

Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik Dan Rasch Model

Sri Jumini^{1,2*}, Sutikno Madnasri³, Edy Cahyono⁴, P Parmin⁵

¹Mahasiswa Program Doktor Pendidikan IPA, Pasca Sarjana, Universitas Negeri Semarang, Sekaran, Gunung pati, Kota Semarang, 50229, Central Java, Indonesia

²Pendidikan Fisika, Fakultas Ilmu Tarbiyah dan Keguruan (FITK), Universitas Sains Al-Qur'an, Jln. Raya Kalibeber, Km. 03, Mojotengah, Wonosobo, Central Java, Indonesia

³Jurusan Fisika, FMIPA, Universitas Negeri Semarang, Semarang, Indonesia, Sekaran, Gunung pati, Kota Semarang, 50229, Central Java, Indonesia

⁴Jurusan Kimia, FMIPA, Universitas Negeri Semarang, Semarang, Indonesia, Sekaran, Gunung pati, Kota Semarang, 50229, Central Java, Indonesia

⁵IPA Terpadu, FMIPA, Universitas Negeri Semarang, Sekaran, Gunung pati, Kota Semarang, 50229, Central Java, Indonesia

*Corresponding Author: srijumini@unsiq.ac.id

Abstrak. Tujuan dari penelitian ini untuk memberikan gambaran hasil perbandingan analisis kualitas soal tes kemampuan literasi sains melalui pendekatan teori tes klasik dan Rasch Model ditinjau dari sisi *validity*, *reliability*, kesulitan soal, dan daya diskriminan. Penelitian ini merupakan penelitian deskriptif kuantitatif dan pengambilan data dilakukan pada bulan Februari 2021. Data diperoleh dengan menggunakan metode dokumenter sebanyak 15 soal pilihan ganda dan diujikan kepada 30 mahasiswa Program Studi Pendidikan Fisika UNSIQ Jawa Tengah Wonosobo. Data dianalisis menggunakan teori uji klasik dengan *Microsoft Excel* dan *Rasch Model* dengan software *Winsteps*. Kualitas soal dari segi validitas dengan menggunakan pendekatan teori tes klasik terungkap bahwa 3 soal diklasifikasikan valid dan 12 soal diklasifikasikan tidak valid, sedangkan pada Rasch Model 6 soal diklasifikasikan valid dan 9 soal diklasifikasikan tidak valid. Skor *reliability Cronbach Alpha (KR-20)* pada butir soal dengan pendekatan teori tes klasik adalah 0,40 (sedang) dan skor *reliability Cronbach Alpha (KR-20)* pada model Rasch adalah 0,43 (sedang) dengan Nilai person *reliability* 0,54 (buruk) dan keandalan produk 0,91 (sangat baik). Tingkat kesukaran butir soal menurut teori tes klasik, butir soal terdistribusi dalam tiga kelompok, yakni mudah 10 soal, sedang 3 soal, dan sukar 2 soal. Sedangkan pada Rasch Model butir soal terdistribusi dalam empat kelompok, yakni mudah 3 soal, sedang 7 soal, sukar 3 soal, dan sangat sukar 2 soal. Pada aspek daya beda baik dengan pendekatan teori tes klasik maupun Rasch Model, nilai $H = 1,57$ dibulatkan menjadi 2, yakni buruk 14 soal dan baik 1 soal.

Kata kunci: Analisis soal; Literasi sains; teori tes klasik; Rasch Model

Abstract. The purpose of this study is to provide an overview of the results of a comparative analysis of the quality of the scientific literacy ability test items through the classical test theory approach and the Rasch Model in terms of validity, reliability, item difficulty, and discriminant power. This research is a quantitative descriptive research and data collection was carried out in February 2021. Data was obtained using the documentary method of 15 multiple choice questions and tested on 30 students of the Physics Education Study Program UNSIQ Central Java Wonosobo. Data were analyzed using classical test theory with Microsoft Excel and Rasch Model with Winsteps software. The quality of the questions in terms of validity using the classical test theory approach revealed that 3 questions were classified as valid and 12 questions were classified as invalid, while in the Rasch Model 6 items were classified as valid and 9 questions were classified as invalid. The Cronbach Alpha (KR-20) reliability score on the items using the classical test theory approach is 0.40 (moderate) and the Cronbach Alpha (KR-20) reliability score on the Rasch model is 0.43 (moderate) with a person reliability score of 0,54 (poor) and product reliability 0.91 (very good). The difficulty level of the items according to classical test theory, the items are distributed in three groups, namely 10 easy questions, 3 moderate questions, and 2 difficult questions. Whereas in the Rasch Model the items were distributed in four groups, namely 3 easy questions, 7 moderate questions, 3 difficult questions, and 2 very difficult questions. On the aspect of differentiating power, both with the classical test theory approach and the Rasch Model, the value of $H = 1.57$ is rounded to 2, namely 14 bad questions and 1 good one.

