TEMATIK: Jurnal Konten Pendidikan Matematika
Volume 4, Number 1, Tahun 2026, pp. 91-104
ISSN: 2985-8844
Open Access: http://doi.org/10.55210/jkpm

Pengembangan dan Analisis Rasch Butir Soal Literasi Matematis pada Materi Persamaan Garis Lurus untuk Siswa SMP/MTs

Dzurotun Ni'mah[1*], Sintha Sih Dewanti[2], Sumbaji Putranto[3], Iwan Kuswidi[4]

[1,2,3,4] Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
Email : dzurotun.nimah123@gmail.com, sintha.dewanti@uin-suka.ac.id, sumbaji.putranto@uin-suka.ac.id, iwan.kuswidi@uin-suka.ac.id

INFORMASI ARTIKEL
Tersedia Online pada:
31 Juli, 2026

Kata Kunci:
Literasi Matematis; Model Rasch; Pengembangan Tes; Persamaan Garis Lurus; Winsteps

Keywords:
Mathematical Literacy; Rasch Model; Straight-Line Equations; Test Development; Winsteps

This is an open access article under the CC BY 4.0 license.
Copyright ©2026 by Author. Published by ***Universitas Islam Zainul Hasan Genggong***

Abstrak

Kemampuan literasi matematis siswa Indonesia masih tergolong rendah, sementara ketersediaan instrumen yang memiliki bukti kualitas psikometrik yang memadai untuk mengukurnya masih terbatas. Penelitian ini bertujuan mengembangkan, memvalidasi, dan menganalisis karakteristik butir tes literasi matematis pada materi persamaan garis lurus untuk siswa SMP/MTs menggunakan Model Rasch. Instrumen berupa 18 soal pilihan ganda dikembangkan berdasarkan proses literasi matematis dan konten persamaan garis lurus. Validitas isi dinilai oleh tiga ahli menggunakan koefisien Aiken's V, sedangkan respons empiris dari 238 siswa kelas VIII di MTs Zainul Hasan Probolinggo dianalisis menggunakan perangkat lunak Winsteps versi 3.73. Hasil penelitian menunjukkan bahwa seluruh butir memiliki validitas isi yang baik (Aiken's V = 0,78 - 1,00), reliabilitas item sangat tinggi (0,98), memenuhi asumsi unidimensionalitas, serta seluruh butir memenuhi kriteria fit Model Rasch. Tingkat kesukaran butir berada pada rentang -2,49 hingga 2,07 logit yang menunjukkan variasi tingkat kesukaran yang memadai, meskipun reliabilitas person masih berada pada kategori sedang (0,51). Secara keseluruhan, instrumen memiliki kualitas psikometrik yang memadai dan layak digunakan sebagai asesmen diagnostik awal untuk mengukur literasi matematis siswa pada materi persamaan garis lurus. Namun, penyempurnaan beberapa butir dan peningkatan reliabilitas person masih diperlukan agar instrumen dapat digunakan secara lebih akurat dalam pengambilan keputusan individual yang bersifat berisiko tinggi.

Abstract

Indonesian students continue to demonstrate low levels of mathematical literacy, while psychometrically sound instruments for assessing this competency remain limited. This study aimed to develop and validate a mathematical literacy test and examine its item characteristics on the topic of linear equations for junior secondary school students using the Rasch Model. The instrument consisted of 18 multiple-choice items developed based on the mathematical literacy processes and the content domain of linear equations. Content validity was evaluated by three experts using Aiken's V coefficient, while empirical responses from 238 eighth-grade students at MTs Zainul Hasan Probolinggo were analyzed using Winsteps version 3.73. The results showed that all items demonstrated good content validity (Aiken's V = $0.78 - 1.00$*), very high Rasch item reliability (0.98), satisfactory unidimensionality, and acceptable fit to the Rasch Model. Item difficulty ranged from −2.49 to 2.07 logits, indicating an adequate distribution of item difficulty, although Rasch person reliability remained at a moderate level (0.51). Overall, the instrument demonstrated satisfactory psychometric properties and is suitable for use as an initial diagnostic assessment of students' mathematical literacy on the topic of linear equations. However, refinement of several items and improvement of person reliability are still required before the instrument can be confidently used for high-stakes individual decision-making.*

PENDAHULUAN

Dalam pendidikan matematika, literasi matematis dipandang sebagai kemampuan penting yang memungkinkan siswa menggunakan konsep, prosedur, fakta, alat, dan penalaran matematika untuk merumuskan, menyelesaikan, serta menafsirkan permasalahan dalam berbagai konteks kehidupan nyata. Kemampuan ini tidak hanya berkaitan dengan ketepatan prosedur hitung, tetapi juga mencakup kecakapan memilih representasi, membangun model matematika, menghubungkan informasi kontekstual, dan mengambil keputusan berdasarkan pertimbangan matematis (OECD, 2023a; Stacey & Turner, 2015).

Konsep tersebut juga telah banyak diadopsi dalam konteks pendidikan Indonesia sebagai dasar pengembangan pembelajaran dan asesmen matematika yang menekankan kemampuan

*Corresponding author.
E-mail addresses: dzurotun.nimah123@gmail.com

menggunakan matematika untuk memecahkan permasalahan kontekstual (Kuswidi, 2015). Dalam konteks pendidikan abad ke-21, literasi matematis menjadi indikator penting untuk menilai sejauh mana pembelajaran matematika mampu mempersiapkan siswa menghadapi persoalan pribadi, sosial, pekerjaan, dan ilmiah yang semakin kompleks (NCTM, 2014).

Urgensi penguatan literasi matematis semakin nyata jika dikaitkan dengan capaian siswa Indonesia dalam asesmen internasional. Laporan PISA 2022 menunjukkan bahwa hanya 18% siswa Indonesia yang mencapai sekurang-kurangnya Level 2 dalam kemampuan matematika, jauh di bawah rata-rata OECD sebesar 69% (OECD, 2023b). Pada level ini, siswa mulai mampu mengenali bagaimana situasi sederhana dapat direpresentasikan secara matematis. Capaian tersebut menunjukkan bahwa banyak siswa Indonesia masih mengalami kesulitan dalam menerapkan konsep matematika pada situasi kontekstual, memilih informasi relevan, membangun representasi matematis, dan menafsirkan solusi secara tepat. Temuan ini sejalan dengan penelitian yang dilakukan oleh Wijaya yang menunjukkan bahwa siswa Indonesia sering mengalami hambatan dalam menyelesaikan tugas matematika berbasis konteks karena kesulitan memahami informasi, memilih data penting, dan mentransformasi situasi dunia nyata ke dalam model matematika (Wijaya et al., 2014).

Salah satu materi yang relevan untuk mengembangkan dan mengukur literasi matematis adalah persamaan garis lurus. Materi ini merepresentasikan hubungan linear antara dua variabel dan banyak digunakan untuk memodelkan fenomena sehari-hari, seperti hubungan jarak dan waktu, tarif dan biaya, harga dan jumlah barang, keuntungan, serta perubahan nilai secara tetap (Arumsari & Adirakasiwi, 2023). Secara konseptual, persamaan garis lurus merupakan bagian penting dalam aljabar sekolah karena mendukung pemahaman siswa terhadap variabel, gradien, titik potong, tabel, grafik, dan hubungan fungsional antarvariabel (NCTM, 2000; Usiskin, 1988). Dalam perspektif literasi matematis, materi ini menuntut siswa tidak hanya mampu menentukan gradien atau menyusun persamaan secara prosedural, tetapi juga mampu menerjemahkan situasi kontekstual ke dalam representasi matematis, menggunakan model linear untuk menyelesaikan masalah, dan menafsirkan makna parameter sesuai konteks (Rayhan et al., 2025).

