

**Jurnal Ilmiah Pendidikan Citra Bakti**

p-ISSN 2355-5106 || e-ISSN 2620-6641

<http://jurnalilmiahcitrabakti.ac.id/jil/index.php/jil>**PENGUJIAN KUALITAS BUTIR TES PEMECAHAN MASALAH MAHASISWA BERBASIS PEMODELAN RASCH**Sri Astuti Iriyani¹⁾, Elyakim N.S Patty²⁾, Vera Latifah Laila³⁾^{1,2}Pendidikan Teknologi Informasi, Universitas Bumigora³Manajemen Pendidikan, Universitas Negeri Jakarta¹⁾ Sri.astuti@universitasbumigora.ac.id, ²⁾ elyakim@universitasbumigora.ac.id,³⁾ Vera.latifahlaila93@gmail.com**Histori artikel***Received:*
9 Mei 2025*Accepted:*
23 Juni 2025*Published:*
5 Agustus 2025**Abstrak**

Penelitian ini bertujuan untuk menganalisis kualitas butir tes pemecahan masalah mahasiswa menggunakan pemodelan Rasch. Instrumen terdiri dari 15 butir soal pilihan ganda yang dikembangkan berdasarkan kategori Higher Order Thinking Skills untuk mengukur kemampuan berpikir tingkat tinggi secara komprehensif. Sampel penelitian berjumlah 34 mahasiswa yang dipilih melalui teknik purposive sampling. Data dianalisis secara kuantitatif deskriptif untuk mengevaluasi validitas, reliabilitas, dan kesesuaian butir soal terhadap parameter model Rasch. Hasil menunjukkan nilai validitas sebesar 0,788, menandakan bahwa tes memiliki kualitas baik dalam mengukur konstruk yang dituju. Reliabilitas instrumen sebesar 0,886 tergolong sangat tinggi, mencerminkan konsistensi dan keandalan pengukuran. Pemetaan butir soal melalui item map menunjukkan distribusi tingkat kesulitan yang merata dan representatif terhadap kemampuan mahasiswa. Seluruh butir soal memenuhi kriteria kelayakan berdasarkan nilai infit dan outfit dalam rentang yang diterima. Dengan demikian, instrumen ini layak digunakan sebagai alat ukur objektif untuk menilai kemampuan pemecahan masalah mahasiswa.

Kata-kata Kunci: Pemecahan Masalah, Kualitas Tes, Pemodelan Rasch, Keterampilan Berpikir Tingkat Tinggi

*Corresponding author: Sri Astuti Iriyani (Sri.astuti@universitasbumigora.ac.id)

Abstract. This study aims to analyze the quality of student problem solving test items using Rasch modeling. The instrument consisted of 15 multiple choice items developed based on the Higher Order Thinking Skills category to comprehensively measure higher order thinking skills. The research sample amounted to 34 students selected through purposive sampling technique. The data were analyzed quantitatively descriptively to evaluate the validity, reliability, and conformity of the items to the Rasch model parameters. The results showed a validity value of 0.788, indicating that the test has good quality in measuring the intended construct. Instrument reliability of 0.886 was classified as very high, reflecting the consistency and reliability of measurement. Item mapping through the item map shows an even distribution of difficulty levels and is representative of student abilities. All items meet the eligibility criteria based on the infit and outfit values within the accepted range. Thus, this instrument is feasible to use as an objective measurement tool to assess students' problem solving ability.

Keywords: Problem Solving, Test Quality, Rasch Modeling, Higher Order Thinking Skills

Latar Belakang

Pembelajaran abad ke-21 menuntut mahasiswa untuk memiliki keterampilan berpikir kritis dan analitis guna menghadapi tantangan yang semakin kompleks (Celik, Gedrimiene, Siklander, & Muukkonen, 2024; King & McCall, 2024; Suwistika, Ibrohim, & Susanto, 2024). Keterampilan abad ke-21 berorientasi pada konsep 4C, yang mencakup *Critical Thinking and Problem Solving, Creativity and Innovation, Communication*, serta *Collaboration* (Adeoye, Prastikawati, & Abimbowo, 2024; Haryani, Cobern, Pleasants, & Feters, 2024; Maker & Bahar, 2024). Kemampuan berpikir kritis diakui sebagai faktor utama dalam mencapai kesuksesan akademik dan profesional di era modern (Aslan & Aybek, 2024; Jia & Tu, 2024; Yefang, Sereerat, Songsiengchai, & Thongkumsuk, 2024). Seiring dengan perkembangan tren global dalam pendidikan tinggi, dosen semakin menekankan pengembangan keterampilan berpikir kritis dalam proses pembelajaran. Untuk mengukur kemampuan berpikir kritis secara efektif, diperlukan instrumen penilaian yang valid dan reliabel. Berbagai metode evaluasi, seperti tes pilihan ganda, tes keterampilan, dan tes uraian, dapat digunakan sebagai alat ukur untuk menilai tingkat berpikir kritis mahasiswa (Illene, Feranie, & Siahaan, 2023; Maryani, Prasetyo, Wilujeng, Purwanti, & Fitriawanati, 2021; Shavelson, Zlatkin-Troitschanskaia, Beck, Schmidt, & Marino, 2019).

