

Pengembangan Instrumen Literasi Matematis Model PISA dengan Pemodelan Rasch untuk Siswa Sekolah Dasar

Mayu Syahwela*, Nurul Qomariyah Ahmad, Awaluddin Tjalla, dan Iva Srifah

Program Studi Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta, Jakarta, Indonesia

**E-mail: mayu.syahwela@student.unj.ac.id*

ABSTRACT. This study aims to develop an instrument used to measure the mathematical literacy skills of elementary school students. This study uses a 4D development model (Define, Design, Develop, and Disseminate) which is modified to the Develop stage only. This study took a sample of elementary school students spread across the Riau region totaling 113 people and DKI Jakarta totaling 176 students (62.97% female and 37.03% male) from 4 different elementary schools. The data collection technique in this study was in the form of a questionnaire and a test. There are two instruments used in this study, namely a question validation sheet with a Likert scale and mathematical literacy questions in the form of multiple choice and essays. Analysis of the results of the question validation sheet using the Aiken V index while the item analysis uses the Rasch model. Based on the results of the data analysis, it was obtained that the 10 questions were declared valid and reliable based on item fit, wright map, person separation index and item information function. The findings show that this instrument can accurately measure the mathematical literacy skills of elementary school students, especially at high and low ability levels, while for moderate ability it is less relevant because there are no items that have a moderate level of difficult.

Keywords: mathematical literacy; measurement instrument; primary school students; rasch model

ABSTRAK. Penelitian ini bertujuan untuk mengembangkan sebuah instrumen yang digunakan untuk mengukur kemampuan literasi matematis siswa SD. Penelitian ini menggunakan model pengembangan 4D (*Define, Design, Develop, and Disseminate*) yang dimodifikasi sampai pada tahap Develop saja. Penelitian ini mengambil sampel siswa SD yang tersebar di wilayah Riau yang berjumlah 113 orang dan DKI Jakarta yang berjumlah 176 orang siswa (62,97 % perempuan dan 37,03% laki-laki) yang berasal dari 4 SD yang berbeda. Teknik pengumpulan data dalam penelitian ini adalah berupa kuisisioner dan tes. Terdapat dua instrumen yang digunakan dalam penelitian ini yaitu lembar validasi soal dengan skala likert dan soal literasi matematis berbentuk pilihan ganda dan *essay*. Teknik analisis yang digunakan yakni untuk Analisis hasil lembar validasi soal menggunakan indeks V Aiken sedangkan analisis butir soal menggunakan model Rasch. Berdasarkan hasil analisis data, diperoleh bahwa ke-10 soal dinyatakan valid dan reliabel berdasarkan *item fit, wright map, person separation index* dan fungsi informasi butir. Hasil temuan menunjukkan bahwa instrumen ini dapat secara akurat untuk mengukur kemampuan literasi matematis siswa SD khususnya pada tingkat kemampuan tinggi dan rendah sedangkan unutm kemampuan *moderate* kurang relevan karena tidak ada butir yang memiliki taraf sukar sedang.

Kata kunci: instrumen pengukuran; literasi matematis; model rasch; siswa sekolah dasar

PENDAHULUAN

Matematika merupakan ilmu universal yang kegunaannya banyak sekali dalam pemecahan masalah sehari-hari. Matematika membantu kita menganalisis dan memecahkan berbagai masalah, baik yang bersifat praktis maupun teoretis (Sitopu dkk., 2024). Matematika bukan hanya angka, tetapi juga cara berpikir yang berguna di berbagai bidang kehidupan (Saputra, 2024). Oleh sebab itu, matematika menjadi salah satu mata pelajaran wajib untuk dikuasai oleh setiap siswa sejak dini.

Pemahaman matematika yang baik sejak dini memiliki peranan yang sangat penting dalam mengembangkan keterampilan kognitif yang lebih luas (Nuryati & Darsinah, 2021).

Salah satu tujuan mempelajari matematika adalah agar siswa memiliki kemampuan literasi matematis. Menurut OECD (2018), literasi matematis adalah kemampuan seseorang untuk merumuskan, menerapkan, dan menafsirkan matematika dalam berbagai konteks, termasuk kemampuan untuk mengenali peran matematika dalam dunia dan membuat penilaian yang bijaksana sebagai warga negara yang reflektif. Senada dengan hal tersebut Geiger dkk. (2015) mendefinisikan literasi matematika sebagai kompetensi yang diperlukan untuk memenuhi tuntutan matematika dalam kehidupan pribadi dan kehidupan sosial dan untuk berpartisipasi dalam masyarakat sebagai warga negara yang berpengetahuan, reflektif, dan berkontribusi. Literasi matematis melatih individu untuk berpikir secara analitis, mempertanyakan asumsi, dan mengevaluasi hasil berdasarkan fakta (NCTM, 2000). Oleh karena itu, literasi matematis merupakan keterampilan dasar yang diperlukan untuk menjadi individu yang mandiri, produktif, dan mampu menghadapi tantangan di dunia modern.

Dalam konteks Kurikulum Merdeka, literasi matematis semakin mendapat perhatian karena relevansinya dengan pembelajaran abad ke-21. Kemampuan ini mendukung siswa untuk menghadapi tantangan dunia yang semakin kompleks, di mana data dan informasi berbasis matematika menjadi bagian integral dalam pengambilan keputusan. Namun, hasil penelitian dan evaluasi internasional, seperti *Programme for International Student Assessment* (PISA) di tahun 2022, menunjukkan bahwa literasi matematis siswa Indonesia, masih relatif rendah dibandingkan negara lain. Hal ini menunjukkan perlunya penguatan kemampuan literasi matematis sejak dini.

