

Analisis Karakteristik dan Kualitas Butir Soal Informatika Menggunakan Model Rasch pada Siswa SMP

Ghulamun Jaizun*, Muhammad Basyam Walidani, Fitria Ekarini, Djuniadi

Program Studi Pendidikan Teknik Informatika, Fakultas Teknik Elektro, Universitas Negeri Semarang, Jl. Raya Banaran, Kota Semarang, Jawa Tengah 50229, Indonesia

*Corresponding Author Email: ghulamunjaizun28@students.unnes.ac.id

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas 20 butir soal Informatika pada siswa kelas VII SMP Negeri 2 Kedungwuni menggunakan Model Rasch melalui perangkat lunak JMETRIX 4.1.1. Analisis difokuskan pada pemeriksaan tingkat kesukaran butir, kecocokan item (INFIT dan OUTFIT MNSQ), reliabilitas person dan item, korelasi antara skor mentah dan estimasi kemampuan (theta), serta pemetaan person-item (Wright Map). Pendekatan kuantitatif diterapkan melalui pengolahan respons dikotomis yang diberikan oleh 33 siswa. Hasil penelitian menunjukkan bahwa sebagian besar butir berada dalam kategori fit, meskipun terdapat beberapa item dengan pola respons tidak sesuai model sehingga memerlukan revisi. Rentang tingkat kesukaran yang diperoleh cukup bervariasi, dengan beberapa butir sangat mudah dan beberapa lainnya tergolong sulit. Reliabilitas person dan item berada pada kategori lemah, menandakan konsistensi pengukuran yang perlu ditingkatkan. Korelasi antara skor mentah dan theta sangat kuat, menunjukkan keselarasan keduanya dalam menggambarkan kemampuan peserta. Wright Map memperlihatkan bahwa sebagian besar butir telah sesuai dengan distribusi kemampuan siswa, meskipun beberapa butir berada di luar rentang kemampuan mayoritas peserta. Secara keseluruhan, instrumen dapat digunakan namun memerlukan penyempurnaan pada butir yang tidak fit serta peningkatan variasi tingkat kesukaran agar pengukuran lebih akurat dan komprehensif.

Kata Kunci: evaluasi pembelajaran, jmetrix, model rasch

Abstract

This study aims to evaluate the quality of 20 multiple-choice Informatics test items administered to Grade VII students of SMP Negeri 2 Kedungwuni using the Rasch Model with JMETRIX 4.1.1. The research employs a quantitative approach to analyze item difficulty, item fit, reliability, and the alignment between students' abilities and item difficulty levels. Responses from 33 students were analyzed using dichotomous Rasch modeling. The results show that most items meet the fit criteria, with INFIT and OUTFIT MNSQ values within the acceptable range of 0.5–1.5. Two items require revision due to indications of underfit and overfit, while difficulty estimates range from -1.99 to 2.05 logits, reflecting adequate variation across items. Item reliability (0.6454) and person reliability (0.3104) fall into the low

Received : 12 Januari 2026

Revised : 19 Mei 2026

Online : 18 Juni 2026

Edisi : 30 Juni 2026



category, suggesting that the instrument has limited stability in distinguishing item difficulty and student ability. The person-item map reveals a moderate alignment between the distribution of student abilities and item difficulty, although several items are positioned far above or below the students' ability range. The correlation between raw scores and theta ability estimates is very strong ($r = 0.9428$), indicating that raw scores adequately represent student performance, although theta provides a more accurate measurement. Overall, the instrument is feasible for use but requires refinement in terms of misfit items and enhancement in terms of difficulty distribution. These results highlight the importance of Rasch-based evaluation in improving the quality and fairness of Informatics assessments at the junior high school level.

Keywords: rasch model, informatics assessment, jmetrix

Pendahuluan

Instrumen tes tertulis merupakan salah satu bentuk evaluasi yang sangat penting dalam pendidikan, karena dengan instrumen yang baik, evaluasi dapat menghasilkan data objektif dan valid sebagai dasar pengambilan keputusan dalam proses pembelajaran (Berbar et al., 2022). Oleh karena itu instrumen tes perlu disusun dan diuji secara sistematis agar memiliki validitas, reliabilitas, dan karakteristik psikometrik yang memadai, sehingga mampu mengukur kemampuan peserta didik secara akurat dan adil (misalnya pada tingkat sekolah menengah) (Adeleke et al., 2025). Instrumen tes merupakan alat utama dalam evaluasi hasil belajar yang berfungsi menghasilkan data kuantitatif untuk pengambilan keputusan pembelajaran; oleh karena itu setiap butir soal harus diuji karakteristiknya (validitas isi, tingkat kesukaran, daya beda, dan fungsi distraktor) agar instrumen mampu mengukur konstruk yang dimaksud secara akurat dan adil (Rafi et al., 2023; Misra, 2026). Analisis kualitas butir secara sistematis penting diterapkan pada berbagai jenjang pendidikan karena dapat menyingkap butir-butir bermasalah yang tidak terdeteksi oleh metode klasik saja (Rotar, 2023).

