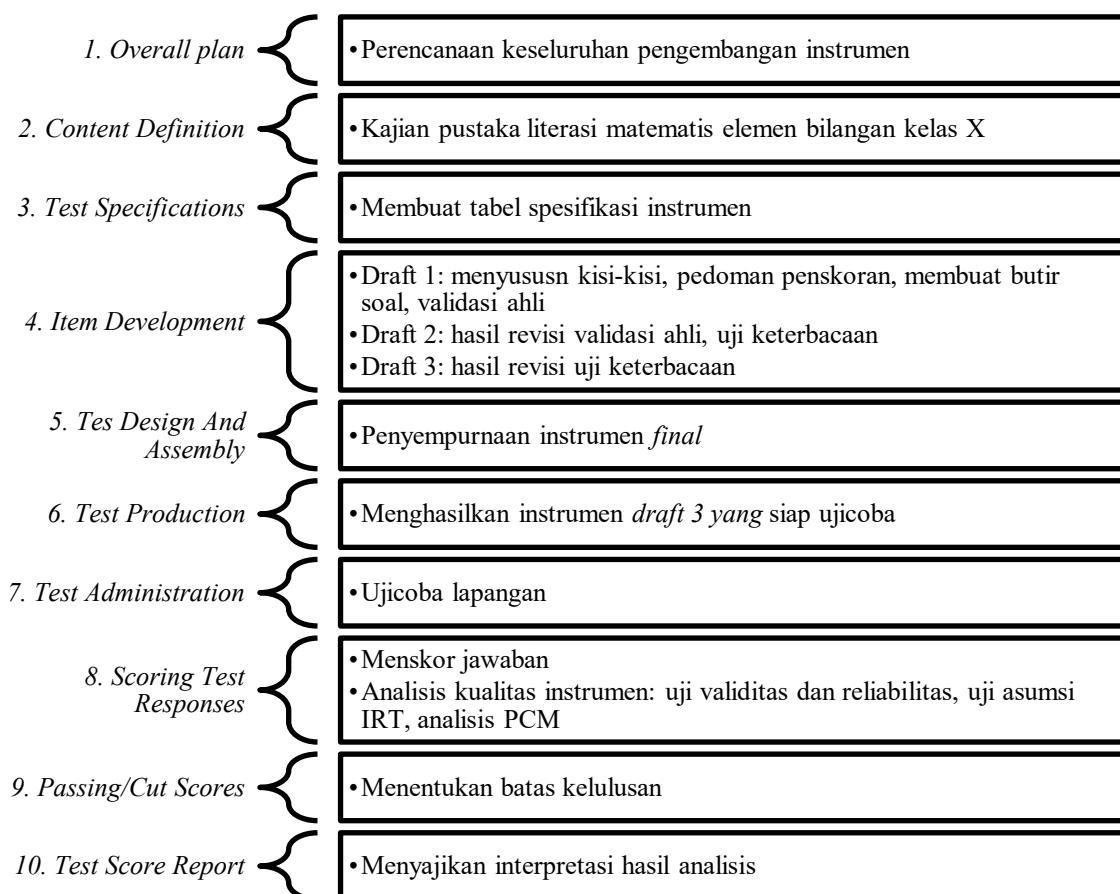
IRT dapat mengestimasi kemampuan menggunakan penskoran dikotomus dan politomus. Penskoran dikotomus biasanya digunakan pada soal pilihan ganda dan penskoran pada tes uraian biasanya menggunakan model penskoran politomi karena skor butir lebih dari dua kategori ([Purnama, 2023](#)). Penelitian ini menggunakan model *Partial Credit Model* (PCM) untuk analisis butir literasi matematis. Hal ini dikarenakan PCM cocok digunakan pada tes prestasi soal perhitungan yang mengharuskan tahap identifikasi permasalahan hingga solusi akhir ([Widhiarso, 2010](#)).

PCM merupakan pengembangan model rasch dikotomus menjadi politomus dengan satu parameter logistik yaitu tingkat kesulitan dan daya beda diasumsikan sama (Istiyono et al., 2014). Tingkat kesulitan butir ( $b$ ) merepresentasikan tingkat kemampuan minimum yang diperlukan untuk dapat menjawab butir dengan benar (Baker, 2001). Rentang tingkat kesulitan biasanya adalah (-2, +2) (Hambleton et al., 1991). Parameter daya beda ( $a$ ) mengukur sejauh mana butir mampu mengidentifikasi perbedaan responden berdasarkan tingkat kemampuannya (tinggi atau rendah), umumnya berada pada interval (0, 2) (Hambleton et al., 1991). Parameter Tingkat kesulitan juga dapat dilihat melalui kurva karakteristik butir soal (*Item Characteristic Curve*). Kurva ini mewakili karakteristik suatu soal yang menunjukkan hubungan kemungkinan seorang peserta tes dengan kemampuan tertentu ( $\theta$ ) dapat menjawab dengan benar (Retnawati, 2014).

Pengembangan instrumen literasi matematis yang dilakukan oleh Azizah et al. (2024), Auliya (2021), dan Wibowo et al. (2020) menunjukkan bahwa analisis instrumen selama ini masih didominasi menggunakan CTT yang bersifat *group dependent*, sehingga kurang mampu menggambarkan kemampuan peserta didik yang sesungguhnya. Instrumen yang dianalisis menggunakan IRT dapat memberikan informasi karakteristik butir dan kemampuan individu secara lebih akurat, sehingga data yang didapatkan benar-benar mencerminkan kemampuan sesungguhnya dan dapat dipertanggungjawabkan dalam pengambilan keputusan. Oleh karena itu, pengembangan instrumen literasi matematis elemen bilangan berbasis IRT perlu dilakukan sehingga dapat memperoleh hasil yang lebih akurat. Hasil pengukuran diharapkan dapat membantu pendidik untuk mendesain pembelajaran yang tepat sehingga dapat meningkatkan literasi matematis peserta didik.

## METODE

Penelitian ini menggunakan metode *Research & Development* (R&D). Subjek uji coba penelitian ini yaitu 195 peserta didik kelas X di salah satu sekolah di Yogyakarta tahun ajaran 2024/2025. Kelas X dipilih sebagai subjek penelitian karena sesuai dengan sasaran produk yang telah dikembangkan. Tahapan penelitian menggunakan prosedur pengembangan tes oleh Downing dan Haladyna (2006) yang terdiri dari: 1) *Overall plan*, 2) *content definition*, 3) *test specifications*, 4) *item development*, 5) *tes design and assembly*, 6) *test production*, 7) *test administration*, 8) *scoring test responses*, 9) *passing/cut scores*, 10) *test score report*. Gambar 1 merupakan rincian tahapan pengembangan instrumen.