Meskipun literasi matematis menjadi fokus penting dalam pendidikan matematika, praktik asesmen di sekolah masih sering didominasi oleh soal-soal rutin dan prosedural (Nopriyanti et al., 2024). Soal semacam ini umumnya menekankan penggunaan rumus, manipulasi simbol aljabar, atau penyelesaian langkah teknis, tetapi belum sepenuhnya menilai kemampuan siswa dalam merumuskan masalah kontekstual, menggunakan representasi, membangun model, dan menafsirkan hasil matematika dalam situasi nyata (Fauziah & Pandra, 2023). Putranto menunjukkan bahwa sebagian siswa masih memandang matematika hanya sebatas konsep dasar yang digunakan dalam kehidupan sehari-hari, karena mereka belum merasakan keterhubungan antara konsep matematika yang dipelajari di sekolah dan pemecahan masalah nyata (Putranto & Ratnasari, 2022). Pada materi persamaan garis lurus, asesmen di kelas juga masih cenderung berfokus pada kemampuan menghitung gradien, menentukan persamaan garis, atau menggambar grafik, sedangkan kemampuan siswa memahami hubungan linear dalam konteks kehidupan nyata belum banyak diukur secara mendalam (Arumsari & Adirakasiwi, 2023). Kondisi tersebut menunjukkan kebutuhan terhadap instrumen asesmen yang secara khusus dirancang untuk mengukur literasi matematis pada materi persamaan garis lurus, bukan sekadar penguasaan prosedural siswa.

Pengembangan instrumen asesmen literasi matematis perlu disertai proses validasi dan analisis butir yang memadai agar hasil pengukuran dapat dipercaya. Instrumen yang baik harus memiliki kesesuaian dengan konstruk yang diukur, tingkat kesukaran yang proporsional, serta informasi diagnostik yang bermakna bagi guru dan siswa (AERA et al., 2014). Dalam konteks ini, analisis butir tidak cukup hanya melihat skor benar dan salah secara umum, tetapi juga perlu mengkaji bagaimana setiap butir berfungsi dalam mengukur kemampuan siswa. Pendekatan analisis butir modern, seperti Model Rasch, diperlukan karena mampu mengestimasi karakteristik butir dan kemampuan peserta didik secara independen pada skala logit yang sama. Dengan demikian, kualitas setiap butir dapat dievaluasi secara lebih objektif dibandingkan

analisis klasik yang hasilnya sangat dipengaruhi oleh karakteristik sampel dan perangkat tes yang digunakan (Bond et al., 2021; Boone et al., 2014).

Model Rasch merupakan salah satu pendekatan dalam teori respons butir yang relevan untuk menganalisis kualitas instrumen secara mendalam karena menempatkan kemampuan siswa dan tingkat kesukaran butir pada skala logit yang sama (Kean et al., 2018). Melalui analisis Rasch, peneliti dapat mengidentifikasi kesesuaian butir dengan model, tingkat kesukaran butir, reliabilitas person, reliabilitas item, indeks separasi, serta peta distribusi antara kemampuan siswa dan kesukaran butir (Bond et al., 2021; Boone et al., 2014). Dengan demikian, Model Rasch relevan digunakan untuk mengevaluasi kualitas butir soal literasi matematis pada materi persamaan garis lurus.

Penelitian sebelumnya telah mengembangkan instrumen literasi matematis dan melakukan validasi empiris terhadap alat asesmen matematika. Suciati mengembangkan dan memvalidasi instrumen literasi matematis untuk mendukung assessment for learning di Indonesia dengan melibatkan bukti validitas isi, validitas konstruk, dan reliabilitas (Suciati et al., 2020). Dewanti juga menunjukkan bahwa pendekatan teori respons butir dapat digunakan untuk menganalisis karakteristik butir tes literasi matematis, termasuk kesesuaian model, parameter butir, distribusi kemampuan siswa, fungsi informasi tes, dan kesalahan pengukuran (Dewanti et al., 2021). Utami mengembangkan butir soal berbasis aljabar untuk mengukur literasi matematis siswa dan melaporkan bahwa butir yang dikembangkan memenuhi kriteria validitas dan reliabilitas (Utami et al., 2025). Namun, penelitian-penelitian tersebut masih cenderung berfokus pada pengembangan instrumen literasi matematis dalam domain yang luas atau pada konten aljabar secara umum. Sejauh penelusuran literatur, pengembangan instrumen yang secara khusus mengukur literasi matematis pada materi persamaan garis lurus bagi siswa SMP/MTs masih sangat terbatas. Selain itu, penelitian yang menganalisis instrumen tersebut menggunakan Model Rasch untuk memperoleh bukti psikometrik yang komprehensif, seperti kesesuaian butir dengan model, tingkat kesukaran butir, reliabilitas person dan item, indeks separasi, serta distribusi kemampuan siswa dan kesukaran butir, juga masih relatif sedikit.

Berdasarkan kesenjangan tersebut, penelitian ini berfokus pada pengembangan dan analisis butir soal literasi matematis pada materi persamaan garis lurus untuk siswa SMP/MTs. Kebaruan penelitian terletak pada penyusunan butir yang mengintegrasikan konteks kehidupan nyata, representasi hubungan linear, dan proses literasi matematis pada materi persamaan garis lurus, kemudian dianalisis menggunakan Model Rasch. Penelitian ini bertujuan mengembangkan, memvalidasi, dan menganalisis karakteristik butir soal literasi matematis pada materi persamaan garis lurus berdasarkan Model Rasch, meliputi validitas isi, kesesuaian butir, tingkat kesukaran, reliabilitas, indeks separasi, dan distribusi kemampuan siswa.

METODE

Desain Penelitian

Penelitian ini merupakan penelitian pengembangan (research and development) dengan pendekatan kuantitatif. Pengembangan instrumen mengacu pada model pengembangan tes yang dikemukakan oleh Lane et al. (2016), yang meliputi tahap perencanaan tes (test planning), pendefinisian konstruk (construct definition), pengembangan butir (item development), telaah butir (item review), uji coba empiris (field testing), analisis empiris (empirical analysis), dokumentasi, dan interpretasi hasil pengukuran. Tahapan tersebut diadaptasi sesuai tujuan penelitian, yaitu mengembangkan instrumen tes literasi matematis pada materi persamaan garis lurus untuk siswa SMP/MTs.

Secara berurutan, penelitian dilaksanakan melalui beberapa tahap, yaitu: (1) menyusun kisi-kisi instrumen berdasarkan proses literasi matematis dalam kerangka PISA dan capaian pembelajaran materi persamaan garis lurus; (2) mengembangkan butir soal sesuai indikator yang telah ditetapkan; (3) melakukan validasi isi oleh tiga orang ahli; (4) merevisi butir soal berdasarkan masukan para validator; (5) melaksanakan uji coba instrumen kepada siswa; (6) menganalisis karakteristik butir menggunakan Model Rasch melalui perangkat lunak Winsteps; dan (7) menetapkan instrumen akhir berdasarkan hasil analisis psikometrik.