Pendekatan pengukuran modern seperti Rasch model menawarkan keunggulan dalam analisis instrumen karena memungkinkan estimasi kemampuan siswa dan kesukaran butir dalam satu metrik (Denovan et al., 2026), serta menghasilkan ukuran yang bebas butir dan bebas sampel sehingga hasil kalibrasi tetap relevan di populasi berbeda. Studi tentang kualitas instrumen dengan Rasch model pada berbagai mata pelajaran menunjukkan bahwa metode ini efektif untuk menilai validitas dan reliabilitas butir soal. Model Rasch, sebagai salah satu pendekatan dalam Item Response Theory (IRT) satu parameter, memodelkan probabilitas respons benar sebagai fungsi perbedaan antara kemampuan peserta (θ) dan kesukaran butir (b). Keunggulan Rasch meliputi specific objectivity (estimasi yang relatif bebas sampel) dan kemampuan menghasilkan parameter butir serta kemampuan person pada skala logit yang sama, sehingga memudahkan pemetaan person-item (Wright Map) dan perbandingan antar-butir. Penggunaan Rasch telah terbukti efektif untuk menilai kualitas instrumen pendidikan dan memberikan informasi tambahan (INFIT/OUTFIT, point-measure correlation, item/person reliability) yang tidak diberikan oleh Teori Tes Klasik (Marji et al., 2026). Dalam praktik analisis Rasch, beberapa indikator utama yang biasa dilaporkan meliputi; Statistik kecocokan butir (INFIT dan OUTFIT MNSQ) untuk mendeteksi butir yang berperilaku menyimpang; Reliabilitas dan indeks separation untuk item dan person sebagai ukuran konsistensi dan kemampuan pemisahan strata; Korelasi antara skor mentah dan estimasi kemampuan (Theta); Peta person-item (Wright Map) untuk melihat keselarasan antara distribusi kemampuan siswa dan tingkat kesukaran butir; Kurva karakteristik butir (ICC) atau fungsi informasi butir untuk mengetahui rentang kemampuan di mana butir paling informatif.

Namun demikian, penelitian tentang analisis butir soal pada mata pelajaran seperti Informatika terutama di jenjang SMP masih sangat terbatas. Sebagian besar penelitian dengan pendekatan Rasch fokus pada mata pelajaran sains, matematika, atau vokasional. Misalnya, penelitian analisis soal Biologi di sekolah menengah atas menunjukkan bahwa Rasch model mampu mengevaluasi tingkat kesukaran soal, fit item, serta reliabilitas secara komprehensif (Prativi et al., 2024). Dengan kondisi demikian, terdapat kekosongan penelitian mengenai kualitas butir soal Informatika di tingkat SMP menggunakan Rasch model padahal mata pelajaran Informatika dan topik berpikir komputasional semakin penting di era digital sekarang. Oleh karena itu, penelitian ini mengusulkan untuk menerapkan Rasch model menggunakan perangkat lunak JMETRIX (atau program sejenis) pada soal Informatika untuk siswa kelas VII di SMP Negeri 2 Kedungwuni, dengan tujuan memperoleh informasi tentang tingkat kesukaran butir, kecocokan butir, reliabilitas instrumen, dan distribusi kemampuan siswa. Penelitian ini diharapkan dapat memperkaya kajian tentang penerapan Rasch model dalam evaluasi pembelajaran Informatika, serta mendorong penggunaan analisis psikometrik yang lebih sistematis dalam penilaian pendidikan di era digital. Dengan demikian, penelitian ini tidak hanya berkontribusi pada peningkatan kualitas evaluasi pembelajaran di sekolah, tetapi juga mendukung upaya peningkatan mutu pendidikan Informatika secara berkelanjutan.

Metode

Penelitian ini menggunakan pendekatan kuantitatif untuk mengevaluasi kualitas 20 butir soal pilihan ganda pada mata pelajaran Informatika. Pendekatan kuantitatif dipilih karena sesuai untuk menghasilkan data numerik yang diperlukan dalam proses generalisasi dan pengujian karakteristik psikometrik instrumen. Analisis data dilakukan menggunakan Model Rasch yang memungkinkan pemeriksaan parameter kesukaran butir, kecocokan butir, serta reliabilitas instrumen secara lebih komprehensif. Penggunaan pendekatan ini sejalan dengan temuan Yu (2020) yang menunjukkan bahwa Model Rasch mampu mengidentifikasi butir bermasalah dan ketidakwajaran pola respons secara lebih akurat dibanding metode klasik, sehingga efektif untuk meningkatkan kualitas instrumen evaluasi pendidikan. Proses analisis dalam penelitian ini dilakukan menggunakan perangkat lunak JMETRIX versi 4.1.1, yang menyediakan estimasi parameter Rasch serta visualisasi pendukung seperti Wright Map dan statistik INFIT/OUTFIT. Data respons siswa dikodekan dalam format dikotomis (1 = benar, 0 = salah) sebelum diolah lebih lanjut. Hasil analisis kemudian digunakan untuk menilai karakteristik setiap butir soal dan menentukan kelayakan instrumen dalam mengukur kemampuan siswa secara tepat dan adil.

Data dan Pengumpulan Data

Informasi mengenai sumber data, karakteristik partisipan, instrumen penelitian, prosedur pengumpulan data, serta teknik analisis yang digunakan dalam penelitian. Penyajian ini bertujuan untuk memberikan gambaran yang jelas mengenai bagaimana data penelitian diperoleh dan diproses, sehingga analisis kualitas butir soal Informatika dengan model Rasch menggunakan JMETRIX dapat dilakukan secara sistematis, transparan, dan dapat direplikasi. Setiap langkah pengumpulan dan pengolahan data dirancang untuk memastikan akurasi hasil analisis serta kesesuaian dengan standar metodologi dalam penelitian evaluasi pendidikan.

Demografi Partisipan

Partisipan dalam penelitian ini adalah seluruh siswa kelas VII SMP Negeri 2 Kedungwuni yang mengikuti mata pelajaran Informatika. Teknik yang digunakan adalah total sampling, sehingga seluruh siswa pada kelas tersebut diikutsertakan dalam pengumpulan data. Komposisi partisipan disajikan berdasarkan kategori gender sebagai berikut.

Tabel 1. Partisipan Penelitian

Kategori	Keterangan	Jumlah
Total Partisipan	Jumlah keseluruhan siswa yang mengikuti tes	33 Siswa/Responden
Gender	Laki-laki	14 Siswa/Responden
	Perempuan	19 Siswa/Responden

Berdasarkan Tabel 1, partisipan pada penelitian ini terdiri dari 33 siswa kelas VII SMP Negeri 2 Kedungwuni. Dari keterangan tabel tersebut bahwa terdapat 14 siswa/responden yang ber gender laki – laki dan 19 siswa/responden perempuan.

