



## Analisis Instrumen Tes Materi Dinamika Rotasi dan Keseimbangan Benda Tegar Menggunakan Model Rasch Dan Teori Tes Klasik

Shovian Maulana Pratama, Siti Annisa Zahra, Iyon Suyana, Heni Rusnayati

Artikel ini telah dipresentasikan pada kegiatan Seminar Nasional Fisika (Sinafi XI)

Universitas Pendidikan Indonesia, Bandung, Indonesia

16 Agustus 2025

### Abstract

*This study aimed to analyze the quality of physics evaluation instruments on the topic of Rotational Dynamics and Rigid Body Equilibrium through item testing. A descriptive quantitative method was employed. Data were obtained through document analysis from a test instrument consisting of 15 multiple-choice questions administered to 33 eleventh-grade students at a senior high school in Bandung. The data were analyzed using Classical Theory Test with Microsoft Excel and the Rasch Model with Winsteps software. Based on validity analysis using Classical Theory Test, 4 items were valid and 11 were invalid, while the Rasch Model classified 11 items as valid and 4 as invalid. The Cronbach's Alpha (KR-20) reliability score based on Classical Theory Test was 0.5551, categorized as moderate, while the Rasch Model yielded a KR-20 reliability of 0.56 (low), with a person reliability of 0.50 (low) and item reliability of 0.91 (very good). In terms of difficulty level, Classical Test Theory identified 8 easy items, 5 moderate, and 2 difficult, while the Rasch Model categorized 4 items as easy, 4 moderate, 5 difficult, and 2 very difficult. Based on the discrimination index using both Classical Test Theory and the Rasch Model, the H value was 1.653 (rounded to 2), indicating that 11 items were poorly discriminating and 4 were fair.*

**Keywords:** Rasch Model · Classical theory test · Instruments

### PENDAHULUAN

Pendidikan yang berkualitas itu meliputi sebuah proses pelaksanaan kegiatan belajar mengajar dengan proses evaluasi yang baik (Alfarisa & Purnama, 2019). Asesmen dari hasil belajar siswa atau penilaian adalah salah satu cara yang digunakan agar bisa melihat sejauh mana pencapaian belajar siswa (Abi Hamid, 2016; Magdalena dkk.,2020). Asesmen hasil belajar berguna untuk mengukur pemahaman dan penguasaan materi yang telah diberikan oleh guru kepada siswa, sehingga asesmen ini dapat memberikan Gambaran informasi berupa prestasi dan pencapaian kompetensi oleh setiap peserta didik yang diperoleh dari proses pembelajaran (Marjiastuti & Wahyuni, 2015). Asesmen dapat membantu siswa dalam pembelajaran, yaitu untuk mencapai tujuan yang diciptakan oleh guru, meningkatkan motivasi belajarnya, mendorong kebiasaan belajar yang baik, dan memberikan umpan balik melalui identifikasi kelemahan dan kelebihan siswa, sehingga asesmen memiliki peranan yang sangat penting (Pisca, 2014). Evaluasi adalah sebuah proses untuk mengumpulkan data dengan melihat bagaimana tujuan pendidikan sudah tercapai (Arikunto, 2013). Evaluasi pembelajaran sangat penting karena untuk melihat

✉ Shovian Maulana Pratama  
[shovianmaulanapratama9@upi.edu](mailto:shovianmaulanapratama9@upi.edu)  
Siti Annisa Zahra  
[stansa@upi.edu](mailto:stansa@upi.edu)

Iyon Suyana  
[iyons@upi.edu](mailto:iyons@upi.edu)  
Heni Rusnayati  
[heni@upi.edu](mailto:heni@upi.edu)

*Physics Education Study Program, Universitas Pendidikan Indonesia, Bandung, Indonesia.*

**How to Cite:** Pratama, S. M., Zahra S. A., Suyana, I. & Rusnayati, H. (2025). Analisis Instrumen Tes Materi Dinamika Rotasi dan Keseimbangan Benda Tegar Menggunakan Model Rasch Dan Teori Tes Klasik. *Prosiding Seminar Nasional Fisika*, 4(1), 241 – 252 .  
<https://proceedings.fisikaupi.id/index.php/sinafi/>