Instrumen dikembangkan berdasarkan tiga proses literasi matematis dalam kerangka PISA, yaitu merumuskan situasi secara matematis (*formulate*), menggunakan konsep, fakta, prosedur, dan penalaran matematika (*employ*), serta menafsirkan dan mengevaluasi hasil matematika dalam konteks masalah (*interpret*)(OECD, 2023a). Ketiga proses tersebut diintegrasikan dengan materi persamaan garis lurus kelas VIII. Kerangka pengembangan instrumen disajikan pada Tabel 1.

Tabel 1. Kerangka Pengembangan Instrumen Literasi Matematis pada Materi Persamaan Garis Lurus

Proses literasi matematis	Fokus kemampuan yang diukur	Contoh tuntutan pada materi persamaan garis lurus
Formulate	Mengidentifikasi informasi penting, mengenali hubungan linear, dan menerjemahkan konteks ke bentuk matematika.	Menentukan variabel, membaca hubungan dari tabel/grafik, dan menyusun model hubungan linear dari situasi nyata.
Employ	Menggunakan konsep, prosedur, da representasi matematika untu memperoleh solusi.	Menggunakan gradien, titik potong, persamaan garis, tabel, atau grafik untuk menyelesaikan masalah.
Interpret	Menafsirkan hasil matematika kembali ke konteks masalah dan mengevaluasi kewajaran jawaban.	Menjelaskan makna gradien, titik potong, perubahan tetap, atau hasil perhitungan sesuai situasi yang diberikan.

Subjek dan Pengumpulan Data

Data penelitian dikumpulkan melalui uji coba instrumen tes literasi matematis pada materi persamaan garis lurus. Instrumen berupa tes pilihan ganda yang terdiri atas 18 butir soal dengan penskoran dikotomi, yaitu skor 1 untuk jawaban benar dan skor 0 untuk jawaban salah. Populasi sasaran penelitian adalah siswa kelas VIII SMP/MTs yang telah mempelajari materi persamaan garis lurus, sedangkan populasi terjangkau adalah seluruh siswa kelas VIII MTs Zainul Hasan Probolinggo Tahun Pelajaran 2025/2026. Subjek dipilih menggunakan teknik purposive sampling dengan kriteria telah menyelesaikan pembelajaran materi persamaan garis lurus, sehingga diperoleh 238 siswa sebagai responden. Uji coba instrumen dilaksanakan secara tatap muka pada semester genap Tahun Pelajaran 2025/2026 selama 60 menit di bawah pengawasan guru mata pelajaran dan peneliti. Sebelum tes dimulai, siswa diberikan petunjuk pengerjaan, kemudian mengerjakan soal secara mandiri tanpa menggunakan buku, catatan, maupun perangkat elektronik. Jumlah sampel sebanyak 238 siswa dinilai memadai untuk analisis Rasch karena mampu menghasilkan estimasi parameter butir yang stabil pada tahap pengembangan instrument (Wright & Stone, 1979).

Validitas Isi

Validitas isi dilakukan untuk menilai kesesuaian butir instrumen dengan konstruk literasi matematis, indikator materi persamaan garis lurus, konteks masalah, kejelasan bahasa, ketepatan stimulus, serta kesesuaian alternatif jawaban. Penilaian dilakukan oleh tiga validator yang memiliki kompetensi di bidang pendidikan matematika, terdiri atas satu dosen Program Studi Pendidikan Matematika bergelar doktor dengan kepakaran pada evaluasi pembelajaran dan pengembangan instrumen, serta dua guru matematika MTs yang memiliki pengalaman mengajar, aktif mengembangkan perangkat pembelajaran, dan sedang menempuh studi Magister Pendidikan Matematika. Setiap validator menilai seluruh butir instrumen menggunakan skala Likert empat tingkat, yaitu 1 = tidak relevan, 2 = kurang relevan, 3 = relevan, dan 4 = sangat relevan.

Hasil penilaian validator dianalisis menggunakan koefisien Aiken's V untuk mengukur tingkat validitas isi butir berdasarkan penilaian ahli terhadap relevansi butir dengan indikator yang diukur (Aiken, 1980). Rumus Aiken's V yang digunakan adalah

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

dengan:

V : validitas isi
s : r – l_0
l_0 : skor terendah pada skala penilaian
n : jumlah validator
c : jumlah kategori pada skala penilaian
r : skor dari validator.
Interpretasi nilai Aiken's V disajikan pada Tabel 2.

Tabel 2. Kategori Validitas Isi

Koefisien validitas	Kategori
V < 0,40	Validitas rendah
0,40 ≤ V ≤ 0,80	Validitas sedang
0,80 < V ≤ 1,00	Validitas tinggi

Butir dengan kategori validitas tinggi dinyatakan layak digunakan dalam uji coba empiris. Butir dengan kategori validitas sedang direvisi berdasarkan saran validator, sedangkan butir dengan kategori validitas rendah direvisi secara menyeluruh atau dipertimbangkan untuk tidak digunakan. Keputusan revisi juga mempertimbangkan catatan kualitatif validator, terutama berkaitan dengan kesesuaian konteks, kejelasan redaksi, ketepatan konsep matematika, dan kualitas alternatif jawaban.

Analisis Rasch

Analisis empiris butir dilakukan menggunakan Model Rasch dikotomi dengan bantuan aplikasi Winsteps versi 3.73. Perangkat lunak tersebut dipilih karena telah mendukung seluruh prosedur analisis yang diperlukan dalam penelitian ini, meliputi estimasi parameter Rasch, analisis kecocokan butir, reliabilitas, *separation*, *Wright Map*, serta pengujian asumsi unidimensionalitas dan independensi lokal. Data respons siswa disusun dalam format *person-by-item*, dengan 238 baris responden dan 18 kolom butir. Setiap respons dikodekan 1 untuk jawaban benar dan 0 untuk jawaban salah. Secara matematis, Model Rasch dikotomi menyatakan bahwa peluang siswa menjawab benar suatu butir ditentukan oleh selisih antara kemampuan siswa dan tingkat kesukaran butir: $\log\left(\frac{P_{ni}}{1-P_{ni}}\right) = \theta_n - \delta_i$ dengan P_{ni} adalah peluang siswa n menjawab benar butir i, θ_n adalah kemampuan siswa, dan δ_i adalah tingkat kesukaran butir.

Output Winsteps yang digunakan meliputi *Summary Statistics*, *Item Statistics*: *Misfit Order*, *Item Statistics*: *Measure Order*, *Wright Map*, *PCA of standardized residuals*, dan *Largest Standardized Residual Correlations*. *Summary Statistics* digunakan untuk menafsirkan *person reliability*, *item reliability*, *person separation*, *item separation*, dan KR-20. *Item Statistics*: *Misfit Order* digunakan untuk mengevaluasi kecocokan butir berdasarkan nilai *Infit MNSQ*, *outfit MNSQ*, *ZSTD*, dan *point-measure correlation*. *Item Statistics*: *Measure Order* digunakan untuk menafsirkan tingkat kesukaran butir berdasarkan nilai measure pada skala logit. *Wright Map* digunakan untuk membaca kesesuaian distribusi kemampuan siswa dan tingkat kesukaran butir dalam satu kontinum pengukuran. *PCA of standardized residuals* digunakan untuk mengevaluasi unidimensionalitas, sedangkan korelasi residual terstandar digunakan untuk meninjau indikasi dependensi lokal antarbutir.