Instrumen Penelitian

Instrumen penelitian berupa tes pilihan ganda yang disusun berdasarkan kompetensi dasar mata pelajaran Informatika kelas VII. Setiap butir terdiri dari satu jawaban benar dan tiga distraktor. Instrumen ini digunakan untuk mengumpulkan data respons siswa yang selanjutnya dianalisis menggunakan model Rasch pada aplikasi JMETRIX.

Tabel 2. Spesifikasi Instrumen Penelitian

Kategori	Keterangan	Jumlah
Jenis Instrumen	Tes pilihan ganda (multiple choice)	-
Jumlah Butir Soal	Total item dalam instrumen	20 Butir Soal
Jumlah Opsi Jawaban	Banyaknya pilihan pada setiap butir	4 Opsi Pilihan
Format Respons	Dikotomi (1 = benar, 0 = salah)	-

Tabel 2 memberikan gambaran singkat mengenai karakteristik instrumen yang digunakan dalam penelitian. Instrumen berbentuk tes pilihan ganda, yang dipilih karena format ini memungkinkan analisis Rasch secara optimal melalui respons dikotomi (benar–salah). Jumlah butir soal yang disajikan harus cukup untuk mewakili cakupan materi namun tetap proporsional dengan kemampuan siswa.

Setiap butir memiliki empat opsi jawab, yang merupakan standar umum dalam tes objektif sehingga mampu mengurangi peluang menebak. Format penskoran dikotomi diperlukan untuk analisis Rasch karena memudahkan estimasi parameter kemampuan siswa (person ability) dan tingkat kesukaran butir (item difficulty) secara konsisten dalam skala logit. Jumlah butir soal akan diisi sesuai instrumen aktual yang digunakan dalam penelitian.

Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan untuk memperoleh respons siswa terhadap instrumen tes Informatika yang telah disusun. Teknik pengumpulan data menggunakan tes tertulis yang diberikan kepada seluruh siswa kelas VII SMP Negeri 2 Kedungwuni melalui metode total sampling, sehingga seluruh siswa yang mengikuti mata pelajaran Informatika menjadi partisipan penelitian.

Pelaksanaan tes dilakukan pada jam pelajaran Informatika sesuai jadwal sekolah. Setiap siswa diberikan waktu tertentu untuk menyelesaikan seluruh butir soal secara mandiri. Instrumen diberikan dalam bentuk pilihan ganda, dan siswa diminta mengisi jawaban pada halaman kuisioner yang telah disediakan. Selama proses tes berlangsung, pengawasan dilakukan oleh guru mata pelajaran dan peneliti untuk memastikan bahwa pengerjaan dilakukan secara jujur, tertib, dan sesuai prosedur.

Teknik dan Analisis Data

Dalam analisis Rasch, terdapat beberapa tahapan penting yang dilakukan untuk memastikan instrumen layak digunakan serta menghasilkan estimasi kemampuan (person ability) dan tingkat kesukaran butir (item difficulty) yang akurat. Statistik Kecocokan Butir (INFIT dan OUTFIT MNSQ) Statistik fit digunakan untuk mengidentifikasi butir soal yang menunjukkan pola respons tidak sesuai dengan prediksi model Rasch. Nilai INFIT dan OUTFIT Mean Square (MNSQ) dievaluasi terhadap rentang nilai yang direkomendasikan untuk menentukan apakah suatu butir perlu direvisi atau dipertimbangkan untuk dihapus; Reliabilitas dan Indeks Separation (Person dan Item) Reliabilitas person menunjukkan konsistensi kemampuan siswa dalam menjawab butir soal, sedangkan reliabilitas item menunjukkan stabilitas tingkat kesukaran butir. Indeks separation digunakan untuk mengetahui kemampuan model dalam memisahkan strata kemampuan siswa maupun strata kesukaran item. Nilai separation yang lebih tinggi menunjukkan daya pembeda yang lebih baik; Korelasi antara Skor Mentah dan Estimasi Kemampuan (Theta) Korelasi ini digunakan untuk memeriksa apakah skor mentah (raw score) sejalan dengan estimasi kemampuan yang dihasilkan model Rasch. Korelasi positif menunjukkan bahwa model bekerja konsisten dalam mengonversi respons menjadi ukuran logit; Peta Person–Item (Wright Map) Wright Map digunakan untuk memvisualisasikan distribusi kemampuan siswa dan distribusi kesukaran butir pada skala yang sama. Analisis ini membantu menentukan apakah tingkat kesukaran soal sudah sesuai dengan rentang kemampuan siswa serta apakah terdapat mismatch antara person dan item; Kurva Karakteristik Butir (Item Characteristic Curve / ICC) ICC digunakan untuk melihat pola probabilitas respon benar pada berbagai level kemampuan. Fungsi informasi butir menjelaskan rentang kemampuan di mana suatu butir memberikan kontribusi informasi terbesar. Analisis ini dapat menunjukkan butir yang terlalu mudah, terlalu sulit, atau butir yang kurang informatif. Proses analisis dilakukan secara berurutan, dimulai dari pemeriksaan kelengkapan dan konsistensi data, pemodelan Rasch, interpretasi setiap indikator, hingga penarikan kesimpulan mengenai kualitas instrumen. Hasil analisis ini menjadi dasar evaluasi terhadap kualitas butir soal Informatika yang diberikan kepada siswa kelas VII.

Hasil dan Analisis

Analisis Rasch dilakukan terhadap 20 butir soal pilihan ganda menggunakan perangkat lunak JMETRIX. Output analisis meliputi statistik kecocokan butir, tingkat kesukaran, reliabilitas skala, pemetaan person–item, serta kurva karakteristik butir (ICC). Hasil penelitian disajikan dalam bentuk tabel dan grafik untuk memperjelas pemaknaan hasil secara komprehensif.

