



Kelayakan Instrumen Kemampuan Pemecahan Masalah Matematika Berbasis Kontekstual untuk Siswa SMP Ditinjau dari Rasch Model

Aminah Ekawati¹, Isna Wardiah², Dina Huriaty³

^{1,3} Universitas PGRI Kalimantan, Indonesia

² Politeknik Negeri Banjarmasin, Indonesia

Corresponding Author: isnawardiah@poliban.ac.id

ABSTRACT

Kemampuan pemecahan masalah merupakan kompetensi penting dalam pembelajaran matematika dan menjadi salah satu fokus utama Kurikulum Merdeka. Namun, instrumen yang digunakan untuk menilai kemampuan tersebut masih didominasi oleh soal prosedural dan kurang mencerminkan konteks autentik kehidupan siswa, sehingga pengukuran menjadi kurang akurat. Penelitian ini bertujuan menilai kelayakan instrumen kemampuan pemecahan masalah matematika berbasis kontekstual untuk siswa SMP menggunakan Model Rasch. Instrumen terdiri dari empat soal uraian pada materi perbandingan, bilangan bulat, pecahan, dan aritmetika sosial yang dikembangkan berdasarkan level kognitif C3 dan C4. Proses validasi dilakukan melalui penilaian ahli terhadap aspek isi, teknis, dan bahasa, kemudian diuji empiris pada 60 siswa kelas VIII SMP menggunakan analisis Rasch melalui RStudio. Hasil validasi menunjukkan seluruh butir memenuhi kriteria I-CVI dan koefisien kappa, meskipun satu butir direvisi sesuai masukan ahli. Hasil analisis Rasch menunjukkan nilai Infit dan Outfit MNSQ berada dalam rentang 0,5-1,5, yang mengindikasikan validitas empiris yang baik. Reliabilitas Cronbach's Alpha sebesar 0,709, WLE reliability 0,678, dan EAP reliability 0,772 menunjukkan konsistensi pengukuran yang memadai. Asumsi unidimensionalitas terpenuhi berdasarkan nilai eigenvalue PCAR (< 3), dan parameter kesulitan menunjukkan variasi logit yang proporsional dari mudah hingga relatif sulit. Instrumen juga memiliki kualitas targeting yang baik, ditunjukkan oleh kesesuaian distribusi kemampuan siswa dengan tingkat kesulitan butir. Dengan demikian, instrumen ini layak digunakan untuk mengukur kemampuan pemecahan masalah matematika berbasis kontekstual pada siswa SMP.

ARTICLE INFO

Article history:

Received

01 January 2026

Revised

10 January 2026

Accepted

21 January 2026

Key Word

Pemecahan Masalah Matematika, Instrumen Berbasis Kontekstual, Model Rasch, Validitas Dan Reliabilitas, Penilaian Pembelajaran SMP

How to cite

<https://pusdikra-publishing.com/index.php/jsr>



This work is licensed under a
[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

PENDAHULUAN

Kemampuan pemecahan masalah merupakan salah satu kompetensi esensial dalam pembelajaran matematika yang tidak hanya dituntut dalam kurikulum nasional

tetapi juga menjadi bagian penting dari kecakapan abad 21. Pada jenjang SMP, pemecahan masalah berfungsi sebagai fondasi bagi siswa untuk mengembangkan kemampuan berpikir kritis, logis, dan kreatif dalam menghadapi situasi nyata (Khoiriyah & Husamah, 2018; Mulyati* et al., 2024). Namun, berbagai studi menunjukkan bahwa kemampuan ini masih tergolong rendah karena siswa cenderung hanya menghafal rumus tanpa memahami konteks permasalahan (Gunawan et al., 2023; Mardiansyah et al., 2025). Kondisi ini mengindikasikan bahwa sebagian besar siswa baru mencapai kemampuan berpikir tingkat rendah, khususnya pada level memahami (C2) dan menerapkan (C3), sementara kemampuan berpikir analitis (C4) belum berkembang secara optimal.

Kurikulum merdeka menegaskan bahwa pembelajaran matematika harus diarahkan pada pengembangan kemampuan berpikir kritis, kreatif, dan reflektif melalui kegiatan pemecahan masalah yang bermakna. (Kemendikbudristek, 2022). Dalam konteks tersebut, kemampuan menerapkan (C3) dan menganalisis (C4) menempati posisi strategis dalam membangun keterampilan berpikir matematis yang utuh, di mana C3 berperan sebagai tahap transisi dari pemahaman konseptual menuju kemampuan berpikir tingkat tinggi (HOTS) yang mulai terefleksi pada level C4. Namun, hasil PISA dan AKM menunjukkan bahwa banyak siswa hanya mampu menyelesaikan soal rutin, tetapi mengalami kesulitan ketika menghadapi masalah kontekstual yang menuntut penerapan konsep dalam situasi baru dan analisis informasi. Tanudjaya & Doorman (2020), menemukan bahwa siswa kesulitan mentransfer pengetahuan, memberikan alasan matematis, dan mengevaluasi informasi secara kritis. Temuan ini sejalan dengan Santriyani (2024), yang menunjukkan bahwa capaian pemecahan masalah pada soal AKM masih berada pada kategori sedang, khususnya pada tahap memahami konteks, membangun model, dan menyelesaikan masalah secara tepat.