Tabel 3. Kriteria Interpretasi Utama dalam Analisis Rasch

Indikator	Kriteria operasional	Interpretasi
Infit/Outfit MNSQ	0,50 ≤ MNSQ ≤ 1,50	Butir masih produktif untuk pengukuran.
ZSTD	-2,00 ≤ *ZSTD* ≤ +2,00	Indikator pendukung; sensitif terhadap ukuran sampel.
Point-measure correlation	Bernilai positif	Butir bergerak searah dengan konstruk yang diukur.
Person reliability	Semakin tinggi semakin baik	Menunjukkan kemampuan instrumen membedakan kemampuan siswa.

Indikator	Kriteria operasional	Interpretasi
Item reliability	Semakin tinggi semakin baik	Menunjukkan stabilitas hierarki tingkat kesukaran butir.
Raw variance explained by measures	≥ 20%	Mendukung keberadaan dimensi utama yang diukur.
Unexplained variance in 1st contrast	< 2,00 eigenvalue	Tidak menunjukkan dimensi sekunder yang dominan.
Residual correlation antarbutir	< 0,30	Tidak mengindikasikan dependensi lokal yang kuat.

Butir dikategorikan sesuai dengan Model Rasch apabila nilai *Infit* dan *outfit MNSQ* berada pada rentang produktif, nilai *ZSTD* berada pada rentang yang diharapkan sebagai indikator pendukung, dan *point-measure correlation* bernilai positif (Bond et al., 2021; Boone et al., 2014). Butir yang memiliki salah satu indikator di luar kriteria tidak langsung dihapus, tetapi ditinjau kembali berdasarkan kesesuaian isi, kejelasan redaksi, ketepatan kunci jawaban, fungsi pengecoh, dan pola respons siswa. Kategori tingkat kesukaran butir pada skala logit disajikan pada Tabel 4.

Tabel 4. Kategori Tingkat Kesukaran Butir pada Skala Logit

Nilai measure (b)	Kategori
b < -2,00	Sangat mudah
-2,00 ≤ b < -1,00	Mudah
-1,00 ≤ b ≤ 1,00	Sedang
1,00 < b ≤ 2,00	Sulit
b > 2,00	Sangat sulit

Analisis dalam penelitian ini difokuskan pada validitas isi, reliabilitas, separasi, unidimensionalitas, indikasi independensi lokal, kecocokan butir, tingkat kesukaran, dan *Wright Map*. Analisis *differential item functioning* dan *person fit* secara rinci belum dilaporkan. Selain itu, fungsi distraktor belum dapat dievaluasi secara penuh karena data yang dianalisis pada Winsteps telah dikodekan secara dikotomi, sehingga informasi pilihan pengecoh tidak tersedia secara terpisah. Keterbatasan ini dijelaskan secara eksplisit pada bagian akhir naskah agar interpretasi hasil tetap proporsional.

HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan instrumen tes literasi matematis pada materi persamaan garis lurus untuk siswa SMP/MTs. Instrumen terdiri atas 18 butir soal pilihan ganda dengan penskoran dikotomi. Butir soal dikembangkan berdasarkan proses literasi matematis yang meliputi merumuskan situasi secara matematis, menggunakan konsep dan prosedur matematika, serta menafsirkan hasil matematika dalam konteks masalah. Ketiga proses tersebut diintegrasikan dengan materi persamaan garis lurus melalui konteks kehidupan nyata, seperti tarif ojek online, desain tangga, dan tanjakan kendaraan.

Validitas Isi Instrumen

Validitas isi dianalisis menggunakan koefisien Aiken's V berdasarkan penilaian tiga validator ahli. Hasil analisis menunjukkan bahwa nilai Aiken's V berada pada rentang 0,78 sampai 1,00 dengan rata-rata 0,94. Dari 18 butir soal yang dikembangkan, sebanyak 16 butir berada pada kategori validitas tinggi dan 2 butir berada pada kategori validitas sedang. Tidak terdapat butir yang termasuk kategori validitas rendah.

Tabel 5. Hasil Validitas Isi Butir Soal

Nomor butir	Aiken's V	Kategori	Keputusan
S1	1,00	Tinggi	Layak digunakan
S2	0,80	Sedang	Direvisi ringan

Nomor butir	Aiken's V	Kategori	Keputusan
S3	0,82	Tinggi	Layak digunakan
S4	0,98	Tinggi	Layak digunakan
S5	0,78	Sedang	Direvisi ringan
S6	0,87	Tinggi	Layak digunakan
S7	1,00	Tinggi	Layak digunakan
S8	1,00	Tinggi	Layak digunakan
S9	0,98	Tinggi	Layak digunakan
S10	1,00	Tinggi	Layak digunakan
S11	1,00	Tinggi	Layak digunakan
S12	0,84	Tinggi	Layak digunakan
S13	0,96	Tinggi	Layak digunakan
S14	1,00	Tinggi	Layak digunakan
S15	1,00	Tinggi	Layak digunakan
S16	1,00	Tinggi	Layak digunakan
S17	0,87	Tinggi	Layak digunakan
S18	0,98	Tinggi	Layak digunakan

Hasil tersebut menunjukkan bahwa secara substansial butir soal telah sesuai dengan konstruk literasi matematis, indikator materi persamaan garis lurus, konteks masalah, serta kejelasan bahasa dan alternatif jawaban. Dua butir yang memperoleh kategori sedang, yaitu S2 dan S5, tetap dapat digunakan setelah dilakukan revisi ringan berdasarkan catatan validator. Revisi terutama diarahkan pada perbaikan redaksi, kejelasan stimulus, dan ketepatan alternatif jawaban agar butir lebih mudah dipahami siswa tanpa mengurangi tuntutan literasi matematis yang diukur.

Reliabilitas, Separasi, dan Ringkasan Pengukuran Rasch

Analisis empiris butir dilakukan menggunakan Model Rasch dikotomi dengan bantuan Winsteps versi 3.73. Data yang dianalisis melibatkan 238 siswa dan 18 butir soal. Hasil *Summary Statistics* menunjukkan bahwa rata-rata skor siswa adalah 9,4 dari skor maksimum 18. Rata-rata kemampuan siswa berada pada 0,13 logit dengan simpangan baku 0,84. Nilai ini menunjukkan bahwa kemampuan rata-rata siswa sedikit berada di atas rata-rata tingkat kesukaran butir yang dikalibrasi pada 0,00 logit.

Tabel 6. Ringkasan Statistik Rasch

Komponen	Nilai
Jumlah siswa	238
Jumlah butir	18
Rata-rata skor siswa	9,4
Rata-rata kemampuan siswa	0,13 logit
SD kemampuan siswa	0,84
Person separation	1,01
Person reliability	0,51
Model *person reliability*	0,55
Cronbach Alpha/KR-20	0,54
Rata-rata measure butir	0,00 logit
SD measure butir	1,17
Item separation	7,18
Item reliability	0,98

Nilai *person reliability* sebesar 0,51, *person separation* sebesar 1,01, dan KR-20 sebesar 0,54 menunjukkan bahwa konsistensi pengukuran kemampuan siswa berada pada kategori sedang. Nilai ini lebih baik dibandingkan hasil analisis sebelumnya, tetapi instrumen masih belum optimal untuk membedakan siswa ke dalam banyak strata kemampuan. Kondisi tersebut dapat disebabkan oleh sebaran kemampuan responden yang relatif homogen, jumlah butir yang

masih terbatas, atau sebagian butir yang belum cukup sensitif terhadap variasi kemampuan literasi matematis siswa.