Dosen berperan sebagai fasilitator, mentor, dan inspirator bagi mahasiswa dalam lingkungan akademik perguruan tinggi. Dosen memiliki peran strategis dalam mengevaluasi dan menilai efektivitas mata kuliah yang diajarkan. Proses evaluasi dan penilaian bertujuan untuk mengukur perkembangan akademik mahasiswa, meningkatkan kualitas pembelajaran, serta memperoleh data empiris mengenai tingkat pemahaman mahasiswa terhadap materi perkuliahan (Garay-Rondero et al., 2024; Oguguo et al., 2021). Evaluasi dilakukan melalui berbagai metode, seperti tes tertulis, diskusi akademik, ujian tengah semester, dan ujian akhir semester, memastikan tercapainya kompetensi yang diharapkan (Hursen, 2021; Kinay & Bagceci, 2016). Pengukuran kemampuan mahasiswa harus dilakukan secara sistematis

dan akurat agar dapat merefleksikan kompetensi yang sebenarnya. Proses penilaian memerlukan instrumen yang tepat dengan validitas dan reliabilitas tinggi guna memperoleh data yang akurat mengenai kemampuan individu. Namun, dalam praktiknya, masih banyak dosen yang menggunakan instrumen tes tanpa terlebih dahulu menganalisis kualitas butir soal yang digunakan. Hal ini dapat berdampak pada hasil penilaian yang kurang objektif, sehingga kemampuan mahasiswa yang sebenarnya tidak dapat teridentifikasi dengan baik.

Analisis butir soal berperan penting dalam mengevaluasi dan meningkatkan kualitas soal agar sesuai dengan tujuan pembelajaran (Hu & Valdivia, 2024; Panigrahi et al., 2024; Beerepoot, 2023). Soal yang berkualitas tidak hanya berfungsi sebagai alat evaluasi, tetapi juga sebagai instrumen diagnostik untuk mengidentifikasi pemahaman mahasiswa terhadap materi ajar (Grájeda, Burgos, Córdova, & Sanjinés, 2023; Svanes & Andersson-Bakken, 2023; Wang et al., 2024). *Item Response Theory* (IRT) merupakan salah satu pendekatan yang digunakan dalam analisis butir soal yang memungkinkan pengukuran lebih akurat dengan memodelkan probabilitas keberhasilan jawaban berdasarkan tingkat kemampuan seseorang. Seseorang dengan kemampuan tinggi memiliki peluang lebih besar untuk menjawab soal dengan benar. IRT dapat memberikan skor yang mencerminkan kemampuan seseorang serta menghubungkan karakteristik butir soal dengan tingkat kemampuan peserta tes (Gyamfi & Acquaye, 2023; Saepuzaman, Istiyono, & Haryanto, 2022; Wallmark, Ramsay, Li, & Wiberg, 2023). Salah satu model IRT yang paling sederhana adalah model Rasch dengan satu parameter. Model Rasch digunakan untuk menganalisis butir soal dengan mengukur tingkat kesulitan soal dan kemampuan peserta tes secara independen (Bakytbekovich et al., 2023; Fitrah, Sofroniou, Ofianto, Judijanto, & Widiastuti, 2024; Sayin & Bulut, 2024). Model ini memastikan bahwa peluang menjawab benar hanya bergantung pada selisih antara kemampuan dan kesulitan soal. Model Rasch sering diterapkan dalam asesmen pendidikan, seleksi, dan evaluasi instrumen psikometrik untuk memastikan tes yang adil dan konsisten.

Penelitian ini memiliki relevansi dengan beberapa penelitian terdahulu yang juga menerapkan Model Rasch dalam analisis instrumen penilaian. (Munali, Liberna, & Mutakin, 2024) penerapan Model Rasch memungkinkan evaluasi komprehensif terhadap berbagai aspek, seperti validitas responden, kemampuan mahasiswa, reliabilitas instrumen, reliabilitas mahasiswa, tingkat kesulitan soal relatif terhadap kemampuan mahasiswa, serta ketepatan jawaban mahasiswa pada program studi Sistem Informasi. Analisis kemampuan berpikir kritis mahasiswa, dengan mengevaluasi reliabilitas instrumen, validitas konstruk, serta distribusi jawaban mahasiswa dalam menilai kemampuan berpikir kritis pada topik perubahan iklim (Ridwan et al., 2023). Selanjutnya, penelitian lain dilakukan oleh (Yudha,

2023) mengevaluasi validitas dan reliabilitas instrumen penilaian *Higher Order Thinking Skills* (HOTS) dalam pembelajaran matematika.