Pengembangan instrumen untuk mengukur literasi matematis siswa SD merupakan langkah penting untuk memahami sejauh mana siswa mampu menguasai kemampuan ini. Instrumen yang baik dan valid dapat membantu guru dan peneliti dalam mengidentifikasi tingkat literasi matematis sehingga guru dapat memberikan intervensi pembelajaran yang tepat, serta merancang strategi pengajaran yang sesuai dengan kebutuhan siswa (Dewi dkk., 2024). Namun, berdasarkan kajian literatur terhadap artikel penelitian yang sudah dipublikasi sebelumnya, belum banyak instrumen yang dirancang secara spesifik untuk mengukur literasi matematis siswa SD model PISA dengan memperhatikan karakteristik perkembangan kognitif mereka. Sebagian besar instrumen yang ada dirancang untuk jenjang yang lebih tinggi, sehingga kurang sesuai untuk siswa SD. Seperti halnya penelitian yang dilakukan oleh Shaumi & Syamsuri (2024), Hayati & Yuhana (2024) yang mengembangkan instrumen literasi matematis untuk siswa SMP. Sedangkan penelitian oleh Sari & Wijaya (2017) menggunakan instrumen untuk mengukur kemampuan literasi matematis untuk siswa SMA. Oleh karena itu, diperlukan upaya untuk mengembangkan instrumen literasi matematis untuk siswa SD yang valid, reliabel, dan sesuai dengan konteks pendidikan di Indonesia, serta mampu mencakup aspek-aspek literasi matematis. Terdapat penelitian yang mengembangkan instrumen literasi matematis untuk siswa SD yakni penelitian oleh Widia dkk. (2024).

Adapun yang membedakan penelitian ini dengan penelitian sebelumnya adalah penelitian ini mencoba menggabungkan dua pendekatan yaitu *Classical test theory* dan *item respons theory* yang pada penelitian sebelumnya tidak menggunakan teori modern dalam analisis butir soalnya. Penelitian sebelumnya hanya menggunakan rumus korelasi *Product moment* untuk menghitung validitas dan *Alpha Cronbach* untuk pengujian reliabilitas. Analisis butir pada penelitian ini dilakukan lebih mendalam dengan menganalisis sifat psikometrik butirnya dengan menggunakan *item respons theory*. Model Rasch adalah salah satu model *item response theory* yang digunakan untuk mengevaluasi kualitas instrumen pengukuran dengan cara menganalisis karakteristik psikometrik secara lebih rinci. Model ini membantu memastikan bahwa instrumen memiliki validitas dan reliabilitas yang tinggi dengan menganalisis hubungan antara kemampuan responden dan karakteristik butir (Suryani, 2018). Model ini dapat mendeteksi item yang tidak sesuai (*misfit*) dengan pola jawaban responden, seperti item yang ambigu atau tidak relevan (Fauzi dkk., 2022). Hal ini membantu memperbaiki kualitas instrumen dan memastikan bahwa hanya item yang benar-benar mengukur konstruk yang digunakan. Selain itu, model Rasch juga mempertimbangkan pendekatan probabilistik untuk

mengevaluasi atribut yang diukur, sehingga memungkinkan untuk memberikan informasi yang sangat akurat ketika menguji instrumen (Indihadi dkk., 2022). Berdasarkan hal tersebut, instrumen yang dikembangkan dalam penelitian ini mampu secara tepat dan layak digunakan untuk mengukur kemampuan literasi matematis siswa SD. Selain itu, Instrumen yang dihasilkan ini juga mampu menilai sejauh mana siswa dapat memahami, menganalisis, dan memecahkan masalah matematis dalam konteks dunia nyata, bukan hanya sekadar menguasai konsep atau rumus matematika. Mengacu pada 4 aspek literasi matematis dalam pengukuran PISA, instrumen ini memungkinkan siswa SD untuk dikenalkan sejak dini dengan standar penilaian internasional, sehingga mempersiapkan mereka untuk mengikuti penilaian serupa di tingkat pendidikan yang lebih tinggi.

METODE

Jenis penelitian ini adalah penelitian pengembangan dengan menggunakan model 4D (*Define, Design, Develop, and Disseminate*) (Thiagarajan, 1974) tetapi di dalam penelitian ini dimodifikasi sampai pada tahap *Develop* saja. Adapun produk yang dikembangkan adalah sebuah instrumen yang digunakan untuk mengukur kemampuan literasi matematis siswa SD khususnya di Indonesia.

Adapun teknik pengambilan sampel dilakukan secara *purposive*. Menurut Arikunto (2010), teknik *purposive sampling* digunakan untuk memilih subjek penelitian yang dianggap memiliki informasi sesuai dengan kebutuhan penelitian. Cara ini memungkinkan peneliti untuk langsung memilih individu yang relevan tanpa harus melakukan survei pada populasi yang lebih besar. Hal ini menghemat waktu, biaya, dan sumber daya penelitian. Dalam mengembangkan instrumen pada penelitian ini, peneliti menggunakan sampel uji coba sebanyak 289 (62,97 % perempuan dan 37,03 % laki-laki) siswa SD kelas 5 yang tersebar di Provinsi DKI Jakarta dan Riau. Untuk lebih jelasnya dapat dilihat pada Tabel 1.

