

Pengembangan Instrumen Kemampuan Pemecahan Masalah Matematika Siswa Menggunakan Model Rasch

Nurul Qomariyah Ahmad^{1*}, Riyan Arthur², Wardani Rahayu³ dan Elfi Rahmadhani⁴

^{1,2,3}Program Studi Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta, Jakarta, Indonesia

⁴Program Studi Tadris Matematika, Institut Agama Islam Negeri Takengon, Takengon, Indonesia

*E-mail: qomariyahnr1@gmail.com

ABSTRACT . Mathematical problem solving skills are important for all students, including fifth grade elementary school students, because they can help in facing the challenges of the times and everyday life that are increasingly complex. This study aims to determine the process of developing and validating students' mathematical problem solving ability instruments psychometrically. The method used is a mixed method with a sequential exploratory type preceded by a qualitative method and then strengthened by a quantitative method. The sampling technique used is purposive sampling. Data analysis in this study used the Rasch model which is part of the item response theory (IRT) in modern test theory through the Winstep application. Based on the results of the data analysis, it can be seen that in the qualitative development process, this instrument is in accordance with the indicators of mathematical problem solving with the achievements and objectives of learning. Then quantitatively according to experts, the mathematical problem solving ability instrument is valid, which means it is suitable for use. Then continued with a field trial on 163 fifth grade student respondents in East Jakarta which resulted in the instrument meeting the requirements of the IRT Rasch model and can be used on average student abilities of around -1 logit to 0 logit.

Keywords: instrument; problem solving skills; rasch model; winstep

ABSTRAK. Kemampuan pemecahan masalah matematika penting dimiliki oleh seluruh siswa termasuk siswa kelas V sekolah dasar, karena dapat membantu dalam menghadapi tantangan zaman dan kehidupan sehari-hari yang semakin kompleks. Penelitian ini bertujuan untuk mengetahui proses pengembangan dan validasi instrumen kemampuan pemecahan masalah matematika siswa secara psikometrika. Metode yang digunakan yaitu *mixed method* dengan jenis *sequential exploratory* yang didahului metode kualitatif dan setelah itu dikuatkan dengan metode kuantitatif. Teknik sampel yang dilakukan adalah *purposive sampling*. Analisis data pada penelitian ini menggunakan model rasch yang merupakan bagian dari teori respon butir (IRT) pada teori tes modern melalui aplikasi *winstep*. Berdasarkan hasil analisis data, dapat diketahui bahwa pada proses pengembangannya secara kualitatif, instrumen ini sudah sesuai antara indikator pemecahan masalah matematika dengan capaian dan tujuan pembelajarannya. Kemudian secara kuantitatif menurut pakar instrumen soal kemampuan masalah matematika sudah valid yang berarti telah layak untuk digunakan. Kemudian dilanjutkan dengan ujicoba lapangan pada 163 responden siswa kelas V di Jakarta Timur yang menghasilkan bahwa instrumen telah memenuhi persyaratan model rasch IRT dan dapat digunakan pada kemampuan siswa rata-rata sekitar -1 logit sampai 0 logit.

Kata kunci: instrumen; kemampuan pemecahan masalah; rash model; winstep

PENDAHULUAN

Menjalankan kehidupan sehari-hari yang penuh tantangan memerlukan usaha agar dapat bertahan di era globalisasi dan persaingan saat ini. Kemampuan pemecahan masalah merupakan salah satu aspek yang mendasar dari matematika yang dapat membantu dalam mengatasi tantangan zaman (Liljedahl dkk, 2016). Kemampuan tersebut merupakan penerapan pengetahuan matematika

untuk memecahkan masalah kehidupan nyata, termasuk yang melibatkan situasi kompleks (OECD, 2023). Mental yang lebih baik ketika menghadapi tantangan dan memecahkan masalah dapat dibentuk apabila siswa dilatih mengenai pola berpikir dan kemampuan pemecahan masalah (Cahyani & Setyawati, 2016). Di sekolah dasar matematika sebagai salah satu mata pelajaran wajib untuk melatih siswa berpikir logis, analitis, sistematis, kritis, dan kreatif yang salah satunya melalui pengerjaan soal-soal pemecahan masalah (Yayuk & As'Ari, 2020).

Kurikulum merdeka di Indonesia juga meminta diasahnya kemampuan ini melalui capaian pembelajaran (CP) yang terdapat pada mata pelajaran matematika. misalnya peserta didik kelas V SD dapat menyelesaikan masalah yang berkaitan dengan uang melalui penjumlahan dan pengurangan bilangan cacah sampai 100.000 (Kepka BSKP, 2024). Tetapi harapan ini berbeda dengan hasil PISA 2022 yang menunjukkan bahwa kemampuan siswa Indonesia dalam matematika peringkat ke 70 dari 81 peserta. Pemecahan masalah matematika, masih jauh tertinggal dibandingkan rata-rata negara OECD. Terjadi penurunan rata-rata skor matematika sebanyak 12 angka dari 379 ke 366 pada tahun 2022, padahal rata-rata OECD sampai 472. Hanya 18% siswa Indonesia yang mencapai tingkat kompetensi minimum (level 2 atau lebih tinggi), sedangkan rata-rata OECD adalah 69%. Bahkan, hampir tidak ada siswa Indonesia yang mencapai level 5 atau 6 (OECD, 2023). Salah satu penyebabnya, ketika dihadapkan soal yang berkaitan dengan pemecahan masalah, peserta didik hanya terprogram mengikuti cara yang rutin, biasa dan terstruktur secara turun temurun tanpa menemukan cara-cara baru yang lebih luas dan lugas. Situasi ini menjadi sangat merugikan ketika benar-benar dihadapkan dengan persoalan matematika yang lebih nyata dalam kehidupan (Ahmad, dkk, 2023).

Pada tujuan pembelajaran matematika terdapat metode, prosedur, dan strategi yang ada juga diarahkan ke pemecahan masalah agar dapat membantu siswa meningkatkan kemampuan matematika (Hendriana, dkk., 2018). Belajar matematika melalui pemecahan masalah berarti siswa bukan hanya belajar memecahkan masalah yang ada di soal, tetapi membantu dalam penerapannya sehari-hari dapat membantu siswa untuk mengembangkan pemahaman, penalaran bahkan sampai pada solusi daripada sekedar mengingat serta menerapkan prosedur (Szabo dkk, 2020). Pengaplikasian kemampuan pemecahan masalah matematika dapat dilakukan pada kehidupan sehari-hari dikarenakan berkaitan juga dengan pengetahuan dari mata pelajaran yang lain (Salwida dkk., 2022). Penggunaannya pula membutuhkan proses berkaitan dengan pengetahuan masa lalu dengan yang dialaminya sekarang (Samosir, dkk., 2020).

