

Implementasi Model Rasch pada Analisis Kualitas Butir Soal Jaringan Dasar pada Jenjang Pendidikan Vokasi

Arjuna Noor Aliffiansyah^{a*}, Fitria Ekarini^b, Djuniadi Djuniadi^c

^{a, b, c} Program Studi Pendidikan Teknik Informatika dan Komputer, Universitas Negeri Semarang, Semarang, Indonesia

* Corresponding author: nooraliffiansyah321@students.unnes.ac.id

Informasi Artikel

Histori Artikel

Submission: 05/12/2025

Accepted: 20/12/2025

Published: 31/12/2025

Kata Kunci

Analisis Butir Soal;
HOTS;
Model Rasch

Abstrak

Fokus utama pendidikan vokasi saat ini adalah mencetak lulusan dengan Keterampilan Berpikir Tingkat Tinggi (HOTS), namun objektivitas kualitas instrumen penilaian yang digunakan sering kali belum terverifikasi secara empiris. Studi ini difokuskan pada evaluasi karakteristik psikometrik serta mutu butir instrumen asesmen harian pada materi Jaringan Dasar Komputer untuk siswa kelas XI dengan memanfaatkan kerangka kerja *Item Response Theory* (IRT), spesifiknya Model Rasch. Mengadopsi menerapkan desain deskriptif kuantitatif, penelitian melibatkan 32 siswa kelas XI SIJA SMK Negeri 7 Semarang sebagai subjek. Instrumen yang diuji terdiri atas 20 item tes objektif tipe pilihan ganda yang dirancang dengan orientasi HOTS, dimana data tersebut kemudian diolah menggunakan perangkat lunak jMetrik untuk menghasilkan data tingkat kesukaran, Peta Wright, serta statistik kecocokan butir (*item fit*). Temuan riset memperlihatkan reliabilitas estimasi butir berada pada kategori cukup (0,72), sedangkan konsistensi responden tergolong lemah (0,61). Berdasarkan analisis Peta Wright, teridentifikasi adanya *ceiling effect* signifikan, dimana kompetensi mayoritas siswa melampaui tingkat kesulitan soal yang disajikan. Mengacu pada kriteria MNSQ, 90% butir soal dinyatakan valid (sesuai model), sementara dua butir terdeteksi *misfit*. Secara keseluruhan, instrumen ini memiliki validitas konstrukt yang memadai dan lebih tepat difungsikan sebagai tes penguasaan (*mastery test*) dibandingkan instrumen seleksi. Meski demikian, revisi diperlukan dengan menaikkan level kesulitan soal guna meningkatkan daya pembeda dan akurasi pengukuran bagi siswa berkemampuan tinggi.

Abstract

Vocational education currently focuses on producing graduates equipped with comprehensive Higher Order Thinking Skills (HOTS). Nevertheless, the empirical objectivity and psychometric quality of the assessment instruments used often remains empirically unverified. This study focuses on evaluating the psychometric characteristics and item quality of daily assessments for the Grade XI Basic Computer Network subject using the modern psychometric framework known as Item Response Theory (IRT), with a specific focus on the 1-parameter logistic model or Rasch Model.. Applying a quantitative descriptive design, the research involved 32 Grade XI SIJA students at SMK Negeri 7 Semarang as subjects. The tested instrument consisted of 20 HOTS-based multiple-choice items, processed using jMetrik software to generate difficulty levels, Wright Maps, and item fit statistics. Research findings indicate item estimation reliability in the sufficient category (0.72), while respondent consistency is classified as weak (0.61). Based on the Wright Map analysis, a significant ceiling effect was identified, where the competence of the majority of students exceeded the difficulty level of the presented items. Referring to MNSQ criteria, 90% of the items were declared valid (model-fit), while two items were detected as misfits. Overall, this instrument possesses adequate construct validity and is more appropriately functioned as a mastery test rather than a selection instrument. Nevertheless, revision is required by increasing the item difficulty level to enhance discriminating power and measurement accuracy for high-ability students.

©2025 The Author's

This is an open-access article under the [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) license.



 [10.37058/metaedukasi](https://doi.org/10.37058/metaedukasi).

Pendahuluan

Sekolah Menengah Kejuruan (SMK) sebagai representasi pendidikan vokasi memegang mandat krusial dalam mencetak tenaga kerja yang kompeten serta adaptif terhadap fluktuasi kebutuhan dunia kerja. Saat ini, sektor industri menempatkan penguasaan *Higher Order Thinking Skills* (HOTS) sebagai prioritas kompetensi yang melampaui kemampuan hafalan dasar (LOTS). Mengacu pada studi [Yassir et al. \(2022\)](#), penguasaan HOTS kini menjadi kualifikasi fundamental bagi tenaga kerja modern, utamanya untuk menunjang kecakapan dalam analisis kritis, pengambilan keputusan strategis, serta penyelesaian masalah yang rumit. Kendati demikian, fakta empiris memperlihatkan bahwa penyerapan lulusan SMK di sektor industri masih belum maksimal, yang disinyalir akibat minimnya kompetensi berpikir tingkat tinggi ([Yassir et al., 2021](#)). Merespons kondisi tersebut, orientasi pembelajaran di SMK seyogianya tidak sekadar menitikberatkan pada keahlian teknis (*hard skills*), melainkan wajib menstimulasi perkembangan kognitif tingkat lanjut. Oleh karena itu, pembelajaran di SMK tidak boleh hanya berfokus pada aspek teknis semata, tetapi juga harus mengembangkan kemampuan kognitif level tinggi, seperti menganalisis, mengevaluasi, dan mencipta, agar siswa mampu menyelesaikan permasalahan nyata yang akan mereka hadapi di lingkungan kerja ([Dewanto et al., 2018](#)).