Tabel 1. Karakteristik Sampel Penelitian

Nama Sekolah	Banyak Sampel	
	Perempuan	Laki-laki
SDN 036 Pekanbaru	68	45
MI Riyadlus Shibani	24	14
SDN 038 Cibubur	56	36
MI AL 19	34	12
Total	182	107
Persentase	62,97%	37,02%

Adapun prosedur penelitian ini diadaptasi dari Mardapi (2008) : (1) identifikasi kebutuhan, menentukan tujuan dan konsep yang akan diukur; (2) kajian literatur, mengumpulkan referensi terkait untuk menyusun indikator literasi matematis; (3) penyusunan instrumen awal, merancang item berdasarkan indikator; (4) validasi ahli, menguji isi dan konstruksi instrumen oleh pakar; (5) uji coba instrumen: mengujikan pada sampel kecil untuk evaluasi awal; (6) analisis validitas dan reliabilitas, menggunakan model Rasch; (7) revisi instrumen, memperbaiki berdasarkan hasil uji coba dan analisis; (8) finalisasi instrumen, menyusun instrumen siap pakai.

Teknik pengumpulan data dalam penelitian ini adalah berupa kuisioner dan tes. Selaras dengan teknik pengumpulan datanya, maka terdapat dua instrumen yang digunakan dalam penelitian ini yaitu lembar validasi soal dengan *skala likert* dan soal literasi matematis berbentuk pilihan ganda dan *essay*. Analisis hasil lembar validasi soal menggunakan indeks V Aiken sedangkan analisis butir soal menggunakan model Rasch.

HASIL DAN PEMBAHASAN

Define

Pada tahap ini, peneliti mengidentifikasi kebutuhan dalam mengembangkan produk instrumen pengukuran literasi matematis siswa SD. Berdasarkan hasil wawancara di lapangan, selama ini belum ada pihak sekolah maupun guru yang mengembangkan instrumen literasi matematis model PISA untuk siswa SD. Selama ini guru hanya melakukan *assessment* dengan menggunakan soal-soal matematika yang sudah tersedia di buku-buku cetak dan internet. Oleh karena itu, peneliti beranggapan bahwa pengembangan instrumen literasi matematis siswa SD layak untuk dilakukan. Ditambah model PISA terbaru tahun 2022 menggunakan indikator yang berbeda dari PISA 2018.

Berdasarkan kajian literatur, maka indikator literasi matematis yang digunakan dalam penelitian ini terdiri dari empat indikator yaitu: (1) *Formulate*, (2) *Employ*, (3) *Interpret/ Evaluate*, and (4) *Reasoning*. Soal matematika yang dikembangkan juga disesuaikan dengan materi yang dipelajari siswa SD kelas 5 dengan mempertimbangkan aspek bahasa dan tingkat kognitif siswa SD.

Design

Berdasarkan kisi-kisi soal literasi matematis, maka peneliti mulai merancang instrumen awal terkait dengan literasi matematis. Peneliti juga membuat pedoman penskoran untuk menilai hasil tes literasi matematis. Soal matematika yang dikembangkan sebanyak 10 soal yang terdiri dari soal pilihan ganda dan juga essay. Kisi kisi instrumen dapat dilihat pada Tabel 2.

Tabel 2. Kisi-Kisi Instrumen Literasi Matematis

Aspek Literasi Matematis	No soal	Bentuk Soal	Level	Model Pengukuran
<i>Formulate</i> (menerjemahkan permasalahan ke dalam istilah matematika, mengidentifikasi variable dan hubungan yang relevan)	3, 6	<i>Simple Multiple Choice</i>	C2	Dikotomi
<i>Employ</i> (menggunakan konsep matematika dalam menyelesaikan permasalahan)	1, 9, 10	<i>Simple Multiple Choice</i>	C3	Dikotomi
<i>Interpret/ Evaluate</i> (memaknai data, menarik kesimpulan logis dari informasi yang diperoleh dari grafik/ tabel maupun cerita)	2, 5, 8	<i>Complex Multiple Choice</i>	C5	<i>Partial Credit</i>
<i>Reasoning</i> (memberikan argumen yang logis dan koheren untuk mendukung solusi yang dipilih)	4, 7	Uraian	C6	<i>Partial Credit</i>

Develop

Ini merupakan tahapan inti dari proses pengembangan instrumen. Pada tahap ini, peneliti melakukan analisis terhadap instrumen yang dikembangkan baik secara kualitatif maupun kuantitatif. Analisis kualitatif dilakukan pada tahap awal pembuatan butir instrumen. Dimana pada tahapan ini peneliti meminta penilaian dari pakar untuk mengecek butir-butir yang telah dibuat apakah sesuai untuk level kelas 5 SD atau tidak. Setelah dicek dan dinilai kesesuaian materi, konstruksi dan Bahasa, maka tahapan selanjutnya adalah analisis butir secara kuantitatif dengan menggunakan model Rasch. Analisis butir dengan model Rasch dilakukan dengan menggunakan aplikasi *Winstep* (v 4.6.0).