Pada hakikatnya kemampuan pemecahan masalah matematika merupakan kemampuan untuk memahami masalah, merancang strategi penyelesaian, menerapkan strategi tersebut, dan memeriksa kembali solusi yang telah diperoleh yang memungkinkan siswa untuk menggunakan konsep matematika dalam situasi yang beragam dan tidak terstruktur (Liljedahl dkk, 2016). Agar dapat menyelesaikan masalah pada pemecahan masalah siswa harus memiliki pemahaman konsep matematika sehingga paham terhadap masalah yang dihadapi (Virgia dkk, 2019). Untuk lebih rinci, Polya (1973), mengemukakan empat tahapan penting yang perlu dilakukan yaitu: 1) memahami masalah, meliputi memahami permasalahan yang ditanyakan, mengetahui data yang diperlukan untuk memecahkan masalah; 2) menentukan rencana, meliputi menemukan kaitan masalah satu dan yang lainnya lalu merencanakan cara penyelesaiannya; 3) melaksanakan rencana, menyelesaikan permasalahan sesuai dengan tahapannya; dan 4) memeriksa kembali, meliputi menguji kembali tahapan yang sudah dilaksanakan dan memeriksa kembali hasil. Pada pemecahan masalah matematika, siswa dilatih mengembangkan kemampuannya dengan menambahkan pengetahuan matematika, dapat memecahkan masalah sehari-hari terutama yang terkait dengan konsep matematika dengan strategi yang sudah dikembangkan.

Untuk dapat mengetahui kemampuan pemecahan matematika siswa, perlu disusun instrument evaluasi yang valid dan reliabel serta layak untuk dijadikan alat evaluasi. Oleh karena itu, perlu dikembangkan sebuah instrument evaluasi yang khusus mengandung indikator-indikator pemecahan masalah. Instrumen tersebut haruslah melalui proses validasi agar dapat dinyatakan layak untuk digunakan baik dari segi uji kelayakan isi maupun dari segi uji empiris.

Pengembangan instrumen menggunakan model rasch adalah proses penyusunan alat ukur atau tes yang didasarkan pada teori respons butir (*Item Response Theory, IRT*) atau teori tes modern, untuk menganalisis data hasil pengukuran, dengan tujuan memastikan bahwa instrumen yang dikembangkan memiliki validitas dan reliabilitas yang tinggi, kemampuannya untuk mengidentifikasi butir yang tidak sesuai (*misfit items*) sehingga dapat dilakukan perbaikan instrumen (Bond dkk, 2020). Teori respon butir (IRT) sebagai salah satu alat yang penting untuk menyelidiki makna yang melekat pada respons ketika butir bersifat kategoris dan mendapatkan skor sifat laten yang mendasari responden (de Ayala, 2022). IRT sendiri merupakan teori tes modern yang agak berbeda dengan teori tes klasik, karena pada IRT secara eksplisit membedakan karakteristik butir dari karakteristik peserta tes (Kondratek, 2022). IRT dapat mengestimasi tingkat kesulitan butir soal dan kemampuan peserta tes secara terpisah, serta mengevaluasi parameter-parameter tersebut dalam skala yang sama, sedangkan teori tes klasik untuk mengevaluasi butir tergantung dari responden. Oleh karena itu, kita dapat memperoleh informasi mengenai kemampuan peserta ujian dan tingkat kesulitan butir soal secara terpisah (Hori dkk, 2022). Dengan kata lain, kita dapat melakukan pengukuran bebas butir dan pengukuran bebas sampel secara bersamaan dengan menggunakan IRT.

IRT dapat membantu mengatasi masalah teknis seperti analisis butir soal, reliabilitas skor, penyalarsan skala, dll., yang terkait dengan keadilan, kualitas, dan validitas yang melekat pada pertanyaan yang terkait dengan pengembangan, administrasi, pemeliharaan, penafsiran, dan penggunaan tes. Fitur-fitur yang menonjol dari model IRT termasuk penggunaan variabel laten untuk mewakili konstruk yang sedang dinilai seperti penguasaan, pencapaian, sikap, atau tingkat keparahan suatu gangguan; penggunaan langsung data tingkat butir (terutama kategorikal) yang dikumpulkan dari interaksi peserta tes atau responden dengan contoh penilaian; serta ketergantungan pada kemandirian bersyarat sebagai struktur dasar dan resep untuk membangun model (Cai, dkk, 2016). Rasch model itu sendiri memperlihatkan mekanisme yang mentransmisikan variasi atribut ke hasil yang diamati dan dapat memberikan spesifikasi untuk pembuatan instrumen tetapi juga menghasilkan kalibrasi yang dapat digunakan untuk mengubah hitungan menjadi jumlah (Stenner, dkk., 2013).

Kurang optimalnya instrumen pembelajaran yang digunakan di sekolah, dibutuhkan pengembangan instrumen penilaian yang sudah teruji menjadi kebutuhan mendesak untuk mendukung pengukuran kemampuan pemecahan masalah siswa secara akurat. Adapun tujuan dari penelitian ini untuk mengetahui proses pengembangan dan kualitas instrumen kemampuan pemecahan masalah matematika siswa kelas V sekolah dasar secara psikometrika yang telah dikembangkan.

METODE

Metode penelitian yang digunakan *mixed method* (Creswell & Creswell, 2023), yaitu adanya metode kualitatif terlebih dahulu yang digunakan setelah itu dilanjutkan dengan metode kuantitatif. Pada tahap satu, metode kualitatif dengan melakukan kajian pustaka tentang kemampuan pemecahan masalah matematika siswa untuk mendapatkan indikatornya dan melakukan analisis kurikulum untuk mendapatkan capaian, elemen dan tujuan pembelajaran yang sesuai untuk siswa kelas V sekolah dasar. Selanjutnya mengembangkan instrumen berdasarkan indikator dan tujuan pembelajaran yang telah didapat.

Pada tahap kedua, dilakukan metode kuantitatif yang diawali dengan melakukan validasi isi dan konstruk melalui ahli, lalu dianalisis dengan melihat saran dari para ahli dan menghitung menggunakan rumus Aiken V agar mendapatkan instrumen yang valid. Dilanjutkan tahap terakhir yaitu uji coba instrumen yang dilakukan di beberapa sekolah dasar untuk siswa kelas V di daerah Jakarta pada akhir bulan November 2024. Metode pengambilan sampel yang digunakan adalah *purposive sampling* untuk mendapatkan data yang lebih memastikan data yang dikumpulkan sesuai dengan tujuan penelitian dan mendapatkan wawasan spesifik untuk memberikan pemahaman

mendalam terhadap fenomena yang diteliti (Cohen dkk., 2018). Setelah itu dilakukan perhitungan hasil uji coba untuk mengetahui validitas secara psikometri dengan menggunakan *rasch model* sebagai bagian dari IRT (teori tes modern) dengan aplikasi *winstep*, dimana antara butir pada instrumen tidak tergantung pada responden (Hori dkk., 2022)

HASIL DAN PEMBAHASAN

Tahap Kualitatif

Pada tahap ini dilakukan kajian kepustakaan mengenai kemampuan pemecahan masalah matematika siswa, didapatkan 4 indikator (Polya, 1973) yaitu , a) memahami masalah seperti memahami pertanyaan yang dimaksud, mengetahui data yang ada dan yang diperlukan untuk menyelesaikan permasalahan, b) menentukan rencana dengan menemukan kaitan masalah atau data yang satu dengan yang lainnya lalu menentukan langkah yang akan diambil untuk menyelesaikan masalah, c) seluruh langkah pemecahan dilaksanakan melalui pelaksanaan rencana dan terakhir d) hasil yang didapat di interpretasikan untuk mengetahui yang sudah dilakukan sesuai atau tidak dengan memeriksa kembali. Setelah itu dilakukan juga analisis kurikulum yang meliputi analisis elemen, capaian, dan tujuan pembelajaran pada siswa kelas V sekolah dasar yang sesuai seperti tabel di bawah ini.