Keywords: Analysis of questions; scientific literacy; classical test theory; Rasch Model

How to Cite: Jumini, S., Madnasri, S., Cahyono, E., Parmin, P. (2023). Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik Dan Rasch Model. *Prosiding Seminar Nasional Pascasarjana*, 2023, 758-765.

PENDAHULUAN

Pendidikan berkualitas tidak terlepas dari

proses pelaksanaan kegiatan belajar mengajar yang disertai dengan proses evaluasi yang baik (Alfarisa & Purnama, 2019). Evaluasi merupakan

sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagaimana tujuan pendidikan sudah tercapai (Arikunto, 2012). Evaluasi pembelajaran menjadi sangat penting, karenanya dapat menjadi gambaran keberhasilan dalam mewujudkan tujuan pendidikan. Evaluasi bukan sekedar mengukur sejauh mana tujuan tercapai, tetapi digunakan untuk membuat keputusan. Evaluasi meliputi mengukur dan menilai. Mengukur lebih pada membandingkan sesuatu dengan satu ukuran, sedangkan menilai adalah mengambil suatu keputusan terhadap suatu dengan ukuran baik buruk. Penilaian atau *assessment* merupakan kegiatan menafsirkan data hasil pengukuran berdasarkan kriteria atau aturan-aturan tertentu (S. E. P. Widoyoko, 2017). Penilaian merupakan kegiatan kualitatif sedangkan mengukur adalah kegiatan kuantitatif.

Penilaian bagi peserta didik sangat membantu dalam melihat sejauh mana telah berhasil mengikuti proses pembelajaran, dan dapat memberikan motivasi pada peserta didik (Alfarisa & Purnama, 2019; Azizah & Wahyuningsih, 2020). Sedangkan bagi guru, penilaian dapat memberikan gambaran progres keberhasilan peserta didik, ketepatan materi yang diajarkan, maupun metode yang digunakan dalam pembelajaran. Hasil evaluasi bagi sekolah dapat menjadi dasar pengambilan keputusan dan kebijakan, serta dapat menjadi tolak ukur kualitas dari suatu satuan pendidikan (Purniasari et al., 2021). Kemampuan peserta didik biasanya ditunjukkan dengan jumlah skor yang dicapai. Jika siswa mencapai kriteria ketuntasan minimal atau melebihi rata-rata dianggap memiliki kemampuan yang tinggi, begitu juga sebaliknya. Akan tetapi skor mentah ini tidak dapat digunakan untuk menunjukkan ukuran prestasi siswa (Ayub et al., 2020; Purniasari et al., 2021). Jadi keputusan yang di dasarkan pada skor yang dimiliki peserta didik, kuranglah tepat.

Dalam proses evaluasi pembelajaran di sekolah/satuan pendidikan, guru adalah pihak yang paling bertanggungjawab. Peran guru/pendidik tidak hanya merencanakan dan melaksanakan pembelajaran, akan tetapi juga bertugas melaksanakan *assessment/penilaian* (Republik Indonesia, 2005). Hasil penilaian menunjukkan bagaimana kualitas pembelajaran yang telah dilakukan. Untuk itu guru harus memiliki bekal keterampilan yang baik dalam melakukan evaluasi pembelajaran, baik dalam perencanaan, penyusunan instrumen penilaian, analisis, dan penafsiran. Guru juga harus

memiliki keterampilan dalam menyusun butir-butir soal yang baik serta menganalisisnya.

Instrumen penilaian yang baik harus memenuhi kriteria keandalan. Kriteria keandalan meliputi validitas, reliabilitas, tingkat kesukaran, dan daya beda/diskriminasi. Untuk menguji keandalan/kualitas instrumen penilaian hasil belajar dilakukan uji coba instrumen. Uji coba instrumen dapat dilakukan secara internal dan eksternal. Uji coba internal dilakukan kepada ahli untuk melihat validitas content dan construct, serta tata bahasa (E. P. Widoyoko, 2012). Penilaian ahli diperlukan untuk mempertimbangkan struktur instrumen sudah benar atau belum dengan struktur atau susunan keilmuan yang digunakan dalam menyusun instrumen itu. Untuk memantapkan validitas instrumen penilaian perlu dilakukan uji coba eksternal, yakni uji coba lapangan (Arikunto, 2017). Uji coba lapangan dapat dilakukan pada subyek yang sejenis/setara dengan subyek yang akan dilakukan penilaian. Kemudian setelah uji coba perlu dilakukan analisis hasil uji coba, meliputi analisis validitas, reliabilitas, tingkat kesukaran, dan analisis daya beda soal.