keberhasilan tujuan Pendidikan. Evaluasi tidak hanya mengukur sejauh mana tujuan sudah tercapai tetapi bisa digunakan untuk membuat Keputusan. Evaluasi ini meliputi mengukur dan menilai. Untuk mengukur yaitu membandingkan sesuatu dengan satu ukuran, sedangkan menilai lebih ke mengambil Keputusan terhadap suatu dengan ukuran baik dan buruk. Analisis kualitas soal merupakan proses yang harus dijalani dan dilaksanakan oleh seorang guru/fasilitator untuk mengetahui kualitas soal yang diberikannya (Farida dan Musyarofah, 2021). Penilaian atau *assessment* adalah kegiatan untuk menggambarkan data hasil pengukuran dengan berdasarkan kriteria aturan tertentu (S. E. P. Widoyoko, 2017). Penilaian adalah proses kegiatan kualitatif sedangkan mengukur itu adalah proses kegiatan kuantitatif.

Penilaian untuk peserta didik sangat membantu dalam keberhasilan mengikuti proses pembelajaran, dan dapat memberikan motivasi pada peserta didik (Alfarisa & Purnama, 2019; Azizah & Wahyuningsih, 2020). Hasil evaluasi bagi sekolah bisa menjadi dasar untuk pengambilan keputusan, kebijakan, dan menjadi tolak ukur kualitas dari suatu satuan Pendidikan (Purniasari et al., 2021). Kualitas dari hasil evaluasi bisa dipengaruhi oleh beberapa hal, diantaranya tes yang sedang digunakan. Tes sebagai alat atau prosedur yang sering digunakan untuk mengukur ataupun mengetahui sesuatu, dengan aturan-aturan dan cara yang sudah ditentukan (Arikunto, 2013). Tes berperan sangat penting dalam peningkatan kualitas pendidikan karena tes merupakan prosedur yang sistematis yang dilakukan untuk menilai atau mengamati perilaku seseorang yang dijelaskan dalam bentuk skor dengan kategori tertentu. Tes yang dikatakan baik untuk alat pengukur, harus memenuhi syarat tes, yaitu adanya validitas, reliabilitas, objektivitas, praktibilitas, dan ekonomis (Arikunto, 2013).

Instrumen penilaian yang memenuhi kriteria kualitas maka itu instrumen penilaian yang baik. Kriteria kualitas diantaranya validitas, reliabilitas, tingkat kesukaran, dan daya beda/diskriminasi. Untuk menguji kualitas instrumen harus dilakukan uji coba instrumen penilaian hasil belajar. Uji coba instrumen dapat dilakukan secara eksternal dan internal. Uji coba internal bisa dilakukan kepada ahli untuk melihat validitas *content* dan *construct*, serta tata bahasa (E. P. Widoyoko, 2012). Untuk memaksimalkan dari validitas instrumen penilaian maka perlu dilakukan uji coba eksternal, yaitu uji coba lapangan (Arikunto, 2017). Kemudian setelah uji coba harus dilakukan analisis hasil dari uji coba, yang meliputi analisis validitas, reliabilitas, tingkat kesukaran, dan analisis daya beda soal/diskriminasi.

Validitas atau kesahihan berasal dari kata *validity* yang mengacu pada sejauh mana ketepatan dan kecermatan suatu instrumen penilaian dalam mengukur apa yang seharusnya diukur (Sudaryono et al., 2013). Validitas Adalah uji yang dilakukan untuk mengukur sejauh mana akurasi suatu tes, sedangkan reliabilitas mengukur sejauh mana hasil suatu instrument dapat dipercaya (Syahrul, 2010; Zulpan & Rusli, 2020). Dengan kata lain, instrumen dikatakan valid apabila butir-butir soal di dalamnya benar-benar merepresentasikan kompetensi yang menjadi sasaran evaluasi. Sementara itu, reliabilitas mengacu pada konsistensi hasil penilaian, yakni sejauh mana hasil yang diperoleh dapat dipercaya apabila dilakukan pengukuran berulang dalam kondisi yang relatif sama. Baik buruknya kualitas instrumen tes bisa dilihat dari hasil validitas dan reliabilitas instrumen tersebut. Semakin tinggi nilai dari validitas dan reliabilitas suatu instrument, maka akan semakin tepat data yang didapat dari suatu penelitian yang kita lakukan (Hayati & Lailatussaadah, 2016). Instrumen yang reliabel akan memberikan hasil yang stabil meskipun digunakan pada waktu yang berbeda atau oleh penilai yang berbeda.