Dalam lingkup pendidikan teknologi informasi, materi Jaringan Dasar Komputer menjadi landasan kecakapan yang mutlak dimiliki peserta didik, terkhusus pada konsentrasi Sistem Informatika Jaringan dan Aplikasi (SIJA) serta Teknik Komputer dan Jaringan (TKJ). [Nurhasanah dan Kurniawan \(2023\)](#) mengungkapkan bahwa konsentrasi keahlian tersebut menjadi salah satu pilihan favorit karena memiliki relevansi yang sangat kuat dengan kebutuhan industri saat ini, di mana siswa dipersiapkan untuk menguasai perakitan sistem, administrasi server, hingga konfigurasi perangkat jaringan. Penguasaan materi ini sangat krusial mengingat lulusan diproyeksikan untuk mengisi peran strategis seperti teknisi jaringan, administrator, maupun analis sistem yang membutuhkan ketelitian tinggi ([Nurhasanah & Kurniawan, 2023](#)). Dengan demikian, evaluasi pembelajaran pada materi ini harus dirancang secara cermat untuk mengukur kedalaman pemahaman siswa secara akurat. Perspektif ini didukung oleh riset [Nugroho et al. \(2016\)](#), yang menggarisbawahi urgensi pengembangan instrumen penilaian di SMK harus memenuhi kriteria validitas dan reliabilitas agar mampu mengukur kompetensi siswa secara keseluruhan.

Meskipun siswa menunjukkan perolehan nilai rata-rata yang tinggi, evaluasi terhadap kualitas instrumen soal tetap mutlak diperlukan untuk memastikan validitas pengukuran tersebut. [Quaigrain dan Arhin \(2017\)](#) menegaskan bahwa analisis butir soal esensial untuk mengidentifikasi item yang menyesatkan atau tidak berfungsi, di mana skor tinggi belum tentu mencerminkan penguasaan konsep jika butir soal tersebut memiliki konstruksi yang buruk. Prinsip-prinsip dalam *Classical Test Theory* (CTT) atau yang juga disebut sebagai Teori Tes Klasik tetap menjadi landasan utama yang sering digunakan dalam mengevaluasi hasil belajar, tak terkecuali pada jenjang vokasi.. Studi terdahulu yang dilakukan [Saputro dan Sumbawati \(2017\)](#) serta [Mustofa \(2020\)](#) mengadopsi kerangka kerja ini guna mengestimasi parameter butir yang meliputi validitas, reliabilitas, serta daya pembeda sebagai indikator ketercapaian program pembelajaran. Namun, [Said dan Marpanaji \(2025\)](#) menyoroti kelemahan mendasar pendekatan klasik ini, yaitu ketergantungannya pada kelompok sampel (*sample-dependent*); artinya, tingkat kesukaran soal dapat berubah-ubah tergantung pada kepandaian kelompok siswa yang mengerjakan tes tersebut, sehingga skor mentah yang dihasilkan sering kali sulit untuk dibandingkan secara setara.

Item Response Theory (IRT) menawarkan perspektif analisis yang lebih maju untuk menyelesaikan kendala tersebut. Berbeda dengan metode lama, pendekatan ini menyoroti performa instrumen secara granular di tingkat butir, bukan sekadar melihat total capaian tes. [Okoye dan Nduka \(2022\)](#) menjelaskan

bahwa IRT, atau yang juga dikenal sebagai model sifat laten (*latent trait models*), mendeskripsikan interaksi probabilitas antara tingkat kompetensi peserta didik dan respon mereka terhadap masing-masing item. Nilai tambah fundamental dari model ini terletak pada prinsip invariansinya, yang memastikan bahwa penaksiran parameter butir soal (seperti tingkat kesukaran) tidak bergantung pada siapa yang mengerjakan tes, dan sebaliknya, estimasi kemampuan siswa tidak bergantung pada paket soal mana yang dikerjakan. Model ini menggunakan parameter logistik yang dapat mencakup tingkat kesukaran (1PL), daya pembeda (2PL), hingga tebakan semu (3PL) untuk memberikan gambaran yang lebih presisi mengenai kualitas instrumen.