Penelitian ini bertujuan menganalisis kualitas tes untuk mengukur kemampuan pemecahan masalah mahasiswa pada ujian akhir semester genap mata kuliah Perkembangan Peserta Didik menggunakan pendekatan Model Rasch. Pendekatan ini memungkinkan evaluasi objektif terhadap tingkat kesulitan soal, daya pembeda, dan kecocokan butir soal dengan kemampuan mahasiswa. Evaluasi kualitas tes didasarkan pada beberapa indikator, seperti kesesuaian butir soal dengan Model Rasch, tingkat kesulitan soal, dan reliabilitas instrumen. Pengembangan instrumen tes dilakukan untuk mengidentifikasi butir soal yang memenuhi kriteria model (fit) dan yang tidak sesuai (misfit). Reliabilitas butir soal diuji dengan koefisien Cronbach's Alpha untuk memastikan bahwa instrumen tes yang digunakan memiliki konsistensi yang baik.

Metode

Penelitian ini bertujuan untuk menganalisis kemampuan pemecahan masalah mahasiswa berdasarkan hasil ujian akhir mata kuliah Perkembangan Peserta Didik. Instrumen penelitian berupa 15 butir soal pilihan ganda yang mengukur level kognitif C4 (analisis), C5 (evaluasi), dan C6 (kreasi). Jawaban yang benar diberi skor 1, sedangkan jawaban salah diberi skor 0. Analisis instrumen menggunakan Model Rasch, yang memungkinkan evaluasi lebih objektif terhadap karakteristik butir soal serta kemampuan peserta tes. Teknik pengambilan sampel dilakukan dengan *purposive sampling*, dengan subjek penelitian terdiri dari 34 mahasiswa Semester genap tahun kedua di Universitas Bumigora Mataram.

Metode penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk mendeskripsikan distribusi kemampuan mahasiswa dalam menyelesaikan soal-soal berbasis pemecahan masalah. Analisis data hasil tes dilakukan menggunakan perangkat lunak RStudio untuk mengukur parameter butir soal sesuai dengan Model Rasch, yang memungkinkan evaluasi tingkat kesulitan dan daya pembeda soal secara lebih akurat. Validasi isi butir soal dianalisis menggunakan formula Aiken, sementara validasi konstruk dilakukan melalui uji prasyarat dan Teori Respons Butir (IRT). Karakteristik butir soal ditinjau berdasarkan validitas konstruk, reliabilitas tes, dan tingkat kesukaran butir soal, sehingga dapat memastikan bahwa instrumen yang digunakan memiliki kualitas pengukuran yang baik dan mampu merepresentasikan kemampuan pemecahan masalah mahasiswa secara objektif.

Hasil dan Pembahasan

Validitas dan Realibilitas Instrumen Tes

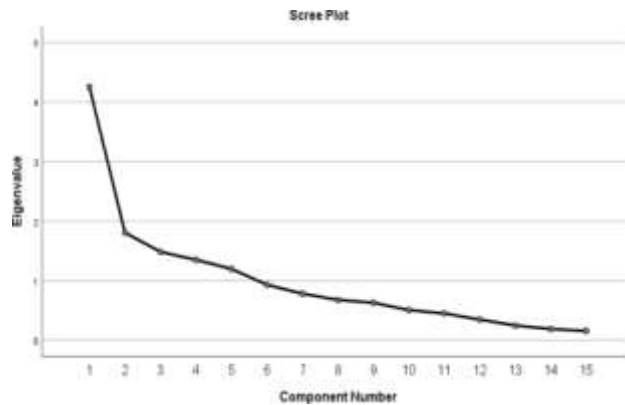
Validitas mengacu pada sejauh mana data yang dikumpulkan sesuai atau tepat dengan realitas yang sebenarnya terjadi pada objek yang diteliti. Untuk memastikan instrumen memiliki validitas tinggi, setiap butir soal harus dirancang secara cermat dan spesifik guna mengukur indikator yang telah ditetapkan, tanpa menyimpang dari tujuan pengukuran, sehingga memberikan gambaran akurat tentang kemampuan atau konstruk yang diukur. Pertanyaan juga perlu memiliki konstruk yang konsisten dan stabil, terkait langsung dengan kemampuan yang diukur, agar hasilnya andal dan bermakna. Prinsip unidimensionalitas harus dipenuhi, yaitu setiap item hanya mengukur satu dimensi sesuai tujuan pengukuran, yang dalam model Rasch dapat diidentifikasi dan diverifikasi melalui analisis nilai yang dihasilkan, memastikan setiap item mengukur satu konstruk yang dimaksud. Model Rasch juga memungkinkan evaluasi kesesuaian item dan identifikasi ketidakonsistenan, sehingga memperkuat validitas dan meningkatkan kualitas pengukuran secara keseluruhan.

Validasi isi menggunakan Aiken memenuhi kriteria valid butir soal sebesar 0,788 dengan kategori sangat baik. Pengujian validitas konstruk dilakukan untuk mengevaluasi nilai Kaiser-Meyer-Olkin (KMO) dan Bartlett's Test of Sphericity, serta menganalisis muatan faktor setiap butir soal. Nilai KMO digunakan untuk mengukur kecukupan sampel dalam analisis faktor, di mana data dengan kriteria $KMO > 0,5$ dan signifikansi Bartlett's Test of Sphericity $< 0,05$ menunjukkan bahwa data memenuhi syarat untuk dilanjutkan ke analisis faktor. Hasil analisis KMO dapat dilihat pada Tabel 1.