Validitas

Validitas mengacu pada sejauh mana suatu instrumen mengukur apa yang dimaksudkan untuk diukur (Heale & Twycross, 2015), dan apakah hasil pengukurannya benar-benar mencerminkan variabel yang ingin diteliti (Fraenkel & Wallen, 2009). Sedangkan menurut Kerlinger dkk., (2000) adalah derajat yang menunjukkan sejauh mana alat ukur dapat memberikan hasil yang tepat dan sesuai dengan yang diharapkan berdasarkan tujuan pengukuran. Oleh sebab itu, instrumen yang dikembangkan haruslah memenuhi unsur valid. Untuk menentukan validitas butir instrumen, peneliti meminta bantuan 6 orang pakar yang terdiri dari 4 dosen matematika dan 2 ahli pengukuran untuk memvalidasi setiap butir. Hasil Validasi yang berupa saran dari validator dapat dilihat pada Tabel 3 berikut.

Tabel 3. Hasil Pengujian Pakar

Pakar	Komentar & Saran	Keterangan
Pakar 1	<ul style="list-style-type: none"> Item 6: soal yang dikemas kurang berkenaan dengan indikator 'formulate', soal tersebut merupakan soal rutin biasa pada topik mengurutkan bilangan Item 9 dan 10: sesuai dengan indikatornya, hanya saja konteks bolu yang disajikan bahan-bahanya itu, hanya mengambil bagian gulanya saja yang kemudian digunakan konsep kaidah perkalian untuk menghitung banyaknya gula untuk 3 bolu. Sementara info bahan-bahan bolu lainnya hanya sebagai wawasan siswa barangkali, tidak digunakan untuk matematika siswa. Menurut saya ada informasi yang tidak dibutuhkan sehingga dapat dihilangkan atau mencari konteks lain yang lebih sesuai dengan indikator 	Sudah Diperbaiki
Pakar 2	Tidak ada komentar	
Pakar 3	seharusnya, tabel diberikan keterangan misal tabel berapa atau tabelnya diberi nama/ keterangan, lalu soal kalimatnya dirancang mulai dari subjek-predikat-objek	Sudah Diperbaiki
Pakar 4	Tidak ada komentar	
Pakar 5	Saran: gambar jangan kabur dan dibuat jelas	Sudah Diperbaiki
Pakar 6	<ul style="list-style-type: none"> Butir 1 bahasa melonjak apakah dipahami siswa? harap dipertimbangkan lagi pilih bahasanya. Pernyataan pada butir 1 kurang bisa dipahami siswa, bisa dibuat kalimat yang lebih mudah lagi Butir 5 masih tergolong mudah (belum mengukur C5). Butir 7 masih tergolong mudah (belum mengukur C5). Butir 8 mudah sekali (belum mengukur c6) secara konstruksi belum terpenuhi, bisa diganti soal atau angka atau cerita yang lebih meminta siswa berpikir pada level 5 dan 6 Butir 9 dan 10, option tidak diurut? Apa ada alasan sehingga di buat tidak urut? Butir 10 optionnya kurang homogen, karena hasil kali 3 dan 8 = 24, angka belakang 4 Cuma 1 jawaban jadi mudah ditebak, sebaiknya direvisi option nya. 	Sudah Diperbaiki

Hasil pengujian pakar berupa komentar dan saran terkait dengan butir tersebut dijadikan sebagai bahan evaluasi. Butir butir tersebut direvisi sebelum digunakan untuk uji coba ke skala yang lebih luas. Setelah butir dievaluasi secara kualitatif, maka selanjutnya dianalisis secara kuantitatif

yaitu dengan cara menentukan nilai indeks V Aiken. Adapun hasil perhitungan nilai indeks V Aiken dapat dilihat pada Tabel 4.

Tabel 4. Hasil Perhitungan Validitas Butir Dengan Menggunakan V Aiken

Item	Materi	Konstruksi	Bahasa	Keterangan
Item 1	0.792	0.708	0.625	Diperbaiki
Item 2	0.875	0.833	0.833	Valid
Item 3	0.833	0.833	0.833	Valid
Item 4	0.792	0.792	0.833	Valid
Item 5	0.833	0.917	0.875	Valid
Item 6	0.792	0.833	0.875	Valid
Item 7	0.792	0.833	0.875	Valid
Item 8	0.792	0.875	0.833	Valid
Item 9	0.792	0.792	0.875	Valid
Item 10	0.833	0.792	0.917	Valid

Kriteria item yang valid berdasarkan indeks Aiken adalah > 0.7 (Naga, 2023) Berdasarkan tabel 4, hanya item 1 yang memiliki indeks V di bawah 0,7 dalam kategori “Bahasa” (0,625). Dengan demikian peneliti kembali memperbaiki Bahasa pada item 1 sebelum diujicoba ke sampel yang lebih luas.

Unidimensi

Pengujian unidimensi bertujuan untuk memastikan bahwa suatu instrumen atau skala pengukuran hanya mengukur satu konstruk atau dimensi utama yang menjadi fokus penelitian (Andriani, 2023). Unidimensi dilihat pada "raw variance explained by measure" terletak di kolom "observed". Hasil yang diperoleh tersebut sebesar 72,0%. Asumsi unidimensionalitas terpenuhi apabila besaran *Raw variance explained by measures* $> 40\%$ (Holster & Lake, 2016). Terdapat kriteria kedua bahwa unexplained variance in 1st contrast nilainya tidak lebih besar dari 2.0 (Smith Jr., 2002). Berdasarkan hasil pengujian, diperoleh nilainya sebesar 1,9006. Dengan demikian asumsi unidimensi telah terpenuhi.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations =	35.6952	100.0%	100.0%
Raw variance explained by measures =	25.6952	72.0%	72.3%
Raw variance explained by persons =	5.1081	14.3%	14.4%
Raw Variance explained by items =	20.5871	57.7%	58.0%
Raw unexplained variance (total) =	10.0000	28.0%	100.0%
Unexplned variance in 1st contrast =	1.9006	5.3%	19.0%
Unexplned variance in 2nd contrast =	1.4299	4.0%	14.3%
Unexplned variance in 3rd contrast =	1.2906	3.6%	12.9%
Unexplned variance in 4th contrast =	1.1776	3.3%	11.8%
Unexplned variance in 5th contrast =	1.0786	3.0%	10.8%