Perbandingan antara kedua pendekatan tersebut menunjukkan bahwa IRT memiliki keunggulan signifikan dibandingkan CTT dalam menghasilkan penilaian yang adil dan akurat. [Mutiawani et al. \(2022\)](#) membuktikan bahwa penggunaan CTT sering kali memberikan bobot yang sama untuk setiap soal (biasanya 1 poin), sedangkan IRT memberikan bobot yang berbeda sesuai dengan karakteristik tingkat kesukaran soal tersebut. Implikasinya, siswa dengan jumlah jawaban benar yang sama bisa memiliki tingkat kemampuan (skor) yang berbeda dalam analisis IRT, tergantung pada bobot soal yang berhasil mereka jawab. Hal ini menyebabkan terjadinya perubahan peringkat yang signifikan antara metode klasik dan modern, di mana IRT dianggap lebih sensitif dalam membedakan kemampuan siswa yang sebenarnya. Oleh karena itu, penggunaan IRT, khususnya model yang sesuai untuk karakteristik data yang ada, menjadi solusi metodologis yang lebih tepat untuk mengevaluasi instrumen tes secara mendalam dan objektif.

Metode

Guna menelaah karakteristik psikometrik dan mutu butir soal pada instrumen kognitif, studi ini menerapkan desain deskriptif dengan pendekatan kuantitatif. Pemilihan pendekatan tersebut bertujuan untuk menyajikan potret objektif terkait kinerja instrumen tes yang berbasis pada data empiris respons peserta didik. Partisipan yang dilibatkan dalam riset ini berjumlah 32 siswa kelas XI pada konsentrasi keahlian Sistem Informatika Jaringan dan Aplikasi (SIJA) 2 di SMK Negeri 7 Semarang. Penentuan subjek diterapkan menggunakan metode *purposive sampling* yang mensyaratkan siswa telah merampungkan materi Jaringan Dasar Komputer secara utuh (teori maupun praktik), sehingga diasumsikan memiliki kompetensi yang memadai untuk menjalani tes tersebut.

Sebagai alat pengumpul data, penelitian mengaplikasikan 20 butir serangkaian tes pilihan ganda (*multiple choice*) yang pengembangannya berorientasi pada aspek *Higher Order Thinking Skills* (HOTS). Tes ini difungsikan sebagai evaluasi harian guna mengukur sejauh mana penguasaan materi Jaringan Dasar oleh siswa. Pengambilan data dilakukan melalui mekanisme tes tertulis, dengan skema penilaian yang memberikan bobot 5 poin bagi setiap jawaban tepat, sedangkan jawaban yang keliru tidak mendapatkan nilai (skor 0). Berdasarkan data mentah yang diperoleh, statistik deskriptif awal menunjukkan capaian yang relatif tinggi dengan rata-rata skor 87, median 95, dan rentang skor 45 hingga 100, yang mengindikasikan perlunya analisis mendalam untuk memastikan apakah instrumen tersebut memiliki tingkat kesulitan yang sesuai atau terlalu mudah.

Penelitian ini mengimplementasikan model logistik satu parameter (1PL) atau yang lazim disebut Model Rasch sebagai kerangka kerja *Item Response Theory* (IRT) dalam mengolah data, dengan bantuan perangkat lunak jMetrik dan mengubah nilai dari hasil tes murid-murid menjadi nilai, yaitu pemberian bobot 1 untuk respon benar, sedangkan respon salah tidak mendapatkan nilai (0). [Kean et al. \(2018\)](#) menjelaskan bahwa Model Rasch menghitung probabilitas siswa menjawab benar sebagai fungsi dari selisih yang menghubungkan level kompetensi peserta didik (θ) dengan taraf kesukaran butir (b), di mana

parameter diskriminasi dianggap konstan untuk seluruh butir soal. Penggunaan Model Rasch dipilih karena kemampuannya mentransformasi skor mentah yang bersifat ordinal menjadi ukuran linear berskala logit (*log-odds unit*), sehingga sehingga menciptakan analisis sejajar terkait posisi kompetensi peserta didik bersandingan dengan bobot kesulitan butir di atas dimensi metrik yang identik (Tesio et al., 2024). Selain itu, pendekatan IRT menawarkan estimasi standard error yang bervariasi untuk setiap tingkat kemampuan, berbeda dengan pendekatan Klasik (*Classical Test Theory*) yang mengasumsikan error pengukuran sama untuk semua individu, sehingga IRT memberikan diagnosis yang lebih akurat pada level individu (Jabrayilov et al., 2016).

Prosedur analisis dimulai dengan pengujian asumsi dasar IRT, yaitu unidimensionalitas dan independensi lokal, untuk memastikan bahwa instrumen hanya mengukur satu dominan kemampuan dan respon jawaban terhadap suatu item tidak dipengaruhi oleh item lainnya (Kılıç et al., 2023). Setelah asumsi terpenuhi, dilakukan kalibrasi butir soal untuk mendapatkan parameter tingkat kesukaran (*item difficulty*) dan statistik kecocokan butir (*item fit*). Evaluasi kesesuaian butir mengacu pada indikator *Infit* dan *Outfit Mean Square* (MNSQ) sebagai acuan untuk mendeteksi butir yang menyimpang (*misfit*) dari model. Secara komprehensif, analisis ini bertujuan mendiagnosis karakteristik ke-20 soal HOTS tersebut, sekaligus mengklasifikasikannya ke dalam kategori layak pakai, perlu perbaikan, atau wajib dieliminasi. Upaya ini mutlak dilakukan demi memastikan instrumen memiliki validitas isi serta reliabilitas yang kokoh dalam menakar kompetensi Jaringan Dasar peserta didik.