Tabel 1. KMO and Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,750
Bartlett's Test of Sphericity	Approx. Chi-Square	304,262
	df	105
	Sig.	0,000

Tabel 1 menunjukkan bahwa instrumen tes yang diujikan pada 34 mahasiswa memperoleh nilai Measure of Sampling Adequacy (MSA) sebesar $0,750 > 0,5$ dan signifikansi Bartlett's Test of Sphericity sebesar $0,000 < 0,05$, sehingga data memenuhi persyaratan analisis faktor dan 15 butir soal dinyatakan layak. Tes memenuhi asumsi unidimensionalitas jika terdapat satu nilai Eigenvalue dominan, yang dapat dilihat melalui Scree Plot dengan satu faktor utama yang menonjol, menunjukkan bahwa instrumen mengukur satu konstruk tunggal secara konsisten dan hasil pengukuran dapat dianggap valid serta reliabel. Nilai eigen soal perkembangan peserta didik pada Gambar 1.



Gambar 1. Scree Plot Eigen Value

Gambar 1 menunjukkan bahwa nilai eigen pada komponen 1 sebesar 4,252, komponen 2 sebesar 1,807 dan komponen 3 sebesar 1,484. Selisih nilai eigen antara komponen pertama dan kedua cukup besar, sedangkan perbedaan antara komponen kedua dan komponen berikutnya relatif kecil, maka instrumen tersebut cenderung mengukur satu faktor dominan. Asumsi unidimensionalitas terpenuhi apabila nilai eigen komponen pertama jauh lebih besar dibandingkan komponen kedua, serta nilai eigen pada komponen-komponen berikutnya relatif konstan. Tabel 2 menyajikan hasil estimasi reliabilitas instrumen Perkembangan Peserta Didik dengan nilai sebesar 0,886. Nilai tersebut menunjukkan bahwa instrumen tersebut memiliki tingkat reliabilitas yang tinggi, sesuai dengan kriteria koefisien reliabilitas $> 0,7$.

Tabel 2. Reliabilitas Instrumen Tes	
Cronbach's Alpha	N of Items
0,886	15

Kecocokan Butir Dengan Model Rasch

Butir soal yang dikembangkan pada instrumen pengukuran kemampuan mahasiswa dalam pemecahan masalah perkembangan peserta didik dikembangkan berdasarkan respon peserta tes. Nilai outfit ideal berada pada rentang 0,5 hingga 1,5, yang menunjukkan bahwa memenuhi persyaratan uji kesesuaian. Sebanyak 15 butir soal dinyatakan memenuhi uji kecocokan dengan persentase 100%, berdasarkan hasil ujian akhir semester mata kuliah Perkembangan Peserta Didik, seperti terlihat pada Tabel 2.

Tabel 2. Item Fit Model		
Item	OutFit	Description
1	0,886	Fits the model
2	0,825	Fits the model
3	0,831	Fits the model
4	0,989	Fits the model
5	0,886	Fits the model
6	0,892	Fits the model
7	1,149	Fits the model
8	1,158	Fits the model
9	1,199	Fits the model
10	0,948	Fits the model

Item	OutFit	Description
11	0,808	Fits the model
12	0,876	Fits the model
13	0,845	Fits the model
14	1,003	Fits the model
15	0,960	Fits the model

Karakteristik Butir Soal

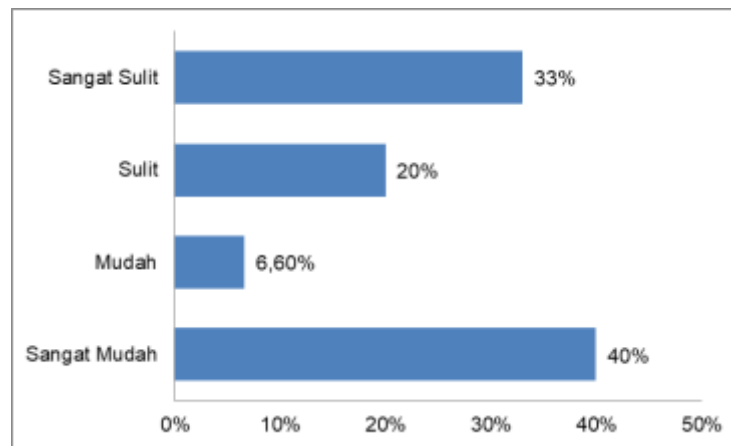
Instrumen tes yang dikembangkan dirancang untuk mengukur kemampuan mahasiswa pada level kognitif Higher Order Thinking Skills (HOTS) C4 (Analisis), C5 (Evaluasi), dan C6 (Kreasi) melalui pendekatan kontekstual pembelajaran abad 21. Berdasarkan informasi yang disajikan dalam Tabel 3, tingkat kesulitan butir soal dalam instrumen ini dikategorikan ke dalam empat tingkatan, yaitu Sangat Mudah, Mudah, Sulit, dan Sangat Sulit.