Sebaliknya, *item reliability* sebesar 0,98 dan *item separation* sebesar 7,18 menunjukkan bahwa hierarki tingkat kesukaran butir sangat stabil. Artinya, jumlah dan karakteristik responden dalam uji coba ini sudah memadai untuk membedakan butir berdasarkan tingkat kesukarannya. Dengan kata lain, instrumen telah mampu menghasilkan urutan tingkat kesukaran butir yang jelas dan konsisten, meskipun kemampuan instrumen dalam membedakan kemampuan siswa masih berada pada kategori sedang.

Unidimensionalitas dan Independensi Lokal

Unidimensionalitas instrumen dianalisis melalui *PCA of standardized residuals* pada output Winsteps. Hasil analisis menunjukkan bahwa total raw variance in observations sebesar 25,3 eigenvalue. *Raw variance explained by measures* sebesar 7,3 eigenvalue atau 28,8%, yang menunjukkan bahwa varians yang dijelaskan oleh ukuran Rasch telah melampaui batas operasional minimal 20%. Varians yang dijelaskan oleh persons sebesar 2,4 eigenvalue atau 9,4%, sedangkan varians yang dijelaskan oleh items sebesar 4,9 eigenvalue atau 19,4%.

Tabel 7. Hasil PCA Residual dan Korelasi Residual Antarbutir

Komponen	Nilai empiris	Interpretasi
Total raw variance in observations	*25,3 eigenvalue (100,0%)*	Varians total observasi dalam analisis residual.
Raw variance explained by measures	*7,3 eigenvalue (28,8%)*	Mendukung keberadaan dimensi utama karena melampaui batas operasional 20%.
Raw variance explained by persons	*2,4 eigenvalue (9,4%)*	Kontribusi variasi kemampuan siswa terhadap varians terjelaskan.
Raw variance explained by items	*4,9 eigenvalue (19,4%)*	Kontribusi variasi tingkat kesukaran butir terhadap varians terjelaskan.
Raw unexplained variance	*18,0 eigenvalue (71,2%)*	Varians residual yang belum dijelaskan oleh ukuran Rasch.
Unexplained variance in 1st contrast	*1,6 eigenvalue (6,5%; 9,1% of unexplained)*	Tidak menunjukkan dimensi sekunder dominan.
Unexplained variance in 2nd contrast	*1,6 eigenvalue (6,1%; 8,6% of unexplained)*	Tidak menunjukkan dimensi sekunder dominan.
Unexplained variance in 3rd contrast	*1,5 eigenvalue (5,8%; 8,1% of unexplained)*	Tidak menunjukkan dimensi sekunder dominan.
Unexplained variance in 4th contrast	*1,4 eigenvalue (5,6%; 7,9% of unexplained)*	Tidak menunjukkan dimensi sekunder dominan.
Unexplained variance in 5th contrast	*1,3 eigenvalue (5,0%; 7,0% of unexplained)*	Tidak menunjukkan dimensi sekunder dominan.
Largest standardized residual correlation	*0,28 absolut (S8-S11)*	Tidak menunjukkan dependensi lokal yang kuat karena masih di bawah 0,30.

Nilai unexplained variance pada first contrast sebesar 1,6 eigenvalue dan second contrast sebesar 1,6 eigenvalue menunjukkan bahwa tidak terdapat kontras residual yang cukup kuat untuk membentuk dimensi sekunder dominan. Dengan demikian, hasil PCA residual mendukung asumsi unidimensionalitas instrumen, meskipun proporsi varians yang dijelaskan oleh measures masih berada pada kategori cukup dan dapat ditingkatkan melalui penyempurnaan butir.

Indikasi independensi lokal ditinjau melalui *Largest Standardized Residual Correlations*. Nilai residual correlation terbesar secara absolut ditemukan pada pasangan S8 dan S11 sebesar 0,28 dengan arah negatif. Karena nilai tersebut masih berada di bawah batas 0,30, tidak ditemukan indikasi dependensi lokal yang kuat antarbutir. Temuan ini menunjukkan bahwa

respons terhadap satu butir tidak secara berlebihan bergantung pada respons terhadap butir lain setelah kemampuan siswa dikendalikan oleh Model Rasch.

Kecocokan Butir dengan Model Rasch

Kecocokan butir dianalisis berdasarkan nilai *Infit MNSQ*, *outfit MNSQ*, *ZSTD*, dan *point-measure correlation*. Berdasarkan kriteria MNSQ, seluruh butir masih berada pada rentang produktif 0,50 sampai 1,50, sehingga secara umum semua butir dapat dipertahankan untuk pengukuran. Namun, lima butir perlu dicermati secara substantif, yaitu S1, S18, S12, S9, dan S4. Kelima butir tersebut tidak harus langsung dihapus, tetapi perlu diperiksa kembali karena memiliki *point-measure correlation* rendah atau indikator *ZSTD* yang kurang ideal.

Tabel 8. Hasil Kecocokan Butir Berdasarkan Model Rasch

Butir	Measure	*Infit MNSQ*	*Infit ZSTD*	*Outfit MNSQ*	Outfit *ZSTD*	PT-Measure Corr.	Keterangan
S1	-2,49	1,04	0,3	1,42	1,5	0,14	Fit; dicermati
S2	-1,32	0,92	-0,9	0,79	-1,7	0,44	Fit
S3	-1,86	0,97	-0,2	0,81	-1,0	0,35	Fit
S4	0,44	1,13	2,6	1,13	1,9	0,22	Fit; dicermati
S5	-0,87	0,94	-0,9	0,89	-1,1	0,43	Fit
S6	0,36	0,98	-0,4	1,00	0,0	0,38	Fit
S7	-0,01	0,90	-2,1	0,87	-2,3	0,49	Fit
S8	0,90	1,03	0,6	1,01	0,2	0,31	Fit
S9	0,99	1,15	2,5	1,26	2,6	0,14	Fit; dicermati
S10	0,36	0,97	-0,5	1,00	0,1	0,38	Fit
S11	1,51	1,03	0,3	1,03	0,3	0,27	Fit
S12	2,07	1,08	0,7	1,39	1,9	0,11	Fit; dicermati
S13	-0,94	0,91	-1,2	0,92	-0,8	0,44	Fit
S14	-0,52	0,91	-1,6	0,89	-1,5	0,47	Fit
S15	-0,80	0,97	-0,4	0,95	-0,5	0,39	Fit
S16	0,57	0,99	-0,2	0,96	-0,6	0,38	Fit
S17	0,42	0,92	-1,7	0,92	-1,2	0,45	Fit
S18	1,21	1,10	1,4	1,40	3,4	0,16	Fit; dicermati

Butir S1 merupakan butir termudah dengan nilai measure -2,49 logit dan dijawab benar oleh 217 dari 238 siswa. Nilai *outfit MNSQ* S1 masih berada dalam rentang produktif, tetapi *point-measure correlation* hanya sebesar 0,14. Kondisi ini menunjukkan bahwa S1 terlalu mudah dan memiliki hubungan yang lemah dengan konstruk literasi matematis yang diukur. Oleh karena itu, S1 sebaiknya direvisi agar tetap berfungsi sebagai butir dasar, tetapi tidak sekadar menuntut pembacaan informasi langsung.