Validitas atau kesahihan berasal dari kata *validity* yang berarti sejauh mana ketepatan dan kecermatan suatu instrumen penilaian melakukan fungsi penilaian, sejauh mana tes telah mengukur apa yang seharusnya diukur (Sudaryono et al., 2013). Reliabilitas adalah sejauh mana hasil penilaian dapat dipercaya, selama beberapa kali pengukuran pada subyek yang sama diperoleh hasil penilaian yang relatif sama. Tingkat kesukaran butir penilaian menunjukkan sejauh mana ketepatan subyek dalam menjawab soal. Soal yang baik bukanlah soal yang terlalu sulit atau pun juga tidak terlalu mudah. Sedangkan daya beda butir penilaian menunjukkan kemampuan siswa tinggi maupun rendah. Untuk dapat mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya beda soal maka perlu dilakukan analisis butir soal penilaian hasil belajar.

Analisis butir soal penilaian hasil belajar dari uji coba lapangan/empiris dapat dilakukan dengan cara klasik yang dikenal dengan teori tes klasik, *classical test theory* (CTT) maupun modern dengan teori respons butir (*Item Response Theory*, IRT). Teori tes klasik di dasarkan pada skor amatan yang merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Kualitas butir soal ditentukan dari tingkat kesukaran dan daya pembeda (Hardianti, 2021), Namun karakteristik

butir soal inkosisten tergantung dari kemampuan peserta didik (Erfan et al., 2020). Cara modern dengan Rasch Model yang dikemukakan pertama oleh Dr. Georg Rasch seorang ahli matematika dari Denmark, hadir untuk mengatasi kelemahan cara klasik. Rasch Model menggunakan skor mentah dengan cara yang berbeda untuk menghasilkan skala pengukuran dengan interval yang sama, sehingga dapat memberikan informasi yang akurat tentang peserta tes maupun kualitas soal (Sumintono & Widhiarso, 2015).

Artikel ini mencoba mendeskripsikan hasil analisis perbandingan kualitas instrumen tes pada aspek validitas, reliabilitas, tingkat kesukaran, dan daya beda melalui dua cara seperti yang telah di paparkan diatas, yakni cara klasik dan modern. Instrumen tes yang di analisis dalam penelitian ini adalah instrumen tes pengukuran kompetensi literasi sains. Aspek kerangka penilaian literasi sains meliputi aspek konteks, pengetahuan, kompetensi, dan sikap (OECD., 2018). Kompetensi literasi sains ditunjukkan dengan kemampuan 1) menjelaskan fenomena secara ilmiah; 2) mengevaluasi dan merancang penyelidikan ilmiah; 3) menafsirkan data dan bukti secara ilmiah (Jumini, Sutikno, et al., 2022). Ketiga kompetensi ini dilakukan dalam konteks pribadi, lokal/nasional, maupun global yang dipengaruhi oleh pengetahuan dan sikap. Pengetahuan meliputi konten, prosedural, epistemik. Sedangkan sikap meliputi Aspek literasi sains area sikap meliputi bagaimana memberikan respons terhadap isu sains baik skala personal, nasional, maupun global. Area ini meliputi ketertarikan terhadap sains, menghargai/menilai pendekatan ilmiah, kesadaran dan kepedulian terhadap lingkungan (Jumini, Hidayah, et al., 2022).

METODE

Penelitian ini dilakukan dengan pendekatan kuantitatif jenis deskriptif, dengan desain penelitian konstruksi dan validasi (Bungin, 2021; Creswell, 2017). Instrumen pengukuran kompetensi literasi sains dikonstruksi berdasarkan aspek pengetahuan, konteks, kompetensi, dan sikap. Untuk aspek pengetahuan dan konteks instrumen soal tes berupa soal pilihan ganda materi suhu, kalor, dan tekanan sejumlah 15 soal. Aspek kompetensi dan sikap di ukur dengan angket. Dalam konteks pembelajaran, sikap ilmiah dapat diukur dari materi yang dipelajari mahasiswa. Format pertanyaan dapat dikembangkan dalam bentuk angket atau

pertanyaan yang menuntut mahasiswa untuk memberikan jawaban dalam bidang pengetahuan (setuju atau tidak setuju), afeksi (senang atau tidak senang), dan kecenderungan psikomotorik/aktivitas (Abidin et al., 2018). Dalam penelitian ini hanya dideskripsikan hasil analisis instrumen pengukuran literasi sains pada soal tes. Teknik pengambilan sampel dengan sampling jenuh, dimana ada satu kelas di ambil sebagai subyek penelitian. Subyek penelitian pada uji coba lapangan dilakukan pada mahasiswa fisika semester 6 dengan mata kuliah IPA terpadu sejumlah 30 mahasiswa. Analisis data dilakukan dengan mendeskripsikan hasil olah data dari teori tes klasik menggunakan *Microsoft Excel* dan *Rasch Model* dengan *software Winsteps*.