Tabel 3. Karakteristik Butir Soal

Item	Subject Matter	Cognitive Level	Level Of Difficulty	Description
1	Perkembangan dan Pertumbuhan	C4	-2,581	Sangat Mudah
2	Perkembangan dan Pertumbuhan	C5	-1,077	Sangat Mudah
3	Perkembangan dan Pertumbuhan	C6	0,996	Sangat Sulit
4	Perkembangan Fisik dan Motorik Peserta Didik	C4	-0,047	Sulit
5	Perkembangan Fisik dan Motorik Peserta Didik	C5	-1,	Sangat Mudah
6	Perkembangan Fisik dan Motorik Peserta Didik	C6	0,996	Sangat Sulit
7	Perkembangan Kecerdasan dan Kreativitas	C4	-1,406	Sangat Mudah
8	Perkembangan Kecerdasan dan Kreativitas	C5	0,996	Sangat Sulit
9	Perkembangan Kecerdasan dan Kreativitas	C6	0,860	Sulit
10	Perkembangan Peserta Didik dalam Pembelajaran Abad 21	C4	0,723	Sulit
11	Perkembangan Peserta Didik dalam Pembelajaran Abad 21	C5	0,439	Sulit
12	Perkembangan Peserta Didik dalam Pembelajaran Abad 21	C6	1,130	Sangat Sulit
13	Upaya mengatasi masalah dan kesulitan belajar Peserta didik	C4	-1,406	Sangat Mudah
14	Upaya mengatasi masalah dan kesulitan belajar Peserta didik	C5	1,264	Sangat Sulit
15	Upaya mengatasi masalah dan kesulitan belajar Peserta didik	C6	0,723	Sulit

Gambar 2 menampilkan persentase kemampuan pemecahan masalah mahasiswa berdasarkan hasil tes yang diberikan. Dari 15 butir soal yang diujikan, 40% termasuk dalam kategori sangat mudah, 6,60% dalam kategori mudah, 20% dalam kategori Sulit dan 33%

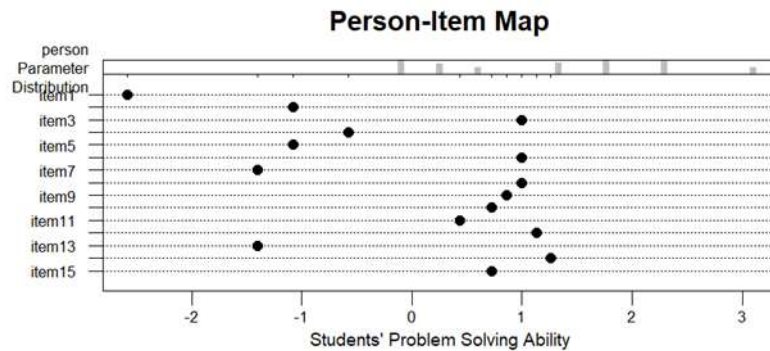
dalam Kategori Sangat Sulit. Hasil ini menunjukkan bahwa sebagian besar soal memiliki tingkat kesulitan yang tinggi, sehingga dapat memberikan tantangan bagi mahasiswa dalam mengembangkan keterampilan pemecahan masalah.