Tabel 1. Elemen, Capaian, dan Tujuan Pembelajaran Matematika Kelas V Sekolah Dasar

Elemen	Capaian Pembelajaran	Tujuan Pembelajaran
Bilangan	Pada akhir kelas V, peserta didik dapat menunjukkan pemahaman dan intuisi bilangan (number sense) pada bilangan cacah sampai 1.000.000. Mereka dapat membaca, menulis, menentukan nilai tempat, membandingkan, mengurutkan, serta melakukan komposisi dan dekomposisi bilangan tersebut. Mereka juga dapat menyelesaikan masalah yang berkaitan dengan uang. Mereka dapat melakukan operasi penjumlahan dan pengurangan bilangan cacah sampai 100.000. Mereka juga dapat menyelesaikan masalah yang berkaitan dengan KPK dan FPB.	Menentukan nilai tempat bilangan cacah sampai 100.000. Menghitung penjumlahan dan pengurangan bilangan cacah sampai 100.000 Menghitung penjumlahan bilangan cacah sampai 100.000 Menjumlahkan dengan menyamakan pecahan biasa Menghitung pengurangan bilangan cacah sampai 100.000 FPB
Aljabar	Pada akhir kelas V, peserta didik dapat mengisi nilai yang belum diketahui dalam sebuah kalimat matematika yang berkaitan dengan operasi penjumlahan, pengurangan, perkalian, dan pembagian bilangan cacah sampai 1.000 (contoh: $10 \times \dots = 900$ dan $900 : \dots = 10$).	Menghitung perkalian bilangan cacah sampai 1.000 Menghitung pembagian bilangan cacah sampai 1.000 Menghitung perkalian dan pembagian bilangan cacah sampai 1.000

Selanjutnya dibuat instrumen dari indikator dan tujuan pembelajaran yang sudah didapatkan pada tahap defenisi sebelumnya. Instrumen terdiri dari kisi-kisi, soal, kunci jawaban dan kartu soal dengan jumlah soal 15 butir.

Tabel 2. Kisi-Kisi Instrumen Kemampuan Pemecahan Masalah Matematika Siswa

Elemen	Tujuan Pembelajaran	Indikator Kemampuan Pemecahan Masalah	Nomor Butir
Bilangan	Menentukan nilai tempat bilangan cacah sampai 100.000.	Memahami masalah	1
	Menghitung penjumlahan dan pengurangan bilangan cacah sampai 100.000	Menentukan rencana	6

Elemen	Tujuan Pembelajaran	Indikator Kemampuan Pemecahan Masalah	Nomor Butir
Aljabar	Menghitung penjumlahan bilangan cacah sampai 100.000	Melaksanakan rencana	15
	Menghitung penjumlahan dan pengurangan bilangan cacah sampai 100.000	Memeriksa kembali	3
	Menjumlahkan dengan menyamakan pecahan biasa	Memahami masalah	12
	Menghitung pengurangan bilangan cacah sampai 100.000	Menentukan rencana	14
	FPB	Melaksanakan rencana	2
	Menghitung penjumlahan dan pengurangan bilangan cacah sampai 100.000	Memeriksa kembali	5
	Menghitung perkalian bilangan cacah sampai 1.000	Memahami Masalah	9
	Menghitung pembagian bilangan cacah sampai 1.000	Merencanakan masalah	10
	Menghitung perkalian bilangan cacah sampai 1.000	Melaksanakan rencana	11
	Menghitung pembagian bilangan cacah sampai 1.000	Memeriksa kembali	4
	Menghitung perkalian dan pembagian bilangan cacah sampai 1.000	Memahami masalah	8
	Menghitung perkalian dan pembagian bilangan cacah sampai 1.000	Merencanakan masalah	13
	Menghitung pembagian bilangan cacah sampai 1.000	Melaksanakan rencana	7

Tahap Kuantitatif

Pada tahapan ini instrumen yang telah dibuat divalidasi secara isi dan konstruk oleh 3 orang ahli yang terdiri dari dua ahli matematika dengan gelar doktor dan satu orang ahli pengukuran bergelar doktor. Ahli juga memberikan saran untuk butir soal nomor 13, bahasa yang digunakan disesuaikan dengan siswa kelas V SD dan indikator kemampuan pemecahan masalah yang terdiri dari 4 indikator, nampaknya harus benar-benar terlihat perbedaan antara satu dengan yang lainnya, pada indikator menentukan rencana dan melaksanakan rencana, jangan sampai redaksi soalnya mirip. Adapun validasi isi dan konstruk dihitung menggunakan rumus Aiken V (Hastuti dkk, 2023):

$$V = \frac{\sum s}{[n(c - 1)]}$$

$$s = r - lo$$

Keterangan:

- V = rata-rata validasi secara keseluruhan
- $\sum s$ = hasil dari skor penilai dikurangi dengan skor terendah
- lo = skor penilaian validitas terendah
- c = skor penilaian validitas tertinggi
- r = skor yang diberikan oleh penilai

Menurut Aiken apabila koefisien validitas isi, dikatakan baik jika koefisien validitas minimum adalah 0,82 (Hastuti dkk, 2023).

Tabel 3. Hasil Perhitungan Aiken V

No	Indikator	Aiken V	Kriteria
1	Materi	0,89	Valid
2	Konstruk	0,86	Valid
3	Bahasa	0,83	Valid
	Rata-rata	0,86	Valid

Berdasarkan tabel 3, hasil validasi isi dan kontruk dari para ahli didapatkan instrumennya valid dengan rata-rata 0,86 yang berarti bisa digunakan untuk tahap uji coba. Pensampelan *purposive* didapatkan beberapa sekolah SD di Jakarta Timur yaitu SDN 01 Cibubur, SDN 05 Pulogebang dan MI Al Wathoniyah 19 dengan seluruh responden berjumlah 163 siswa yang dikerjakan dengan waktu 2 jam pelajaran atau 70 menit untuk 15 butir soal pilihan ganda. Data yang didapat dianalisis menggunakan rasch model melalui aplikasi *winstep*. Berdasarkan hasil analisisnya didapatkan hasil uji prasyarat berupa:

Unidimensionalitas

Memastikan unidimensionalitas sangat penting untuk analisis psikometrik yang akurat, karena memastikan konstruk tunggal pada instrumen (Strachan dkk, 2022). Jika tes tidak unidimensional, estimasi parameter model, seperti kesulitan item dan kemampuan orang, dapat terpengaruh, yang mengarah ke hasil yang tidak dapat diandalkan (Meyer, 2014). Ini sejalan dengan konsep bahwa unidimensi membutuhkan satu faktor umum untuk menjelaskan korelasi antar item. Jika ada beberapa dimensi, sebagaimana dibuktikan dengan nilai eigen yang tinggi, asumsi model satu faktor dilanggar, yang mengarah pada kesimpulan bahwa pengujian mengukur lebih dari satu konstruk yang mendasarinya. Dengan demikian, unidimensionalitas tes dikompromikan (Falissard, 1999).