**Gambar 1.** Tahapan Pengembangan Instrumen

Teknik pengumpulan data meliputi validasi ahli, uji keterbacaan, serta pemberian tes literasi matematis. Pengumpulan data pada penelitian ini menggunakan instrumen yang terdiri dari: 1) lembar validasi, 2) lembar uji keterbacaan, dan 3) instrumen tes literasi matematis elemen bilangan kelas X. Data yang didapat dari instrumen tersebut digunakan untuk menganalisis kualitas butir tes yang telah dikembangkan yang terdiri dari validitas isi, validitas konstruk, estimasi reliabilitas, karakteristik butir tes, dan mengukur kemampuan peserta didik.

Teknik analisis data terdiri dari analisis validitas isi, analisis validitas konstruk, reliabilitas konstruk, serta karakteristik butir dianalisis menggunakan analisis teori respons butir model PCM. Validitas isi dihitung menggunakan indeks Aiken dengan bantuan *Microsoft Excel*. Nilai V minimum untuk instrumen yang baik ditentukan berdasarkan tabel Aiken sesuai dengan jumlah rater dan jumlah kategori yang digunakan.

Selanjutnya, analisis validitas konstruk menggunakan Faktor Konfirmatori atau *Confirmatory Factor Analysis* (CFA) berbantuan software lisrel 8.80. Sebelum uji CFA, terlebih dahulu dilakukan uji kelayakan sampel *Kaiser Meyer Olkin* (KMO) dan uji Bartlett's menggunakan software SPSS 25. Setelah semua uji asumsi terpenuhi, maka dapat dilanjutkan dengan uji CFA untuk mengkonfirmasi apakah model pengukuran yang disusun telah sesuai dengan teori yang ada (Retnawati, 2016). Analisis CFA dilakukan untuk mencapai unidimensionalitas, reliabilitas, validitas, serta indeks kecocokan yang dibutuhkan. Reliabilitas konstruk dapat diketahui

menggunakan CR atau *Construct Reliability*. CR dikatakan baik apabila memenuhi  $CR \geq 0.7$  (Hair et al., 2019).

Sebelum dilakukan analisis IRT, terlebih dahulu dilakukan uji asumsi yang terdiri dari uji unidimensi, independensi lokal, dan invariansi parameter (Hambleton et al., 1991). Selanjutnya yaitu analisis butir item menggunakan model PCM berbantuan software *RStudio* 4.4.0. Beberapa hal yang dianalisis menggunakan PCM pada penelitian ini yaitu 1) tingkat kesulitan butir tes; 2) kurva karakteristik butir atau *Item Characteristic Curve* (ICC); 3) fungsi informasi dan SEM; dan 4) *item fit* dan *person fit*.

## **HASIL DAN PEMBAHASAN**

Penelitian pengembangan ini bertujuan untuk mengembangkan suatu instrumen literasi matematis elemen bilangan kelas X yang memiliki kualitas serta karakteristik butir yang baik. Kualitas dan karakteristik butir instrumen literasi matematis dianalisis menggunakan teori respons butir. Adapun tahapan penelitian pengembangan ini sebagai berikut.

### ***Overall Plan***

Pada tahap ini dilakukan perencanaan secara keseluruhan pengembangan instrumen literasi matematis. Pertama, dilakukan kajian mengenai definisi beserta indikator-indikator literasi matematis serta materi elemen bilangan kelas X. Selanjutnya yaitu pembuatan kisi-kisi instrumen, alternatif penyelesaian, pedoman penskoran, dan *blueprint* instrumen. Setelah semua perencanaan tersebut selesai, dilakukan uji validasi oleh ahli. Instrumen yang sudah valid kemudian diujicobakan kepada peserta didik kelas X. Data yang didapatkan kemudian digunakan untuk analisis kualitas instrumen yang meliputi validitas konstruk, reliabilitas konstruk serta analisis karakteristik butir menggunakan teori respon butir. Model analisis yang digunakan yaitu PCM.

### ***Content Definition***

Pada tahap ini dilakukan kajian mengenai domain kognitif literasi matematis dan materi elemen bilangan kelas X. Penentuan indikator literasi matematis mengacu pada *framework PISA* 2022 yang diterbitkan oleh OECD dan juga dari sumber lain misalnya, jurnal ilmiah dan buku. Untuk penentuan capaian pembelajaran dan cakupan materi elemen bilangan kelas X mengacu pada dokumen resmi dari kemendikbud tentang Capaian Pembelajaran pada PAUD, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah pada kurikulum merdeka.

### ***Test Specifications***

Pada tahap ini dibuat rancangan spesifikasi instrumen yang akan dibuat. Jenis soal yang dikembangkan yaitu soal uraian dengan jumlah kategori jawaban sebanyak empat yakni, nol sampai empat. Soal yang dikembangkan berjumlah 12 soal uraian dengan durasi penggerjaan 120 menit. Terdapat 4 indikator literasi matematis yang digunakan dalam penelitian ini yaitu 1) *formulate* yaitu peserta didik dapat merumuskan konteks masalah nyata secara matematis, 2) *employ* yaitu peserta didik dapat menggunakan konsep, fakta, dan prosedur matematika untuk menemukan solusi, 3) *interpret* yaitu peserta didik dapat menafsirkan solusi, hasil atau kesimpulan

matematika yang didapat dalam konteks masalah nyata, 4) *reasoning*/ penalaran yaitu peserta didik dapat memberikan argumen dan bukti untuk mendukung dan menjelaskan solusi (OECD, 2023a).

Capaian Pembelajaran elemen bilangan untuk Fase E (kelas X) yang digunakan dalam penelitian ini yaitu “Peserta didik menggeneralisasi dapat sifat-sifat bilangan berpangkat. Mereka dapat menerapkan barisan dan deret aritmetika dan geometri, termasuk masalah yang terkait bunga tunggal dan bunga majemuk.” Adapun materinya meliputi barisan dan deret (aritmetika dan geometri); serta bilangan berpangkat. Materi barisan dan deret (aritmetika dan geometri); serta bilangan berpangkat dibagi menjadi 3 sub materi yaitu: barisan dan deret aritmetika; barisan dan deret geometri; dan bilangan berpangkat. Masing-masing sub materi terdapat 4 soal untuk mewakili masing-masing indikator literasi matematis.

### **Item Development**

Penyusunan instrumen tes literasi matematis mengacu pada distribusi indikator dengan materi yang telah dibuat sebelumnya. Penskoran soal pada instrumen ini menggunakan penskoran politomus dengan empat kategori skor (nol sampai tiga) di mana masing-masing kategori memiliki kriteria jawaban yang harus terpenuhi. Setelah kisi-kisi, alternatif penyelesaian, dan pedoman penskoran selesai dibuat, selanjutnya dilakukan validasi instrumen. Instrumen literasi matematis yang terdiri dari 12 soal divalidasi oleh 3 dosen pendidikan matematika serta 2 guru matematika. Saran-saran perbaikan dari validator dapat dilihat pada [Tabel 1](#).