Analisis Kecocokan Butir

Analisis kecocokan butir dilakukan menggunakan nilai INFIT dan OUTFIT Mean Square (MNSQ) yang menunjukkan sejauh mana pola respons peserta sesuai dengan prediksi Model Rasch. Sejalan dengan Nuridayanti et al. (2024), rentang nilai INFIT dan OUTFIT yang dianggap memenuhi kriteria kecocokan adalah 0,5–1,5, di mana nilai di atas batas mengindikasikan underfit (respons tidak konsisten dengan model), sedangkan nilai yang lebih rendah mengindikasikan overfit (respons terlalu terduga sehingga informasi butir rendah). Data yang telah diolah tersaji pada Tabel 3 berikut

Tabel 3. Statistik Kecocokan Butir (Fit Statistics)

Item	Difficulty	INFIT	OUTFIT	Interpretasi	Keputusan
in1	-1.99	1.12	0.69	Fit	Dipertahankan
in2	0.22	0.71	0.53	Fit (cenderung overfit)	Dipertahankan / cek redundansi

Item	Difficulty	INFIT	OUTFIT	Interpretasi	Keputusan
in3	-0.70	0.92	1.11	Fit	Dipertahankan
in4	-0.03	1.10	1.46	Fit (mendekati underfit ringan)	Dipertahankan
in5	0.64	1.06	1.26	Fit	Dipertahankan
in6	0.44	1.12	1.25	Fit	Dipertahankan
in7	-0.03	1.31	1.35	Underfit ringan	Revisi
in8	-0.33	0.88	0.80	Fit	Dipertahankan
in9	-0.33	0.80	0.58	Fit (cenderung overfit)	Dipertahankan / cek kemudahan
in10	-1.19	0.77	0.64	Fit	Dipertahankan
in11	1.16	1.08	1.09	Fit	Dipertahankan
in12	-1.19	0.77	0.64	Fit	Dipertahankan
in13	-0.03	0.91	0.98	Fit	Dipertahankan
in14	0.83	0.91	0.87	Fit	Dipertahankan
in15	-1.99	0.61	0.11	Overfit kuat	Dipertahankan / cek perbedaan fungsi
in16	0.22	1.20	1.02	Fit	Dipertahankan
in17	0.44	1.16	1.15	Fit	Dipertahankan
in18	0.64	0.96	0.86	Fit	Dipertahankan
in19	1.16	1.08	1.19	Fit	Dipertahankan
in20	2.05	1.01	0.98	Fit	Dipertahankan

Hasil analisis pada tabel menunjukkan bahwa mayoritas butir berada dalam kategori fit, karena nilai INFIT dan OUTFIT masih berada dalam rentang 0,5–1,5. Hal ini mengindikasikan bahwa sebagian besar item bekerja sesuai dengan prediksi Model Rasch dan mampu mengukur konstruk secara konsisten. Beberapa item menunjukkan kondisi khusus. Item 7 memiliki nilai INFIT dan OUTFIT yang mendekati batas atas (1.31 dan 1.35), sehingga dikategorikan sebagai underfit ringan. Pola respons pada item ini tidak sepenuhnya konsisten dengan model, sehingga perlu direvisi, khususnya pada redaksi atau fungsi distraktor. Item 16 juga menunjukkan indikasi kecenderungan overfit ringan dengan nilai INFIT 1.20, meskipun OUTFIT masih berada dalam batas toleransi.

Item ini tetap dapat dipertahankan namun perlu dilakukan pemeriksaan ulang agar tidak terlalu mudah ditebak. Sementara itu, item 15 teridentifikasi sebagai overfit kuat (INFIT 0.61; OUTFIT 0.11), yang berarti respons peserta terlalu mudah diprediksi, kemungkinan akibat opsi jawaban yang tidak berfungsi seimbang atau butir yang terlalu sederhana.

Pada parameter tingkat kesukaran (difficulty), nilai negatif menunjukkan butir lebih mudah sedangkan nilai positif menunjukkan butir lebih sulit. Berdasarkan hasil analisis, item termudah adalah item 1 dan item 15 (difficulty -1.99), sedangkan item tersulit adalah item 20 (difficulty 2.05). Penyebaran nilai kesukaran yang lebar ini menunjukkan bahwa variasi tingkat kesulitan soal cukup memadai, meskipun butir yang sangat mudah dapat memunculkan pola overfit karena sebagian siswa menjawab benar tanpa melibatkan proses kognitif yang diharapkan. Sebaliknya, butir yang terlalu sulit berpotensi menghasilkan respons tidak terduga sehingga memunculkan underfit.

Temuan misfit pada beberapa butir sejalan dengan penjelasan Schmucker dan Moore (2026) bahwa ketidaksesuaian butir dalam analisis Rasch umumnya disebabkan oleh distraktor yang tidak berfungsi, redaksi soal yang ambigu, atau perbedaan pemahaman konsep antar peserta. Oleh karena itu, revisi terhadap item 7 dan item 15 menjadi langkah penting untuk meningkatkan kualitas instrumen, sedangkan butir lainnya dapat dipertahankan dengan penyesuaian minor sesuai kebutuhan.

Reliabilitas Rasch

Analisis reliabilitas dilakukan untuk menilai konsistensi pengukuran baik pada level item maupun responden (person). Reliabilitas Rasch mencakup dua indikator utama, yaitu reliability dan separation index, yang menunjukkan sejauh mana instrumen mampu membedakan tingkat kesukaran butir maupun kemampuan peserta. Sejalan dengan Avinç dan Doğan (2024), reliabilitas yang rendah menunjukkan bahwa instrumen belum stabil dalam membedakan kemampuan peserta maupun kesukaran butir, sehingga diperlukan peningkatan kualitas instrumen untuk menghasilkan estimasi yang lebih konsisten. Selain itu, dalam kerangka Model Rasch, nilai reliabilitas juga sangat dipengaruhi oleh jumlah butir dan jumlah responden yang terlibat dalam pengukuran, di mana instrumen dengan item yang terbatas dan sampel yang kecil cenderung menghasilkan reliabilitas yang rendah (O'Neill et al., 2020). Tabel berikut menyajikan statistik reliabilitas hasil analisis menggunakan JMETRIX.