Tabel 4. Kriteria Unidimensionalitas

Nilai Persentase <i>Raw Variance Explained by Measures "Observed"</i>	Kriteria
>60 %	Sangat baik
40 % - 60 %	Baik
20 % - 40 %	Cukup
≥ 20 %	Minimum
< 20 %	Buruk
< 15 %	Varian Tidak Terduga

Asumsi unidimensionalitas terpenuhi apabila besaran *raw variance explained by measures* > 40% (Perera dkk., 2018; Sumintono, 2018).

	Eigenvalue	Observed	Expected
Total raw variance in observations =	25.4421	100.0%	100.0%
Raw variance explained by measures =	10.4421	41.0%	41.3%
Raw variance explained by persons =	2.8634	11.3%	11.3%
Raw Variance explained by items =	7.5787	29.8%	29.9%
Raw unexplained variance (total) =	15.0000	59.0%	58.7%
<u>Unexplned</u> variance in 1st contrast =	1.9004	7.5%	12.7%
<u>Unexplned</u> variance in 2nd contrast =	1.6817	6.6%	11.2%
<u>Unexplned</u> variance in 3rd contrast =	1.5183	6.0%	10.1%
<u>Unexplned</u> variance in 4th contrast =	1.3248	5.2%	8.8%
<u>Unexplned</u> variance in 5th contrast =	1.2972	5.1%	8.6%

Gambar 1. Unidimensionalitas pada Eigenvalue

Unidimensionalitas dilihat pada *raw variance explained by measure* terletak di kolom *observed*. Hasil yang diperoleh tersebut sebesar 41%. Selain itu terdapat kriteria kedua bahwa *unexplained variance in 1st contrast* pada eigenvalue nilainya tidak lebih besar dari 2.0 karena keberadaan nilai eigen di atas 2 dapat diartikan sebagai bukti multidimensi dalam struktur uji (Burg, 2007). Pada gambar 1, secara berturut-turut *unexplained variance in 1st – 5th contrast eigenvalue* 1, 9004, 1,5183, 1,3248 dan 1,2972 yang berarti kurang dari 2 membuktikan unidimensional. Selanjutnya *unexplained variance in 1st – 5th contrast observed* yang kriterianya dapat dilihat pada tabel 5. (Perera, dkk., 2018).

Tabel 5. Kriteria Variansi *Unexplained*

Variansi Unexplained pada 1st - 5 th Residual PCA	Kriteria
< 3 %	Sangat baik
3 – 5 %	Baik
5 – 10 %	Cukup
10 - 15 %	Minimum
> 15 %	Buruk

Berdasarkan gambar 1. variansi yang tidak dapat dijelaskan pada kolom *observed* secara berurutan, yaitu: 7,5%, 6,6%, 6,0%, 5,2% dan 5,1%. Bila dibandingkan dengan tabel 5. maka masih mendapatkan kriteria cukup. Dengan terpenuhinya unidimensionalitas berarti secara empirik instrumen termasuk unidimensi dan membangun validitas konstruk.

1. Lokal Independen

CORRELATION	ENTRY NUMBER	ENTRY NUMBER	ENTRY NUMBER	ENTRY NUMBER
.22	14	14	15	15
.20	3	3	5	5
-.39	12	12	13	13
-.34	1	1	3	3
-.32	10	10	11	11
-.27	3	3	6	6
-.22	7	7	15	15
-.22	1	1	2	2
-.21	12	12	14	14
-.21	5	5	9	9
-.21	2	2	4	4
-.21	12	12	15	15
-.20	9	9	10	10
-.20	4	4	10	10
-.20	6	6	8	8
-.20	4	4	12	12
-.20	6	6	7	7
-.18	5	5	6	6
-.18	9	9	11	11
-.17	9	9	12	12

Gambar 2. Korelasi pada Lokal Independen

Berdasarkan gambar 2, mengacu pada kriteria indeks Q3 yang telah ditentukan, ditemukan bahwa korelasi residual antar pasangan item tidak ada yang > 0,30 (das Nair dkk., 2011), artinya tidak terdapat pasangan item yang mengalami lokal dependen. Item yang memiliki korelasi residual tertinggi adalah item 14 dan item 15 sebesar 0,22, item 3 dan item 5 sebesar 0,20. Ini mengacu pada gagasan bahwa item tes tidak boleh terkait erat, karena ini dapat menyebabkan analisis statistik yang tidak dapat diandalkan (Ha, 2022). Independensi lokal menyiratkan bahwa satu-satunya hal yang menyebabkan item-item menjadi bervariasi adalah variabel laten yang dimodelkan. Pelanggaran terhadap asumsi ini, yang secara tepat dinamakan ketergantungan lokal, dapat menimbulkan konsekuensi serius terhadap parameter yang diestimasi (Edwards dkk., 2018). Dengan kata lain, asumsi lokal independen pada penelitian ini telah terpenuhi. Temuan ini sejalan dengan terpenuhinya asumsi unidimensionalitas.

Sifat Monotonisasi

CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	INFINIT AVRGE	OUTFIT EXPECT	COHERENCE MNSQ	M->C	C->M	ESTIM RMSR	DISCR	
0	0	1116	46	-1.074	-1.07	1.00	.95	79%	74%	.4162	0
1	1	1329	54	1.335	1.334	1.01	.97	79%	84%	.3517	1

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

Gambar 3. Monotonisasi

Monotonisasi (invarinitas kelompok) merupakan rentang kemampuan pada kelompok responden. Pemenuhan prasyarat pada IRT, sifat monotonisasi nilainya harus cenderung naik dari kecil ke besar (Linacre, 2012). Pada gambar 3, monotonisasi terlihat pada kolom *observed average* yang bergerak dari -1,074 naik ke 1,335, yang menandakan adanya kenaikan, walaupun dari negatif menuju pada positif. Langkah selanjutnya adalah mengetahui validitas instrumen secara psikometri berupa tingkat kesulitan butir, *person fit*, *item fit*, reliabilitas, *separation*, *precision of measurement* dan fungsi informasi dari gambar yang dijelaskan sebagai berikut:

Tingkat Kesulitan Butir

Instrumen yang sudah dibuat setelah di ujicobakan harus dilihat tingkat kesulitan pada setiap butirnya, dapat dikelompokkan menjadi tingkat kesulitan sangat mudah, mudah, sulit, dan sangat sulit berdasarkan kemampuan yang dibutuhkan untuk menjawabnya. Kriteria pengelompokkannya dapat dilihat pada tabel 6.

Tabel 6. Kriteria Tingkat Kesulitan Butir

Nilai Logit	Kriteria
$Measure\ logit < -SD$	Sangat Mudah
$-SD \leq Measure\ logit < 0,00$	Mudah
$0,00 \leq Measure\ logit \leq SD$	Sulit
$Measure\ logit > SD$	Sangat Sulit

Terkait tingkat kesulitan pada hasil penelitian dapat dilihat pada gambar 4 berikut.