**Tabel 1.** Saran Perbaikan Validator

No	Saran Perbaikan
1	Ciri khas soal literasi menggunakan stimulus fakta, satu fakta bisa untuk beberapa soal. Sajikan fakta dari sumber asli
2	Stimulus fakta biaya taksi bukan merupakan barisan dan deret. Suku pada barisan dan deret berupa bilangan bulat bukan diskrit seperti jarak tempuh pada taksi.
3	Suku bunga bank biasanya dihitung tiap tahun, pastikan suku bunga yang diperoleh berlaku tiap bulan atau tahun
4	Tabel suku bunga didetaikan. Tabel 1 bulan, 3 bulan, 6 bulan, 12 bulan, 24 bulan tambahkan baris di atasnya lama deposito.
5	Tabel suku bunga dibuat lebih kontras dan jelas.

Instrumen yang telah direvisi berdasarkan saran kemudian divalidasi. Hasil validasi kemudian dihitung menggunakan indeks Aiken. Batas nilai V minimum untuk instrumen yang valid menggunakan indeks Aiken dengan 5 rater, jumlah kategori 5, dan taraf signifikansi 5% adalah  $\geq 0,80$ . Oleh karena itu, nomor butir instrumen pada aspek materi, konstruksi, dan bahasa dikatakan valid apabila nilai V pada rumus Aiken  $\geq 0,80$ . Berdasarkan hasil nilai V setiap butir pada setiap aspek bernilai  $\geq 0,80$ . Hal ini berarti bahwa setiap butir dapat dikatakan valid secara isi. Oleh karena itu, secara keseluruhan, instrumen literasi matematis yang dibuat dapat dikatakan valid secara isi.

### **Test Design and Assembly**

Instrumen tes yang telah divalidasi kemudian diperbaiki berdasarkan saran yang diberikan. [Tabel 2](#) merupakan kisi-kisi instrumen tes *final* yang digunakan dalam penelitian.

**Tabel 2.** Kisi-Kisi Instrumen Tes Literasi Matematis

Capaian Pembelajaran (CP)	Indikator Soal	Indikator KLM	Level	No. Soal
Peserta didik dapat menerapkan barisan dan deret aritmetika	Diberikan permasalahan kontekstual, peserta didik dapat merumuskan konteks masalah nyata secara matematis	<i>Formulate</i>	1	1
	Diberikan permasalahan kontekstual, peserta didik dapat menggunakan konsep, fakta, dan prosedur matematika untuk menemukan solusi.	<i>Employ</i>	2	4
	Diberikan permasalahan kontekstual, peserta didik dapat menafsirkan solusi, hasil atau kesimpulan matematika dalam konteks masalah nyata	<i>Interpret</i>	2	2
	Diberikan permasalahan kontekstual, peserta didik dapat memberikan argumen dan bukti untuk mendukung dan menjelaskan solusi	<i>Reasoning</i>	2	3
	Diberikan permasalahan kontekstual, peserta didik dapat merumuskan konteks masalah nyata secara matematis	<i>Formulate</i>	2	5
	Diberikan permasalahan kontekstual, peserta didik dapat menggunakan konsep, fakta, dan prosedur matematika untuk menemukan solusi.	<i>Interpret</i>	3	6
	Diberikan permasalahan kontekstual, peserta didik dapat menafsirkan solusi, hasil atau kesimpulan matematika dalam konteks masalah nyata	<i>Employ</i>	1	7
	Diberikan permasalahan kontekstual, peserta didik dapat memberikan argumen dan bukti untuk mendukung dan menjelaskan solusi	<i>Reasoning</i>	3	8
	Diberikan permasalahan kontekstual, peserta didik dapat merumuskan konteks masalah nyata secara matematis	<i>Formulate</i>	2	9
	Diberikan permasalahan kontekstual, peserta didik dapat menggunakan konsep, fakta, dan prosedur matematika untuk menemukan solusi.	<i>Employ</i>	3	10
	Diberikan permasalahan kontekstual, peserta didik dapat menafsirkan solusi, hasil atau kesimpulan matematika dalam konteks masalah nyata	<i>Interpret</i>	1	11
	Diberikan permasalahan kontekstual, peserta didik dapat memberikan argumen dan bukti untuk mendukung dan menjelaskan solusi	<i>Reasoning</i>	1	12

### Test Production

Lembar instrumen soal yang diujicobakan kepada peserta didik berisi instrumen *final* yang disusun berdasarkan konteks permasalahan dan materi yang digunakan.

### Test Administration

Uji coba lapangan dilakukan pada 195 peserta didik kelas X di salah satu sekolah di Yogyakarta. Instrumen yang diujicobakan merupakan hasil revisi dari validator ahli, yaitu instrumen literasi matematis yang terdiri dari 12 butir soal uraian.

### Scoring Test Responses

Data hasil uji coba lapangan yang telah diperoleh kemudian digunakan untuk analisis validitas konstruk dan reliabilitas konstruk. Selanjutnya dilakukan analisis asumsi IRT yaitu, unidimensi, independensi lokal, dan invariansi parameter. Langkah selanjutnya dilakukan analisis IRT model PCM untuk mengetahui karakteristik butir instrumen literasi matematis.

### Passing/Cut Scores

*Cut scores* instrumen literasi matematis ditentukan menggunakan metode Angoff. Melalui metode ini *rater* menetapkan *cut scores* berdasarkan perkiraan peluang menjawab benar terhadap butir yang dikelompokkan berdasarkan tingkat kesulitan mudah, sedang, sulit (Retnawati, 2014). Penentuan *cut scores* pada metode ini melibatkan 3 *rater*. *Rater* diminta mengisi proporsi menjawab benar dan banyaknya butir yang termasuk dalam tiap tingkat kesulitan. Rangkuman hasil pengukuran oleh rater disajikan pada [Tabel 3](#).

**Tabel 3.** Rangkuman Hasil Pengukuran

Rater	Tingkat kesulitan		
	Mudah	Sedang	Susah
1	4 butir (90%)	2 butir (75%)	2 butir (30%)
2	4 butir (90%)	2 butir (70%)	2 butir (30%)
3	3 butir (90%)	3 butir (70%)	2 butir (30%)

*Cut scores* dari masing-masing rater ditentukan dengan menjumlahkan hasil perkalian presentase peluang menjawab benar dengan banyak butir. *Cut scores* untuk *rater* 1 yaitu  $4 \times 90\% + 2 \times 70\% + 2 \times 30\% = 5,7$ , *cut score* untuk *rater* 2 yaitu  $4 \times 90\% + 2 \times 70\% + 2 \times 30\% = 5,6$ , *cut score* untuk *rater* 3 yaitu  $3 \times 90\% + 3 \times 70\% + 2 \times 30\% = 5,55$ . *Cut scores* akhir merupakan rata-rata dari hasil *cut scores* 3 *rater*. Hasil penentuan *cut scores* dari 3 *rater* sebesar 5,6 jika berupa skor atau sama dengan 23,4 jika berupa nilai pada skala 100.