Tabel 4. Statistik Reliabilitas Rasch

Statistik	Items	Persons
Observed Variance	10.165	0.6416
Observed Std. Dev.	10.082	0.8010
Mean Square Error	0.3605	0.4425
Root MSE	0.6004	0.6652
Adjusted Variance	0.6560	0.1991
Adjusted Std. Dev.	0.8099	0.4462
Separation Index	13.490	0.6708
Number of Strata	21.319	12.278

Reliability	0.6454	0.3104
-------------	--------	--------

Hasil analisis reliabilitas menunjukkan bahwa reliabilitas butir sebesar 0.6454, yang menurut kategori reliabilitas Rasch termasuk kategori lemah, karena berada di bawah batas minimal 0.67. Reliabilitas yang rendah menunjukkan bahwa konsistensi butir dalam membedakan tingkat kesukaran belum optimal, sehingga estimasi parameter butir masih mudah berubah apabila menggunakan sampel yang berbeda. Separation index butir sebesar 1.3490, yang ekuivalen dengan sekitar dua strata kesukaran, menunjukkan bahwa butir-butir dalam tes ini hanya mampu memetakan dua tingkat kesukaran yang berbeda secara jelas. Temuan ini sejalan dengan Chinnarasri et al. (2021), yang menyatakan bahwa separation index rendah mengindikasikan kualitas instrumen yang belum memadai dalam memisahkan kemampuan peserta atau tingkat kesulitan butir.

Pada sisi responden, reliabilitas sebesar 0.3104 juga tergolong lemah, menunjukkan bahwa kemampuan peserta belum terpetakan dengan stabil dalam kerangka Model Rasch. Nilai separation index responden sebesar 0.6708, yang hanya menghasilkan sekitar 1,22 strata kemampuan, mengindikasikan bahwa peserta tes cenderung memiliki kemampuan yang homogen atau jumlah peserta yang terlibat belum cukup untuk menghasilkan pemisahan kemampuan yang jelas.

Secara keseluruhan, nilai reliabilitas butir maupun responden menunjukkan bahwa instrumen ini masih memiliki keterbatasan dalam kestabilan pengukuran. Peningkatan reliabilitas dapat dilakukan dengan cara menambah jumlah butir, merevisi butir yang tidak fit, memperbaiki kualitas distraktor, atau melibatkan jumlah responden yang lebih besar dan lebih heterogen. Upaya tersebut penting agar instrumen mampu memberikan estimasi kemampuan yang lebih konsisten dan akurat dalam analisis selanjutnya.

Korelasi antara Skor Mentah dan Estimasi Kemampuan (Theta)

Tabel berikut menyajikan korelasi antara skor mentah (sum score) dan estimasi kemampuan peserta (theta) pada Model Rasch. Korelasi yang sangat tinggi menunjukkan bahwa kedua ukuran tersebut selaras dalam menggambarkan kemampuan peserta.

Tabel 5. Korelasi antara Skor Mentah dan Estimasi Kemampuan Rasch

Variabel	Sum	Theta
Sum	1.0000	0.9428
Theta	0.9428	1.0000

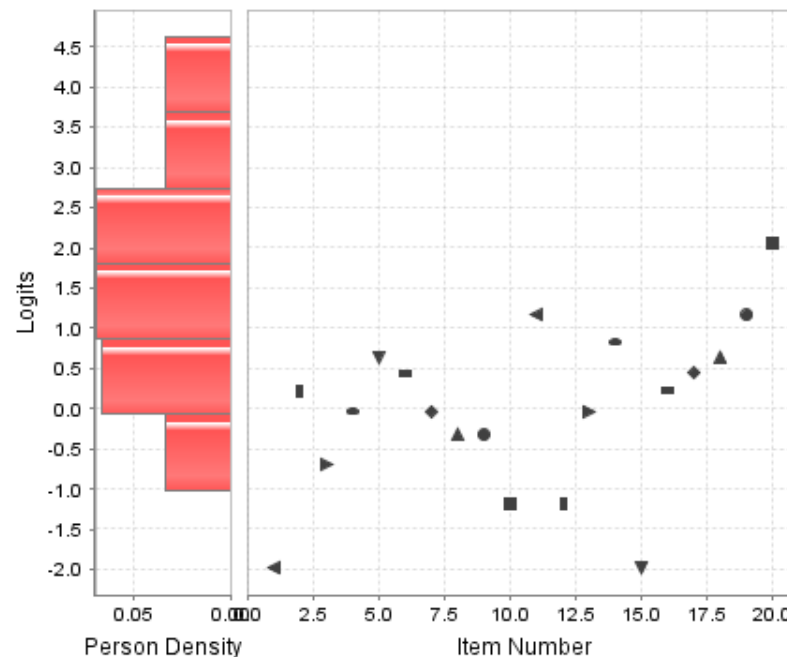
Nilai korelasi sebesar 0.9428 menunjukkan hubungan yang sangat kuat antara skor mentah dan estimasi kemampuan (theta). Hal ini selaras dengan prinsip dasar Model Rasch bahwa kedua indikator tersebut mengukur konstruk kemampuan yang sama, namun estimasi theta memberikan hasil yang lebih akurat karena memperhitungkan tingkat kesukaran setiap butir. Dengan demikian, meskipun skor mentah tetap memiliki nilai informatif, nilai theta memberikan estimasi kemampuan yang lebih stabil dan bebas bias.

Korelasi yang tinggi ini juga menunjukkan bahwa instrumen mampu memetakan kemampuan peserta secara konsisten. Skor mentah cenderung merepresentasikan jumlah jawaban benar tanpa mempertimbangkan perbedaan kesulitan butir, sementara theta menggambarkan kemampuan sebenarnya dengan menyesuaikan bobot kesukaran tiap item. Oleh karena itu, temuan ini memperkuat validitas internal instrumen dan menunjukkan bahwa proses pengukuran telah bekerja secara tepat dalam kerangka Model Rasch.

Peta Person-Item (Wright Map)

Peta person-item (Wright Map) digunakan untuk menggambarkan hubungan antara distribusi kemampuan peserta dengan tingkat kesukaran butir. Grafik Wright Map yang dihasilkan dari analisis Rasch

ditampilkan pada Gambar 1.