Gambar 1. Hasil Pengujian Asumsi Unidimensi dengan Eigen Value

Independen Lokal

Asumsi independen lokal penting untuk memastikan bahwa interpretasi kemampuan individu dalam model Rasch valid, serta untuk menjaga kualitas dan keandalan pengukuran. Pengujian independen lokal bertujuan untuk memastikan bahwa respons individu terhadap suatu item dalam instrumen pengukuran tidak saling bergantung dengan respons terhadap item lainnya (Debelak & Koler, 2019), setelah mempertimbangkan faktor atau dimensi utama yang diukur. Jika asumsi ini tidak terpenuhi, artinya ada hubungan antaritem yang tidak sepenuhnya dijelaskan oleh kemampuan individu. Kriteria independent lokal terlanggar apabila korelasi residual antar pasangan butir nilainya positif dan > 0.30 (Christensen dkk., 2017). Pada Gambar 2, tidak ada pasangan butir

yang memiliki korelasi residual dengan arah positif dan negatif dan lebih besar dari 0.30. Oleh karena itu, dapat disimpulkan bahwa asumsi *local independence* terpenuhi

CORRELATION	ENTRY NUMBER	IT	ENTRY NUMBER	ITE
.19	9	B9	10	B10
.14	3	B3	9	B9
-.27	5	B5	8	B8
-.26	2	B2	8	B8
-.26	2	B2	3	B3
-.23	4	B4	5	B5
-.23	2	B2	4	B4
-.23	8	B8	10	B10
-.23	8	B8	9	B9
-.21	5	B5	7	B7
-.21	3	B3	7	B7
-.20	7	B7	9	B9
-.19	5	B5	9	B9
-.19	4	B4	7	B7
-.19	4	B4	8	B8
-.17	2	B2	10	B10
-.16	1	B1	2	B2
-.15	6	B6	7	B7
-.14	4	B4	6	B6
-.14	2	B2	9	B9

Gambar 2. Hasil Pengujian Asumsi Lokal Independen

Monotonisasi

Pengujian monotonisasi bertujuan untuk memastikan bahwa hubungan antara respons peserta pada suatu item dan kemampuan atau atribut laten yang diukur bersifat monoton, yaitu ketika kemampuan laten meningkat, probabilitas untuk memberikan respons positif atau skor lebih tinggi juga meningkat. Menguji monotonisasi membantu memastikan bahwa data empiris konsisten dengan struktur matematis model Rasch (Linacre, 2002). Hal ini diperlukan untuk menghasilkan estimasi parameter yang akurat, baik untuk kemampuan peserta maupun tingkat kesulitan item. Penerapan prinsip monotonisasi membantu memastikan bahwa model Rasch memberikan estimasi kemampuan yang akurat, konsisten, dan valid. Sifat monotonisasi dalam winstep dapat dilihat dari kolom *observed average* di mana nilai harus cenderung monoton naik. *Tren Observed Average* yang tidak monoton dapat menjadi indikasi adanya masalah seperti ketidakcocokan model (*misfit*) atau kategori respons yang tidak berfungsi dengan baik (Bond & Fox, 2015). Berdasarkan tabel tersebut dapat diketahui bahwa nilai *observed average* pada data ini digunakan adalah cenderung naik dari angka -2,858 sampai 1,769. Dengan demikian, sifat monotonisasi pada instrumen ini telah terpenuhi.

SUMMARY OF CATEGORY STRUCTURE. Model="R"										
CATEGORY LABEL	SCORE	OBSERVED COUNT	OBSVD %	SAMPLE AVRG	INFINIT EXPECT	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE		
0	0	637	24	-2.858	-3.06	1.88	1.34	NONE	(-3.92)	0
1	1	876	33	-2.094	-1.79	.66	.95	-2.78	-1.47	1
2	2	338	13	.667	.199	.76	.63	.16	.15	2
3	3	581	22	1.107	1.193	1.03	.86	.24	1.52	3
4	4	226	9	1.769	1.649	.90	.94	2.38	(3.56)	4
MISSING		2	0	1.822						

Gambar 3. Hasil Pengujian Asumsi Monotonisasi

Person fit

Person fit dalam model Rasch merujuk pada sejauh mana pola jawaban seorang individu konsisten dengan ekspektasi model Rasch. Evaluasi person fit digunakan untuk mendeteksi responden yang memberikan jawaban yang tidak sesuai dengan model, yang dapat disebabkan oleh berbagai faktor, seperti ketidakjujuran, kebingungan terhadap butir, atau pola menjawab acak. Person fit atau tidak terhadap model berdasarkan statistik Outfit, nilai yang digunakan adalah Outfit ZSTD. Jika nilainya > 1.96 menunjukkan bahwa person tidak fit terhadap model (Bond & Fox, 2015). Adapun person yang tidak fit adalah 268, 289, 67, 255, 223, 284, 263, 66, 251, 220, 279, 258, 64, 276, 63, 274, 253, 241, 212, 270, 249, 238, 209.