ITEM STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
13	19	163	2.560	.257	1.03	.23	.92	-.18	.26	.29	88.3	88.7	13
3	22	163	2.372	.242	1.04	.27	1.08	.39	.24	.30	87.7	87.0	3
2	23	163	2.315	.238	.85	-.93	.72	-1.15	.46	.30	85.9	86.4	2
8	27	163	2.102	.224	1.08	.60	1.15	.70	.22	.32	84.0	84.1	8
4	61	163	.834	.175	1.04	.57	1.05	.48	.33	.37	66.3	68.9	4
6	62	163	.804	.174	.89	-1.72	.82	-1.80	.49	.37	75.5	68.5	6
9	92	163	-.070	.171	1.05	.83	1.03	.35	.33	.38	65.6	66.7	9
10	116	163	-.821	.186	1.12	1.41	1.17	1.17	.23	.36	73.0	74.3	10
11	118	163	-.891	.188	1.02	.22	1.04	.29	.34	.35	73.0	75.2	11
5	121	163	-1.000	.192	1.04	.45	1.10	.69	.30	.35	74.8	76.6	5
12	121	163	-1.000	.192	1.06	.60	.97	-.11	.31	.35	72.4	76.6	12
1	126	163	-1.192	.200	1.14	1.31	1.31	1.63	.18	.34	77.3	78.8	1
15	131	163	-1.402	.210	.80	-1.78	.63	-2.03	.54	.32	85.3	81.3	15
14	136	163	-1.636	.224	.81	-1.40	.65	-1.63	.50	.31	84.7	83.8	14
7	154	163	-2.974	.353	1.01	.13	.75	-.45	.23	.21	94.5	94.5	7
MEAN	88.6	163.0	.000	.215	1.00	.05	.96	-.11			79.2	79.4	
P. SD	46.4	.0	1.667	.045	.10	.99	.20	1.06			8.3	7.8	

Gambar 4. Tingkat Kesulitan Butir

Berdasarkan gambar 4. terlihat butir yang paling sulit atau sukar pada nomor 13, dengan tingkat kesukaran 2,560 logit yang artinya siswa dapat menjawab dengan benar butir tersebut apabila memiliki kemampuan minimal 2,560 logit dan butir yang paling mudah pada nomor butir 7, dengan nilai logitnya -2, 974 memiliki makna siswa dengan kemampuan minimal -2,974 logit dapat menjawab butir tersebut. Selanjutnya butir-butir pada instrumen dapat diklasifikasikan tingkat kesukaran dari paling mudah sampai paling sukar berdasarkan standar deviasinya (SD) yang dirangkum pada tabel 7.

Tingkat kesulitan merupakan ukuran atau parameter yang digunakan untuk menggambarkan seberapa sulit atau mudah (Rush dkk, 2016), mengetahui tingkat komplektesitas butir soal. Pada penentuan tingkat kesulitan akan dapat menentukan kualitas butir soal dan memastikan bahwa soal tersebut sesuai dengan tingkat kemampuan siswa yang menjadi targetnya.

Tabel 7. Tingkat Kualitas Butir

Nilai Logit	Kate	Nomor Butir	Jumlah	Persentase
$Measure\ logit > 1.67$	Sangat Sukar	13, 3, 2, 8	4	26,67%
$0 \leq measure\ logit \leq 1.67$	Sukar	4,6	2	13,33%
$-1.67 \leq measure\ logit \leq 0$	Mudah	9, 10, 11, 5, 12, 1, 15, 14	8	53,33%
$Measure\ logit < -1.67$	Sangat Mudah	7	1	6,67%
	Total		15	100

Person Fit

Person fit bila dilihat pada konsep pada psikometri, digunakan untuk mengevaluasi sejauh mana pola jawaban individu dalam tes konsisten dengan model pengukuran seperti rasch model.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT OBS%	MATCH EXP%	PERSON	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
85	9	15	.574	.680	1.20	.60	2.93	2.40	A	.45	.62	73.3	80.4	85
157	12	15	2.127	.774	1.24	.71	2.18	1.13	B	.35	.52	73.3	82.8	157
111	8	15	.126	.660	1.62	1.58	1.97	1.60	C	.31	.61	66.7	79.1	111
143	6	15	-.712	.642	1.12	.51	1.90	1.33	D	.45	.56	73.3	74.9	143
25	9	15	.574	.680	1.68	1.57	1.89	1.40	E	.31	.62	73.3	80.4	25
101	12	15	2.127	.774	1.22	.67	1.85	.97	F	.37	.52	86.7	82.8	101
117	6	15	-.712	.642	1.38	1.34	1.84	1.27	G	.36	.56	60.0	74.9	117

Gambar 5. Person Fit

Pada gambar 5. *person fit* atau tidak terhadap model berdasarkan statistik outfit, nilai yang digunakan yaitu outfit ZSTD, Jika nilainya > 1.96 menunjukkan bahwa person tidak fit terhadap model (Linacre, 2012). Diperoleh person 85, person tidak fit bearti tidak cocok terhadap model, maka *person* ini dibuang. Bisa saja *person* (peserta) 85 mungkin menjawab soal mudah dengan salah tetapi menjawab soal sulit dengan benar atau menjawab soal dengan cara yang tidak biasa, seperti tebak-tebakan, tidak serius, atau mengalami kesulitan teknis. Metode *person fit* yang baik telah diturunkan untuk model Rasch. Untuk model IRT lainnya, distribusi empiris dan teoritis berbeda untuk sebagian besar statistik person fit ketika digunakan dengan tes pendek dan sedang. Tingkat deteksi statistik *person fit* bergantung pada jenis pola skor item yang tidak sesuai, panjang tes, dan tingkat sifat. Kegunaan statistik kecocokan person untuk meningkatkan pengukuran tergantung pada aplikasi. kesesuaian skor tes ditentukan oleh kecocokan pola skor butir soal dengan model tes. Metode *person-fit* di sini mengacu pada metode statistik untuk mengevaluasi ketidaksesuaian kinerja tes individu dengan model IRT atau pola skor item lainnya dalam sampel orang (Meijer & Sijtsma, 2001). Dengan kata lain, person fit mengukur "kecocokan" atau kesesuaian antara respons peserta terhadap soal dan pola yang diharapkan berdasarkan kemampuan mereka.

Item Fit

Salah satu cara untuk menilai ketepatan model IRT yang dipilih adalah melalui pemeriksaan kecocokan data. Pemeriksaan ini dapat dilakukan pada tingkat keseluruhan tes, maupun pada tingkat butir soal atau *item fit*, menunjukkan butir soal bekerja sebagaimana yang diharapkan dalam mengukur kemampuan peserta berdasarkan asumsi model. Kualitas item secara empiris dapat digunakan ketentuan dari sebagai berikut: a) nilai outfit MNSQ antara 0,5 sampai 2,0; b) Nilai *Outfit Z-Standard* (ZSTD) lebih besar dari -2,0 tetapi kurang dari +2,0; dan c) Nilai *Outfit Measure Correlation* (Pt Mean Corr), lebih besar dari 0,32 logit tetapi kurang dari 0,8 logit. Butir soal dapat dikatakan tidak fit apabila hanya memenuhi satu kriteria di atas bahkan harus diganti. Butir soal dapat dipertahankan dan dikatakan fit apabila minimal memenuhi dua kriteria tersebut (Sumintono & Widhiarso, 2013).