Hasil dan Pembahasan

Data empiris yang bersumber dari respon 32 peserta didik terhadap 20 item soal HOTS pada mata pelajaran Jaringan Dasar Komputer menjadi basis analisis utama. Pengolahan data tersebut dilakukan melalui penerapan kerangka kerja *Item Response Theory* (IRT) spesifiknya Model Rasch (1PL) yang diproses menggunakan aplikasi jMetrik. Pemaparan hasil analisis akan diuraikan secara sekuensial, diawali dengan statistik deskriptif, diikuti parameter butir, Peta Wright (*Wright Map*), hingga evaluasi kualitas skala.

Statistik Deskriptif

Ringkasan statistik yang menggambarkan sebaran data hasil tes siswa terlampir pada Tabel 1.

Tabel 1. Statistik Deskriptif

No.	Nama Bagian	Nilai
1.	Total Soal	20
2.	Total Siswa	32
3.	Nilai Minimum	9
4.	Nilai Maksimum	20
5.	Nilai Rata-rata (Mean)	17,56
6.	Nilai Tengah (Median)	19
7.	Standar Deviasi	2,9889
8.	Rentang Interkuartil	1,75
9.	Kemiringan (Skewness)	-1,7867
10.	Kurtosis	2,5086

Merujuk pada data Tabel 1, terlihat gambaran performa 32 responden dalam mengerjakan 20 butir soal. Secara umum, capaian siswa tergolong tinggi, ditandai dengan perolehan rerata (*mean*) di angka 17,56 serta median 19. Variasi skor terbentang dari nilai terendah 9 hingga nilai sempurna 20. Lebih lanjut, indikator *skewness* berada pada poin -1,7867, yang mengonfirmasi bentuk kurva distribusi negatif

(condong ke kiri). Fenomena statistik ini mengindikasikan bahwa materi tes cenderung mudah dikuasai oleh mayoritas siswa, sehingga terjadi penumpukan skor pada level tinggi.

Statistik Butir Soal (*Item Statistics*)

Parameter tingkat kesukaran (*difficulty*) dan kecocokan butir (*item fit*) hasil analisis Rasch dipaparkan secara rinci melalui Tabel 2.

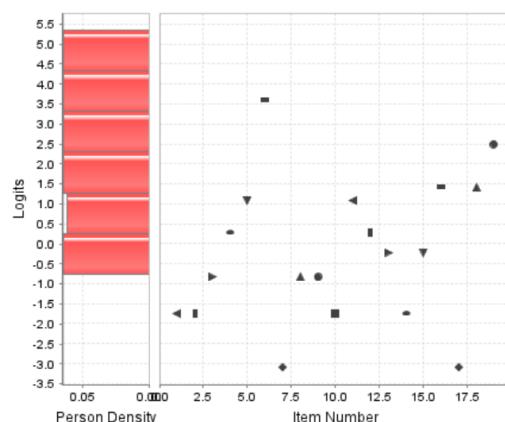
Tabel 2. Statistik Butir/Item

Butir	Kesukaran	Std. Error	WMS	Std. WMS	UMS	Std. UMS
in1	-1.75	1.10	1.13	0.42	0.33	-0.18
in2	-1.75	1.10	0.77	-0.06	0.14	-0.54
in3	-0.84	0.85	1.17	0.48	4.80	1.93
in4	0.28	0.67	1.43	1.07	1.07	0.37
in5	1.08	0.59	1.14	0.50	0.87	-0.05
in6	3.61	0.47	1.30	1.80	1.27	0.61
in7	-3.10	1.88	0.06	-0.81	0.01	-1.15
in8	-0.84	0.85	0.81	-0.21	0.23	-0.35
in9	0.84	0.85	1.03	0.24	0.47	0.00
in10	-1.75	1.10	1.13	0.42	0.33	-0.18
in11	1.08	0.59	0.82	-0.41	0.60	-0.62
in12	0.28	0.67	0.95	0.02	0.57	-0.29
in13	-0.21	0.74	0.56	-1.06	0.21	-0.69
in14	-1.75	1.10	0.77	-0.06	0.14	-0.54
in15	1.08	0.74	1.10	0.11	0.37	-0.36
in16	1.41	0.56	0.54	-1.52	0.41	-1.42
in17	-3.10	1.88	0.06	-0.81	0.01	-1.15
in18	1.41	0.56	1.83	2.06	2.10	1.89
in19	2.50	0.49	0.74	-1.09	0.63	-0.98
in20	-1.75	1.10	1.13	0.42	0.33	-0.18

Merujuk paparan data pada Tabel 2, terlihat sebaran parameter tingkat kesulitan (*difficulty*) yang bervariasi dalam skala logit. Analisis mengidentifikasi Item 6 (3,61 logit) dan Item 19 (2,50 logit) sebagai butir soal yang paling menantang atau sukar bagi siswa. Sebaliknya, Item 7 dan Item 17 menempati posisi terendah dengan nilai -3,10 logit, yang mengategorikannya sebagai soal termudah. Selanjutnya, evaluasi terhadap aspek kecocokan butir (*item fit*) mendeteksi adanya anomali pada dua soal, yakni Item 3 (UMS 4,80) serta Item 18 (UMS 2,10). Nilai *Outfit Mean Square* (UMS) yang melampaui ambang batas toleransi pada kedua butir tersebut mengindikasikan adanya ketidaksesuaian pola respon (*misfit*) terhadap model pengukuran.