*Cut scores* digunakan sebagai batas kelulusan untuk dasar menafsirkan hasil tes dari responden tes literasi matematis. Responden yang mendapatkan skor sama atau lebih dari 5,6 dinyatakan lulus tes literasi matematis, sedangkan responden yang mendapatkan skor lebih rendah dari batas kelulusan dinyatakan belum lulus tes literasi matematis. Dari 195 responden dalam penelitian ini, terdapat 15 responden yang belum lulus tes literasi matematis.

### Test Score Report.

Data hasil uji coba lapangan yang telah diperoleh kemudian digunakan untuk analisis validitas konstruk dan reliabilitas konstruk. Selanjutnya dilakukan analisis asumsi IRT yaitu, unidimensi, independensi lokal, dan invariansi parameter. Langkah selanjutnya dilakukan analisis IRT model PCM untuk mengetahui karakteristik butir instrumen literasi matematis.

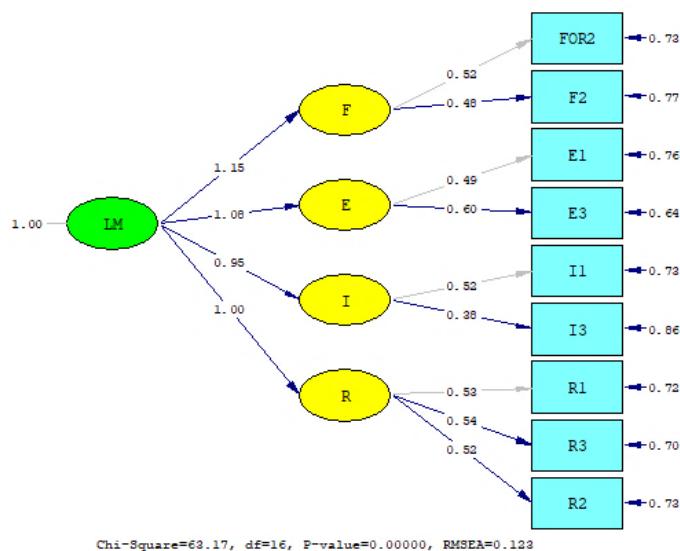
Validitas isi memastikan bahwa setiap butir telah sesuai dengan indikator dan materi pembelajaran yang biasanya dinilai oleh ahli (Sudaryono, 2012). Berdasarkan hasil validitas isi menggunakan indeks V Aiken, setiap butir instrument literasi matematis pada setiap aspek memiliki nilai  $V \geq 0,80$ . Hal ini menunjukkan bahwa 12 butir tes literasi matematis yang dikembangkan dinyatakan valid dan layak digunakan. Selanjutnya dilakukan uji KMO dan Barlett's untuk mengetahui kecukupan sampel.

**Tabel 4.** Output Uji KMO dan Bartlett's

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,789
Bartlett's Test of Sphericity	Approx. Chi-Square	493,892
	df	66
	Sig.	,000

Output hasil uji KMO dan Barlett's pada [Tabel 4](#) menunjukkan  $KMO \geq 0,5$  yang menunjukkan bahwa sampel sebanyak 195 pada penelitian ini sudah cukup. Signifikansi uji Barlett's 0,000 yang menunjukkan  $sig. < 0,05$  artinya data membentuk matriks korelasi dengan adanya hubungan antar variabel. Selanjutnya uji *anti-image correlation* untuk menentukan variabel yang layak digunakan dalam analisis faktor. Nilai *Measures of Adequate Sampling (MSA)* pada interval  $0,692 \leq MSA \leq 0,859$  dengan nilai minimal 0,692 pada butir F1. Hal ini menunjukkan bahwa semua butir layak masuk dalam analisis faktor.

Selanjutnya dilakukan validitas konstruk untuk mengukur sejauh mana butir dapat merepresentasikan konstruk secara teoritis melalui analisis faktor konfirmatori (CFA). Berdasarkan CFA diperoleh *standardized solution* konstruk instrumen literasi matematis yang disajikan pada [Gambar 2](#).



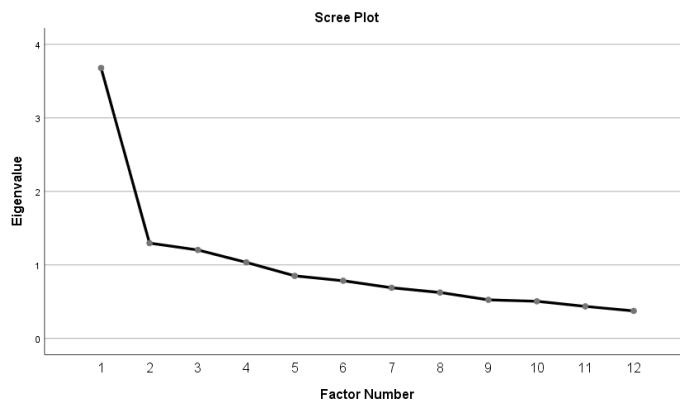
**Gambar 2.** Path Diagram CFA

Validitas konstruk instrumen literasi matematis dapat dilihat dari besaran SLF (*Standardized Loading Factor*). Kelayakan validitas dalam analisis CFA yaitu apabila  $SLF > 0,45$  untuk sampel sebanyak 150 ([Hair et al., 2019](#)). Hasil analisis awal didapatkan bahwa butir F1, E2, dan I2 memiliki nilai  $SLF > 0,45$  sehingga dikelurkan dari jalur dan dilakukan analisis ulang. Gambar 2 merupakan hasil analisis akhir yang menunjukkan bahwa butir FOR2, F2, E1, E3, I1, R1, R2, R3 memiliki nilai  $SLF > 0,45$ , sedangkan nilai  $SLF$  butir I3 < 0,45. Oleh karena itu, butir FOR2, F2, E1, E3, I1, R1, R2, R3 dikatakan valid secara konstruknya.

Sementara itu, CR (*Construct Reliability*) digunakan untuk mengukur keajegan suatu instrument tes. Hasil perhitungan CR instrumen literasi matematis yaitu 0,75. Hal ini menunjukkan bahwa  $CR > 0,7$  sehingga konstruk tes literasi matematis dapat dikatakan reliabel ([Hair et al., 2019](#)). Dengan demikian, dapat disimpulkan bahwa kualitas instrumen literasi matematis secara keseluruhan memenuhi syarat baik dari segi validitas isi, validitas konstruk, maupun reliabilitas konstruk. Selanjutnya, karakteristik butir tes yang dianalisis meliputi estimasi tingkat kesulitan (*b*);

kurva karakteristik butir (ICC); fungsi informasi dan SEM; kecocokan item (*item fit*); dan kecocokan respon (*person fit*).