HASIL DAN PEMBAHASAN

1) Validitas

Validitas merupakan tingkat kesahihan suatu instrumen, kemampuan suatu tes untuk mengukur apa yang seharusnya diukur (Arikunto, 2017). Validitas terdiri dari tiga tipe, yakni validitas isi, validitas konstruk, dan validitas kriteria (Azwar, 2019). Validitas isi terdiri dari validitas kenampakan dan validitas logika. Validitas isi menunjukkan bagaimana suatu instrumen penilaian menggambarkan keseluruhan apa yang akan diukur, berupa analisis domain yang akan diukur. Validitas logika dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah keseluruhan butir merupakan sampel representatif dari keseluruhan butir yang mungkin dibuat. Validitas kriteria merupakan validitas prediktif, yang dapat meramalkan keberhasilan peserta didik di masa yang akan datang, dapat di ketahui dengan cara mengkorelasikan instrumen dengan kriteria tertentu. Validitas konstruk menunjukkan seberapa jauh suatu instrumen tes mengukur sifat atau kemampuan peserta didik (Sudaryono et al., 2013). Validitas instrumen dalam pendekatan teori tes klasik dapat dihitung dengan menggunakan rumus korelasi product moment atau bisa juga dengan teknik korelasi point biserial (r_{pbi}). Soal dikatakan valid jika koefisien korelasi $r_{xy} > r$ tabel, dan jika $r_{xy} \leq r$ maka soal dapat dianggap tidak valid. Pada Model Rasch, untuk melihat kualitas butir soal dari aspek validitas adalah jika memenuhi beberapa kriteria berikut ini (Sumintono & Widhiarso, 2015).

- a) Nilai Outfit MNSQ (Mean Square) yang diterima adalah: $0,5 < \text{Outfit} - \text{MNSQ} < 1,5$
- b) Nilai Outfit ZSTD (Z - Standard) yang diterima adalah: $-2,0 < \text{ZSTD} < +2,0$
- c) Nilai Pt Measure Corr (Point Measure Correlation): $0,4 < \text{Point Measure Corr} < 0,85$

Butir soal dikatakan valid, jika telah memenuhi minimal 2 kriteria dan diperbaiki jika memenuhi salah satu dari ketiga kriteria tersebut, serta dibuang bila tidak ada yang memenuhi kriteria. Setelah dilakukan analisis butir soal tes kompetensi literasi sains baik secara teori tes klasik dan model Rasch, diperoleh hasil perbandingan hasil analisis validitas butir soal tes seperti ditunjukkan pada tabel 3.1 sebagai berikut.

Tabel 1. Perbandingan Hasil Analisis Validitas Butir Soal Melalui Pendekatan Teori Tes Klasik dan Model Rasch

No.	Hasil	Nomor Soal	
		Teori Tes Klasik	Model Rasch
1.	Valid	2,3,14	1, 2, 4, 5, 11, 13
2.	Tidak valid	1,4,5, 6,7,8,9,10,11,12,13,15	3, 6, 7, 8, 9, 10, 11, 12, 14, 15

Berdasarkan hasil analisis butir soal tes literasi sains pada tabel 3.1 diperoleh informasi bahwa secara teori tes klasik terdapat 3 soal yang dinyatakan valid dan 12 soal termasuk dalam kategori tidak valid. Sedangkan pada analisis kualitas soal dengan permodelan Rasch diperoleh 6 butir soal yang dikatakan valid dan 9 soal dikatakan tidak valid karena tidak memenuhi syarat outfit MNSQ, Outfit ZSTD, serta Point Measure Correlation (Pt Measure Corr). Ada satu soal yang tidak memenuhi ketiga kriteria yakni no 12, soal ini di buang, tidak dapat di perbaiki. Sedangkan soal tidak valid yang lainnya diperbaiki. Terdapat perbedaan hasil analisis dari dua pendekatan klasik dan modern. Pada pengujian model rasch di dapatkan soal yang valid lebih banyak daripada klasik. pendekatan Model Rasch tampaknya lebih benar karena jika item pertanyaan memenuhi 3 (tiga) kriteria yaitu nilai Outfit MNSQ, nilai Outfit ZSTD, serta nilai Point Measure Correlation maka pertanyaan tersebut dianggap valid.