Butir S18, S12, S9, dan S4 juga perlu dicermati. S18 memiliki outfit *ZSTD* sebesar 3,4 dan *point-measure correlation* sebesar 0,16; S12 memiliki *point-measure correlation* sebesar 0,11 dan menjadi butir paling sulit; S9 memiliki *Infit ZSTD* sebesar 2,5, outfit *ZSTD* sebesar 2,6, dan *point-measure correlation* sebesar 0,14; sedangkan S4 memiliki *Infit ZSTD* sebesar 2,6 dan *point-measure correlation* sebesar 0,22. Karena nilai MNSQ keempat butir masih produktif, keputusan revisi sebaiknya tidak hanya didasarkan pada statistik, tetapi juga mempertimbangkan kejelasan konteks, ketepatan kunci jawaban, kualitas pengecoh, serta kesesuaian proses literasi matematis yang dituntut.

Tingkat Kesukaran Butir

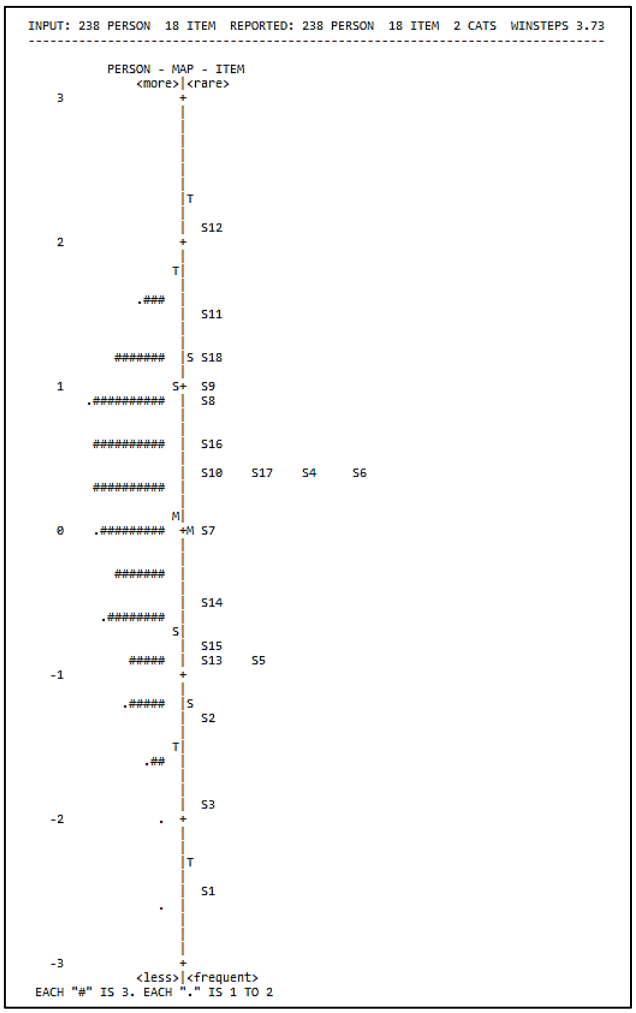
Tingkat kesukaran butir dianalisis berdasarkan nilai measure pada skala logit. Nilai measure yang lebih tinggi menunjukkan butir yang lebih sulit, sedangkan nilai measure yang lebih rendah menunjukkan butir yang lebih mudah. Hasil analisis menunjukkan bahwa tingkat kesukaran butir berada pada rentang -2,49 sampai 2,07 logit. Butir tersulit adalah S12 dengan nilai 2,07 logit, sedangkan butir termudah adalah S1 dengan nilai -2,49 logit.

Tabel 9. Kategori Tingkat Kesukaran Butir

Kategori	Rentang measure	Butir	Jumlah
Sangat sulit	b > 2,00	S12	1
Sulit	1,00 < b ≤ 2,00	S11, S18	2
Sedang	-1,00 ≤ b ≤ 1,00	S4, S5, S6, S7, S8, S9, S10, S13, S14, S15, S16, S17	12
Mudah	-2,00 ≤ b < -1,00	S2, S3	2
Sangat mudah	b < -2,00	S1	1

Sebaran tingkat kesukaran menunjukkan bahwa sebagian besar butir berada pada kategori sedang, yaitu 12 dari 18 butir. Dua butir berada pada kategori sulit, satu butir sangat sulit, dua butir mudah, dan satu butir sangat mudah. Komposisi ini menunjukkan bahwa instrumen cukup sesuai untuk mengukur siswa dengan kemampuan menengah. Namun, distribusi butir masih kurang proporsional pada area kemampuan ekstrem karena hanya satu butir berada pada kategori sangat mudah dan satu butir berada pada kategori sangat sulit.

Analisis *Wright Map*



Gambar 1. *Wright Map (Person–Item Map)* Distribusi Kemampuan Siswa dan Tingkat Kesukaran

Berdasarkan Gambar 1, Wright Map menunjukkan distribusi kemampuan siswa (person measure) dan tingkat kesukaran butir (item measure) pada skala logit yang sama. Sebaran kemampuan siswa berada pada rentang sekitar −2,49 hingga 2,07 logit, sedangkan tingkat kesukaran butir tersebar dari sekitar −2,56 hingga 2,28 logit. Pada sisi butir, S12 menempati posisi logit tertinggi sehingga merupakan butir dengan tingkat kesukaran paling tinggi,

sedangkan S1 berada pada posisi logit terendah sehingga merupakan butir yang paling mudah. Sebagian besar butir terkonsentrasi pada rentang −1 hingga 1 logit, yang menunjukkan bahwa instrumen didominasi oleh butir dengan tingkat kesukaran menengah dan sesuai dengan distribusi kemampuan mayoritas responden.

Rata-rata kemampuan siswa berada pada 0,13 logit, sedikit lebih tinggi dibandingkan rata-rata tingkat kesukaran butir yang secara konvensi ditetapkan pada 0 logit. Hal ini menunjukkan bahwa secara umum instrumen memiliki targeting yang baik karena tingkat kesukaran butir relatif sesuai dengan kemampuan responden. Meskipun demikian, Wright Map juga memperlihatkan bahwa pada rentang kemampuan sangat tinggi masih terdapat sedikit butir yang mampu mengukur kemampuan tersebut secara optimal. Oleh karena itu, pengembangan instrumen selanjutnya dapat mempertimbangkan penambahan butir dengan tingkat kesukaran yang lebih tinggi agar cakupan pengukuran menjadi lebih merata pada seluruh rentang kemampuan siswa.

Pembahasan

Hasil penelitian menunjukkan bahwa instrumen tes literasi matematis pada materi persamaan garis lurus memiliki bukti validitas isi yang baik. Nilai Aiken's V berada pada rentang 0,78–1,00, yang menunjukkan bahwa sebagian besar butir dinilai relevan oleh validator terhadap konstruk literasi matematis, indikator materi, konteks soal, serta kejelasan bahasa. Temuan ini mengindikasikan bahwa butir yang dikembangkan telah memiliki kesesuaian dengan konstruk yang diukur sehingga layak digunakan untuk tahap uji empiris. Validitas isi merupakan tahap yang krusial dalam pengembangan instrumen karena memastikan setiap butir merepresentasikan domain kemampuan yang hendak diukur sebelum dilakukan pengujian psikometrik (AERA et al., 2014; Aiken, 1980).