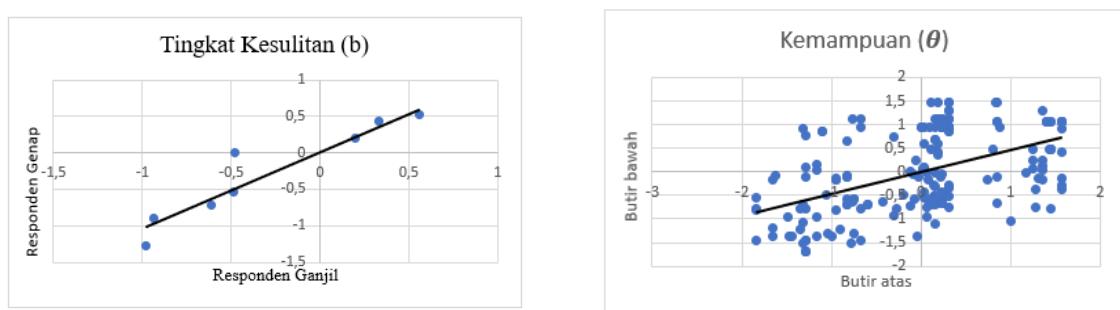
Sebelum karakteristik butir dianalisis dengan PCM, yang pertama dilakukan uji asumsi IRT. Asumsi dalam IRT yang perlu diujii meliputi unidimensi, independensi lokal, serta invariansi parameter (Hambleton et al., 1991). Pembuktian asumsi unidimensi dapat dilihat dari *scree-plot* dan nilai eigen pada [Gambar 3](#).



**Gambar 3.** Output Scree-plot

[Gambar 3](#) memperlihatkan adanya satu curaman utama, kemudian grafik melandai pada nilai eigen kedua hingga berikutnya. Hal ini menunjukkan bahwa tes literasi matematis mengukur 1 faktor dominan. Hasil ini diperkuat dengan persentase nilai eigen komponen 1 yang cukup besar dengan persentase varians sebesar 30,658%. Jika persentase varians melebihi 20%, maka instrument dianggap mengukur dimensi utama sehingga asumsi unidimensi terpenuhi (Retnawati, 2017). Ketika asumsi unidimensi terpenuhi maka secara tidak langsung membuktikan bahwa asumsi independensi lokal juga terpenuhi (Hambleton et al., 1991).

Selanjutnya dilakukan uji asumsi invariansi parameter butir dan invariansi parameter kemampuan. Asumsi invariansi parameter butir dilakukan dengan mengestimasi parameter butir pada responden nomor ganjil dan nomor genap. Invariansi parameter kemampuan dibuktikan dengan mengestimasi parameter kemampuan pada separuh butir bagian atas dan separuh butir bagian bawah. Diagram pencar untuk tingkat kesulitan yang diestimasi pada responden ganjil dan genap disajikan pada [gambar](#) dan diagram pencar untuk tingkat kesulitan yang diestimasi pada separuh butir bagian atas dan separuh butir bagian bawah disajikan pada [Gambar 4](#).



**Gambar 4.** Diagram Pencar Invariansi Parameter Butir dan Kemampuan

Berdasarkan [Gambar 4](#) dapat dilihat bahwa posisi titik-titik relatif mendekati garis  $y = x$ . Hal ini mengindikasikan bahwa invariansi parameter butir yang diestimasi pada responden ganjil dan genap terpenuhi serta invariansi parameter kemampuan yang diestimasi pada separuh butir bagian atas dan separuh butir bagian bawah terpenuhi. Oleh karena itu, ketiga asumsi IRT telah terpenuhi, sehingga dapat dilakukan analisis butir menggunakan PCM.

Tingkat kesulitan butir diestimasi menggunakan model IRT politomus PCM (*Partial Credit Model*). PCM hanya mengestimasi parameter kesulitan ( $b$ ) untuk setiap batas kategori skor dalam satu butir. Sedangkan untuk parameter diskriminasi ( $a$ ) diasumsikan sama untuk setiap butir yaitu  $a = 1$  yang berarti kemampuan membedakan responden antara satu butir dan lainnya dianggap sama. Hasil estimasi tingkat kesulitan ( $b$ ) instrumen literasi matematis menunjukkan bahwa setiap butir berada dalam kategori baik karena berada pada interval  $-2 < b < 2$  ([Hambleton & Swaminathan, 1985](#)). Hasil analisis estimasi tingkat kesulitan butir ( $b$ ) dapat dilihat pada [Tabel 5](#).

**Tabel 5.** Estimasi Tingkat Kesulitan ( $b$ )

Butir Soal	$a$	Nilai			Lokasi
		$b_1$	$b_2$	$b_3$	
FOR2	1	0,657	-1,368	-1,270	-0,660
F2	1	-0,912	-0,373	0,561	-0,241
E1	1	-0,458	-0,746	-2,105	-1,103
E3	1	-0,693	2,550	-0,720	0,378
5I1	1	0,337	-1,215	-1,864	-0,914
R1	1	0,812	0,798	-1,008	0,200
R2	1	-1,254	-0,677	0,404	-0,509
R3	1	-0,574	1,671	0,532	0,543

Keterangan:

$a$  = indeks daya beda

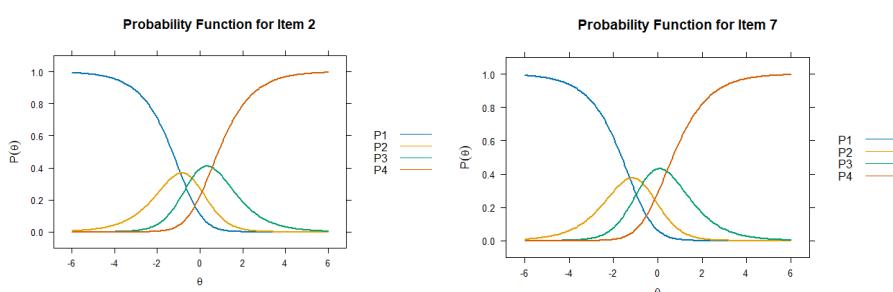
$b_1$  = titik perpotongan kategori 0 dan 1

$b_2$  = titik perpotongan kategori 1 dan 2

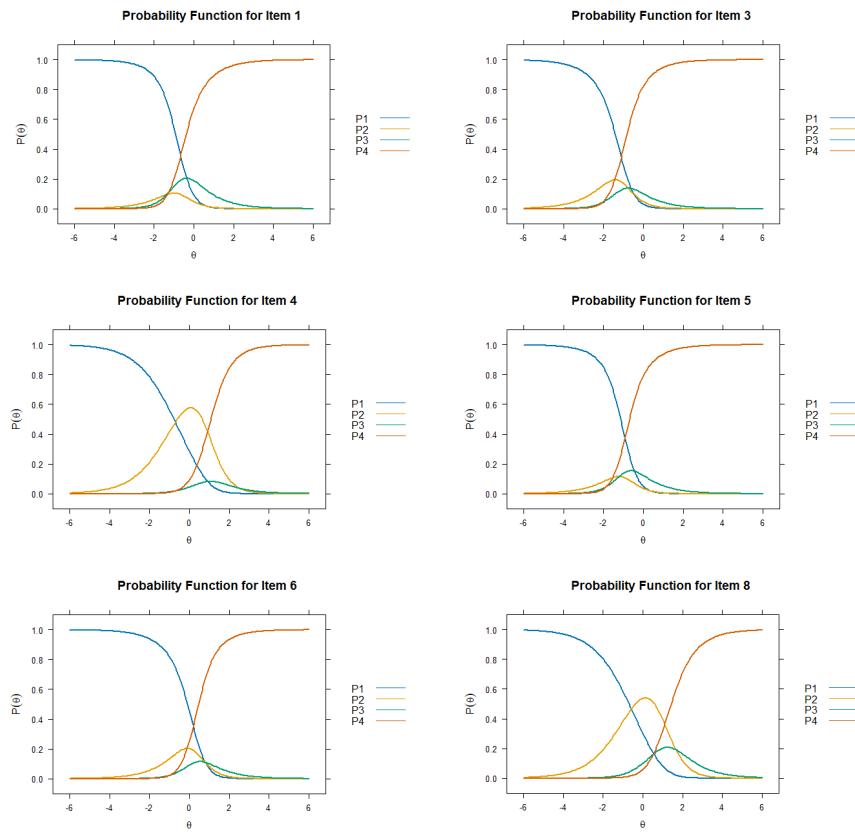
$b_3$  = titik perpotongan kategori 2 dan 3