2) Reliabilitas

Reliabilitas/keandalan merupakan konsistensi dari serangkaian alat ukur (Arikunto, 2017). Reliabilitas selalu menunjukkan keadaan instrumen penelitian dalam berbagai bentuk,

yakni hasil pengujian yang sama jika dilakukan oleh orang yang berbeda (inter-penilai), hasil pengujian yang sama jika dilakukan oleh orang yang sama dalam waktu berbeda (pengetesan ulang), hasil pengujian yang sama jika dilakukan oleh orang yang berbeda dalam waktu yang bersamaan dengan tes yang berbeda (bentuk paralel), dan hasil pengujian yang sama dengan menggunakan berbagai pertanyaan-pertanyaan membangun (konsistensi internal). Kualitas butir soal dari aspek reliabilitas butir soal pada pendekatan teori tes klasik menggunakan rumus Alpha Cronbach (KR-20) (Sugiyono, 2021). Butir soal dianggap reliabel jika memenuhi kriteria koefisien korelasi reliabilitas instrumen yang ditunjukkan pada Tabel 3.2 sebagai berikut.

Tabel 2. Kriteria Koefisien Korelasi Reliabilitas Instrumen

Koefisien Korelasi	Korelasi	Interpretasi/Reliabilitas
$0,90 \leq r \leq 1,00$	Sangat Tinggi	Sangat tetap/sangat baik
$0,70 \leq r < 0,90$	Tinggi	Tetap/baik
$0,40 \leq r < 0,70$	Sedang	Cukup tetap/cukup baik
$0,20 \leq r < 0,40$	Rendah	Tidak tetap/buruk
$r < 0,20$	Sangat Rendah	Sangat tidak tetap/sangat buruk

Sedangkan dalam Rasch Model Kriteria dalam menentukan nilai *Item Reliability* dan *Person Reliability* berdasarkan kriteria ditunjukkan pada tabel 3.3 sebagai berikut (Sumintono & Widhiarso, 2015).

Tabel 3. Kriteria Reliabilitas pada Pemodelan Rasch

Nilai Reliability (Person/Item)	Interpretasi
$>0,94$	Istimewa
$0,91 - 0,94$	Bagus sekali
$0,81 - 0,90$	Bagus
$0,67 - 0,80$	Cukup
$< 0,67$	lemah

Hasil analisis reliabilitas butir soal pengukuran kompetensi literasi sains dengan cara klasik maupun modern ditunjukkan pada tabel 3.4 berikut.

Tabel 4. Hasil Analisis Reliabilitas Soal dengan Cara Klasik dan Modern

Teori Tes Klasik		Model Rasch	
Kriteria Alpha Cronbach = 0,4	Kategori sedang	Kriteria Person Reliability=0,54	Kategori Lemah
		Item Reliability=0,91	Bagus sekali

Dari tabel 3.4 terlihat bahwa menurut teori tes klasik kualitas butir soal tes pengukuran kompetensi literasi sains pada aspek reliabilitas secara keseluruhan memiliki konsistensi sedang.

Pada Rasch Model Nilai person reliability 0,54 dalam kategori lemah, ini menunjukkan konsistensi mahasiswa saat menjawab soal dalam kondisi lemah/inkonsisten. Nilai item reliability 0,91 dalam kategori bagus sekali, ini menunjukkan kualitas butir-butir soal dalam instrumen memiliki aspek reliabilitas bagus sekali.

3) Tingkat Kesukaran

Tingkat kesulitan butir soal menunjukkan besarnya kemungkinan seberapa banyak responden yang dapat menjawab suatu butir soal dengan benar (Arikunto, 2017). Tingkat kesulitan butir soal menyatakan apakah butir soal dalam kategori sukar, sedang, atau mudah. Soal yang baik adalah soal yang tidak terlalu susah dan juga tidak terlalu mudah. Pada teori tes klasik tingkat kesukaran butir soal ditentukan berdasarkan kriteria pada tabel 3.5 sebagai berikut (Sudaryono et al., 2013).

Tabel 5. Kriteria Tingkat Kesukaran Butir Soal

No.	Nilai Indeks Kesukaran	Interpretasi Indeks Kesukaran
1.	IK = 0,00	Sangat Sukar
2.	0,00 < IK ≤ 0,30	Sukar
3.	0,30 < IK ≤ 0,70	Sedang
4.	0,70 < IK ≤ 1,00	Mudah
5.	IK=1	Terlalu mudah

Pada Rasch Model, tingkat kesukaran butir soal dikategorikan berdasarkan Measure logit dan nilai Simpangan Baku (SD) logit item dan dibagi menjadi empat kategori sebagai berikut (Sumintono & Widhiarso, 2015):