Instrumen yang dikembangkan juga sejalan dengan kerangka literasi matematis PISA yang menekankan tiga proses utama, yaitu formulate, employ, dan interpret (Stacey & Turner, 2015). Hal tersebut tercermin pada penggunaan konteks autentik, seperti tarif ojek online, desain tangga, dan tanjakan kendaraan, yang dirancang untuk mengukur kemampuan siswa dalam merumuskan situasi secara matematis, menggunakan konsep persamaan garis lurus, serta menafsirkan hasil penyelesaian dalam konteks kehidupan nyata. Secara empiris, penggunaan konteks tersebut didukung oleh hasil validitas isi dengan koefisien Aiken's V sebesar 0,78 − 1,00, yang menunjukkan bahwa para validator menilai konteks dan representasi masalah telah sesuai dengan konstruk literasi matematis. Selain itu, hasil analisis Rasch menunjukkan bahwa seluruh butir memenuhi kriteria item fit, sehingga konteks yang digunakan mampu mengukur konstruk yang sama secara konsisten tanpa menunjukkan adanya penyimpangan pengukuran. Temuan ini mengindikasikan bahwa penggunaan konteks autentik tidak hanya meningkatkan relevansi soal terhadap situasi kehidupan sehari-hari, tetapi juga menghasilkan instrumen yang memiliki karakteristik psikometrik yang baik. Hasil tersebut sejalan dengan penelitian-penelitian pengembangan instrumen literasi matematis yang melaporkan bahwa penggunaan konteks kehidupan nyata dapat meningkatkan representativitas konstruk sekaligus kualitas psikometrik instrumen setelah dianalisis menggunakan Model Rasch.

Hasil analisis Rasch menunjukkan bahwa seluruh butir masih bekerja secara produktif berdasarkan kriteria MNSQ. Hal ini berarti pola respons siswa terhadap butir secara umum masih sesuai dengan ekspektasi Model Rasch, yaitu siswa dengan kemampuan lebih tinggi cenderung memiliki peluang lebih besar untuk menjawab benar dibandingkan siswa dengan kemampuan lebih rendah (Bond et al., 2021; Boone et al., 2014). Meskipun demikian, beberapa butir tetap memerlukan pemeriksaan substantif karena kualitas butir tidak hanya ditentukan oleh MNSQ, tetapi juga oleh korelasi butir, *ZSTD*, kejelasan konstruksi soal, dan ketepatan fungsi pengecoh.

Bukti tambahan dari PCA residual memperkuat argumentasi bahwa instrumen terutama mengukur satu konstruk utama, yaitu literasi matematis pada materi persamaan garis lurus. *Raw variance explained by measures* sebesar 28,8% menunjukkan bahwa dimensi Rasch menjelaskan bagian varians yang memadai, sementara first contrast sebesar 1,6 eigenvalue

tidak menunjukkan adanya dimensi sekunder yang dominan. Selain itu, korelasi residual antarbutir yang berada di bawah 0,30 menunjukkan bahwa tidak terdapat indikasi dependensi lokal yang kuat. Dengan demikian, respons siswa terhadap butir-butir instrumen dapat ditafsirkan relatif independen setelah dikontrol oleh kemampuan laten yang diukur.

Lima butir, yaitu S1, S18, S12, S9, dan S4, perlu dicermati lebih lanjut. S1 terlalu mudah dan memiliki *point-measure correlation* rendah sehingga berpotensi kurang efektif dalam membedakan kemampuan siswa. S18 dan S9 memiliki *ZSTD* yang tinggi serta korelasi butir rendah, sehingga perlu diperiksa dari aspek kejelasan konteks, ketepatan kunci, dan kemungkinan ambiguitas penalaran. S12 menjadi butir tersulit dan hanya dijawab benar oleh 36 dari 238 siswa, sehingga perlu dipastikan bahwa kesulitannya berasal dari tuntutan literasi matematis, bukan dari redaksi atau informasi konteks yang kurang jelas. S4 juga perlu diperiksa karena memiliki *Infit ZSTD* tinggi dan korelasi butir yang relatif rendah. Dalam interpretasi Rasch, MNSQ tetap diprioritaskan sebagai indikator utama, sedangkan *ZSTD* dan *point-measure correlation* digunakan sebagai bukti pendukung untuk memutuskan revisi butir.

Hasil reliabilitas menunjukkan pola yang perlu dibaca secara hati-hati. *Item reliability* sebesar 0,98 dan *item separation* sebesar 7,18 menunjukkan bahwa hierarki tingkat kesukaran butir sangat stabil. Artinya, sampel siswa dalam penelitian ini mampu membedakan butir berdasarkan tingkat kesukarannya dengan baik. Namun, *person reliability* sebesar 0,51, *person separation* sebesar 1,01, dan KR-20 sebesar 0,54 menunjukkan bahwa kemampuan instrumen dalam membedakan tingkat kemampuan siswa masih berada pada kategori sedang. Kondisi ini dapat disebabkan oleh rentang kemampuan siswa yang relatif homogen, jumlah butir yang belum banyak, atau distribusi tingkat kesukaran butir yang belum sepenuhnya menjangkau seluruh variasi kemampuan siswa.

Dari aspek tingkat kesukaran, rentang measure dari -2,49 sampai 2,07 logit menunjukkan bahwa instrumen memiliki variasi tingkat kesukaran, tetapi masih terkonsentrasi pada kategori sedang. Dalam perspektif literasi matematis, butir yang lebih sulit umumnya menuntut proses berpikir yang lebih kompleks, seperti generalisasi, pemodelan, penggunaan informasi kontekstual, dan penafsiran hasil matematika dalam situasi nyata (OECD, 2023a; Stacey & Turner, 2015). Posisi S12 sebagai butir tersulit dapat dipahami karena butir tersebut menuntut generalisasi dan pemodelan hubungan linear yang lebih abstrak dibandingkan butir lain. Akan tetapi, *point-measure correlation* S12 yang rendah menunjukkan perlunya pemeriksaan lebih lanjut agar tingkat kesulitan butir tetap berkaitan kuat dengan konstruk literasi matematis yang diukur.

Secara keseluruhan, hasil penelitian menunjukkan bahwa instrumen memiliki validitas isi yang baik, seluruh butir masih produktif berdasarkan kriteria MNSQ, hierarki kesukaran butir stabil, mendukung asumsi unidimensionalitas, dan tidak menunjukkan dependensi lokal yang kuat. Namun, instrumen masih perlu disempurnakan agar lebih sensitif dalam membedakan kemampuan siswa. Revisi terutama perlu difokuskan pada S1, S18, S12, S9, dan S4, serta penambahan butir pada area kemampuan tinggi dan rendah agar cakupan pengukuran menjadi lebih seimbang.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil menghasilkan instrumen tes literasi matematis pada materi persamaan garis lurus untuk siswa SMP/MTs yang memiliki kualitas psikometrik yang memadai berdasarkan hasil validasi dan analisis Rasch. Hasil validitas isi menunjukkan bahwa seluruh butir memiliki tingkat relevansi yang baik terhadap konstruk literasi matematis, indikator materi, konteks soal, dan aspek kebahasaan sehingga layak digunakan pada tahap pengujian empiris.