Peta Wright (*Wright Map*)

Visualisasi Peta Wright (*Wright Map*) dapat dicermati pada Gambar 1. Grafik ini memetakan sebaran kapabilitas siswa (*person*) bersandingan dengan tingkat kesukaran butir (*item*) dalam dimensi pengukuran logit yang sama dan sebanding.



Gambar 1. Wright Map

Interpretasi terhadap Gambar 1 memperlihatkan ketimpangan distribusi yang mencolok. Posisi logit kemampuan siswa (lajur kiri) secara mayoritas menempati area yang lebih tinggi ketimbang sebaran tingkat kesulitan soal (lajur kanan). Fenomena ini menegaskan bahwa kompetensi rata-rata kelompok responden telah melampaui tantangan yang ditawarkan oleh instrumen tes. Tercatat hanya satu butir soal, yakni Item 6, yang memiliki tingkat kesulitan di atas rerata kemampuan siswa.

Statistik Kualitas Skala

Rekapitulasi data statistik yang merepresentasikan kualitas pengukuran, khususnya terkait aspek reliabilitas estimasi baik pada level item maupun responden, terangkum secara rinci dalam Tabel 3.

Tabel 3. Statistik Kualitas Skala

Statistik	Butir	Responden
Varians Teramati	2.3987	2.2961
Standar Deviasi Teramati	1.5488	1.5153
Rerata Kesalahan Kuadrat (MSE)	0.6681	0.8868
Akar MSE	0.8173	0.9417
Varians Disesuaikan	1.7306	1.4093
Standar Deviasi Disesuaikan	1.3155	1.1872
Indeks Pemisahan	1.6095	1.2607
Jumlah Strata	2.4794	2.0142
Reliabilitas	0.7215	0.6138

Mengacu pada hasil komputasi Tabel 3, instrumen ini mencatatkan koefisien reliabilitas butir (*Item Reliability*) di angka 0,7215, sementara konsistensi respon siswa (*Person Reliability*) berada pada level 0,6138. Selanjutnya, parameter daya pisah atau *Separation Index* menunjukkan nilai 1,6095 untuk butir soal dan 1,2607 untuk responden. Statistik ini menjadi indikator krusial dalam memvalidasi tingkat keajegan pengukuran yang dihasilkan oleh instrumen terhadap subjek penelitian.

Pembahasan

Distribusi Kemampuan Siswa dan Tingkat Kesukaran Soal

Merujuk pada paparan Tabel 1, performa akademik siswa tergolong superior. Hal ini terbukti dari capaian rerata (*mean*) yang menyentuh angka 17,56 dari total skor 20, dengan median berada di titik 19. Lebih jauh, koefisien kemiringan (*skewness*) sebesar -1,7867 mengonfirmasi terbentuknya kurva

distribusi negatif ekstrem. Fenomena statistik ini menyiratkan bahwa sebaran skor didominasi oleh siswa yang meraih nilai sempurna atau hampir sempurna. Ditinjau dari perspektif psikometrika, situasi di mana alat ukur kehilangan sensitivitasnya untuk mendiferensiasi kompetensi siswa level atas akibat tingkat kesulitan soal yang terlalu rendah didefinisikan sebagai *ceiling effect* (efek atap).

Tingginya performa atau hasil pembelajaran siswa ini dapat diraih karena materi pembelajaran yang telah mereka lalui. Siswa kelas XI SIJA 2 telah mendapatkan pembelajaran materi Jaringan Dasar Komputer yang mencakup teori dan praktikum dengan perangkat fisik. Hasil ini berkorelasi positif dengan riset [Habibullah dan Ekohariadi \(2019\)](#) yang menyebutkan bahwa pemanfaatan media pembelajaran berbasis simulasi (seperti GNS3) dan praktikum pada materi jaringan (LAN dan mikrotik) terbukti secara signifikan meningkatkan prestasi belajar siswa dibandingkan metode konvensional, karena siswa dapat memvisualisasikan konsep abstrak secara nyata. Karena subjek penelitian ini telah terpapar dengan metode pembelajaran serupa, penguasaan kompetensi teknis mereka menjadi sangat matang, sehingga butir soal yang diujikan terasa kurang menantang.

Selain itu, meskipun instrumen ini dirancang berbasis HOTS (*Higher Order Thinking Skills*), kesiapan kognitif siswa yang tinggi memungkinkan mereka menyelesaikan soal-soal analisis tersebut dengan baik. [Rosidin et al. \(2019\)](#) menjelaskan bahwa penerapan model pembelajaran yang mengintegrasikan aspek sains dan teknologi (STEM) bersama dengan asesmen berbasis HOTS secara efektif dapat meningkatkan keterampilan berpikir siswa, mulai dari menganalisis hingga mengevaluasi. Dalam penelitian ini, tingginya skor siswa bukan indikasi kegagalan instrumen dalam mengukur HOTS, melainkan bukti bahwa kompetensi berpikir tingkat tinggi siswa pada materi jaringan dasar telah terbentuk secara optimal melalui proses pembelajaran sebelumnya. Namun, dari perspektif evaluasi, instrumen ini memerlukan penyesuaian tingkat kesulitan agar distribusi skor tidak menumpuk di batas atas.