Selain validitas dan reliabilitas, kualitas instrumen tes juga dapat dilihat dari analisis butir soal, khususnya melalui indikator tingkat kesukaran dan daya beda. Tingkat kesukaran menunjukkan proporsi siswa yang menjawab suatu soal dengan benar, yang merefleksikan apakah soal tersebut tergolong mudah, sedang, atau sulit. Soal yang baik idealnya berada pada tingkat kesukaran sedang, karena terlalu mudah atau terlalu sulit tidak dapat secara optimal membedakan kemampuan siswa. Daya pembeda merupakan suatu kemampuan antara setiap butir soal untuk dapat membedakan antara siswa yang menguasai materi yang diujikan dan siswa yang belum menguasai materi yang diujikan, sedangkan Tingkat kesukaran merupakan kesanggupan siswa pada saat menjawab soal yang akan diujikan (Fatimah & Alfath, 2019; Hanifah, 2017). Daya beda atau indeks diskriminasi merupakan indikator yang menunjukkan sejauh mana suatu butir soal mampu membedakan antara siswa yang memiliki kemampuan tinggi dan siswa yang memiliki kemampuan rendah. Butir soal yang memiliki daya beda tinggi akan lebih efektif dalam menilai kompetensi siswa secara menyeluruh.

Oleh karena itu, untuk memastikan bahwa suatu instrumen penilaian layak digunakan dalam mengukur hasil belajar siswa, diperlukan proses analisis butir soal secara menyeluruh. Analisis ini dapat dilakukan menggunakan pendekatan teori tes klasik dan model pengukuran modern seperti Rasch Model. Uji validitas, reliabilitas, Tingkat kesukaran dan daya pembeda pada instrument tes mutlak harus dilakukan untuk mendapatkan hasil instrumen tes yang berkualitas (Erfan et al., 2020).

Analisis butir soal penilaian hasil belajar dari uji coba lapangan bisa dilakukan dengan menggunakan cara klasik atau yang dikenal dengan teori tes klasik, Classical Test Theory (CTT) bisa juga menggunakan cara modern dengan teori respon butir (Item Response Theory, IRT). Kualitas butir soal dilihat dari tingkat kesukaran dan daya beda/diskriminasi soal (Hardianti, 2021), tetapi karakteristik butir soal inkosisten tergantung dari kemampuan peserta didiknya (Erfan et al., 2020). Pendekatan menggunakan teori tes klasik atay Classical Theory Test (CTT) yang paling umum dan banyak digunakan dalam Pendidikan sampai saat ini. Teori ini dikembangkan oleh Charles Spearman pada tahun 1904, dan digunakan untuk memprediksikan hasil dari suatu tes (Sumintono & Widhiarso, 2015). Pada teori ini, aspek yang sangat menentukan baik atau buruknya kualitas butir soal Adalah Tingkat kesukaran dan daya pembeda tetapi, karakteristik butir soal yang dihasilkan oleh teori tes klasik bersifat inkonsisten (berubah) sebab bergantung pada kemampuan peserta tes. Oleh karena itu kekurangan tersebut, maka terciptalah pendekatan dengan cara modern yaitu Rasch Model yang pertama dikemukakan oleh Dr. Georg Rasch seorang ahli matematika yang berasal dari Denmark (Nuryanti dkk., 2018), yang mampu mengatasi kelemahan pada cara klasik. Pada Rasch Model menggunakan skor yang mentah dengan cara berbeda untuk menghasilkan skala pengukuran dengan interval yang sama, sehingga mampu memberikan informasi yang lebih akurat tentang kualitas soal maupun peserta tes (Sumintono & Widhiarso, 2015).

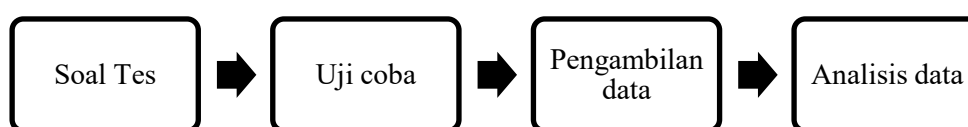
CTT ini memiliki beberapa keterbatasan tertentu seperti sifatnya yang sampel-dependent dan item-dependent. Artinya Tingkat kesulitan soal tergantung pada sampel yang mengerjakan, dan sebaliknya hasil skor yang diperoleh dipengaruhi oleh Tingkat kesulitan yang diberikan. Pada Model rasch bersifat perspektif dengan tujuan memastikan data sesuai dengan model ideal yang telah ditetapkan. Apabila tidak sesuai maka dilakukan diagnosis untuk mengetahui penyebab ketidaksesuaian tersebut dengan mengevaluasi butir soal atau mengidentifikasi pola jawaban responden yang tidak konsisten (Akhtar 2023).