Lokasi = estimasi tingkat kesulitan

Berdasarkan hasil estimasi tingkat kesulitan pada [Tabel 5](#) kolom Lokasi, dapat dilihat bahwa tingkat kesulitan setiap butir berada pada interval  $-1,1034253 \leq b \leq 0,5431551$ . Hasil ini menunjukkan bahwa setiap butir berada dalam kategori baik karena berada pada interval  $-2 < b < 2$  ([Hambleton & Swaminathan, 1985](#)). Untuk lebih mudah dipahami, disajikan plot ICC (*Item Characteristic Curve*). Hasil plot ICC dapat dilihat pada [Gambar 5](#) dan [Gambar 6](#).



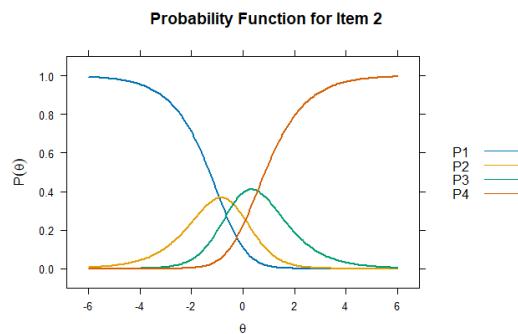
**Gambar 5.** ICC Ideal



Gambar 6. ICC Tidak Ideal

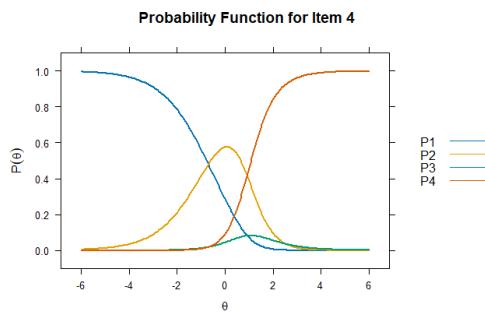
Sumbu x ( $\theta$ ) menunjukkan tingkat kemampuan dari responden. Interval kemampuan umumnya dari -6 hingga 6, dengan 0 sebagai kemampuan rata-rata. Sumbu y ( $P(\theta)$ ) menunjukkan peluang seorang responden dengan kemampuan sebesar ( $\theta$ ) akan memberikan respon pada kategori tertentu. nilai  $b$  merupakan perpotongan antar kategori. P1 merupakan kategori 0, P2 merupakan kategori 1, P3 merupakan kategori 2, dan P4 merupakan kategori 3. Nilai perpotongan P1 dengan P2 adalah  $b_1$ , nilai perpotongan P2 dengan P3 adalah  $b_2$ , dan nilai perpotongan P3 dengan P4 adalah  $b_3$ .

Plot ICC yang ideal ditandai dengan perpotongan kurva antar kategori respon yang semakin bergeser ke kanan atau dengan kata lain semakin ke kanan plot, nilai  $b$  semakin meningkat (Dewanti et al., 2024). Hal ini berarti butir berfungsi baik dalam membedakan responden berdasarkan tingkat kemampuannya. Dari hasil plot ICC, plot yang ideal disajikan pada [Gambar 5](#), sedangkan untuk plot ICC yang tidak ideal disajikan pada [Gambar 6](#). Dari 8 butir tes literasi matematis, yang termasuk kategori ideal yaitu butir F2 dan R2, sedangkan 6 butir lainnya dinyatakan tidak ideal. Butir yang tidak ideal ini disebabkan karena PCM tidak mensyaratkan adanya urutan tertentu dalam langkah penyelesaian maupun kesamaan tingkat kesulitan pada setiap langkah (De Ayala dalam Dewanti et al., 2020). Untuk menganalisis lebih dalam, akan dibahas lebih lanjut salah satu butir ideal yaitu butir F2 dan satu butir tidak ideal yaitu butir I1. ICC butir F2 disajikan pada Gambar 7.



Gambar 7. ICC Butir F2

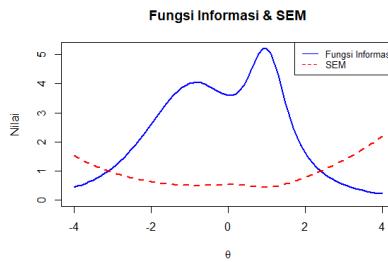
Gambar 7 menunjukkan bahwa titik potong (*threshold*) garis P1 dengan P2 menyatakan bahwa kemampuan minimal untuk mendapatkan skor 1 yaitu -0,912. Selanjutnya, kemampuan minimal untuk mendapatkan skor 2 yaitu -0,373. Kemampuan minimal untuk mendapatkan skor 3 yaitu 0,561. Oleh karena itu, plot ICC butir F2 merupakan plot yang ideal karena menunjukkan bahwa semakin tinggi kategori skor, semakin tinggi pula tingkat kemampuan yang dibutuhkan (Retnawati, 2014). Selanjutnya, ICC butir I1 disajikan pada Gambar 8.



Gambar 8. ICC Butir E3

Gambar 8 menunjukkan bahwa titik potong garis P1 dengan P2 menyatakan bahwa kemampuan minimal untuk mendapatkan skor 1 yaitu -0,693. Selanjutnya, kemampuan minimal untuk mendapatkan skor 2 yaitu 2,550. Selanjutnya, *threshold* ketiga, yang merupakan batas antara skor 2 dan 3, justru turun drastis ke -0,720, yang berarti responden dengan kemampuan rendah justru bisa langsung mendapatkan skor 3. Urutan *threshold* yang tidak berurutan secara logis ini dari -0,693 ke 2,550 lalu turun ke -0,720 menandakan bahwa plot tersebut tidak ideal. Butir yang tidak ideal ini disebabkan karena PCM tidak mensyaratkan adanya urutan tertentu dalam langkah penyelesaian maupun kesamaan tingkat kesulitan pada setiap langkah (De Ayala dalam Dewanti et al., 2020).