Hasil analisis Rasch menunjukkan bahwa instrumen memenuhi karakteristik pengukuran yang baik. Seluruh butir berfungsi secara produktif dalam mengukur kemampuan literasi matematis, mendukung asumsi unidimensionalitas, serta tidak menunjukkan adanya indikasi dependensi lokal yang berarti. Selain itu, reliabilitas butir yang tinggi menunjukkan

bahwa tingkat kesukaran butir tersusun secara stabil sehingga instrumen mampu memberikan pengukuran yang konsisten terhadap konstruk yang diukur.

Meskipun demikian, reliabilitas responden masih berada pada kategori sedang, yang mengindikasikan bahwa kemampuan instrumen dalam membedakan tingkat kemampuan siswa masih dapat ditingkatkan. Oleh karena itu, penelitian selanjutnya disarankan melakukan penyempurnaan instrumen, terutama melalui penambahan dan penyebaran tingkat kesukaran butir yang lebih merata agar instrumen mampu mengukur kemampuan siswa pada berbagai rentang kemampuan secara lebih optimal.

Saran

Instrumen ini dapat digunakan sebagai asesmen diagnostik awal untuk memetakan kemampuan literasi matematis siswa pada materi persamaan garis lurus. Namun, karena person reliability dan KR-20 masih berada pada kategori sedang, instrumen belum disarankan untuk digunakan sebagai dasar pengambilan keputusan individual yang bersifat penting.

Butir S1, S18, S12, S9, dan S4 perlu direvisi sebelum instrumen digunakan secara lebih luas. Revisi dapat difokuskan pada kejelasan redaksi, ketepatan stimulus, kualitas alternatif jawaban, ketepatan kunci, dan kesesuaian butir dengan konstruk literasi matematis. S1 perlu dibuat lebih informatif sebagai butir mudah, sedangkan S12 perlu dipastikan benar-benar mengukur kompleksitas literasi matematis, bukan sekadar kesulitan akibat redaksi atau konteks yang kurang jelas.

Penelitian lanjutan disarankan melibatkan sampel yang lebih luas dan beragam, baik dari sekolah negeri maupun swasta, agar hasil analisis memiliki daya generalisasi yang lebih kuat. Selain itu, pengembangan butir tambahan dengan tingkat kesukaran yang lebih proporsional perlu dilakukan untuk meningkatkan person reliability, KR-20, dan proporsi Raw variance explained by measures. Studi berikutnya juga disarankan melakukan pengujian ulang struktur dimensional pada sampel berbeda, melaporkan analisis differential item functioning dan person fit secara lebih rinci, serta menggunakan data respons pilihan ganda secara lengkap agar fungsi distraktor dapat dianalisis secara komprehensif.

DAFTAR PUSTAKA

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Arumsari, W. P. A., & Adirakasiwi, A. G. (2023). Jurnal Pembelajaran Matematika Inovatif Kemampuan Pemahaman Konsep Matematis Siswa dalam Menyelesaikan Soal Persamaan Garis Lurus. *Jurnal Pembelajaran Matematika Inovatif*, *6*(3), 1257–1268. https://doi.org/10.22460/jpmi.v6i3.17077

Bond, T. G., Yan, Z., & Heene, M. (2021). Applying the Rasch model : fundamental measurement in the human sciences. In *Applying the Rasch Model* (4th ed.). Routledge. https://doi.org/10.4324/9781410614575

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Dewanti, S. S., Yogyakarta, U. N., Sunan, U. I. N., & Yogyakarta, K. (2021). The Application of Item Response Theory in Analysis of Characteristics of Mathematical Literacy Test Items. *İlköğretim Online*, *20*(1), 1226–1237. https://doi.org/10.17051/ilkonline.2021.01.119

Fauziah, A., & Pandra, V. (2023). Developing PISA-like Mathematics Tasks in Musi Rawas Regency Contexts using Lesson Study. *Mathematics Education Journal*, *17*(3), 279–296. https://doi.org/10.22342/jpm.17.3.20063.279-296

Kean, J., Brodke, D. S., Biber, J., & Gross, P. (2018). *U . S . Department of Veterans Affairs the Eating Assessment Tool (EAT-10)*. *19*, 91–102. https://doi.org/10.1017/BrImp.2017.31.An

Kuswidi, I. (2015). *Brain-Based Learning untuk Meningkatkan Literasi Matematis Siswa Iwan*. *6*(2), 195–202.

Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of Test Development* (S. Lane, M. R. Raymond, & T. M. Haladyna, Eds.; Second, Vol. 2). Routledge.

NCTM. (2000). Principles and Standards for School Mathematics: Executive Summary. In *National Council of Teachers of Mathematics*. https://www.nctm.org

NCTM. (2014). *Principles to actions: Ensuring mathematical success for all*. National Council of Teachers of Mathematics.

Nopriyanti, T. D., Zulkardi, Z., Putri, R. I. I., & Aisyah, N. (2024). Soal Matematika Model Programme For International Student Assesment (PISA) Pada Pembelajaran Matematika Sekolah Menengah. *SJME (Supremum Journal of Mathematics Education)*, *8*(1), 1–12. https://doi.org/10.35706/sjme.v8i1.8650

OECD. (2023a). PISA 2022 assessment and analytical framework. In *OECD Publishing*.

OECD. (2023b). PISA 2022 Result: Factsheets Indonesia. *OECD Publishing*, 10. https://www.oecd.org/publication/pisa-2022-results/country-notes/malaysia-1dbe2061/

Putranto, S., & Ratnasari, G. I. (2022). Why I Am Confused To Apply Mathematics Concept: Student Perspective Of Mathematics Role In Life. *Aksioma: Jurnal Program Studi Pendidikan Matematika*, *11*(1), 538–549. https://doi.org/10.24127/ajpm.v11i1.4534

Rayhan, N. C., Suryadi, D., & Hasanah, A. (2025). Investigating Learning Obstacles In Mathematical Literacy: A Study On Linear Equations With One Variable. *Mathline : Jurnal Matematika Dan Pendidikan Matematika*, *10*(1), 193–207. https://doi.org/10.31943/mathline.v10i1.825

Stacey, K., & Turner, R. (2015). Assessing Mathematical Literacy. *Assessing Mathematical Literacy*. https://doi.org/10.1007/978-3-319-10121-7

Suciati, S., Munadi, S., Sugiman, S., & Febriyanti, W. D. R. (2020). Design and validation of mathematical literacy instruments for assessment for learning in Indonesia. *European Journal of Educational Research*, *9*(2), 865–875.

Usiskin, Z. (1988). Conceptions of school algebra and uses of variables. In A. F. Coxford & A. P. Shulte (Eds.), *Defining Algebraic Thinking and an Algebra Curriculum* (pp. 7–13). National Council of Teachers of Mathematics.

Utami, H. B., Salsabila, E., Wiraningsih, E. D., & Hidajat, F. A. (2025). Developing Algebra- Based Test Items for Measuring Students' Mathematical Literacy. *EduMa: Mathematics Education Learning And Teaching*, *14*(1), 330–343.

Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast*, *11*(3), 555–584.

Wright, B. D., & Stone, M. H. (1979). Best Test Design. In *Testing in the Professions*. MESA PRESS.