Kualitas Butir Soal dan Validitas Instrumen

Evaluasi mendalam terhadap kualitas butir soal dilakukan menggunakan statistik *Outfit Mean Square* (MNSQ) untuk memverifikasi kesesuaian data empiris dengan model pengukuran. Mengacu pada kriteria penilaian ideal antara 0,5 hingga 1,5, butir soal dalam instrumen ini terdistribusi ke dalam tiga kategori karakteristik, meliputi: *fit* (sesuai), *underfit* (tidak sesuai/acak), dan *overfit* (terlalu sesuai/prediktif).

Hasil evaluasi berdasarkan kriteria MNSQ (0,5 – 1,5) mengonfirmasi bahwa butir soal nomor 4, 5, 6, 11, 12, dan 19 berada dalam koridor yang layak dan berfungsi optimal sebagai alat ukur. Di sisi lain, analisis mengidentifikasi dua item yang melanggar batas toleransi atas (*underfit*), yakni soal nomor 3 (MNSQ 4,80) dan Item 18 (MNSQ 2,10). Lonjakan nilai yang melebihi ambang batas 1,5 ini mengindikasikan adanya inkonsistensi pola respons (*noise*) yang tidak produktif. Distorsi ini umumnya diasosiasikan dengan perilaku siswa yang menjawab secara spekulatif (*lucky guessing*) atau kurangnya attensi (*careless responding*). Oleh sebab itu, demi menjaga presisi pengukuran, kedua butir tersebut direkomendasikan untuk direkonstruksi ulang atau dieliminasi.

Sebaliknya, fenomena dominan justru terlihat pada kelompok butir yang mencatatkan nilai MNSQ di bawah 0,5 (kategori *overfit*), yang mencakup soal nomor 1, 2, 7, 8, 9, 10, 13, 14, 15, 16, 17, serta 20. Rendahnya nilai statistik ini menyiratkan bahwa pola jawaban siswa bersifat sangat deterministik atau terlalu seragam. Kondisi ini merupakan konsekuensi logis ketika tingkat kesulitan soal jauh di bawah kompetensi siswa yang telah menguasai materi, sehingga variasi jawaban menjadi minim. Walaupun secara statistik berada di luar interval ideal, implikasi butir *overfit* tidak sefatal *underfit*. Sebagaimana dijelaskan oleh [Kean et al. \(2018\)](#), item yang bersifat *overfit* (terlalu prediktif) tidak serta-merta merusak validitas konstruk instrumen, sehingga lazimnya tetap dipertahankan dalam analisis.

Berdasarkan pertimbangan tersebut, ke-12 butir soal yang tergolong *overfit* tetap diperhitungkan sebagai item yang valid dan berfungsi, sedangkan butir soal *underfit* (soal nomor 3 dan 18) dianggap cacat. Dengan demikian, secara keseluruhan instrumen ini memiliki tingkat keberfungsi yang tinggi (90%), di mana butir-butir yang valid layak untuk disimpan dalam bank soal untuk penggunaan masa depan.

Analisis Peta Wright dan Kualitas Skala

Distribusi kemampuan siswa dan tingkat kesukaran soal digambarkan melalui Peta Wright (*Wright Map*). Berdasarkan Gambar 1, terlihat jelas adanya kesenjangan (*gap*) yang signifikan antara distribusi kompetensi peserta didik (lajur kiri) dengan tingkat kesulitan butir (lajur kanan). Mayoritas siswa berada pada posisi logit yang lebih tinggi dibandingkan rata-rata tingkat kesukaran soal.

Visualisasi ini mengonfirmasi temuan pada analisis *item fit* sebelumnya. Penumpukan butir soal di area logit rendah (bagian bawah peta) menjelaskan mengapa banyak butir terdeteksi sebagai *overfit*. Karena tingkat kesulitan soal jauh di bawah kemampuan rata-rata siswa, respon siswa menjadi sangat terpola (deterministik), yang secara statistik terbaca sebagai *overfit*. Sebagaimana dijelaskan oleh [Nugrahnastiti dan Kamaludin \(2024\)](#), posisi logit siswa yang mendominasi di atas logit soal menunjukkan adanya indikasi kuat terjadinya *ceiling effect* (efek atap) pada instrumen tersebut, yang berarti instrumen kurang mampu membedakan variasi kompetensi pada siswa berkemampuan tinggi karena tantangan yang diberikan terlalu rendah.