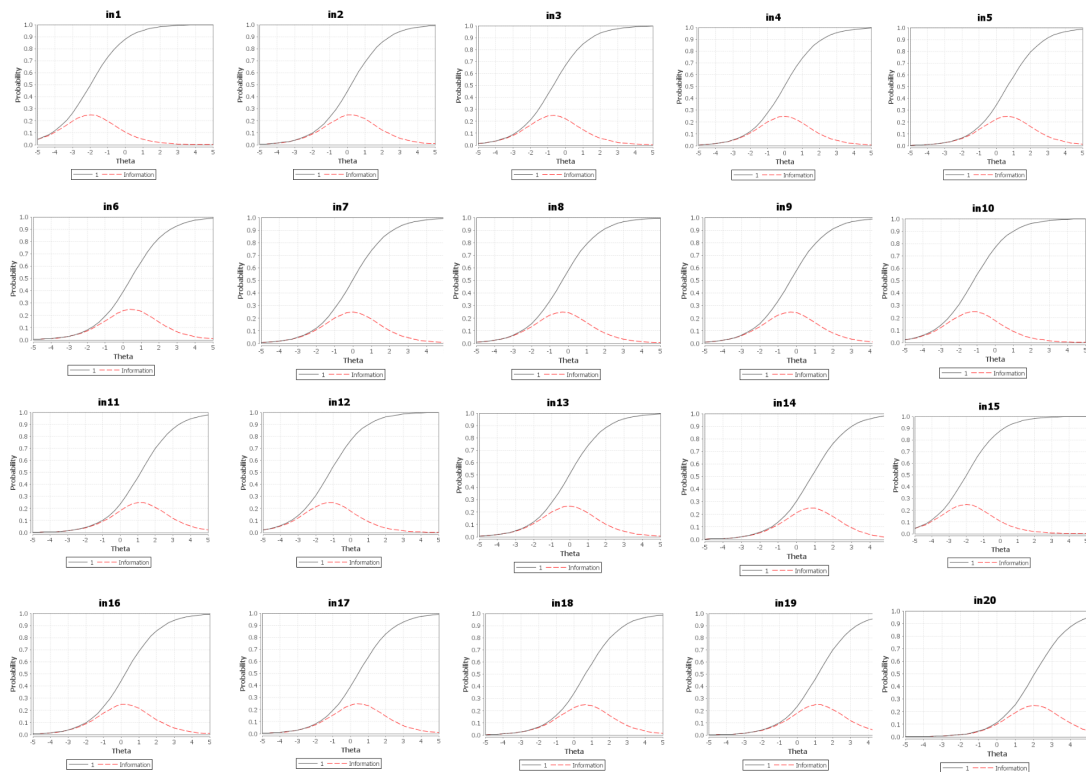
Gambar 1. Peta Person-Item (Wright Map)

Gambar 1 menunjukkan bahwa kemampuan siswa berkisar pada rentang sekitar -2.0 hingga $+4.5$ logit, dengan konsentrasi terbesar pada rentang 0 hingga 2 logit. Distribusi ini menunjukkan bahwa sebagian besar peserta berada pada kemampuan menengah ke atas. Sementara itu, sebaran kesukaran butir berada pada rentang 2.0 hingga $+2.0$ logit, yang berarti cakupan kesukaran item relatif cukup untuk menjangkau sebagian besar peserta. Secara visual, terdapat beberapa item yang berada jauh di atas distribusi kemampuan siswa, terutama Item 20, yang memiliki tingkat kesukaran tertinggi (2.05 logit). Hal ini mengindikasikan bahwa butir tersebut lebih sulit dibanding kemampuan mayoritas siswa. Sebaliknya, beberapa item berada jauh di bawah distribusi kemampuan siswa, seperti Item 1 dan Item 15 (-1.99 logit), menunjukkan bahwa item tersebut sangat mudah bagi sebagian besar peserta.

Kesejajaran antara distribusi kemampuan dan kesukaran item pada peta menunjukkan bahwa sebagian besar butir telah selaras dengan kemampuan peserta. Namun, adanya perbedaan ekstrem pada beberapa butir mengindikasikan perlunya revisi agar pengukuran lebih optimal. Dalam pengukuran berbasis Rasch, kesesuaian antara kemampuan peserta dan tingkat kesukaran butir sangat menentukan ketepatan hasil ukur. Instrumen yang memiliki butir terlalu mudah atau terlalu sulit dibandingkan kemampuan siswa akan menghasilkan informasi yang kurang optimal dan menurunkan daya pembeda instrumen (Courtney et al., 2021).

Kurva Karakteristik Butir (ICC)

Untuk memberikan pemahaman yang komprehensif mengenai performa masing-masing butir dalam instrumen, Gambar 2 menampilkan representasi grafis Kurva Karakteristik Butir (Item Characteristic Curve/ICC) untuk seluruh item dari $in1$ hingga $in20$. Visualisasi ini memungkinkan analisis mengenai bagaimana probabilitas peserta menjawab benar berubah pada berbagai tingkat kemampuan (θ), sehingga pola kerja setiap butir dapat diamati secara lebih mendalam.



Gambar 2. Kurva Karakteristik Butir (ICC) Item in1–in20

Kurva ICC yang ditampilkan menunjukkan pola kenaikan yang halus tanpa adanya fluktuasi atau ketidakstabilan pada setiap item. Dengan demikian, secara visual tidak ditemukan indikasi ketidaksesuaian perilaku butir terhadap model. Pola yang stabil ini memperkuat asumsi unidimensionalitas instrumen, karena semua item menunjukkan respons yang searah terhadap kemampuan.