Salah satu penyebab rendahnya kemampuan tersebut adalah instrumen penilaian di sekolah yang masih bersifat mekanistik, prosedural, dan kurang kontekstual. Soal-soal yang diberikan sering kali tidak menuntut siswa menghubungkan konsep matematika dengan fenomena kehidupan nyata, padahal pembelajaran berbasis kontekstual (*Contextual Teaching and Learning*) yang dapat memberikan peluang besar untuk mengaitkan konsep-konsep matematika dengan pengalaman hidup siswa (C3) dan menganalisis data, hubungan, serta pola yang muncul dalam konteks permasalahan tersebut (C4). Sehingga, pengembangan instrumen penilaian kontekstual menjadi kebutuhan penting agar penilaian mampu menangkap proses berpikir siswa secara lebih autentik.

Di sisi lain, validasi instrumen pemecahan masalah selama ini banyak menggunakan teori tes klasik yang memiliki keterbatasan, seperti ketergantungan pada sampel dan kurang mampu menggambarkan karakteristik butir secara individual.

Model Rasch menawarkan pendekatan yang lebih objektif (Sheptian et al., 2024) dan komprehensif karena mampu menganalisis parameter butir dan kemampuan siswa dalam skala logit yang linear, sekaligus menyediakan informasi mengenai reliabilitas, unidimensionalitas, daya beda, dan kesesuaian item-item (Hope et al., 2024; Yang et al., 2023).

Meskipun penelitian mengenai instrumen pemecahan masalah sudah cukup banyak, masih terbatas penelitian yang secara khusus mengembangkan instrumen berbasis kontekstual untuk siswa SMP kemudian dianalisis menggunakan Rasch Model. Kesenjangan ini menunjukkan perlunya kajian mendalam yang mengombinasikan pengembangan butir kontekstual dengan analisis Rasch untuk memastikan kualitas instrumen secara komprehensif.

Berdasarkan uraian tersebut, penelitian ini bertujuan mengembangkan dan menilai kelayakan instrumen pemecahan masalah matematika berbasis kontekstual untuk siswa SMP menggunakan Rasch Model. Melalui analisis validitas isi, validitas empiris, reliabilitas, unidimensionalitas, dan parameter kesulitan, penelitian ini diharapkan menghasilkan instrumen yang layak digunakan dalam penilaian kemampuan pemecahan masalah siswa secara lebih akurat dan bermakna.

METODE PENELITIAN

Penelitian ini menggunakan desain pengembangan instrumen dengan pendekatan psikometri modern melalui Item Response Theory (IRT) menggunakan Model Rasch. Model ini dipilih karena memiliki prinsip specific objectivity, yaitu kemampuan menghasilkan estimasi kemampuan responden dan parameter butir secara independen tanpa dipengaruhi karakteristik subjek maupun rater (Pratama, 2020; Prayoga et al., 2024; Verdú-Soriano & González-de la Torre, 2024). Selain itu, Model Rasch menghasilkan skor logit pada skala interval yang linier sehingga meningkatkan akurasi pengukuran dan memudahkan analisis psikometrik lanjutan (Darman et al., 2024; Pratama, 2020; Prayoga et al., 2024; Verdú-Soriano & González-de la Torre, 2024). Rasch mendukung pengukuran yang lebih objektif dan terstandarisasi dalam proses validasi instrumen (Darman et al., 2024; Ibrahim, 2025; Verdú-Soriano & González-de la Torre, 2024). Oleh karena itu, dengan menggunakan Model Rasch dapat menjamin validitas butir, reliabilitas, unidimensionalitas, serta pemetaan kemampuan dan kesukaran item dengan lebih akurat dan valid.

Responden penelitian terdiri dari siswa kelas VIII SMPN di kota Banjarmasin yang berada di dua kelas yang berbeda dan dipilih secara acak dan disajikan pada Tabel 1. Siswa pada Tabel 1 telah mempelajari materi pada instrumen kemampuan pemecahan masalah matematika berbasis kontekstual.

Tabel 1. Responden Penelitian

Siswa	Kelas A	Kelas B
Laki-laki	15	15
Perempuan	15	15
Jumlah	30	30

Instrumen yang dikembangkan sebanyak 4 soal yang terdiri dari materi perbandingan, aritmatika sosial, bilangan bulat, dan pecahan, di mana soal-soal tersebut berada pada level kognitif taksonomi bloom C3 dan C4. Pada level C3, siswa dituntut menggunakan rumus, prosedur, dan algoritma matematika untuk menyelesaikan masalah. Sementara pada level C4, siswa perlu menganalisis hubungan antar-informasi, membandingkan hasil, menentukan perubahan, serta memberikan alasan yang logis berdasarkan perhitungan. Keempat soal dirancang sesuai dengan karakteristik penilaian pemecahan masalah dalam Kurikulum Merdeka. Adapun materi dan soal yang digunakan disajikan pada tabel 2.

Tabel 2.