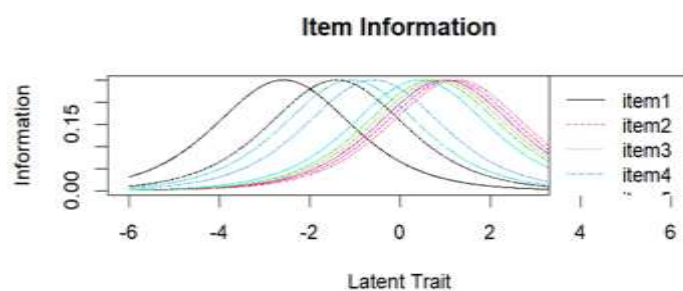
Gambar 2. Tingkat Kesukaran Soal

Tingkat Kemampuan Peserta Tes

Gambar 3 menampilkan hubungan antara tingkat kemampuan peserta tes dan tingkat kesulitan butir soal dalam suatu instrumen pengukuran, serta menggambarkan ambang batas parameter untuk setiap item. Person-Item Map menunjukkan bahwa tingkat kesulitan butir soal bervariasi, dengan beberapa butir berada pada tingkat kesulitan tinggi (terletak di sisi kiri) dan lainnya memiliki tingkat kesulitan yang lebih rendah (terletak di sisi kanan). Distribusi kemampuan siswa menunjukkan bahwa mayoritas peserta berada dalam rentang kemampuan 0 hingga 1, yang relatif sesuai dengan tingkat kesulitan sebagian besar butir soal. Temuan ini mengindikasikan bahwa instrumen tes yang digunakan cukup valid dalam mengukur kemampuan pemecahan masalah mahasiswa. Gambar 4 menampilkan Item Information Function, dengan sumbu X merepresentasikan tingkat kemampuan siswa dalam skala latent trait, dan sumbu Y menunjukkan jumlah informasi yang diberikan oleh 15 butir soal. Kurva ini menggambarkan sejauh mana suatu item membedakan individu dengan tingkat kemampuan berbeda serta menilai efektivitas butir soal dalam mengukur konstruk yang diinginkan. Selain itu, kurva ini mencerminkan plot koefisien korelasi intrakelas, yang digunakan untuk mengevaluasi konsistensi dan kualitas butir soal. Analisis Item Information Function (IIF) terhadap 15 butir soal dalam Model Rasch atau Teori Respons Butir (IRT) menunjukkan bahwa setiap item memiliki tingkat kesulitan dan efektivitas berbeda dalam mengukur kemampuan peserta tes. Beberapa butir lebih informatif bagi peserta berkemampuan rendah, sementara yang lain lebih efektif untuk peserta dengan kemampuan menengah hingga tinggi. Kombinasi seluruh butir soal mencakup spektrum kemampuan yang luas, sehingga meningkatkan akurasi dan reliabilitas pengukuran terhadap kemampuan pemecahan masalah mahasiswa dalam pembelajaran abad ke-21.



Gambar 3. Parameter Kemampuan Peserta Tes dan Kesulitan Soal



Gambar 4. Informasi item

Pembahasan

Penelitian ini bertujuan untuk menganalisis kemampuan pemecahan masalah mahasiswa dalam menghadapi pembelajaran abad ke-21 dengan mengacu pada indikator dalam mata kuliah Perkembangan Peserta Didik. Kemampuan pemecahan masalah merupakan keterampilan esensial dalam dunia pendidikan modern, terutama dalam menghadapi tantangan yang menuntut pemikiran kritis, kreativitas, komunikasi, dan kolaborasi. Untuk mengevaluasi kemampuan mahasiswa serta tingkat kesulitan soal, penelitian ini menerapkan Model Rasch sebagai pendekatan analitis. Model ini memungkinkan analisis objektif dan mendalam terhadap respons mahasiswa, sehingga memberikan gambaran yang lebih akurat mengenai distribusi kemampuan mereka serta validitas dan reliabilitas instrumen yang digunakan.

Dalam konteks pemecahan masalah, kemampuan berpikir kritis mencakup proses analisis dan evaluasi terhadap suatu permasalahan, yang memungkinkan mahasiswa untuk mengidentifikasi, mengkaji, serta menafsirkan informasi secara mendalam. Sementara itu, berpikir kreatif berperan dalam pengembangan solusi inovatif yang tidak hanya didasarkan pada pengetahuan deklaratif maupun prosedural, tetapi juga mempertimbangkan pendekatan yang lebih adaptif dan efektif. Oleh karena itu, soal HOTS dirancang tidak hanya untuk mengukur tingkat pemahaman mahasiswa terhadap materi, tetapi juga untuk menilai

kemampuan mereka dalam mengaplikasikan, menggeneralisasi, serta merekonstruksi konsep dalam berbagai konteks akademik maupun praktis.

Penyusunan butir soal perlu mempertimbangkan variasi tingkat kesulitan guna memastikan asesmen yang komprehensif dan proporsional. Soal yang terlalu mudah hanya mengukur kompetensi dasar, sementara soal yang terlalu sulit dapat menjadi hambatan dalam proses pemecahan masalah. Oleh karena itu, keseimbangan dalam penyusunan instrumen evaluasi menjadi aspek krusial agar mahasiswa dapat mengembangkan keterampilan berpikir kritis dan kreatif secara bertahap. Selain itu, penggunaan berbagai model soal, seperti studi kasus dan pemecahan masalah berbasis skenario, dapat meningkatkan efektivitas asesmen dalam menilai kemampuan berpikir tingkat tinggi secara lebih akurat dan sistematis.

Tingkat kesulitan butir soal dalam mengukur kemampuan pemecahan masalah mahasiswa berkorelasi dengan nilai kesulitan yang diperoleh. Butir soal dengan tingkat kesulitan rendah menunjukkan bahwa sebagian besar mahasiswa mampu menjawabnya dengan benar, sedangkan butir soal dengan tingkat kesulitan tinggi menandakan bahwa hanya mahasiswa dengan kompetensi lebih tinggi yang dapat menjawabnya secara tepat. Dalam Model Rasch, variasi tingkat kesulitan ini dinyatakan dalam satuan logit, di mana nilai logit yang lebih rendah merepresentasikan soal yang lebih mudah, sedangkan nilai logit yang lebih tinggi menunjukkan soal yang lebih sulit. Dengan demikian, pemetaan tingkat kesulitan soal menjadi aspek krusial dalam penyusunan evaluasi yang seimbang dan sesuai dengan kompetensi mahasiswa.