Berdasarkan fungsi informasi butir, terlihat bahwa sebagian besar item memberikan informasi terbesar pada rentang kemampuan sedang, yaitu sekitar $\theta = 0$ hingga $\theta = 1$ logit. Hal ini menunjukkan bahwa instrumen paling efektif dalam membedakan peserta dengan kemampuan menengah, namun kurang sensitif untuk peserta dengan kemampuan sangat rendah maupun sangat tinggi. Dalam teori pengukuran berbasis Rasch, puncak fungsi informasi menunjukkan tingkat kemampuan di mana suatu butir memberikan kontribusi pengukuran paling besar. Jika sebagian besar butir memiliki puncak informasi pada rentang kemampuan yang sama, maka instrumen cenderung kurang optimal dalam mengukur kemampuan ekstrem (Le et al., 2022).

Kecenderungan berkumpulnya puncak informasi pada kemampuan menengah juga mengonfirmasi bahwa rentang tingkat kesulitan item relatif homogen. Keseragaman ini menguatkan hasil analisis kecocokan butir (INFIT dan OUTFIT) yang sebelumnya menunjukkan bahwa keseluruhan butir berada dalam batas toleransi model. Dengan demikian, ICC memberikan bukti tambahan bahwa instrumen bekerja secara stabil, meskipun variasi kesulitan antar-butir belum terlalu luas.

Dari perspektif pengukuran, hasil ICC menegaskan bahwa instrumen ini paling efektif untuk mengukur peserta dengan kemampuan sedang, karena pada rentang tersebut informasi yang dihasilkan butir berada pada titik optimal. Meskipun demikian, distribusi kesulitan yang terlalu terpusat ini menimbulkan keterbatasan, terutama bila instrumen ditujukan untuk populasi dengan kemampuan yang sangat beragam. Untuk meningkatkan daya diskriminasi instrumen pada kemampuan ekstrem, dapat dipertimbangkan untuk menambah (1) butir dengan tingkat kesulitan lebih rendah, agar memberikan informasi yang memadai untuk peserta berkemampuan rendah ($\theta \leq -1$), dan (2) butir dengan tingkat kesulitan lebih tinggi, agar cakupan informasi mencakup peserta berkemampuan tinggi ($\theta \geq +1$).

Dengan demikian, analisis ICC tidak hanya memberikan gambaran mengenai hubungan probabilitas respons dan

kemampuan, tetapi juga berfungsi sebagai dasar diagnostik untuk penyempurnaan instrumen, terutama dalam konteks pengembangan rentang kesulitan yang lebih bervariasi.

Kesimpulan

Dengan Penelitian ini mengevaluasi kualitas 20 butir soal Informatika kelas VII SMP Negeri 2 Kedungwuni menggunakan Model Rasch melalui perangkat lunak JMETRIX 4.1.1. Hasil analisis menunjukkan bahwa sebagian besar butir berada dalam kategori fit dengan nilai INFIT dan OUTFIT yang sesuai rentang ideal 0,5–1,5. Dua butir yaitu item 7 (INFIT 1.31; OUTFIT 1.35) dan item 15 (INFIT 0.61; OUTFIT 0.11) menunjukkan indikasi ketidakcocokan, masing-masing berupa underfit ringan dan overfit kuat sehingga memerlukan revisi. Tingkat kesukaran soal berada pada rentang -1.99 hingga 2.05 logit, dengan mayoritas item berada pada kesukaran rendah hingga sedang. Korelasi antara skor mentah dan estimasi kemampuan theta sebesar 0.9428 , menunjukkan hubungan yang sangat kuat dan konsisten antara kedua indikator kemampuan.

Nilai reliabilitas butir sebesar 0.6454 dan reliabilitas responden sebesar 0.3104 termasuk kategori lemah, menandakan bahwa konsistensi pengukuran masih belum optimal dan instrumen belum mampu memetakan kemampuan siswa secara stabil. Peta Wright menunjukkan bahwa distribusi kemampuan peserta (sekitar -2.0 hingga 4.5 logit) cenderung lebih tinggi dibandingkan rentang kesukaran item (-2.0 hingga 2.0 logit), sehingga terdapat beberapa item yang terlalu mudah dan beberapa item yang lebih sulit dibandingkan kemampuan mayoritas peserta. Secara keseluruhan, instrumen dapat digunakan sebagai alat evaluasi, namun masih memerlukan penyempurnaan terutama pada butir yang tidak fit serta peningkatan variasi tingkat kesukaran.

Berdasarkan temuan tersebut, disarankan agar guru dan penyusun soal melakukan revisi terhadap butir yang teridentifikasi tidak fit, khususnya item 7 dan item 15, dengan memperbaiki redaksi maupun fungsi distraktor agar pola respons siswa lebih sesuai dengan prediksi model. Selain itu, variasi tingkat kesukaran perlu diperluas dengan menambah butir mudah dan sulit sehingga instrumen mampu mengukur kemampuan siswa secara lebih komprehensif. Meningkatkan kualitas distraktor juga penting untuk mengurangi kemungkinan overfit dan memastikan setiap opsi berfungsi optimal.

Penelitian selanjutnya sebaiknya melibatkan jumlah responden yang lebih besar dan lebih heterogen agar estimasi parameter Rasch lebih stabil. Penambahan jumlah butir juga direkomendasikan untuk meningkatkan reliabilitas instrumen dan kedalaman cakupan materi. Selain itu, pengembangan bank soal berbasis Model Rasch dapat menjadi solusi jangka panjang untuk menyediakan instrumen evaluasi yang berkualitas dan berkelanjutan pada pembelajaran Informatika.

Statements dan Declaration

Konflik Kepentingan (Competing Interests). Para penulis menyatakan bahwa tidak terdapat konflik kepentingan dalam penelitian dan publikasi artikel ini.

Kontribusi Penulis (Author Contributions). GJ berkontribusi dalam konseptualisasi penelitian, kajian literatur, perancangan metodologi, pengumpulan data, analisis data menggunakan Model Rasch, dan penyusunan manuskrip awal. MBW berkontribusi dalam pengumpulan dan pengolahan data, analisis karakteristik butir soal, serta penelaahan manuskrip. FE berkontribusi dalam pengembangan instrumen, interpretasi hasil analisis, validasi hasil penelitian, dan penelaahan manuskrip. D berkontribusi dalam pengembangan metodologi, supervisi penelitian, interpretasi hasil, serta peninjauan dan penyempurnaan manuskrip. Seluruh penulis telah membaca dan menyetujui versi akhir manuskrip.