Instrumen Kemampuan Pemecahan Masalah Matematika Berbasis Kontekstual

No.	Pokok Bahasan	Soal	Taksonomi Bloom
1	Perbandingan	Adit dan Iwan memiliki 48 buah mangga. Rasio mangga Adit dan Iwan adalah 5 : 7. Tentukan Selisih banyaknya mangga mereka?	C3
2	Bilangan bulat	Rina memiliki kartu e-money. Di kartu tersebut sudah ada uangnya. Setiap kali ia naik bus saldo e-money berkurang Rp8.000 . Setiap kali ia top up saldo , bertambah Rp20.000 . <ol style="list-style-type: none"> Pada hari Sabtu, Rina melakukan 3 kali perjalanan naik bus dan sekali top up saldo. Pada hari Minggu, ia melakukan 4 kali perjalanan naik bus dan 2 kali top up saldo. Bandingkan perubahan saldo Rina pada hari Sabtu dan Minggu. Pada hari apa saldo akhir Rina berkurang dan pada hari apa bertambah? Jelaskan	C4
3	Pecahan	Bu Susi membuat minuman dengan mencampurkan air lemon $2\frac{1}{2}$ liter, air mineral $6\frac{3}{8}$ liter, dan air gula putih $1\frac{1}{4}$ liter. Campuran tersebut kemudian dimasukkan	C3

ke dalam botol dengan volume 600 ml. Tentukan banyaknya botol yang diperlukan untuk menampung paling sedikit semua minuman tersebut.

4 Aritmatika sosial	<p>Sebuah toko menjual sebuah sepatu dengan harga Rp240.000,00. Ternyata, sepatu tersebut mengalami penurunan harga sebesar 20% karena diskon. Setelah seminggu, sepatu itu dinaikkan kembali harganya sebesar 25% dari harga setelah diskon. Bandingkan harga akhir sepatu dengan harga awalnya. Apakah lebih mahal, lebih murah, atau sama? Jelaskan alasanmu dengan perhitungan yang tepat.</p>	C4
---------------------	--	----

Data dikumpulkan melalui pemberian instrumen kepada responden penelitian. Responden diminta mengerjakan empat soal tersebut dengan waktu penggerjaan maksimal 60 menit. Kedua kelas melaksanakan tes pada waktu yang berbeda, namun peneliti memastikan bahwa soal tidak diketahui oleh kelas yang belum mengikuti tes dengan cara mengumpulkan seluruh lembar soal yang telah dikerjakan sebelum diberikan kepada kelas berikutnya.

Sebelum instrumen digunakan oleh siswa, terlebih dahulu dilakukan validasi oleh ahli, yang terdiri dari 3 orang guru SMPN yang ada di kota Banjarmasin. Guru menilai dari aspek kesesuaian isi, kelayakan teknis, dan bahasa. Validator menilai di lembar validasi dari setiap aspek dengan kategori tidak sesuai sampai sangat sesuai, dengan skor satu sampai empat. Berdasarkan hasil penilaian tersebut, skor tiga atau empat dikategorikan sebagai layak dengan diberi nilai satu, sedangkan butir yang memperoleh skor satu atau dua dikategorikan tidak layak dan diberi nilai nol (Almanasreh et al., 2019). Indeks validitas isi (CVI) dapat dihitung pada tingkat butir (item-level CVI atau I-CVI) maupun pada tingkat keseluruhan instrumen (instrument-level CVI) (Yusoff, 2019). Nilai I-CVI diperoleh dengan membagi jumlah validator yang memberikan kategori layak (nilai 1) dengan jumlah seluruh validator. Dengan jumlah tiga validator, nilai I-CVI sebesar 1 menunjukkan bahwa butir tersebut memenuhi kriteria, sedangkan nilai kurang dari 1 menandakan perlunya revisi (Almanasreh et al., 2019). Selain itu, reliabilitas penilaian validator dihitung menggunakan koefisien kappa, dengan rumus:

$$\kappa = \frac{(I - CVI - PC)}{(1 - PC)}$$

di mana PC dihitung menggunakan rumus:

$$PC = \frac{N!}{(A!(N-A)!)} \times 0.5^n$$

Zamanzadeh et al., (2015), dengan N menunjukkan jumlah validator dan A menunjukkan banyaknya validator yang memberikan skor 3 atau 4. Nilai κ yang dianggap memadai adalah $\geq 0,6$. Setelah dinyatakan layak oleh para validator, instrumen kemudian diuji secara empiris menggunakan Model Rasch melalui aplikasi RStudio untuk menilai kelayakan dan kualitas butir instrumen.

HASIL DAN PEMBAHASAN

Tabel 3 menunjukkan hasil I-CVI dan koefisien Kaffa berdasarkan aspek-aspek yang dinilai oleh validator. Seperti yang terlihat pada Tabel 3, nilai I-CVI untuk soal nomor 1, 2, 3 dan 4 telah memenuhi kriteria yang dipersyaratkan dan dinyatakan layak. Meskipun demikian, validator memberikan masukan pada soal nomor 3 agar konteksnya lebih kontekstual, yaitu dengan mengubah jumlah air mineral menjadi $6\frac{1}{2}$ atau $6\frac{1}{4}$ karena dianggap lebih kontekstual dibandingkan $6\frac{3}{8}$ dan menyarankan agar botol yang digunakan 500 ml untuk memudahkan siswa dalam menghitung.