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mendeskripsikan dan membandingkan kualitas instrumen tes berdasarkan empat indikator utama, yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, melalui dua pendekatan analisis yang berbeda, yakni pendekatan klasik (*Classical Test Theory/CTT*) dan pendekatan modern (Model Rasch). Fokus penelitian ini adalah pada analisis instrumen tes yang dikembangkan untuk mengukur penguasaan siswa terhadap materi Dinamika Rotasi dan Kesetimbangan Benda Tegar, yang merupakan bagian penting dalam kurikulum Fisika tingkat sekolah menengah atas. Pemilihan topik ini dilatarbelakangi oleh pentingnya ketersediaan instrumen evaluasi yang memiliki kualitas baik agar hasil asesmen mampu mencerminkan kemampuan siswa secara objektif dan dapat dipertanggungjawabkan secara ilmiah.

Penelitian ini menggunakan Teknik sampling jenuh karena hanya terdapat satu kelas yang dijadikan sebagai subjek penelitian, dengan jumlah siswa sebanyak 33 orang. Instrumen tes yang digunakan dianalisis menggunakan dua perangkat bantu, yaitu Microsoft Excel untuk analisis berbasis teori tes klasik dan perangkat lunak Winsteps untuk analisis Model Rasch. Hasil analisis dari kedua pendekatan tersebut akan disajikan secara deskriptif dan dibandingkan untuk melihat konsistensi dan perbedaan karakteristik butir soal yang dihasilkan. Temuan dari penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan instrumen evaluasi pembelajaran yang lebih akurat, adil, dan sesuai.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan desain penelitian konstruksi dan validasi instrumen (Bungin, 2021; Creswell, 2017). Pendekatan ini dipilih karena bertujuan untuk mendeskripsikan karakteristik dan kualitas instrumen tes yang dikembangkan, khususnya dalam aspek validitas, reliabilitas, tingkat kesukaran, dan daya beda. Fokus utama penelitian ini adalah pada analisis hasil uji coba instrumen tes yang disusun untuk mengukur penguasaan peserta didik terhadap materi Dinamika Rotasi dan Kesetimbangan Benda Tegar dalam mata pelajaran Fisika.



Gambar 1. Alur Penelitian

Teknik pengambilan sampel yang digunakan adalah *sampling jenuh*, yaitu seluruh populasi yang tersedia dijadikan sebagai sampel penelitian, mengingat hanya terdapat satu kelas yang digunakan sebagai subjek. Subjek penelitian terdiri atas 33 peserta didik jenjang sekolah menengah atas dan penelitian dilaksanakan pada semester genap 2024 yang mengikuti uji coba lapangan instrumen tes.

Pengumpulan data dilakukan melalui pelaksanaan tes tertulis berbentuk pilihan ganda dengan butir soal yang telah disusun dan dikembangkan sebelumnya. Data yang diperoleh dari hasil jawaban siswa kemudian dianalisis menggunakan dua pendekatan, yaitu teori tes klasik dan model pengukuran modern. Analisis teori tes klasik dilakukan menggunakan aplikasi Microsoft Excel untuk memperoleh informasi mengenai validitas empiris, reliabilitas, tingkat kesukaran, dan daya beda setiap butir soal. Sementara itu, analisis Rasch Model dilakukan

dengan menggunakan perangkat lunak Winsteps untuk memperoleh estimasi parameter butir dan kemampuan peserta didik secara simultan dan independen.