Tabel 6. Kriteria Tingkat Kesulitan Butir Soal dengan Pemodelan Rasch

Nilai Measure (logit)	Interpretasi Kesulitan Butir Soal
Measure logit > SD logit	Sangat Sukar
0 ≤ Measure logit ≤ SD logit	Sukar
- SD logit ≤ Measure logit ≤ 0	Sedang
Measure logit < -SD logit	Mudah

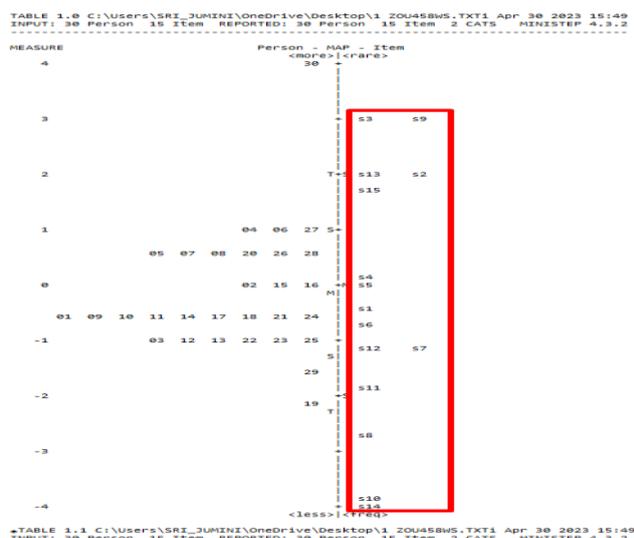
Nilai standar deviasi pada butir soal pengukuran literasi sains ini 2,37.

Hasil analisis tingkat kesukaran butir soal baik dengan cara klasik maupun modern ditunjukkan pada tabel 3.7 berikut.

Tabel 7. Hasil Analisis Butir Soal dengan Cara Klasik dan Modern

Kriteria	Soal	
	Teori Tes Klasik	Model Rasch
Sangat Sukar	-	3, 9
Sukar	2, 14	2,13,15
Sedang	4, 5, 6,	1,4, 5,6,7,11,12
Mudah	1, 3, 7, 8, 9, 10, 11, 12, 13, 15	8,10,14

Menurut data yang ditunjukkan tabel 3.7 terlihat bahwa menurut teori tes klasik butir soal tes pengukuran literasi sains terdistribusi dalam tiga kelompok, yakni mudah 10 soal, sedang 3 soal, dan sukar 2 soal. Sedangkan pada Rasch Model butir soal terdistribusi dalam empat kelompok, yakni mudah 3 soal, sedang 7 soal, sukar 3 soal, dan sangat sukar 2 soal. Pada Rasch Model tingkat kesukaran butir soal tes pengukuran kompetensi literasi sains ini dapat dikonfirmasi pada item map yang ditunjukkan pada gambar 3.1 berikut.



Gambar 1. Hasil Analisis Item Map Analisis Tingkat Kesukaran Butir Soal Literasi Sains

Dari item map ini terlihat bahwa soal nomor 3, dan 9 terletak pada logit tertinggi kategori

sangat sulit, dan tidak ada satupun mahasiswa yang mampu mengerjakan. Urutan dibawahnya

soal nomor 2,13,15 kategori sulit, tidak ada satupun yang dapat menjawab. Kemudian sebagian besar mahasiswa dapat mengerjakan soal dalam kategori sedang yakni soal nomor 1,4, 5,6,7,11,12. Soal urutan palin bawah nomor 8,10,14 kategori mudah, terlihat semua mahasiswa dapat menjawab dengan benar. Sedangkan pada Teori tes klasik dapat diknfirmasi pada jawaban peserta didik secara langsung.

4) Daya Beda

Daya pembeda butir soal merupakan kemampuan soal untuk membedakan peserta didik yang mampu menjawab soal atau peserta didik yang memiliki tingkat kemampuan yang tinggi dengan peserta didik yang memiliki kemampuan menjawab soal yang rendah. Pada analisis butir soal dengan teori tes klasik, dikriteriakan seperti ditunjukkan pada tabel 3.8 berikut.

Tabel 8. Hasil analisis Daya Beda Butir Soal dengan Teori Tes Klasik

Daya Pembeda (DP)	Interpretasi	Jumlah Butir soal
$DP \geq 0,70$	Baik	0
$0,40 \leq DP < 0,70$	Baik	1
$0,20 \leq DP < 0,40$	Cukup	0
$DP < 0,20$	Buruk	14

Hasil analisis dengan teori tes klasik sebagai besar butir soal menunjukkan pada daya pembeda yang buruk.