Tabel 3.
Hasil I-CVI dan Koefisien Kaffa

Indikator Penilaian	Soal							
	1		2		3		4	
	I-CVI	Kap pa						
Aspek Kesesuaian Isi								
1. Soal sesuai dengan kompetensi dasar dan indikator kemampuan pemecahan masalah	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2. Materi dalam soal sesuai dengan tingkat perkembangan kognitif siswa SMP.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3. Soal mencerminkan konteks kehidupan nyata yang relevan (berbasis kontekstual).	1,00	1,00	1,00	1,00	0,67	0,78	1,00	1,00
4. Setiap butir mengukur satu aspek kemampuan pemecahan masalah yang jelas.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
5. Gambar, tabel, atau data dalam soal akurat dan bermakna.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Indikator Penilaian	Soal							
	1		2		3		4	
	I-CVI	Kap pa						
Aspek Konstruksi (Kelayakan Teknis)								
1. Rumusan soal jelas dan tidak menimbulkan makna ganda.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2. Petunjuk pengeraaan soal mudah dipahami siswa.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3. Soal memiliki tingkat kesukaran yang sesuai dengan kemampuan siswa SMP.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
4. Opsi atau langkah jawaban tidak menyesatkan.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
5. Format penulisan, simbol, dan satuan matematis digunakan secara konsisten.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Aspek Bahasa								
1. Bahasa yang digunakan komunikatif dan sesuai dengan kaidah bahasa Indonesia.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2. Kalimat soal tidak terlalu panjang atau berbelit-belit.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3. Istilah matematika digunakan dengan tepat.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
4. Tidak terdapat kesalahan ejaan atau tanda baca.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Rerata	1,00	1,00	1,00	1,00	0,96	0,98	1,00	1,00

Berdasarkan Tabel 3 menurut Almanasreh et al. (2019), soal nomor 3 harus dilakukan revisi. Revisi dilakukan sesuai saran dan menjadi dasar penyempurnaan instrumen sebelum uji empiris yang disajikan pada Tabel 4.

Tabel 4.
Revisi Soal Nomor 3

Soal Sebelum Revisi	Soal Setelah Revisi
Bu Susi membuat minuman dengan mencampurkan air lemon $2\frac{1}{2}$ liter, air mineral $6\frac{3}{8}$ liter, dan air gula putih $1\frac{1}{4}$ liter. Campuran tersebut kemudian dimasukkan ke dalam botol dengan volume 600 ml. Tentukan banyaknya botol yang diperlukan untuk menampung paling sedikit semua minuman tersebut.	Bu Susi membuat minuman dengan mencampurkan air lemon $2\frac{1}{2}$ liter, air mineral $6\frac{1}{2}$ liter, dan air gula putih $1\frac{1}{4}$ liter. Campuran tersebut kemudian dimasukkan ke dalam botol dengan volume 500 ml. Tentukan banyaknya botol yang diperlukan untuk menampung paling sedikit semua minuman tersebut.

Kemudian instrumen diujicobakan kepada siswa SMPN VIII di dua kelas yang berbeda. Selanjutnya diberikan skor dan digunakan rasch model di R studio. Hasil Validitas butir di Rstudio berdasarkan nilai Infit dan Outfit MNSQ, disajikan pada Gambar 1.

item	Mean_Infit	Mean_Outfit	Status
<chr>	<dbl>	<dbl>	<chr>
1 Soal1	1.09	0.935	Layak
2 Soal2	0.991	0.843	Layak
3 Soal3	1.04	1.19	Layak
4 Soal4	1.21	1.24	Layak

Gambar 1. Hasil Mean_Infit dan Mean_Outfit

Nilai Mean_Infit dan Mean_Outfit yang berada dalam rentang 0,5 hingga 1,5 dikategorikan layak (Davis & Boone, 2021). Berdasarkan Gambar 1, nilai Mean_Infit berkisar antara 0,991 hingga 1,21, sedangkan Mean_Outfit berada pada rentang 0,843 hingga 0,935. Sehingga, seluruh butir dinyatakan memiliki validitas yang layak. Selanjutnya Tabel 5. menyajikan hasil reliabilitas.

Tabel 5. Reliabilitas instrumen

Jenis Reliabilitas	Nilai
Croanbach Alpha	0,709
WLE Reliability	0,678
EAP Reliability	0,772

Berdasarkan Lacave et al. (2018) Cronbach's alpha $\geq 0,7$, reliabel, pada Tabel 6 diperoleh nilai 0,709 sehingga dapat dikatakan instrumen reliabel. Selanjutnya, nilai WLE reliability diperoleh sebesar 0,678, yang mencerminkan reliabilitas moderat. Meskipun berada sedikit di bawah batas ideal, hasil ini tetap dapat diterima dalam konteks pengembangan instrumen pendidikan. Hal ini sejalan dengan temuan Davis & Boone (2021), melaporkan WLE reliabilitas berkisar antara 0,60–0,74 tetap menyatakan

instrumen mereka dapat diterima. Rentiana et al. (2024), juga menjelaskan bahwa nilai WLE reliabilitas sekitar 0,65–0,70 tergolong sedang dan masih layak untuk digunakan, dengan nilai ideal berada di atas 0,80. Sementara itu, nilai EAP reliabilitas sebesar 0,772 dapat diterima, sesuai dengan temuan Bean & Bowen, (2021), Wess et al. (2021), dan Kılıç et al. (2023) yang menyatakan bahwa EAP reliabilitas lebih dari 0,7 menunjukkan konsistensi pengukuran yang memadai. Oleh karena itu, secara keseluruhan ketiga indikator reliabilitas memenuhi untuk menyatakan instrumen kemampuan pemecahan masalah yang dikembangkan telah memiliki konsistensi pengukuran yang memadai dan dapat digunakan.