Item Fit

Kriteria Infit dan Outfit MNSQ 0.5-1.5 menunjukkan bahwa item fit terhadap model (Linacre & Wright, 2000). Berdasarkan Gambar 4, 10 item secara keseluruhan telah memenuhi kategori tersebut. Lebih lanjut, untuk mengetahui kualitas item secara empiris dapat digunakan ketentuan dari William P. Fisher, Jr yaitu Nilai *Outfit Measure Correlation* 0.32 logit < *Pt Measure Corr* < 0.8 logit. Berdasarkan tabel diperoleh nilai *Ptmeasure Corr* dari rentang 0,38 sampai 0,62. Dengan demikian ke10 item dinyatakan fit.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	135	266	2.123	.116	.53	-6.34	.53	-4.86	.62	.36	82.7	63.3	B1
3	145	266	1.991	.114	.62	-4.82	.62	-3.85	.51	.37	73.7	62.8	B3
6	166	266	1.728	.110	.59	-4.96	1.19	1.68	.43	.39	76.7	62.7	B6
9	174	266	1.633	.108	.63	-4.27	.67	-3.34	.38	.39	73.7	62.4	B9
10	177	266	1.598	.108	.62	-4.42	.68	-3.23	.38	.40	73.3	62.5	B10
2	611	266	-1.400	.074	1.13	1.56	1.13	1.52	.50	.59	48.5	44.7	B2
8	615	266	-1.422	.074	1.41	4.54	1.36	3.93	.59	.59	47.4	44.8	B8
5	674	264	-1.791	.077	1.07	.90	1.04	.48	.64	.60	51.1	46.3	B5
4	727	266	-2.073	.079	1.23	2.61	1.26	2.77	.53	.60	43.6	48.6	B4
7	775	266	-2.386	.083	1.13	1.44	1.21	2.20	.60	.60	37.2	49.1	B7
MEAN	419.9	265.8	.000	.094	.90	-1.38	.97	-.27			60.8	54.7	
P.SD	264.6	.6	1.840	.017	.31	3.74	.29	3.04			15.8	8.1	

Gambar 4. Hasil Pengujian Item fit Berdasarkan Nilai Outfit dan Infit MNSQ

Reliabilitas

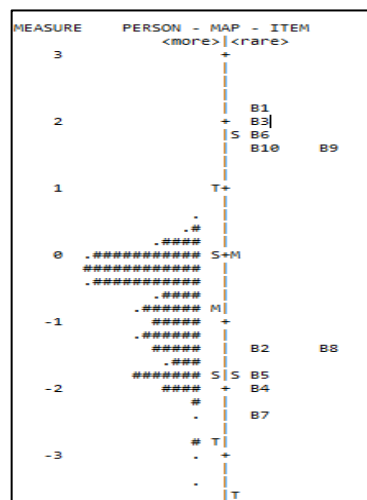
Reliabilitas merujuk pada konsistensi dan keandalan instrumen dalam mengukur suatu variabel atau konstruk (Lowenthal, 2001). Instrumen yang reliabel akan memberikan hasil yang konsisten jika digunakan dalam kondisi yang sama atau diulang pada waktu yang berbeda (Taherdoost, 2016). Salah satu indikator utama reliabilitas dalam model Rasch adalah *Person Separation Index* (PSI). PSI mengukur kemampuan tes untuk membedakan individu dengan tingkat kemampuan yang berbeda. Indeks ini mengukur sejauh mana tes dapat mengidentifikasi perbedaan kemampuan individu, dan nilai yang lebih tinggi menunjukkan reliabilitas yang lebih baik. Kriteria *Separation index* sebesar 1,5 dianggap cukup untuk melakukan analisis tingkat individu dan nilai > 2.5 dianggap cukup untuk melakukan analisis komparatif pada tingkat kelompok (Tennant & Conaghan, 2007). Berdasarkan Gambar 5, diperoleh nilai person separation index sebesar 1,76 > 1,5. Dengan demikian instrumen literasi matematis sudah cukup baik dalam mengestimasi kemampuan individu. Selain daripada itu, kita juga dapat melihat reliabilitas instrumen secara keseluruhan dengan melihat nilai *Cronbach Alpha*, dan dalam penelitian ini nilainya sebesar 0,83 (reliabel).

	TOTAL		MODEL		INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	15.8	10.0	-.896	.469	1.00	-.05	.95	-.17
SEM	.3	.0	.059	.004	.04	.08	.04	.07
P.SD	4.2	.1	.956	.061	.60	1.24	.72	1.06
S.SD	4.2	.1	.958	.061	.60	1.25	.72	1.06
MAX.	23.0	10.0	.697	1.102	4.07	4.14	9.90	3.71
MIN.	1.0	9.0	-6.111	.435	.09	-3.04	.10	-2.97
REAL RMSE	.530	TRUE SD	.796	SEPARATION	1.50	PERSON RELIABILITY	.82	
MODEL RMSE	.473	TRUE SD	.831	SEPARATION	1.76	PERSON RELIABILITY	.84	
S.E. OF PERSON MEAN = .059								
PERSON RAW SCORE-TO-MEASURE CORRELATION = .98 (approximate due to missing data)								
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .83 SEM = 2.36 (approximate due to missing data)								
STANDARDIZED (50 ITEM) RELIABILITY = .94								