Pada pemodelan Rasch digunakan analisis pada tingkat abilitas individu sebagai alat untuk membedakan kemampuan peserta didik yang

mampu menjawab soal dan yang tidak mampu menjawab soal. Nilai daya beda dapat dilihat dari nilai model standar eror. Nilai Model SE kurang dari 0,5 mengindikasikan bahwa daya beda item tersebut bagus, sementara jika nilainya antara 0,5 - 1 daya bedanya dikategorikan cukup mampu untuk membedakan dan jika nilai. Model Standar Eror lebih besar dari 1, maka daya bedanya jelek atau tidak mampu membedakan (Purniasari et al., 2021). Selain itu, dapat pula menggunakan cara mengidentifikasi kelompok responden berdasarkan indeks separasi responden. Semakin besar nilai separasi item maka kualitas instrumen dalam hal keseluruhan responden dan butir soal makin bagus, karena dapat mengidentifikasi kelompok responden dan kelompok butir (Sumintono & Widhiarso, 2015). Persamaan lain untuk mengetahui pengelompokkan secara lebih teliti digunakan persamaan strata (H):

$$H = \frac{[(4 \times Separation) + 1]}{3}$$

Hasil analisis daya beda butir soal pengukuran literasi sains ditunjukkan pada gambar 2 berikut. Berdasarkan Gambar 2 diketahui bahwa nilai separasi butir soal sebesar 3,17 maka nilai H = 4,56 dibulatkan menjadi 5, sehingga terdapat lima kelompok butir soal yang dapat diidentifikasi, dan untuk responden diperoleh nilai separasi sebesar 0,93 dengan H = 1,57 dibulatkan menjadi 2, menunjukkan bahwa kelompok responden dapat dibedakan menjadi dua kelompok berdasarkan nilai separasi responden, yang menyatakan kemampuan mahasiswa.

SUMMARY OF 30 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S. E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	7.7	15.0	-.15	.74	.98	-.08	1.39	.23
SEM	.4	.0	.20	.01	.07	.18	.38	.20
P. SD	1.9	.0	1.09	.07	.37	.99	2.03	1.06
S. SD	1.9	.0	1.11	.07	.38	1.01	2.07	1.08
MAX.	14.0	15.0	3.94	1.12	1.85	1.69	9.90	3.26
MIN.	4.0	15.0	-2.18	.71	.44	-1.73	.30	-.84
REAL RMSE	.80	TRUE SD	.74	SEPARATION	.93	Person RELIABILITY	.46	
MODEL RMSE	.75	TRUE SD	.80	SEPARATION	1.07	Person RELIABILITY	.54	
S. E. OF Person MEAN	= .20							
Person RAW SCORE-TO-MEASURE CORRELATION = .99								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .43 SEM = 1.44								
SUMMARY OF 14 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S. E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	14.3	30.0	.09	.57	.95	-.04	1.48	.43
SEM	2.5	.0	.57	.05	.04	.18	.44	.43
P. SD	9.2	.0	2.05	.19	.16	.65	1.60	1.57
S. SD	9.5	.0	2.13	.20	.17	.68	1.66	1.63
MAX.	29.0	30.0	3.03	1.03	1.23	1.23	6.80	5.54
MIN.	2.0	30.0	-3.92	.40	.65	-1.63	.30	-.94
REAL RMSE	.62	TRUE SD	1.96	SEPARATION	3.17	Item RELIABILITY	.91	
MODEL RMSE	.60	TRUE SD	1.96	SEPARATION	3.25	Item RELIABILITY	.91	
S. E. OF Item MEAN	= .57							

Gambar 2. Summary Statistik Daya Beda Soal Tes LS

Hasil analisis daya beda butir soal secara kalsik maupun modern di dapatkan hasil yang hampir sama dengan berdasrkan nilai separasi dapat dikelompokkan menjadi dua, daya beda buruk 14 soal, dan baik 1 soal.

SIMPULAN

Hasil penelitian menunjukkan kualitas butir soal pengukuran literasi sains pada aspek validitas secara teori tes klasik terdapat 3 soal yang dinyatakan valid dan 12 soal termasuk dalam kategori tidak valid. Sedangkan pada analisis kualitas soal dengan permodelan Rasch diperoleh 6 butir soal yang dikatakan valid dan 9 soal dikatakan tidak valid. Analisis reliabilitas butir soal melalui pendekatan teori tes klasik diperoleh nilai Alpha Cronbach secara keseluruhan memiliki konsistensi sedang. Pada Rasch Model Nilai person reliability 0,54 dalam kategori lemah, ini menunjukkan konsistensi mahasiswa saat menjawab soal dalam kondisi lemah/inkonsisten. Nilai item reliability 0,91 dalam kategori bagus sekali, ini menunjukkan kualitas butir-butir soal dalam instrumen memiliki aspek reliabilitas bagus sekali. Tingkat kesukaran butir soal menurut teori tes klasik, butir soal terdistribusi dalam tiga kelompok, yakni mudah 10 soal, sedang 3 soal, dan sukar 2 soal. Sedangkan pada Rasch Model butir soal terdistribusi dalam empat kelompok, yakni mudah 3 soal, sedang 7 soal, sukar 3 soal, dan sangat sukar 2 soal. Pada aspek daya beda baik dengan pendekatan teori tes klasik maupun Rasch Model, nilai $H = 1,57$ dibulatkan menjadi 2, menunjukkan bahwa kelompok responden dapat dibedakan menjadi dua kelompok berdasarkan nilai separasi responden, yang menyatakan kemampuan mahasiswa, yakni instrumen pengukuran memiliki daya pembeda yang berada pada kategori buruk (14 item), dan hanya 1 item yang dianggap memiliki daya pembeda yang baik.