Analisis selanjutnya yaitu fungsi informasi dan SEM. Fungsi informasi tes diperoleh dari akumulasi fungsi informasi masing-masing butir. Setiap informasi yang dihasilkan pengukuran selalu memiliki kesalahan yang dalam IRT disebut dengan *Standard Error of Measurement* (SEM). Menurut Hambleton et al. (1991) hubungan antara fungsi informasi dengan SEM adalah berbanding terbalik sehingga ketika fungsi informasi meningkat, maka nilai SEM menurun. Hal ini berarti bahwa semakin akurat pengukuran, maka nilai SEM semakin kecil. Hasil analisis fungsi informasi dan SEM disajikan pada Gambar 9.



Gambar 9. Fungsi Informasi dan SEM

Gambar 9 menunjukkan bahwa fungsi informasi memiliki dua puncak. Artinya, tes literasi matematis akurat untuk mengukur kemampuan sedikit di bawah rata-rata dan kemampuan sedikit di atas rata-rata. Titik puncak dari fungsi informasi tidak selalu tunggal karena setiap butir tes mengukur kemampuan yang berbeda (Mulvia et al., 2021).

Fungsi informasi pada teori respon butir dapat digunakan untuk mengetahui koefisien reliabilitas tes (Mardaphi dalam Sumaryanta, 2021). Puncak tertinggi fungsi informasi berada di nilai 5.217 dengan SEM sebesar 0.437 dan theta 0,95, artinya tes literasi matematis akan reliabel pada responden dengan kemampuan sedikit lebih tinggi di atas rata-rata. Puncak kedua berada di nilai 4,045 dengan SEM sebesar 0.497 dan theta -0,85, artinya tes literasi matematis akan reliabel pada responden dengan kemampuan sedikit lebih rendah dari rata-rata.

Puncak kurva fungsi informasi yang pertama dengan theta 0,95 lebih tinggi dibandingkan puncak yang kedua dengan theta -0,85. Hal ini berarti bahwa butir tes memberikan lebih banyak informasi pada theta 0,95 dibandingkan pada theta -0,85. Puncak informasi yang lebih tinggi menunjukkan semakin banyak informasi yang dapat digunakan untuk menggambarkan kemampuan literasi matematis responden. Hal ini berarti bahwa butir tes literasi matematis reliabel apabila diberikan kepada responden dengan kemampuan theta 0 sampai 2 atau pada kemampuan literasi matematis yang lebih tinggi dari rata-rata. Temuan ini sejalan dengan hasil penelitian Fitriana (2022) dan Mulvia et al. (2021) yang menunjukkan bahwa fungsi informasi dalam IRT mampu menggambarkan tingkat ketepatan pengukuran secara lebih akurat pada rentang kemampuan tertentu.

Selanjutnya, kecocokan data dengan model merupakan hal penting dalam penerapan analisis IRT. Terdapat dua jenis kecocokan yaitu *item fit* dan *person fit*. *Item fit* dapat digunakan untuk mengidentifikasi item yang mungkin tidak berfungsi sesuai dengan ekspektasi model, sedangkan *person fit* menunjukkan ada tidaknya penyimpangan pola jawaban responden terhadap ekspektasi model (Reise, 1990). Hasil analisis *item fit* disajikan pada Tabel 6.

Tabel 6. Hasil Analisis Item Fit

Item	MNSQ	ZSTD	Keterangan
FOR2	0,753	-1,426	Fit
F2	0,875	-1,424	Fit
E1	0,641	-1,547	Fit
E3	0,807	-1,791	Fit
I1	0,888	-0,403	Fit
R1	0,740	-2,135	Tidak Fit
R3	0,882	-1,106	Fit
R2	0,874	-1,356	Fit

Berdasarkan [Tabel 6](#), hasil analisis *item fit* menunjukkan bahwa dari 8 butir tes literasi matematis hanya terdapat satu butir yang tidak *fit* terhadap pola jawaban responden yaitu butir R1 dengan indikator reasoning level 1. Hal ini dapat disebabkan karena kesalahan konstruksi butir sehingga butir R1 perlu ditelaah kembali ([Zubairi & Kassim, 2006](#)).

Kecocokan pola respon terhadap instrumen literasi matematis (*person fit*) bertujuan untuk memberikan informasi seberapa besar responden yang memiliki kecocokan sesuai dengan ekspektasi model PCM. Berdasarkan hasil *person fit*, diketahui bahwa terdapat 82 dari 195 responden yang tidak *fit*. Temuan ini sejalan dengan penelitian [Amelia \(2021\)](#) yang juga menemukan adanya *person missfit* dari responden yang ia teliti. Hal ini dapat disebabkan karena adanya *lucky guessing, carelessness*, atau *cheating* ([Wright & Stone, 1999](#)).

## SIMPULAN

Instrumen tes literasi matematis elemen bilangan kelas X terbukti berkualitas baik dengan ditunjukkan dengan hasil validitas isi, validitas konstruk, dan reliabilitas konstruk yang telah memenuhi syarat. Karakteristik butir tes literasi matematis yaitu: (1) tingkat kesulitan (*b*) masing-masing butir berada dalam kategori baik dengan interval  $-2 \leq b \leq 2$ ; (3) kurva karakteristik butir (ICC) menunjukkan terdapat 2 butir dengan plot ICC yang ideal dan terdapat 6 butir yang memiliki plot ICC tidak ideal; (4) kurva fungsi informasi dan SEM menunjukkan bahwa instrumen tes literasi matematis memberikan informasi paling baik dan akurat pada tingkat kemampuan responden sedikit di atas rata-rata; (5) analisis *item fit* menunjukkan bahwa hanya terdapat satu dari delapan butir yang tidak *fit*, sedangkan hasil analisis *person fit* menunjukkan terdapat 15 dari 195 responden yang tidak *fit*.

Instrumen tes literasi matematis ini memiliki kualitas dan karakteristik butir yang baik sehingga dapat dijadikan sebagai alat ukur kemampuan literasi matematis. Hasil pengembangan ini diharapkan dapat memberikan gambaran dalam menyusun instrumen literasi yang valid dan reliabel. Selain itu, penelitian ini diharapkan memperkaya kajian mengenai pengembangan instrumen literasi matematis menggunakan teori respons butir. Penelitian selanjutnya diharapkan untuk mengembangkan instrumen tes pada elemen atau jenjang berbeda dengan melibatkan responden yang lebih luas agar estimasi parameter lebih stabil. Selain itu, penelitian selanjutnya dapat mencoba menggunakan analisis model IRT lain, seperti 2PL atau 3PL.