Pada Gambar 1 juga dapat dianalisis daya beda dari soal yang digunakan. Hampir semua kategori skor pada setiap item fit terhadap model Rasch, artinya pola respons siswa konsisten dengan prediksi model. Item-item dengan *Outfit* mendekati 1,0 menunjukkan daya beda yang baik, sedangkan nilai ekstrem (<0,5 atau >1,5) hanya muncul di beberapa kategori dan tidak signifikan secara keseluruhan. Sehingga, seluruh butir dapat dikatakan memiliki daya beda empiris yang memadai, mampu membedakan kemampuan siswa secara konsisten. Setelah memastikan validitas dan reliabilitas instrumen, langkah selanjutnya adalah menguji asumsi unidimensionalitas.

Hasil uji unidimensionalitas PCAR ditunjukkan pada Gambar 2 dan hasil *loading* masing-masing item ditunjukkan pada Tabel 7.

	Eigenvalue	PropVar	CumVar
1	1.7924	0.4481	0.4481
2	1.2645	0.3161	0.7642
3	0.9028	0.2257	0.9899
4	0.0402	0.0101	1.0000
5	NA	NA	NA
6	NA	NA	NA

Gambar 2.

Hasil uji unidimensionalitas PCAR

Tabel 6.

Hasil Loading masing-masing Item

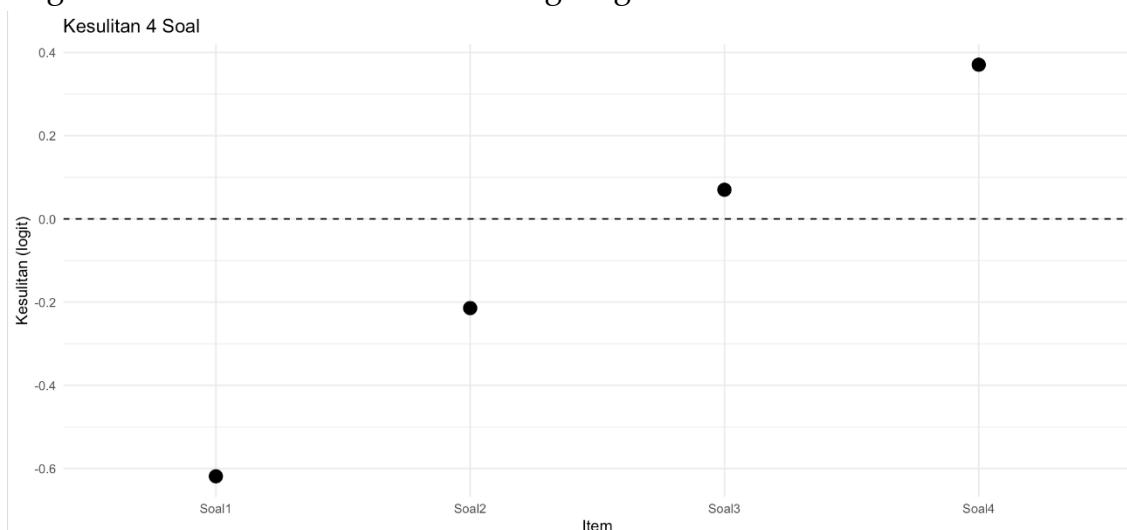
Soal 1	Soal 2	Soal 3	Soal 4
0,53563364	0,02530769	0,70824696	0,45917575

Merujuk pada kriteria yang digunakan oleh (Muttaqin et al., 2024), instrumen dinyatakan unidimensional jika nilai *eigenvalue* pada *first contrast* berada di bawah 3,0, artinya tidak adanya subdimensi substantif dalam residual. Dari Gambar 2, nampak bahwa nilai *eigenvalue first contrast* berada di antara 0,0402 sampai 1,7924, sehingga konstruk kemampuan pemecahan masalah matematika kontekstual bersifat

unidimensional. Temuan ini mengonfirmasi bahwa seluruh butir mengukur satu konstruk utama. Selain itu, pola *loading residual* yang tidak membentuk kelompok butir baru menunjukkan tidak adanya indikasi *local item dependence*. Berdasarkan ini, asumsi unidimensionalitas terpenuhi, dan instrumen layak dianalisis menggunakan Model Rasch.