Fenomena distribusi data tersebut berimplikasi signifikan terhadap mutu pengukuran, yang terindikasi melalui parameter reliabilitas serta indeks pemisahan (*separation index*). Sebagaimana terangkum dalam Tabel 3, hasil analisis mencatatkan koefisien *Item Reliability* di angka 0,72, sedangkan *Person Reliability* berada pada level 0,61. Mengacu pada kriteria yang dipaparkan oleh [Nugrahnastiti dan Kamaludin \(2024\)](#) serta [Rahmah et al. \(2023\)](#), nilai reliabilitas butir 0,72 tergolong dalam kategori "Cukup" (*Fair*) dan dapat diterima secara psikometrik, yang menunjukkan bahwa sampel butir soal konsisten dalam mengukur konstruk. Selain itu, tercatat nilai *Item Separation* sebesar 1,61 dan *Person Separation* sebesar 1,26. [Tesio et al. \(2024\)](#) menyatakan bahwa nilai pemisahan (*separation*) merupakan indikator kemampuan instrumen untuk mengelompokkan subjek atau butir ke dalam strata yang berbeda secara statistik, di mana nilai di bawah 2,0 menunjukkan bahwa instrumen belum mampu memisahkan siswa ke dalam minimal dua tingkat kemampuan yang berbeda (misalnya: tinggi dan rendah).

Rendahnya nilai *Person Separation* (1,26) ini menegaskan bahwa instrumen tes gagal mendekripsi gradasi kemampuan yang halus di antara para siswa. Hal ini merupakan konsekuensi logis dari homogenitas kemampuan siswa yang tinggi (semua siswa mampu menjawab soal mudah) dan kurangnya soal sulit di bagian atas peta untuk membedakan tingkatan kemampuan mereka secara lebih presisi ([Harsi, 2016](#)). Oleh karena itu, penambahan butir soal dengan tingkat kesukaran tinggi sangat diperlukan untuk meningkatkan nilai *separation* dan reliabilitas pengukuran untuk penggunaan di kemudian hari.

Kesimpulan

Berdasarkan hasil analisis, studi ini menarik simpulan bahwa instrumen tes HOTS materi Jaringan Dasar Komputer yang dikembangkan memiliki validitas konstruk yang memadai, di mana 90% butir soal terbukti berfungsi sesuai model Rasch. Namun, temuan utama menunjukkan adanya efek atap (*ceiling effect*) yang signifikan, di mana mayoritas siswa memiliki kemampuan jauh di atas tingkat kesulitan instrumen akibat keberhasilan proses pembelajaran berbasis praktikum yang sudah mereka laksanakan

sebelumnya. Kondisi ini memiliki efek pada rendahnya reliabilitas siswa (*person reliability*), yang menandakan bahwa instrumen saat ini lebih cocok berfungsi sebagai alat ukur penguasaan dasar (*mastery test*) daripada sebagai instrumen seleksi guna mendiferensiasi peserta didik yang memiliki kompetensi unggul dengan mereka yang berkemampuan dasar. Secara praktis dan metodologis, hasil ini menjelaskan pentingnya penyesuaian ulang tingkat kesukaran soal agar sesuai dengan kompetensi siswa yang telah terasah melalui praktik. Sebagai rekomendasi untuk penelitian lanjutan dan pengembangan instrumen, disarankan untuk merevisi butir soal yang terdeteksi *misfit* (nomor soal 3 dan 18) serta meningkatkan proporsi soal dengan tingkat kesukaran tinggi untuk meningkatkan daya pembeda.

Ucapan Terima Kasih

Keterberhasilan penelitian ini didapatkan berkat dukungan dan bantuan dari berbagai pihak di SMK Negeri 7 Semarang. Penulis berterima kasih kepada jajaran pimpinan sekolah atas izin dan bantuan administratifnya, khususnya pada Program Keahlian PPLG. Apresiasi khusus juga diberikan kepada guru pembimbing dari konsentrasi SIJA serta seluruh siswa kelas XI SIJA 2 atas kontribusi aktif mereka selama proses pengambilan data.

Referensi

- Dewanto, S. A., Marpanaji, E., Mahali, M. I., & Wulandari, B. (2018). *Penerapan Problem Based Learning Untuk Higher Order*. 3(May), 52–62. <https://doi.org/10.21831/elinfo.v3i1.19779>
- Habibullah, M., & Ekohariadi. (2019). Pemanfaatan Gns3 Untuk Meningkatkan Hasil Belajar Pada Mata Pelajaran Merancang Bangun Dan Menganalisa Wide Area Network Di Smk Negeri 1 Sidayu Gresik. *IT-Edu : Jurnal Information Technology and Education*, 4(2). <https://doi.org/10.26740/it-edu.v4i2.29885>
- Informatika, J. T., Teknik, F., Surabaya, U. N., Informatika, J. T., Teknik, F., & Surabaya, U. N. (2017). *Penerapan Modul Ajar Praktikum Implementasi Routing Jaringan Dengan Bangun Jaringan Muchamad Ilham Nur Saputro Meini Sondang Sumbawati*. 02, 74–84. <https://doi.org/10.26740/it-edu.v2i1.20574>
- Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). *Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment*. <https://doi.org/10.1177/0146621616664046>
- 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>
- Kılıç, A. F., Koyuncu, İlhan, & Uysal, İbrahim. (2023). Scale Development Based on Item Response Theory: A Systematic Review. *International Journal of Psychology and Educational Studies*, 10(1), 209–223. <https://doi.org/10.52380/ijpes.2023.10.1.982>
- Mustofa Yulian Said, & Eko Mapanaji. (2025). Item Analysis of End Gasal Semester Assessment Questions for Class XI and XII Computer and Network Engineering Department of SMK Muhammadiyah 1 Salam. *Journal of Information Technology and Education (JITED)*, 3(1), 1–9. <https://doi.org/10.21831/jited.v3i1.1043>
- Mustofa, Z. (2020). Evaluasi Penyelenggaraan Ujian Semester Berbasis Komputer (USBK) Menggunakan Model CIPP. *Jurnal Teknодик*, (2), 15–26. <https://doi.org/10.32550/teknodik.v0i2.519>
- Mutiawani, V., Athaya, A. M., Saputra, K., & Subianto, M. (2022). *Implementing Item Response Theory (IRT) Method in Quiz Assessment System*. 11(1), 210–218. <https://doi.org/10.18421/TEM111>