Persetujuan Etik (Ethics Approval). Penelitian ini melibatkan data hasil pengerjaan soal oleh siswa dan dilaksanakan dengan memperhatikan prinsip-prinsip etika dalam penelitian pendidikan. Data yang digunakan dalam penelitian dianalisis untuk kepentingan akademik dan tidak mencantumkan informasi pribadi yang dapat mengidentifikasi peserta secara langsung.

Persetujuan untuk Berpartisipasi (Informed Consent to Participate). Partisipasi dalam penelitian ini dilakukan secara sukarela dan data peserta digunakan untuk kepentingan penelitian. Persetujuan diperoleh sebelum proses pengumpulan data dilakukan, sesuai dengan prosedur yang berlaku dalam pelaksanaan penelitian.

Persetujuan untuk Publikasi (Consent for Publication). Tidak berlaku. Penelitian ini tidak memuat informasi

pribadi atau data yang dapat mengidentifikasi individu.

Ketersediaan Data (Data Availability). Data dan hasil simulasi yang mendukung temuan penelitian ini tersedia dari penulis korespondensi berdasarkan permintaan yang wajar.

Penggunaan Alat Kecerdasan Buatan (Use of Artificial Intelligence Tools). Tidak ada alat OpenAI atau alat kecerdasan buatan generatif lainnya yang digunakan dalam pelaksanaan penelitian, pengembangan atau pelaksanaan simulasi, maupun analisis data. Jika alat pemeriksaan bahasa digunakan selama penyusunan manuskrip, penggunaannya terbatas pada pemeriksaan tata bahasa dan bahasa. Seluruh saran perbaikan bahasa ditinjau oleh para penulis, yang bertanggung jawab penuh atas isi akhir manuskrip.

Daftar Pustaka

- Adeleke, J. O., Balogun, H. A., & Ayanwale, M. A. (2025). Assessment of content and cognitive dimensions of learners' mathematics performance. *STEM Education*, 5(3), 383–400. <https://doi.org/10.3934/steme.2025019>
- Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies*, 29(17), 22895–22941. <https://doi.org/10.1007/s10639-024-12662-7>
- Berbar, H., Lotfi, S., & Talbi, M. (2022). Approaches and tools for quality measurement scales validation in education: An initial systematic literature review collection. *International Journal of Emerging Technologies in Learning (iJET)*, 17(16), 73–91. <https://doi.org/10.3991/ijet.v17i16.31685>
- Chinnarasri, P., Wongpakaran, N., & Wongpakaran, T. (2021). Developing and validating the narcissistic personality scale (NPS) among older Thai adults using Rasch analysis. *Healthcare*, 9(12), 1717. <https://doi.org/10.3390/healthcare9121717>
- Courtney, M. G. R., Chang, K. C. T., Mei, B., Meissel, K., Rowe, L. I., & Issayeva, L. B. (2021). Autopsych: An R shiny tool for the reproducible rasch analysis, differential item functioning, equating, and examination of group effects. *PLOS ONE*, 16(10), e0257682. <https://doi.org/10.1371/journal.pone.0257682>
- Denovan, A., Dagnall, N., Peverell, M., Drinkwater, K. G., Escolà-Gascón, Á., & Carter, D. (2026). Evaluating the psychometric properties of the generic conspiracist beliefs scale: A rasch modelling approach. *Frontiers in Psychology*, 17. <https://doi.org/10.3389/fpsyg.2026.1721668>
- Le, C., Guttersrud, Ø., Sørensen, K., & Finbråten, H. S. (2022). Developing the HLS19-YP12 for measuring health literacy in young people: A latent trait analysis using Rasch modelling and confirmatory factor analysis. *BMC Health Services Research*, 22(1). <https://doi.org/10.1186/s12913-022-08831-4>
- Marji, M. S., Musta'amal, A. H., Che Derasid, N. A., & Ting Chee, L. (2026). Development and validation of visualisation ability test for technical drawing students using rasch models. *Journal Of Technical Education And Training*, 18(1). <https://doi.org/10.30880/jtet.2026.18.01.015>
- Misra, A. (2026). Item analysis of multiple-choice questions in pharmacology for evaluation of undergraduate medical students: A cross sectional study. *Indian Journal of Health Care Medical & Pharmacy Practice*, 7(1), 234–241. <https://doi.org/10.59551/ijhmp/25832069/2026.7.1.128>
- O'Neill, T. R., Gregg, J. L., & Peabody, M. R. (2020). Effect of sample size on common item equating using the dichotomous rasch model. *Applied Measurement in Education*, 33(1), 10–23. <https://doi.org/10.1080/08957347.2019.1674309>
- Prativi, R., Ramli, M., & Rianto, Y. (2024). Frayer model-based assessment for detecting students' conceptual understanding in evolution. *International Journal of Evaluation and Research in Education (IJERE)*, 13(6), 3686–3686. <https://doi.org/10.11591/ijere.v13i6.29423>
- Rafi, I., Retnawati, H., Apino, E., Hadiana, D., Lydiati, I., & Rosyada, M. N. (2023). What might be frequently overlooked is actually still beneficial: Learning from post national-standardized school examination. *Pedagogical Research*, 8(1), em0145. <https://doi.org/10.29333/pr/12657>
- Rotar, O. (2023). Partnership models in online learning design and the barriers for successful collaboration. *Research and Practice in Technology Enhanced Learning*, 19, 010. <https://doi.org/10.58459/rptel.2024.19010>
- Schmucker, R., & Moore, S. (2026). The impact of item-writing flaws on difficulty and discrimination in item response theory. *Computers and Education: Artificial Intelligence*, 11, 100632. <https://doi.org/10.1016/j.caeai.2026.100632>
- Yu, C. H. (2020). Objective measurement: How rasch modeling can simplify and enhance your assessment. *Rasch Measurement*, 47–73. https://doi.org/10.1007/978-981-15-1800-3_4