## DAFTAR RUJUKAN

- Amelia, R. N. (2021). Identifikasi item fit dan person fit dalam pengukuran hasil belajar kimia. *Jurnal Ilmiah WUNY*, 3(1), 13–26. <https://doi.org/10.21831/jwuny.v3i1.40703>
- Auliya, L. (2021). *Pengembangan instrumen tes literasi matematis berbasis budaya betawi untuk siswa SMP/MTs* [UIN Syarif Hidayatullah Jakarta]. <https://repository.uinjkt.ac.id/dspace/handle/123456789/59007>
- Azizah, N. L., Arliani, E., Purbaningrum, M., & Ramadhan, S. (2024). Literasi matematika materi sistem persamaan dan pertidaksamaan linier: pengembangan butir soal. *AKSIOMA: Jurnal*

- Program Studi Pendidikan Matematika, 13(2), 708–719.*  
<https://doi.org/10.24127/ajpm.v13i2.8547>
- Azwar, S. (2019). *Reliabilitas dan validitas*. Pustaka Pelajar.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.
- Dewanti, S. S. (2021). *Model penilaian kompetensi matematika abad xxi pada mahasiswa program studi S1 pendidikan matematika* [Universitas Negeri Yogyakarta].  
<https://eprints.uny.ac.id/73381/>
- Dewanti, S. S., Ayriza, Y., & Setiawati, F. A. (2020). The application of item response theory for development of a students' attitude scale toward mathematics. *New Educational Review*, 60, 108–123. <https://czasopisma.marszalek.com.pl/journals/19/376/6185>
- Dewanti, S. S., Izzah, J. N., & Kiranasari, S. P. (2024). Utilizing item response theory for the analysis of self-regulated learning scale in mathematics. *Jurnal Elemen*, 10(3), 614–629. <https://doi.org/10.29408/jel.v10i3.26618>
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Lawrence Erlbaum Associates, Inc.
- Farida, R. N., Qohar, A., & Rahardjo, S. (2021). Analisis kemampuan literasi matematis siswa SMA kelas x dalam menyelesaikan soal tipe pisa konten change and relationship. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 5(3), 2802–2815.  
<https://doi.org/10.31004/cendekia.v5i3.972>
- Fitriana, D. A. (2022). Validitas dan reliabilitas instrumen penilaian matematika dengan problem solving berbasis penskoran politomus. *Jurnal IKLILA: Jurnal Studi Islam dan Sosial*, 5(1), 13–25. <https://ejournal.iaikhozin.ac.id/ojs/index.php/iklila/article/view/118>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (eighth). Pearson Education, Inc.
- Hambleton, R. K., & Swaminathan, H. (1985). *Items response theory: Principles and application*. Kluwer, Inc.
- Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals of item response theory*. Sage Publications.
- Istiyyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (pysTHOTS) peserta didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 1–12. <https://doi.org/10.21831/pep.v18i1.2120>
- Kurniawan, R., & Djidu, H. (2021). Kemampuan literasi matematis siswa: Sebuah studi literatur. *Jurnal Edumatic: Jurnal Pendidikan Matematika*, 2(01), 24–30.  
<https://doi.org/10.21137/edumatic.v2i01.468>
- Mulvia, R., Ramalis, T. R., & Efendi, R. (2021). Mendekripsi keajegan butir tes dengan fungsi informasi. *Jurnal Pendidikan Indonesia*, 2(1), 72–84.  
<https://doi.org/10.36418/japendi.v2i1.66>
- Nurjanah, U., & Hakim, D. L. (2019). Number sense siswa pada materi bilangan. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika Sesiomadika 2019*, 2(1), 1174–1182.
- OECD. (2023a). *PISA 2022 assessment and analytical framework*. OECD Publishing.
- OECD. (2023b). *PISA 2022 results: The state of learning and equity in education*. OECD Publishing.
- Ogunsakin, I. B., & Shogbesan, Y. O. (2018). Item response theory (IRT): A modern statistical theory for solving measurement problem in 21st century. *International Journal of Scientific*

- Research in Education*, 11(3B), 627–635. <https://www.ijsre.com.ng/assets/vol.%2C-11%283b%29-ogunsakin---shogbesan.pdf>
- Purnama, D. N. (2023). Perbandingan estimasi kemampuan akuntansi siswa smk pada penskoran dikotomus dan politomus. *Jurnal Pendidikan Akuntansi Indonesia*, 20(2), 145–154. <https://doi.org/10.21831/jpai.v20i2.54143>
- Purwanto, A. H., & Jailani. (2024). Pengembangan soal matematika model AKM menggunakan konteks budaya yogyakarta untuk mengetahui profil literasi matematika siswa sekolah menengah atas. *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumian Dan Angkasa*, 2(6), 182–204. <https://doi.org/0.62383/algoritma.v2i6.299>
- Putra, Y. Y., & Vebrian, R. (2019). Pengembangan soal matematika model PISA konteks kain cual Bangka Belitung. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 3(2), 333–340. <https://doi.org/10.31004/cendekia.v3i2.114>
- Reise, S. P. (1990). A comparison of item and person fit methods of assessing model data fit in IRT. *Applied Psychological Measurement*, 14(2), 127–137. <https://doi.org/10.1177/014662169001400202>
- Retnawati, H. (2014). *Teori respons butir dan penerapannya untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Nuha Medika.
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian (panduan peneliti, mahasiswa, dan psikometri)*. Parama Publishing.
- Retnawati, H. (2017). *Validitas, reliabilitas & karakteristik butir (panduan untuk peneliti, mahasiswa, dan psikometri)*. Parama Publishing.
- Stacey, K. (2015). Selected regular lectures from the 12th international congress on mathematical education. In *Selected regular lectures from the 12th international congress on mathematical education* (pp. 771–790). Springer. <https://doi.org/10.1007/978-3-319-17187-6>
- Sudaryono. (2012). *Dasar-dasar evaluasi pembelajaran*. Graha Ilmu.
- Sumaryanta. (2021). *Teori tes klasik dan teori respon butir: konsep dan contoh penerapannya*. CV. Confident.
- Takaria, J., Pattimukay, N., & Kaary, K. M. (2022). Analisis kemampuan literasi numerasi siswa dalam menyelesaikan soal cerita ditinjau dari kemampuan awal matematis (KAM). *Pedagogika: Jurnal Pedagogik dan Dinamika Pendidikan*, 10(2), 318–327. <https://doi.org/10.30598/pedagogikavol10issue2page318-327>
- Wibowo, A. A., Rif'at, M., & Yani, A. (2020). Pengembangan instrumen tes untuk mengukur kemampuan koneksi literasi siswa SMP. *JPPK: Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 9(7), 1–12. <https://doi.org/10.26418/jppk.v9i7.41316>
- Widhiarso, W. (2010). *Model politomi dalam teori respons butir*. Universitas Gadjah Mada.
- Wright, B. D., & Stone, M. H. (1999). Measurement essentials. In *Measurement* (2nd ed.). Wide Range, Inc.
- Zubairi, A. M., & Kassim, N. L. A. (2006). Classical and rasch analyses of dichotomously scored reading comprehension test items. *Malaysian Journal of ELT Research*, 2(1), 1–20. <https://meltjournals.com/index.php/majer/article/view/663>