- Nugrahnastiti, N., & Kamaludin, A. (2024). Development of a Higher-Order Thinking Skills (HOTS) Test Instrument on Electrochemical Material for High School Students. *Jurnal Kependidikan : Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran*, 10(2), 451–466. <https://doi.org/10.33394/jk.v10i2.11370>
- Nugroho, B. S., Djuniadi, D., & Rusilowati, A. (2016). Pengembangan Penilaian Kinerja Menggambar Teknik Potongan di SMK pada Kurikulum 2013. *Journal of Research and Educational Research Evaluation*, 5(1), 1–7. <https://doi.org/10.15294/jrer.v5i1.14874>
- Nurhasanah, T., Kurniawan, E., Indonesia, U. P., & Indonesia, U. P. (2023). *ESP Need Analysis of Computer and Network Engineering in Vocational High School*. 10(2), 139–154. [https://doi.org/10.25299/jshmic.2023.vol10\(2\).13347](https://doi.org/10.25299/jshmic.2023.vol10(2).13347)
- Okoye, R. O., & Nduka, S. V. (2022). *Application of item Response theory in calibrating 2020 neco mathematics multiple-choice questions*. 7(1), 1–5.
- Quaigrain, K., & Arhin, A. K. (2017). a teacher-developed test in educational measurement and evaluation Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2017.1301013>
- Rahmah, I., Edi, I., & Widihastuti. (2023). *English final examination items befitting the criteria: An item Response theory approach*. *Eurasian Journal of Educational Research*, 106, 338–350. <https://doi.org/10.14689/ejer.2023.106.020>
- Rosidin, U., Suyatna, A., & Abdurrahman, A. (2019). *A Combined HOTS-Based Assessment / STEM Learning Model to Improve Secondary Students ' Thinking Skills : A Development and Evaluation Study A Combined HOTS-Based Assessment / STEM Learning Model to Improve Secondary Students ' Thinking Skills : A Development and Evaluation Study*. September.
- Said, M. Y., Drs, I., & Marpanaji, E. (2025). *Analisis Butir Soal Penilaian Akhir Semester Gasal Kelas Xi Dan Xii Jurusan Teknik Komputer Dan Jaringan Smk Muhammadiyah 1 Salam Mustofa Yulian Said 1 , Dr. Ir. Drs. Eko Marpanaji, M.T. 2. 3(1), 1–9*. <https://doi.org/10.21831/jited.v3i1.1043>
- Smk, J., & Yogyakarta, N. (2016). *Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Pemrograman Web Smk Kelas X Teknik Komputr Jaringan The Analysis Quality Of Item Final Examination Semester Test For Web Programming Subject Of X Grade Smk Students The Major Computer Network Academic Year 2015 / 2016 . Oleh : Deradi Harsi , Universitas Negeri Yogyakarta , 1, 1–10*.
- Tesio, L., Caronni, A., Kumbhare, D., Scarano, S., Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 1 . The “ most likely ” measures coming from the model. *Disability and Rehabilitation*, 46(3), 591–603. <https://doi.org/10.1080/09638288.2023.2169771>
- Yassir, M., Syam, H., & Nur, H. (2021). *Model Of Assessment For Learning Based On Higher Order Thinking Skills For Computer Network Learning For Students Of Vocational School*. <https://doi.org/10.4108/eai.27-8-2020.2305784>
- Yassir, M., Syam, H., & Nur, H. (2022). *Higher Order Thinking Skills (HOTS) based Assessment for Learning : A Model for Computer Networks Learning in Vocational School*. 10(1), 9–15. <https://doi.org/10.24203/ajas.v10i1.6876>

Kontribusi Penulis

Penulis pertama memegang peran sentral dalam merumuskan gagasan riset, menyusun desain metodologis, serta mengeksekusi pengambilan data empiris. Selain itu, proses komputasi data via jMetrik hingga penulisan manuskrip awal juga menjadi tanggung jawab penulis pertama. Sementara itu, penulis kedua dan ketiga bertindak sebagai supervisor yang memverifikasi validitas interpretasi data, serta melakukan peninjauan kritis (*critical review*) dan penyuntingan naskah demi penyempurnaan artikel. Naskah final ini telah mendapat persetujuan penuh dari seluruh tim penulis sebagai syarat kelayakan publikasi.