ITEM STATISTICS: MISFIT ORDER														
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT OBS%	MATCH EXP%	ITEM	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
1	126	163	-1.192	.200	1.14	1.31	1.31	1.63	A	.18	.34	77.3	78.8	1
10	116	163	-.821	.186	1.12	1.41	1.17	1.17	B	.23	.36	73.0	74.3	10
8	27	163	2.102	.224	1.08	.60	1.15	.70	C	.22	.32	84.0	84.1	8
5	121	163	-1.000	.192	1.04	.45	1.10	.69	D	.30	.35	74.8	76.6	5
3	22	163	2.372	.242	1.04	.27	1.08	.39	E	.24	.30	87.7	87.0	3
12	121	163	-1.000	.192	1.06	.60	.97	-.11	F	.31	.35	72.4	76.6	12
4	61	163	.834	.175	1.04	.57	1.05	.48	G	.33	.37	66.3	68.9	4
9	92	163	-.070	.171	1.05	.83	1.03	.35	H	.33	.38	65.6	66.7	9
11	118	163	-.891	.188	1.02	.22	1.04	.29	I	.34	.35	73.0	75.2	11
13	19	163	2.560	.257	1.03	.23	.92	-.18	F	.26	.29	88.3	88.7	13
7	154	163	-2.974	.353	1.01	.11	.75	-.45	E	.23	.21	94.5	94.5	7
6	62	163	.804	.174	.89	-1.72	.82	-1.80	D	.49	.37	75.5	68.5	6
2	23	163	2.315	.238	.85	-.93	.72	-1.15	C	.46	.30	85.9	86.4	2
14	136	163	-1.636	.224	.81	-1.40	.65	-1.63	B	.50	.31	84.7	83.8	14
15	131	163	-1.402	.210	.80	-1.78	.63	-2.03	A	.54	.32	85.3	81.3	15
MEAN	88.6	163.0	.000	.215	1.00	.95	.96	-.11				79.2	79.4	
P.SD	46.4	.0	1.667	.045	.10	.99	.20	1.06				8.3	7.8	

Gambar 6. Item Fit

Berdasarkan gambar 6, didapatkan seluruh item nilai outfit MNSQ 0.5-2.0. Walaupun item 15, ZSTD -2.03, nilai outfit MNSQ 0.63 dan *Pt Measure Corr* 0.54, berarti memenuhi dua kriteria maka item 15 masih fit. Pada item 1, 10, 8, 5, 3, 12, 13 dan 7 *Pt Measure Corr* kurang dari 0.32 logit, tetapi outfit MNSQ 0.5 - 2.0 dan ZSTD nya lebih dari -2.0 dan kurang dari 2.0, sehingga item-item tersebut masih fit. Dapat disimpulkan seluruh item Fit terhadap model. Sejumlah prosedur telah dikembangkan untuk evaluasi kecocokan model untuk tes dan butir soal melalui indeks kecocokan butir (*item fit*). Meskipun penilaian kecocokan data model untuk keseluruhan tes memiliki tantangan tersendiri. Secara umum, cara yang paling mudah untuk menilai kecocokan adalah dengan membuat statistik kecocokan yang membandingkan data yang diamati dengan prediksi yang dimodelkan karena kemampuan yang mendasari yang diukur oleh tes didefinisikan sebagai variabel laten (Orlando dkk., 2003).

Reliabilitas

Uji reliabilitas dilakukan untuk melihat keterandalan konsistensi dari instrumen. *Person Reliability* digunakan untuk melihat responden secara konsisten menjawab pertanyaan dengan benar. *Item reliability* (reliabilitas butir) merupakan ukuran yang menunjukkan sejauh mana sebuah butir soal konsisten dalam mengukur apa yang seharusnya diukur. Reliabilitas butir penting untuk memastikan bahwa hasil tes dapat dipercaya dan memberikan informasi yang stabil serta akurat tentang kemampuan peserta atau atribut lain yang diukur. Reliabilitas cronbach alpha mengukur interaksi antara *person* dan item (Muslih, dkk., 2022).

Tabel 8. Kriteria Nilai Reliabilitas Person dan Reliabilitas Item

Nilai	Kriteria
< 0,67	Lemah
0,67 - 0,80	Cukup
0,81 - 0,90	Bagus
0,91 - 0,94	Bagus Sekali
> 0,94	Istimewa

Tabel 9. Kriteria Cronbach Alpha

Nilai	Kriteria
< 0,5	Buruk
0,5 - 0,6	Jelek
0,6 - 0,7	Cukup
0,7 - 0,8	Bagus
> 0,8	Bagus Sekali

Tabel 8 dan Tabel 9, merupakan kriteria nilai reliabilitas person, item dan Cronbach Alpha (Sumintono & Widhiarso, 2013). Reliabilitas person 0,40 (lemah), mungkin tidak secara efektif membedakan antar individu, menyebabkan hasil yang kurang dapat diandalkan. Reliabilitas item 0,98 (istimewa), dapat diandalkan dan dapat memberikan informasi yang akurat tentang kemampuan pemecahan masalah matematis siswa. Cronbach Alpha (KR-20) 0,42 (buruk), item-item tersebut mungkin tidak mengukur konstruk yang sama atau memiliki konsistensi internal yang buruk, instrumen kurang reliabel untuk digunakan. Berdasarkan hasil analisis penelitian reliabilitas person, item dan cronbach alpha dapat dilihat pada gambar 7 berikut.

SUMMARY OF 163 MEASURED PERSON								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	8.2	15.0	.235	.680	1.00	.04	.96	.05
SEM	.2	.0	.074	.004	.02	.07	.04	.06
P.SD	2.0	.0	.938	.049	.31	.91	.49	.79
S.SD	2.0	.0	.941	.049	.31	.91	.49	.79
MAX.	14.0	15.0	3.704	1.099	1.68	1.58	2.93	2.40
MIN.	2.0	15.0	-2.664	.642	.36	-2.09	.28	-1.60
REAL RMSE	.724	TRUE SD	.596	SEPARATION	.82	PERSON RELIABILITY .40		
MODEL RMSE	.682	TRUE SD	.644	SEPARATION	.94	PERSON RELIABILITY .47		
S.E. OF PERSON MEAN = .074								
PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .42 SEM = 1.55								
STANDARDIZED (50 ITEM) RELIABILITY = .75								
SUMMARY OF 15 MEASURED ITEM								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	88.6	163.0	.000	.215	1.00	.05	.96	-.11
SEM	12.4	.0	.445	.012	.03	.26	.05	.28
P.SD	46.4	.0	1.667	.045	.10	.99	.20	1.06
S.SD	48.0	.0	1.725	.047	.11	1.03	.20	1.10
MAX.	154.0	163.0	2.560	.353	1.14	1.41	1.31	1.63
MIN.	19.0	163.0	-2.974	.171	.80	-1.78	.63	-2.03
REAL RMSE	.224	TRUE SD	1.652	SEPARATION	7.38	ITEM RELIABILITY .98		
MODEL RMSE	.220	TRUE SD	1.652	SEPARATION	7.52	ITEM RELIABILITY .98		
S.E. OF ITEM MEAN = .445								
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.99								
Global statistics: please see Table 44.								
UMEAN=.0000 USCALE=1.0000								

Gambar 7. Reliabilitas

Separation

Separation person dan *item* dilakukan untuk mengetahui perbedaan kemampuan siswa. Semakin luas penyebaran butir soal dari butir soal yang mudah ke butir soal yang sulit, dan semakin tinggi indeks pemisahan orang dan butir soal, maka semakin akurat distribusi butir soal merespon butir soal (Linacre, 2010). Pada gambar 7, Person Separation 0,82 tidak diterima, penyebaran siswa merespon item, tidak tersebar. Item separation 7,38, istimewa, luas penyebaran item dari mudah sampai item yang sulit mendukung keterandalan item.