Gambar 5. Hasil Pengujian Reliabilitas berdasarkan *Person Separation Index*

Wright Map

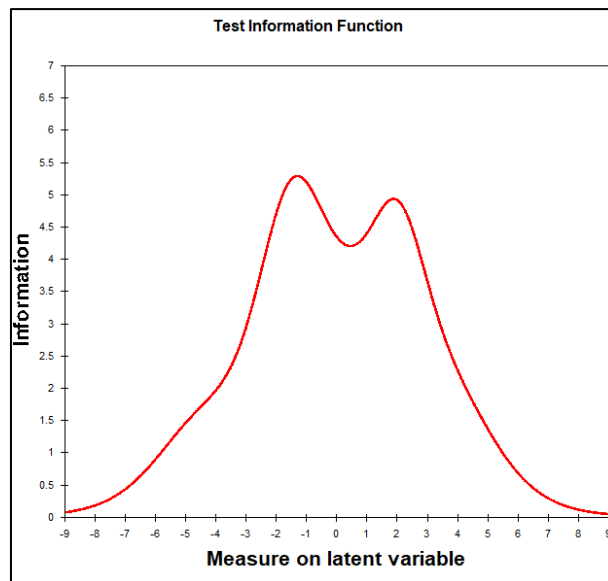
Dari Gambar 6, dapat terlihat jelas penyebaran kemampuan siswa dan tingkat kesulitan butir. Butir berada di tingkat kesukaran tinggi dan rendah. 5 butir tersukar ada di soal no 1, 3, 6, 9, dan 10. Sedangkan 5 butir termudah adalah butir 2, 8, 5, 4 dan 7. Butir yang paling sulit adalah butir 1 dengan nilai logit 2,213 dan butir termudah adalah butir 7 dengan nilai logit -2,386. Sedangkan kemampuan individu yang diukur kebanyakan berada pada level sedang yakni pada logit -2 sampai +1.



Gambar 6. *Wright Map* Dari Tes Literasi Matematika Yang Mewakili Perbandingan Langsung Antara Sebaran Orang Dan Sebaran Butir Soal

Fungsi Informasi Butir

Fungsi informasi butir mengukur seberapa baik suatu item dapat membedakan individu berdasarkan kemampuan mereka pada suatu titik tertentu. TIF memberikan gambaran keseluruhan tentang seberapa efektif tes tersebut dalam mengukur kemampuan pada berbagai tingkat θ . Menurut Doran (2005) fungsi informasi tes (TIF) θ adalah jumlah dari fungsi informasi butir soal secara individual. Adapun TIF dari instrumen literasi matematis dapat dilihat pada Gambar 7 di bawah ini.



Gambar 7. Fungsi Informasi Tes dari Tes Literasi Matematis

Gambar 7 menunjukkan grafik mengenai fungsi informasi pengukuran literasi matematis. Ini menunjukkan informasi *Fisher* untuk tes (set item) pada setiap titik di sepanjang variabel laten. Fungsi informasi tes melaporkan “informasi statistik” dalam data yang sesuai dengan setiap skor atau ukuran pada tes lengkap. Temuan untuk tes literasi matematis menunjukkan bahwa fungsi informasi tertinggi berada pada tingkat kemampuan rendah untuk literasi matematis, dengan informasi yang diperoleh dari hasil pengukuran cukup banyak. Pada tingkat kemampuan yang tinggi untuk literasi matematis, informasi yang diperoleh dari pengukuran relatif lebih tinggi dari kemampuan sedang. Hasil ini menunjukkan bahwa instrumen tes literasi matematis menghasilkan informasi yang optimal ketika diberikan kepada individu dengan tingkat *trait* yang rendah dan tinggi. Sedangkan instrumen literasi matematis ini tidak cocok digunakan untuk mengukur kemampuan literasi matematis dengan tingkat moderat.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka instrumen literasi matematis model PISA yang sudah dikembangkan sudah memenuhi kriteria valid dan reliabel. Valid berdasarkan hasil pengujian pakar terhadap materi, konstruk, maupun bahasa pada setiap butir instrumen. Serta reliabel dapat dilihat melalui item fit, wright map, person separation index dan fungsi informasi butir yang nilainya sdh memenuhi kriteria yang sudah ditetapkan dalam pengukuran model Rasch. Oleh karena itu, instrumen ini dapat secara akurat untuk mengukur kemampuan literasi matematis siswa SD khususnya pada tingkat kemampuan tinggi dan rendah sedangkan untuk kemampuan moderate kurang relevan karena tidak ada butir yang memiliki taraf sukar sedang.

REFERENSI

- Andriani, R. (2023). Sikap Ilmiah Siswa Sekolah Menengah Atas (SMA): Pengembangan Instrumen. *Jurnal Pendidikan MIPA*, 13(1),1-9. <https://doi.org/10.37630/jpm.v13i1.787>
- Arikunto, S. (2010). *Metode penelitian*. Rineka Cipta.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental in the human sciences. Third Edit, 11*.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen’s Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations.