Analisis loading residual pada Tabel 7, memperlihatkan bahwa Soal 3 memiliki loading terbesar (0,708), sedangkan butir lainnya memiliki loading lebih rendah dan tersebar dalam arah yang berbeda (Soal 1 = -0,536; Soal 4 = -0,459; Soal 2 = -0,025). Pola ini menunjukkan bahwa tidak terdapat dua atau lebih butir yang membentuk kelompok residual yang sama, sehingga tidak muncul indikasi local item dependence maupun subdimensi tersembunyi (Marais & Andrich, 2008). Secara keseluruhan, kombinasi nilai eigenvalue yang rendah dan pola loading residual yang terpencar mendukung asumsi unidimensionalitas, dengan kata lain analisis Rasch PCM dapat diterapkan secara tepat pada instrumen ini.

Selanjutnya parameter kesulitan dari masing-masing soal ditunjukkan pada Gambar 3. Merujuk pada Abdellatif (2023), estimasi parameter kesulitan dalam model Rasch dinyatakan dalam satuan logit, di mana nilai logit yang lebih tinggi menunjukkan item yang lebih sulit, sedangkan nilai logit negatif menunjukkan item yang relatif lebih mudah. Selain itu, sebagaimana dijelaskan oleh Ekstrand et al. (2022), perbedaan nilai logit bersifat linier dan dapat digunakan untuk membandingkan tingkat kesulitan antarbutir secara langsung.



Gambar 3.
Hasil Parameter Kesulitan Masing-Masing Soal

Berdasarkan Gambar 3, dari empat soal, dua soal berada di sisi negatif dan dua soal berada di sisi positif. Kemudian, diperoleh estimasi kesulitan soal satu, nilai kesulitan -0,61 logit, artinya soal ini merupakan soal paling mudah di antara semua soal. Soal dua juga berada pada sisi negatif yaitu -0,20 logit, sehingga juga

tergolong relatif mudah, meskipun sedikit lebih sulit daripada Soal satu. Soal tiga memiliki nilai 0,08 logit, berada sedikit di atas rata-rata, sehingga termasuk tingkat kesulitan sedang. Sedangkan soal empat memiliki nilai kesulitan tertinggi, yaitu 0,38 logit. Oleh karena itu, urutan kesulitan dari yang paling mudah hingga paling sulit adalah Soal satu, Soal dua, Soal tiga, dan Soal empat.

Sehingga, dapat dikatakan instrumen memiliki variasi tingkat kesulitan yang proporsional, karena mencakup rentang item mudah, sedang, hingga relatif sulit. Komposisi seperti ini, sebagaimana disarankan Abdellatif (2023), merupakan karakteristik yang baik dalam suatu instrumen karena memungkinkan pengukuran kemampuan siswa pada berbagai level secara lebih akurat. Selain itu, sebaran logit yang tidak terlalu ekstrem selaras dengan temuan Ekstrand et al. (2022) bahwa item dengan rentang kesulitan moderat cenderung menghasilkan pengukuran yang lebih stabil dan informatif.

Gambar 4 menyajikan kecocokan antara kemampuan siswa dan tingkat kesulitan item. Dari Gambar 4, distribusi kemampuan siswa berdasarkan estimasi EAP berkisar antara -0.684 hingga 0.632 logit, dengan nilai median mendekati 0. Rentang ini tumpang tindih dengan tingkat kesulitan butir (-0.61 hingga 0.38 logit), sehingga sebagian besar kemampuan siswa berada dalam cakupan rentang item. Kondisi ini menunjukkan bahwa instrumen memiliki targeting yang baik, karena distribusi person measure selaras dengan distribusi item difficulty pada skala logit yang sama. Menurut Bond & Fox (2015) dan Boone et al. (2014), targeting optimal terjadi ketika kemampuan responden dan kesulitan item saling tumpang tindih, sehingga setiap siswa memiliki peluang yang proporsional untuk dijelaskan oleh butir yang tersedia. Linacre (2021) dan Tennant & Conaghan (2007) juga menegaskan bahwa instrumen yang baik harus mencakup seluruh rentang kemampuan responden untuk memastikan kualitas pengukuran yang akurat dan bebas bias. Oleh karena itu, data ini mengindikasikan bahwa instrumen yang dikembangkan telah memiliki kualitas targeting yang baik dan layak digunakan dalam pengukuran kemampuan pemecahan masalah.

> summary(model_tam\$person\$EAP)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.6839803	-0.2685779	0.0002243	-0.0696469	0.0185006	0.6318729

Gambar 4.

Rangkuman Kecocokan Antara Kemampuan Siswa dan Tingkat Kesulitan Item

Meskipun instrumen ini telah terbukti valid, reliabel, dan memenuhi asumsi Rasch model, namun masih terdapat keterbatasan. Pertama, instrumen yang digunakan terdiri dari empat soal uraian, di mana setiap materi diwakili oleh satu soal. Format uraian memang memungkinkan pengukuran proses berpikir dan strategi pemecahan masalah secara lebih mendalam, namun soal ini berbeda dari praktik asesmen harian di sekolah. Sejumlah penelitian menunjukkan bahwa guru lebih tertarik menggunakan

teknik tes berupa pilihan ganda untuk menilai hasil belajar peserta didik (Wahiah et al., 2023). Selain itu, (Agustin et al., 2023) menegaskan bahwa tes pilihan ganda memudahkan guru dalam proses penilaian dan memungkinkan pemberian umpan balik yang lebih cepat kepada siswa. Sehingga, perbedaan format antara soal uraian dalam instrumen ini dan format asesmen yang umum digunakan di sekolah perlu menjadi pertimbangan dalam pengembangan lebih lanjut. Pengembangan instrumen pada tahap berikutnya dapat mempertimbangkan penambahan jumlah butir dan variasi format soal agar instrumen lebih kompatibel dengan konteks penilaian di sekolah tanpa mengurangi kekuatan pengukuran yang telah dicapai. Kedua, sampel berasal dari dua kelas dalam satu sekolah, mungkin studi lanjutan dapat melibatkan beberapa sekolah dengan berbeda karakteristik untuk memperkuat temuan.