Kesimpulan

Hasil analisis menggunakan pemodelan Rasch menunjukkan bahwa kualitas tes pemecahan masalah mahasiswa dalam mata kuliah Perkembangan Peserta Didik telah memenuhi karakteristik tes yang baik. Butir soal terdistribusi secara merata pada peta butir dan mampu mengukur kemampuan mahasiswa dalam pemecahan masalah secara komprehensif. Validasi isi menunjukkan hasil yang valid dengan nilai 0,788, sementara 15 butir soal memiliki karakteristik kualitas yang sepenuhnya sesuai dengan pemodelan Rasch 100% Fit. Indeks kesukaran butir soal berada dalam kategori baik, dengan reliabilitas instrumen tes sebesar 0,886, yang termasuk dalam kategori reliabel. Selain itu, hasil uji prasyarat unidimensionalitas menggunakan Kaiser-Meyer-Olkin (KMO) sebesar 0,750 menunjukkan bahwa instrumen memenuhi persyaratan konstruk dan memiliki kualitas butir soal yang memadai.

Daftar Pustaka

- Adeoye, M. A., Prastikawati, E. F., & Abimbowo, Y. O. (2024). Empowering Learning: Pedagogical Strategies for Advancing 21st Century Skills and Quality Education. *Journal of Nonformal Education*, 10(1), 10–21. <https://doi.org/https://doi.org/10.15294/jone.v10i1.1451>
- Aslan, S., & Aybek, B. (2024). Development of Critical-Thinking Skills Rubric within the Scope of Multicultural Education. *Educational Process: International Journal*, 13(3), 139–158. <https://doi.org/10.22521/edupij.2024.133.8>
- Bakytbekovich, O. N., Mohammed, A., Alghurabi, A. M. K., Alallo, H. M. I., Ali, Y. M., Hassan, A. Y., ... Afif, A. K. N. S. (2023). Distractor Analysis in Multiple-Choice Items Using the Rasch Model. *International Journal of Language Testing*, 13, 69–78. <https://doi.org/10.22034/IJLT.2023.387942.1236>
- Beerepoot, M. T. P. (2023). Formative and Summative Automated Assessment with Multiple-Choice Question Banks [Research-article]. *Journal of Chemical Education*, 100(8), 2947–2955. <https://doi.org/10.1021/acs.jchemed.3c00120>
- Celik, I., Gedrimiene, E., Siklander, S., & Muukkonen, H. (2024). The affordances of artificial intelligence-based tools for supporting 21st-century skills: A systematic review of empirical research in higher education. *Australasian Journal of Educational Technology*, 40(3), 1–20. <https://doi.org/https://doi.org/10.14742/ajet.9069>
- Fitrah, M., Sofroniou, A., Ofianto, Judijanto, L., & Widiastuti. (2024). Reliability and separation index analysis of mathematics questions integrated with the cultural architecture framework using the Rasch model. *Journal of Education and E-Learning Research*, 11(3), 499–509. <https://doi.org/10.20448/jeelr.v11i3.5861>
- Garay-Rondero, C. L., Castillo-Paz, A., Gijón-Rivera, C., Domínguez-Ramírez, G., Rosales-Torres, C., & Oliart-Ros, A. (2024). Competency-based assessment tools for engineering higher education: a case study on complex problem-solving. *Cogent Education*, 11(1), 1–31. <https://doi.org/10.1080/2331186X.2024.2392424>
- Grájeda, A., Burgos, J., Córdova, P., & Sanjinés, A. (2023). Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2023.2287917>
- Gyamfi, A., & Acquaye, R. (2023). Parameters and Models of Item Response Theory (IRT): A Review of Literature. *Acta Educationis Generalis*, 13(3), 68–78. <https://doi.org/10.2478/atd-2023-0022>
- Haryani, E., Cobern, W. W., Pleasants, B. A.-S., & Feters, M. K. (2024). Exploring Pedagogical Strategies: Integrating 21st-Century Skills in Science Classrooms. *Journal of Education in Science, Environment and Health*, 10(2), 106–119. <https://doi.org/10.55549/jeseh.697>
- Hu, T. L., & Valdivia, D. S. (2024). Assessing the Psychometric Properties of Quality Experience in Undergraduate Research Using Item Response Theory. *Research in Higher Education*, 65(1), 1965–1991. <https://doi.org/10.1007/s11162-024-09814-6>
- Hursen, C. (2021). The Effect of Problem-Based Learning Method Supported by Web 2.0 Tools on Academic Achievement and Critical Thinking Skills in Teacher Education. *Technology, Knowledge and Learning*, 26(3), 515–533. <https://doi.org/10.1007/s10758-020-09458-2>
- Illene, S., Feranie, S., & Siahaan, P. (2023). Create multiple-choice tests based on experimental activities to assess students' 21st century skills in heat and heat transfer topic. *Journal of Education and Learning*, 17(1), 44–57.