UCAPAN TERIMA KASIH

Ucapan terima kasih kepada Kementerian Pendidikan dan Kebudayaan Dikti Ristek yang telah memberikan pendanaan dengan nomor kontrak 10.11.5/UN37/PPK.6.8/2022, sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

Abidin, Y., Tita, A., & Hana, Y. (2018). *PEMBELAJARAN LITERASI Strategi*

meningkatkan Kemampuan Literasi Matematika, Sains, Membaca, dan Menulis (Y. N. Indahsari (ed.); 2nd ed.). Bumi Aksara.

Alfarisa, F., & Purnama, D. N. (2019). *Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Ekonomi SMA Menggunakan RASCH Model*. 11(2).

Arikunto, S. (2012). *Dasar-dasar evaluasi Pendidikan* (Restu Damayanti (ed.); Edisi ke-2). Bumi Aksara.

Arikunto, S. (2017). *Pengembangan Instrumen Penelitian dan Penilaian Program*. Yogyakarta: Pustaka Pelajar.

Ayub, M. R. S. S. N., Istiyono, E., Munadi, S., Permadi, C., Pattiserlihun, A., & Sudjito, D. N. (2020). Analisa Penilaian Soal Fisika Menggunakan Model Rasch Dengan Program R. *Jurnal Sains Dan Edukasi Sains*, 3(2), 46–52. <https://doi.org/10.24246/juses.v3i2p46-52>

Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, 3(1), 45–50. <https://doi.org/10.30598/jupitekvol3iss1pp45-50>

Azwar, S. (2019). *Penyusunan Skala Psikologi*. Yogyakarta: Pustaka Pelajar.

Bungin, B. (2021). *Metode Penelitian Sosial*. Universitas Airlangga Press.

Creswell, J. W. (2017). *RESEARCH DESAIN Pendekatan Kualitatif, Kuantitatif, dan Mixed* (A. Fawaid (ed.); ketiga). Pustaka Pelajar.

Erfan, M., Mauliyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19.

Hardianti, H. (2021). Karakteristik Tes Kemampuan Berpikir Kritis Siswa SMA pada Materi Momentum dan Impuls: Perbandingan Classical Theory Test (CTT) dan Model Rasch. *WaPFI (Wahana Pendidikan Fisika)*, 6(2), 167–173. <https://ejournal.upi.edu/index.php/WapFi/article/view/30958>

Jumini, S., Hidayah, R., & Hamzah, H. (2022). Science Technology Society (STS) Learning Through The Utilization of Potato Skin Waste on Students' Science Literacy. *Radiasi : Jurnal Berkala Pendidikan Fisika*, 15(2), 72–78. <https://doi.org/10.37729/radiasi.v15i2.1828>

- Jumini, S., Sutikno, Cahyono, E., & Parmin. (2022). *IPA TERPADU Berbasis Sciencetechnopreneurship*. Wonosobo: Mangkubumi. <https://jurnal.uns.ac.id/jpkim>
- OECD., K. (2018). *OECD Science, Technology and Innovation Outlook 2018*. oecd-ilibrary.org. https://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/korea_sti_in_outlook-2016-71-en
- Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia Sma N 1 Kutowinangun Tahun Pelajaran 2019/2020 Menggunakan Model Iteman Dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205–214.
- Republik Indonesia, P. (2005). *UU 14-2005 Guru dan Dosen* (p. 17).
- Sudaryono, Margono, G., & Rahayu, W. (2013). *Pengembangan Instrumen Penelitian Pendidikan*. Yogyakarta: Graha Ilmu.
- Sugiyono. (2021). *Metode Penelitian Kualitatif*. Bandung : Alfabeta.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH Pada Assessment Pendidikan* (C. T. Komunikata (ed.)).
- Widoyoko, E. P. (2012). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.
- Widoyoko, S. E. P. (2017). *Evaluasi Program Pelatihan*. Yogyakarta: Pustaka Pelajar.