Secara keseluruhan, terpenuhinya validitas isi, validitas empiris, reliabilitas, unidimensionalitas, daya beda, dan *targeting* yang baik menunjukkan bahwa instrumen ini layak digunakan untuk mengukur kemampuan pemecahan masalah siswa SMP. Instrumen ini berkontribusi secara teoretis dalam memperkuat asesmen yang memadukan konteks autentik dengan pengukuran berbasis model Rasch. Secara praktis, instrumen ini memberikan alat evaluasi yang dapat digunakan guru untuk keperluan penilaian formatif maupun diagnostik, membantu mengidentifikasi kemampuan pemecahan masalah siswa khususnya berbasis kontekstual yang masih menjadi kesulitan siswa.

KESIMPULAN

Penelitian ini menghasilkan instrumen kemampuan pemecahan masalah matematika berbasis kontekstual yang terdiri dari empat soal uraian pada materi perbandingan, bilangan bulat, pecahan, dan aritmatika sosial. Berdasarkan hasil validasi ahli, seluruh butir memenuhi kriteria validitas isi dan satu butir direvisi sesuai saran validator agar lebih kontekstual. Analisis empiris menggunakan Model Rasch menunjukkan bahwa seluruh butir memiliki nilai Infit dan Outfit dalam rentang yang direkomendasikan, sehingga dinyatakan valid secara empiris. Nilai reliabilitas Cronbach's Alpha, WLE reliability, dan EAP reliability berada pada kategori memadai, sehingga instrumen dapat dikatakan konsisten dalam mengukur kemampuan siswa. Asumsi unidimensionalitas turut terpenuhi melalui hasil analisis PCAR, yang menunjukkan bahwa keempat butir mengukur satu konstruk utama. Selain itu, parameter kesulitan item membentuk rentang yang proporsional dari mudah hingga relatif sulit, sehingga instrumen mampu membedakan variasi kemampuan siswa. Distribusi kemampuan siswa yang tumpang tindih dengan tingkat kesulitan item menunjukkan bahwa instrumen memiliki kualitas targeting yang baik. Oleh karena itu, instrumen ini layak digunakan untuk mengukur kemampuan pemecahan masalah

matematika berbasis kontekstual pada siswa SMP. Instrumen ini dapat dimanfaatkan guru untuk keperluan asesmen formatif dan diagnostik, khususnya untuk mengidentifikasi kesulitan siswa dalam menyelesaikan masalah kontekstual. Penelitian selanjutnya disarankan untuk menambah jumlah butir dan memperluas sampel agar cakupan konstruk dan generalisasi hasil menjadi lebih kuat.

DAFTAR PUSTAKA

- Abdellatif, H. (2023). Test results with and without blueprinting: Psychometric analysis using the Rasch model. *Educacion Medica*, 24(3). <https://doi.org/10.1016/j.edumed.2023.100802>
- Agustin, R., Surani, D., Nur Khasanah, A., Septina Pratiwi, K., Nafizah, D., & Intan Mutiara Fajrin, R. (2023). Penggunaan tes pilihan ganda sebagai alat evaluasi di Sekolah Menengah Pertama Negeri 2 Kedawung Sragen. *PANDU: Jurnal Pendidikan Anak Dan Pendidikan Umum*, 1(4), 1-9. <https://doi.org/10.59966/pandu.v1i4.965>
- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Bean, G. J., & Bowen, N. K. (2021). Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work (United States)*, 18(6), 597–618. <https://doi.org/10.1080/26408066.2021.1906813>
- Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and validation of Scientific Inquiry Literacy Instrument (SILI) using Rasch measurement model. *Education Sciences*, 14(3). <https://doi.org/10.3390/educsci14030322>
- Davis, D. R., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, 2. <https://doi.org/10.1016/j.ijedro.2021.100054>
- Ekstrand, J., Westergren, A., Årestedt, K., Hellström, A., & Hagell, P. (2022). Transformation of Rasch model logits for enhanced interpretability. *BMC Medical Research Methodology*, 22(1). <https://doi.org/10.1186/s12874-022-01816-1>
- Gunawan, R. G., Mudjiran, M., Suherman, S., Yerizon, Y., Musdi, E., Armiati, A., & Rozika, E. (2023). An analysis of mathematical problem-solving skills of Junior High School students. *Indonesian Journal of Science and Mathematics Education*, 6(3), 351. <https://doi.org/10.24042/ijjsme.v6i3.18062>
- Hope, D., Kluth, D., Homer, M., Dewar, A., Goddard-Fuller, R., Jaap, A., & Cameron, H. (2024). Exploring the use of Rasch modelling in “common content” items for