<https://doi.org/10.11591/edulearn.v17i1.20540>

- Jia, X. H., & Tu, J. C. (2024). Towards a New Conceptual Model of AI-Enhanced Learning for College Students: The Roles of Artificial Intelligence Capabilities, General Self-Efficacy, Learning Motivation, and Critical Thinking Awareness. *Systems*, 12(3), 1–25. <https://doi.org/10.3390/systems12030074>
- Kinay, I., & Bagceci, B. (2016). The Investigation of the Effects of Authentic Assessment Approach on Prospective Teachers' Problem-Solving Skills. *International Education Studies*, 9(8), 51. <https://doi.org/10.5539/ies.v9n8p51>
- King, C. R. P., & McCall, M. (2024). How the fine arts create the finest students: A design thinking study. *Higher Education Quarterly*, 78(3), 1162–1174. <https://doi.org/10.1111/hequ.12521>
- Maker, J., & Bahar, K. (2024). Talent Development in Inclusive Classrooms: An Analysis of Student Growth! *Australasian Journal of Gifted Education*, 33(2), 28–56. <https://doi.org/10.21505/ajge.2024.0013>
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitrianawati, M. (2021). HOTs Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers. *Journal of Turkish Science Education*, 18(4), 674–690. <https://doi.org/10.36681/tused.2021.97>
- Munali, Liberna, H., & Mutakin, T. Z. (2024). ANALYSIS OF PERSON ' S PROBLEM-SOLVING SKILLS. 8(2), 467–477. <https://doi.org/http://dx.doi.org/10.31000/prima.v8i2.11304>
- Oguguo, B. C. E., Nannim, F. A., Agah, J. J., Ugwuanyi, C. S., Ene, C. U., & Nzeadibe, A. C. (2021). Effect of learning management system on Student's performance in educational measurement and evaluation. *Education and Information Technologies*, 26(2), 1471–1483. <https://doi.org/10.1007/s10639-020-10318-w>
- Panigrahi, R., Nihar, K. L., & Singh, N. (2024). Quality Measurement of the Blended Learning Model in Higher Education: Scale Development and Validation. *Higher Learning Research Communications*, 14(1), 86–102. <https://doi.org/10.18870/hlrc.v14i1.1467>
- Ridwan, I. M., Kaniawati, I., Suhandi, A., Ramalis, T. R., Mahmudah, I. R., & Vilmala, B. K. (2023). Rasch modeling: Developing a critical thinking ability test for students on climate change material. *Momentum: Physics Education Journal*, 7(2), 247–258. <https://doi.org/10.21067/mpej.v7i2.8310>
- Saepuzaman, D., Istiyono, E., & Haryanto. (2022). Characteristics of Fundamental Physics Higher-order Thinking Skills Test Using Item Response Theory Analysis. *Pegem Egitim ve Ogretim Dergisi*, 12(4), 269–279. <https://doi.org/10.47750/pegegog.12.04.28>
- Sayin, A., & Bulut, O. (2024). The difference between estimated and perceived item difficulty: An empirical study. *International Journal of Assessment Tools in Education*, 11(2), 368–387. <https://doi.org/10.21449/ijate.1376160>
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., & Marino, J. P. (2019). Assessment of University Students' Critical Thinking: Next Generation Performance Assessment. *International Journal of Testing*, 19(4), 337–362. <https://doi.org/10.1080/15305058.2018.1543309>
- Suwistika, R., Ibrohim, I., & Susanto, H. (2024). Improving critical thinking and creative thinking skills through POPBL learning in high school student. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 10(1), 115–122. <https://doi.org/https://doi.org/10.22219/jpbi.v10i1.30172>
- Svanes, I. K., & Andersson-Bakken, E. (2023). Teachers' use of open questions: investigating the various functions of open questions as a mediating tool in early literacy

- education. *Education Inquiry*, 14(2), 231–250. <https://doi.org/10.1080/20004508.2021.1985247>
- Wallmark, J., Ramsay, J. O., Li, J., & Wiberg, M. (2023). Analyzing Polytomous Test Data: A Comparison Between an Information-Based IRT Model and the Generalized Partial Credit Model. *Journal of Educational and Behavioral Statistics*, 753–779. <https://doi.org/10.3102/10769986231207879>
- Wang, J., Wang, Y., Moore, Y., Sneed, S., Thacker, B., & Hart, S. (2024). Analyzing the Patterns of Questioning Chains and Their Intervention on Student Learning in Science Teacher Preparation. *International Journal of Science and Mathematics Education*, 22(4), 809–836. <https://doi.org/10.1007/s10763-023-10408-4>
- Yefang, W., Sereerat, B., Songsiangchai, S., & Thongkumsuk, P. (2024). The Development of an Instructional Model Based on Experiential Learning Theory and Six Thinking Hats to Improve the Critical Thinking Ability of Undergraduate Students. *World Journal of Education*, 14(1), 66. <https://doi.org/10.5430/wje.v14n1p66>
- Yudha, R. P. (2023). Higher Order Thinking Skills (HOTS) Test Instrument: Validity and Reliability Analysis With The Rasch Model. *Eduma: Mathematics Education Learning and Teaching*, 12(1), 21. <https://doi.org/10.24235/eduma.v12i1.9468>