Tabel 10. Kriteria Separation Person dan Item

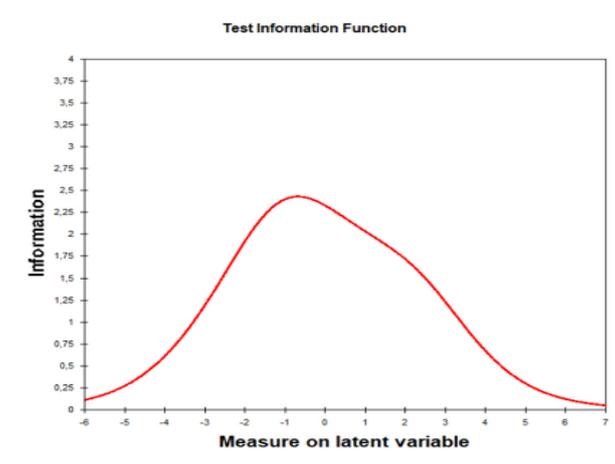
<i>Separation Person dan Item</i>	Kriteria
>5	Istimewa
3 – 4	Sangat Baik
2 – 3	Baik
≥ 1,5	Diterima
< 1,5	Tidak Diterima

Precision of Measurement (Nilai Estimasi Item)

Nilai estimasi item yang diperoleh pada kolom “Model S.E”, standar error yang baik dalam suatu instrumen harus kurang dari 0,5 (Perera et al., 2018). Pada gambar 7. nilai model S.E berturut-turut sebesar 0,224 dan 0,220, berada dibawah 0,5. Pada nilai tersebut *precision of measurement* berarti masih pada indikasi item fit, ajeg dan adanya presisi pengukuran yang baik.

Informasi Fungsi

Semakin tinggi nilai pada informasi fungsi, semakin besar informasi yang diperoleh dari tes tersebut pada tingkat kemampuan tertentu. Tes ini paling informatif untuk peserta dengan kemampuan sedang (theta sekitar 0), yang berarti tes dirancang untuk membedakan peserta di tingkat kemampuan ini.



Gambar 8. Informasi Fungsi pada Instrumen Kemampuan Pemecahan Masalah Matematika Siswa

Pada gambar 8. Didapatkan informasi berupa instrumen ini secara umum diperuntukkan untuk siswa dengan kemampuan -1 sampai 0 . Instrumen ini dirancang untuk memberikan informasi maksimal bagi siswa dengan kemampuan sekitar rata-rata hingga sedikit di bawah rata-rata (sekitar -1 hingga 1). Untuk siswa dengan kemampuan jauh lebih rendah atau lebih tinggi dari rentang tersebut, tes menggunakan instrumen ini memberikan lebih sedikit informasi, sehingga akurasi pengukuran di area tersebut lebih rendah.

KESIMPULAN

Proses pengembangan instrumen kemampuan pemecahan masalah matematika siswa kelas V SD menggunakan model rasch melalui aplikasi *winstep* untuk menjawab tujuan penelitian yang pertama didapatkan melalui tahapan kualitatif yaitu empat indikator yaitu memahami masalah, menentukan rencana, pelaksanaan rencana dan memeriksa kembali. Terkait kurikulum matematika kelas V SD, instrumen menggunakan elemen bilangan dengan capaian dan tujuan pembelajaran berupa nilai tempat, penjumlahan dan pengurangan bilangan cacah sampai 100.000, menjumlahkan dengan penyamaan pecahan biasa dan FPB. Pada elemen aljabar, capaian dan dan tujuan pembelajarannya menyangkut menghitung perkalian dan pembagian bilangan cacah sampai 1.000. Selanjutnya instrumen dibuat dengan 15 butir soal yang mewakili setiap indikator kemampuan pemecahan masalah matematika siswa yang dikaitkan dengan elemen, capaian dan tujuan pembelajaran yang telah ditentukan.

Pada tahapan kuantitatif yang pertama, dilakukan dengan memvalidasi isi dan konstruk instrumen oleh 3 orang ahli yang bergelar doktor yang dianalisis dengan aiken V diperoleh rata-rata dari ketiga kriteria 0,86 dengan kategori valid yang berarti instrumen dapat dilanjutkan dengan uji coba. Tujuan penelitian yang kedua yaitu menjawab validitas psikometri dilakukan dengan tahapan kuantitatif kedua melalui ujicoba beberapa sekolah SD di Jakarta Timur dengan jumlah responden 163 siswa. Jawaban siswa dianalisis dengan model rasch meruapakan bagian dari IRT (teori tes modern) melalui aplikasi *winstep*, hasil uji prasyarat didapatkan instrumen telah memenuhi 3 persyaratan IRT yaitu unidimensional, lokal independen dan monotonisasi.

Adapun analisis validasi instrumen secara psikometrika, didapatkan dari hasil tingkat kesulitan butir yang sangat sukar nomor 13, 2,560 logit dan yang paling mudah butir nomor 7, -2,974 logit. *Person fit* didapatkan ada satu person yang $>$ dari 1,96 yaitu person 85 sebaiknya dibuang. Pada *item fit* seluruh butir fit karena telah memenuhi 2 persyaratan dari 3 persyaratan. Reliabilitas person reliabilitas 0,40 (lemah), reliabilitas item 0,98 (Istimewa) dan *Cronbach alpha* (KR-20) 0,42 pada kriteria buruk. Pada *person separation* didapatkan 0,82 (tidak diterima) dan item separation 7,38 (istimewa). *Precision of measurement* kurang dari 0,5 berarti memenuhi satandar yaitu

0,224 dan 0,220. Berdasarkan hal tersebut instrumen dapat digunakan secara efektif. Pada gambar fungsi informasi didapatkan instrumen ini secara umum diperuntukkan untuk siswa dengan kemampuan rata-rata -1 logit sampai 0 logit.