- multi-site and multi-year assessment. *Advances in Health Sciences Education*, 30, 427–438. <https://doi.org/10.1007/s10459-024-10354-y>
- Ibrahim, N. L. (2025). Analisis literatur penggunaan model Rasch untuk validasi instrumen asesmen pendidikan: Tren, temuan, dan implikasi. *Lencana: Jurnal Inovasi Ilmu Pendidikan*, 3(2), 46–60. <https://doi.org/10.55606/lencana.v3i2.5012>
- Kemendikbudristek. (2022). Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah pada Kurikulum Merdeka.
- Khoiriyah, A. J., & Husamah, H. (2018). Problem-based learning: Creative thinking skills, problem-solving skills, and learning outcome of seventh grade students. *Jurnal Pendidikan Biologi Indonesia*. <https://doi.org/10.22219/jpbi.v4i2.5804>
- Kılıç, A. F., Koyuncu, İ., & Uysal, İ. (2023). Scale development based on item response theory: A systematic review. *International Journal of Psychology and Educational Studies*, 10(1), 209–223. <https://doi.org/10.52380/ijpes.2023.10.1.982>
- Lacave, C., Molina, A. I., & Redondo, M. A. (2018). A preliminary instrument for measuring students' subjective perceptions of difficulties in learning recursion. *IEEE Transactions on Education*, 61(2), 119–126. <https://doi.org/10.1109/TE.2017.2758346>
- Mardiansyah, D., Hendriana, H., Supriatna, T., & Putra, H. D. (2025). Analysis of Junior High School students' mathematical problem-solving skills in solving social arithmetic problems. (*JIML*) *JOURNAL OF INNOVATIVE MATHEMATICS LEARNING*, 8(2), 181–192. <https://doi.org/10.22460/jiml.v8i2.23842>
- Mulyati*, Y., 'Ainunnizar, D., Hasan, S., Wicaksono, A., & Ichsan, M. (2024). Enhancing Students' Critical and Creative Thinking Skills Through Problem-Based Learning. *Indonesian Journal of Biology Education*. <https://doi.org/10.31002/ijobe.v7i2.2069>
- Muttaqin, D., Dermawan, K., & Wibaningrum, G. (2024). Validation of the psychological control scale-youth self-report Indonesian version: Rasch model and factor analysis. *Revista CES Psicología*, 17(2), 30–44. <https://doi.org/10.21615/cesp.6673>
- Pratama, D. (2020). Analisis kualitas tes buatan guru melalui pendekatan Item Response Theory (IRT) model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Prayoga, K. P., Suryana, D., Supriatna, M., & Budiman, N. (2024). Penggunaan Rasch model Untuk menganalisis konstruk instrumen kontrol diri pada siswa sekolah menengah. *G-Couns: Jurnal Bimbingan Dan Konseling*, 9(1), 367–381. <https://doi.org/10.31316/gcouns.v9i1.4459>
- Rentiana, L. H., Dhini, U. R., Yuliawati, L., & Wulandari, E. T. (2024). Item Fit Analysis for Evaluating Academic Writing Performance With Rasch Measurement.

- Attractive : Innovative Education Journal, 6(1).
<https://www.attractivejournal.com/index.php/aj/>
- Santriyani, E. (2024). Analisis kemampuan pemecahan masalah matematis dalam menyelesaikan soal cerita pada asesmen kompetensi minimum siswa SMP. *Jurnal Basicedu*, 8(4), 2819–2830. <https://doi.org/10.31004/basicedu.v8i4.8179>
- Sheptian, R., Arthur, R., & Wardhani, R. (2024). Psychometric analysis of mathematics readiness assessment tool for elementary school students: R Aasch model approach. *LEARNING : Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 4(4), 1213–1220. <https://doi.org/10.51878/learning.v4i4.3725>
- Tanudjaya, C. P., & Doorman, M. (2020). Examining higher order thinking in Indonesian lower secondary mathematics classrooms. In *Journal on Mathematics Education* (Vol. 11, Issue 2, pp. 277–300). Sriwijaya University. <https://doi.org/10.22342/jme.11.2.11000.277-300>
- Verdú-Soriano, J., & González-de la Torre, H. (2024). Rasch analysis implementation in nursing research: A methodological approach. *Enfermeria Clinica*, 34(6), 493–506. <https://doi.org/10.1016/j.enfcli.2024.10.005>
- Wahiah, Z., Prabowo, S. M., & Hana, A. S. (2023). Eksplorasi efektivitas tes pilihan ganda berbasis komputer sebagai evaluasi pembelajaran. *EDUCATIVO: Jurnal Pendidikan*, 2(2), 342–347.
- Wess, R., Klock, H., Siller, H.-S., & Greefrath, G. (2021). International Perspectives on the Teaching and Learning of Mathematical Modelling Measuring Professional Competence for the Teaching of Mathematical Modelling A Test Instrument. Springer. <http://www.springer.com/series/10093>
- Yang, Y., Peng, Y., Li, W., Lu, S., Wang, C., Chen, S., & Zhong, J. (2023). Psychometric evaluation of the academic involution scale for college students in China: An application of Rasch analysis. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1135658>
- Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, 4(2), 165–178. <https://doi.org/10.15171/jcs.2015.017>