- Applied Psychological Measurement*, 41(3), 178–194.
<https://doi.org/10.1177/0146621616677520>
- Debelak, R., & Koller, I. (2019). Debelak, R., & Koller, I. (2019). Testing the Local Independence Assumption of the Rasch Model With Q3-Based Nonparametric Model Tests. *Applied Psychological Measurement*, 014662161983550. <https://doi.org/10.1177/0146621619835501>
- Dewi, N. L. P. E. S., Pramatha, A. A. G. Y., & Wahyuni, L. G. E. (2024). Pelatihan Penyusunan Asesmen Diagnostik Untuk Guru Bahasa Inggris Sekolah Dasar. *Seminar Nasional Pengabdian kepada Masyarakat*, 9(1), 1733-1738.
- Doran, H. C. (2005). The Information Function for the One-Parameter Logistic Model: Is it Reliability? *Educational and Psychological Measurement*, 65(5), 665–675. <https://doi.org/10.1177/0013164404272500>
- Fauzi, A. A., Susongko, P., & Hayati, M. N. (2022). Tes Kemampuan Berpikir Kritis pada Pembelajaran IPA di SMP Berbasis Model Rasch. *PSEJ (Pancasakti Science Education Journal)*, 7(1), 59-67. <https://doi.org/10.24905/psej.v7i1.146>
- Fraenkel, J. R., & Wallen, N. E. (2009). *How to Design and Evaluate Research in Education (7thEd.)*. McGraw-Hill Companies.
- Geiger, V., Forgasz, H., & Goos, M. (2015). A Critical Orientation to Numeracy across the Curriculum. *ZDM: The International Journal on Mathematics Education*, 47(4), 611–624. <https://doi.org/10.1007/s11858-014-0648-1>
- Hayati, R., & Yuhana, Y. (2024). Development Of Mathematical Literacy Skill Instruments On Circle Material For Junior High School Students. *Journal of Authentic Research on Mathematics Education (JARME)*, 6(2), 165-174. <https://doi.org/10.37058/jarme.v6i2.11635>
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 18(3), 66–67. <https://doi.org/10.1136/eb-2015-102129>
- Holster, T. A., & Lake, J. (2016). Guessing and the Rasch Model. *Language Assessment Quarterly*, 13(2), 124–141. <https://doi.org/10.1080/15434303.2016.1160096>
- Indihadi, D., Suryana, D., & Ahmad, A. B. (2022). The analysis of construct validity of Indonesian creativity scale using Rasch model. *Creativity Studies*, 15(2), 560-576. <https://doi.org/10.3846/cs.2022.15182>
- Kerlinger, F. N., Lee, H. B., & Bhanumnavin, D. (2000). Foundations of behavioral research: The most sustainable popular textbook by Kerlinger & Lee (2000). *Journal of Social Development*, 13(2), 131–144.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85–106.
- Linacre, J. M., & Wright, B. (2000). *Winsteps*. URL: <http://www.Winsteps.Com/index.Htm> [accessed 2013-06-27][WebCite Cache].
- Lowenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales. 2nd ed.* Psychology Press.
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes Dan Non Tes*. Mitra Cendekia.
- Naga, D. S. (2023). *Teori sektor pada pengukuran mental*. PT. Ngarani Citrayasa.
- NCTM. (2000). *Principles and Standards for School Mathematics*. The National Council of Teachers of Mathematics, Inc.
- Nuryati, N., & Darsinah, D. (2021). Implementasi Teori Perkembangan Kognitif Jean Piaget dalam Pembelajaran Matematika di Sekolah Dasar. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, 3(2), 153-162. <https://doi.org/10.36232/jurnalpendidikandasar.v3i2.1186>

- OECD. (2018). *PISA 2018 Assessment And Analytical Framework: Mathematics, Reading, Science, Problem Solving And Financial Literacy*. OECD Publishing.
- Saputra, H. (2024). Perkembangan Berpikir Matematis Pada Anak Usia Sekolah Dasar. *Jemari (Jurnal Edukasi Madrasah Ibtidaiyah)*, 6(2), 53–64. <https://doi.org/10.30599/jemari.v6i2>.
- Sari, R. H. N., & Wijaya, A. (2017). Mathematical literacy of senior high school students in Yogyakarta. *Jurnal Riset Pendidikan Matematika*, 4(1), Article 1. <https://doi.org/10.21831/jrpm.v4i1.10649>.
- Shaumi, F. N., & Syamsuri, S. (2024). Pengembangan Instrumen Literasi Matematis dengan Konteks Kue Tradisional Banten pada Tingkat SMP. *Kognitif: Jurnal Riset HOTS Pendidikan Matematika*, 4(1), 585-604. <https://doi.org/10.51574/kognitif.v4i1.1691>
- Sitopu, J. W., Khairani, M., Roza, M., Judijanto, L., & Aslan, A. (2024). The Importance Of Integrating Mathematical Literacy In The Primary Education Curriculum: A Literature Review. *International Journal of Teaching and Learning*, 2(1), 121-134.
- Smith Jr., E. V. (2002). Understanding Rasch measurement: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3(2), 205–231.
- Suryani, Y. E. (2018). Aplikasi Rasch Model dalam Mengevaluasi Intelligenz Structure Test (IST). *Psikohumaniora: Jurnal Penelitian Psikologi*, 3(1), 73. <https://doi.org/10.21580/pjpp.v3i1.2052>
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management (IJARM)*, 5. <https://hal.science/hal-02546799>
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- Thiagarajan, S., & Others, A. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Council for Exceptional Children, 1920 Association Drive, Reston, Virginia 22091 (Single Copy, \$5. <https://eric.ed.gov/?id=ED090725>
- Widia, W., Fuadiah, N. F., & Surmilasari, N. (2024). Pengembangan Instrumen Asesmen Berbasis Literasi Numerasi Pada Materi Skala Perbandingan Pada Kelas V SD. *Pentagon : Jurnal Matematika Dan Ilmu Pengetahuan Alam*, 2(3), 99–107. <https://doi.org/10.62383/pentagon.v2i3.232>.