REFERENSI

- Ahmad, N. Q., Noviani, J., & Sari, A. P. (2023). Analysis of High School Students' Difficulties in Solving Hots Problems Using The Two-Tier Test Method. *Jurnal Math Educator Nusantara: Wahana Publikasi Karya Tulis Ilmiah Di Bidang Pendidikan Matematika*, 9(2), 95–110. <https://doi.org/10.29407/jmen.v9i2.20822>
- Bond, T. G., Yan, Z., & Heene, M. (2020). Applying The Rasch Model: Fundamental Measurement in The Human Sciences. In *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Taylor and Francis. <https://doi.org/10.4324/9780429030499>
- Cahyani, H., & Setyawati, R. W. (2016). Pentingnya Peningkatan Kemampuan Pemecahan Masalah melalui PBL untuk Mempersiapkan Generasi Unggul Menghadapi MEA. In *PRISMA, Prosiding Seminar Nasional Matematika*.
- Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item Response Theory. *Annual Review of Statistics and Its Application*, 3(1), 297–321. <https://doi.org/https://doi.org/10.1146/annurev-statistics-041715-033702>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education*. Routledge.
- Creswell, J. W., & Creswell, J. D. (2023). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Sixth Edition). SAGE Publications.
- Das Nair, R., Moreton, B. J., & Lincoln, N. B. (2011). Rasch Analysis of The Nottingham Extended Activities of Daily Living Scale. *Journal of Rehabilitation Medicine*, 43(10), 944–950. <https://doi.org/10.2340/16501977-0858>
- De Ayala, R. J. (2022). *The Theory and Practice of Item Response Theory* (Second Edition). Guilford Publication.
- Edwards, M. C., Houts, C. R., & Cai, L. (2018). A Diagnostic Procedure to Detect Departures from Local Independence in Item Response Theory Models. *Psychological Methods*, 23(1), 138–149. <https://doi.org/10.1037/met0000121>
- Falissard, B. (1999). The Unidimensionality of A Psychiatric Scale: A Statistical Point of View. *International Journal of Methods in Psychiatric Research*, 8(3), 162–167. <https://doi.org/https://doi.org/10.1002/mpr.66>
- Ha, H. T. (2022). Test Format and Local Dependence of Items Revisited: A Case of Two Vocabulary Levels Tests. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.805450>
- Hastuti, T. A., Sari, I. P. T. P., & Andrianto, S. D. (2023). Analysis of The Aiken Index to Know The Content Validity of The Lesson Plan Evaluation Instrument on Physical Fitness Materials Viewed from Learning Strategies. In *6th Yogyakarta International Seminar on Health, Physical Education and Sport Science (YISHPESS 2023)*, 276–282. https://doi.org/10.2991/978-94-6463-356-6_31
- Hendriana, H., Johanto, T., & Sumarmo, U. (2018). The Role of Problem-Based Learning to Improve Student's Mathematical Problem-Solving Ability and Self Confidence. *Journal on Mathematics Education*, 9(2), 291–300.
- Hori, K., Fukuhara, H., & Yamada, T. (2022). Item Response Theory and Its Applications in Educational Measurement Part I: Item Response Theory and Its Implementation in R. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(2). <https://doi.org/10.1002/wics.1531>
- Kondratek, B. (2022). UIRT: A Command for Unidimensional IRT Modeling. *Stata Journal*, 22(2), 243–268. <https://doi.org/10.1177/1536867X221106368>
- Liljedahl, P., Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). *Problem Solving in Mathematics Education*. Springer Nature. <http://www.springer.com/series/14352>

- Linacre, J. M. (2010). Requests for Reprints Should be Sent to Predicting Responses from Rasch Measures. *Predicting Responses from Rasch Measures*, 11(1), 1–10.
- Linacre, J. M. (2012). *A User's Guide to WINSTEPS Ministep: Rasch-Model Computer Programs*. Winsteps.com.
- Meijer, R. R., & Sijtsma, K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*, 25(2), 107–135. <https://doi.org/10.1177/01466210122031957>
- Muslihah, H. Y., Suryana, D., Ahman, Suherman, U., & Dahlan, T. H. (2022). Analysis of the Reliability and Validity of the Self-Determination Questionnaire Using Rasch Model. *International Journal of Instruction*, 15(2), 207–222. <https://doi.org/10.29333/iji.2022.15212a>
- OECD. (2023). *PISA 2022 Results (Volume I)*. OECD. <https://doi.org/10.1787/53f23881-en>
- Orlando, M., Health Division, R., Monica, S., & Thissen, D. (2003). Further Investigation of the Performance of S - X 2: An Item Fit Index for Use With Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, 27(4), 289–298. <https://doi.org/10.1177/0146621603253011>
- Perera, C. J., Sumintono, B., & Na, J. (2018). The Psychometric Validation of The Principal Practices Questionnaire Based on Item Response Theory. *International Online Journal of Educational Leadership*, 2(1), 21–38. <https://doi.org/https://doi.org/10.22452/iojel.vol2no1.3>
- Polya, G. (1973). *How To Solve It: Vol. Second Edition*. Princeton University Press.
- Rush, B. R., Rankin, D. C., & White, B. J. (2016). The Impact of Item-Writing Flaws and Item Complexity on Examination Item Difficulty and Discrimination Value. *BMC Medical Education*, 16(1). <https://doi.org/10.1186/s12909-016-0773-3>
- Salwida, K., Yuanita, P., & Sakur, S. (2022). Perangkat Pembelajaran Berbasis Model Problem Based Learning untuk Memfasilitasi Kemampuan Pemecahan Masalah Matematis Peserta Didik. *Journal for Research in Mathematics Learning* p, 5(1), 89–098. <https://doi.org/http://dx.doi.org/10.24014/juring.v3i4.10312>
- Samosir, C. M., Solfitri, T., & Armis, A. (2020). Penerapan Model Pembelajaran Berdasarkan Masalah untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa Kelas VII A SMP PGRI Pekanbaru Tahun Pelajaran 2019/ 2020. *Journal for Research in Mathematics Learning* p, 3(4), 403–412. <https://doi.org/http://dx.doi.org/10.24014/juring.v3i4.10312>
- Stenner, A. J., Fisher Jr, W. P., Stone, M. H., & Burdick, D. S. (2013). Causal Rasch models. *Frontiers in Psychology*, 4(AUG). <https://doi.org/10.3389/fpsyg.2013.00536>
- Strachan, T., Cho, U. H., Ackerman, T., Chen, S. H., de la Torre, J., & Ip, E. H. (2022). Evaluation of the Linear Composite Conjecture for Unidimensional IRT Scale for Multidimensional Responses. *Applied Psychological Measurement*, 46(5), 347–360. <https://doi.org/10.1177/01466216221084218>
- Sumintono, B. (2018). Rasch Model Measurements as Tools in Assessment for Learning. In *1st International Conference on Education Innovation (ICEI 2017)*, 38–42.
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Komunikata Publishing House.
- Szabo, Z. K., Körtesi, P., Guncaga, J., Szabo, D., & Neag, R. (2020). Examples of Problem-Solving Strategies in Mathematics Education Supporting The Sustainability of 21st-Century Skills. *Sustainability (Switzerland)*, 12(23), 1–28. <https://doi.org/10.3390/su122310113>
- Virgia, Z., Granita, & Nelson, Z. (2019). Pengaruh Penerapan Strategi Metakognitif terhadap Kemampuan Pemahaman Konsep dan Pemecahan Masalah Matematis Siswa SMP. *Journal for Research in Mathematics Learning* p, 2(4), 371–379. <https://doi.org/http://dx.doi.org/10.24014/juring.v2i4.8566>
- Yayuk, E., & As'Ari, A. R. (2020). Primary School Students' Creative Thinking Skills in Mathematics Problem Solving. *European Journal of Educational Research*, 9(3), 1281–1295. <https://doi.org/10.12973/eu-jer.9.3.1281>