Perlu dicatat bahwa dengan jumlah responden 33 orang, analisis rasch model ini sebagai kajian eksploratif. Hal ini untuk memberikan gambaran awal mengenai karakteristik instrumen pada pendekatan antara CTT dan Model rasch. Perbandingan ini dapat dilakukan, namun interpretasi hasil harus mempertimbangkan keterbatasan jumlah sampel dan sifat eksploratif penelitian ini. Menurut Linarce (1994), untuk mendapatkan hasil analisis rasch yang sangat reliabel dengan Tingkat kepercayaan 99% diperlukan ukuran sampel antara 108 hingga 243 responden. Meskipun jumlah peserta dalam penelitian ini hanya sebanyak 33 orang, penggunaan Rasch Model tetap dianggap layak secara metodologis. Menurut Linacre, (1994), ukuran sampel yang kecil ini dapat digunakan selama data menunjukkan sebaran respons yang memadai dan tidak terdapat kesenjangan yang ekstrem dalam kategori skor. Menurut, Wright & Stone, (1979) juga menegaskan bahwa dalam tahap pengembangan awal instrumen, sampel kecil dapat memberikan informasi yang cukup untuk mengevaluasi kelayakan item, dinyatakan bahwa proses kalibrasi item tidak memerlukan jumlah peserta yang besar, selama respons yang dikumpulkan memiliki distribusi yang bermakna dan bisa mencerminkan variasi kemampuan peserta. Hasil dari kedua pendekatan ini disajikan dalam bentuk deskriptif komparatif guna memberikan gambaran yang menyeluruh mengenai kualitas instrumen yang dikembangkan.

## HASIL DAN PEMBAHASAN

### Validitas

Validitas adalah tingkat kasahihan pada suatu instrumen, kemampuan tes untuk mengukur suatu yang seharusnya diukur (Arikunto, 2017). Validitas ini terdiri dari 3 jenis, yaitu validitas isi, validitas konstruk, dan validitas kriteria (Azwar, 2019). Validitas soal ini berkaitan dengan sejauh mana butir soal dapat mengukur hasil belajar yang akan diukur (Tarmizi, dkk, 2020). Validitas intrumen dengan menggunakan pendekatan teori tes klasik dapat dihitung dengan teknik korelasi *product moment*. Soal yang valid apabila koefisien korelasi  $t_{hitung} > t_{tabel}$  tabel, dan jika  $t_{hitung} < t_{tabel}$  maka soal akan dianggap tidak valid. Pada Model Rasch, kita bisa melihat kualitas butir soal dari aspek validitas jika memenuhi beberapa kriteria berikut ini (Sumintono & Widhiarso, 2015).

Tabel 1. Validitas Butir Soal

Kesesuaian Butir Soal	
Kriteria	Nilai
<i>Outfit</i> MNSQ	$0,5 < \text{MNSQ} < 1,5$
<i>Outfit</i> ZSTD	$-2,0 < \text{ZSTD} < 2,0$
<i>Pt Measure Corr</i>	$0,4 < \text{Point Measure Corr} < 0,85$

Butir soal akan dikatakan valid, jika memenuhi minimal 2 kriteria dan akan diperbaiki apabila hanya memenuhi salah satu dari ketiga kriteria tersebut, akan dibuang apabila tidak ada yang memenuhi 3 kriteria tersebut. Analisis pada butir soal secara teori tes klasik dan Model Rasch diperoleh hasil perbandingan validitas butir soal seperti pada tabel 1 sebagai berikut.

Tabel 2. Perbandingan Hasil Analisis Validitas Butir Soal Melalui Pendekatan Teori Tes Klasik dan Model Rasch

Hasil Validitas	
Teori Tes Klasik	Model rasch
Valid	Valid
Tidak Valid	Tidak Valid
Tidak Valid	Valid
Tidak Valid	Valid
Tidak Valid	Tidak Valid
Tidak Valid	Valid
Tidak Valid	Valid
Tidak Valid	Tidak Valid
Valid	Valid
Valid	Valid
Valid	Valid
Tidak Valid	Tidak Valid
Tidak Valid	Valid
Tidak Valid	Tidak Valid
Tidak Valid	Valid

Berdasarkan hasil analisis pada tabel diatas diperoleh informasi bahwa melalui pendekatan teori tes klasik 5 soal termasuk dalam kategori valid dan 10 soal termasuk dalam kategori tidak valid. Sedangkan pada hasil analisis menggunakan pendekatan Model Rasch 10 soal dinyatakan valid dan 5 soal dinyatakan tidak valid karena tidak memenuhi kriteria *Outfit* MNSQ, *Outfit* ZSTD, serta *Point Measure Correlation* (Pt Measure Corr). Ada dua soal yang tidak memenuhi ketiga kriteria tersebut yaitu no 2 dan no 12, soal ini akan di buang karena tidak dapat di perbaiki. Sedangkan soal yang tidak valid lainnya akan diperbaiki namun tidak di buang. Hasil analisis pada tabel terdapat perbedaan antara pendekatan klasik dan modern. Pada pengujian menggunakan Model Rasch hasil soal yang valid lebih banyak dibandingkan dengan menggunakan teori tes klasik. Menggunakan pendekatan Model Rasch memiliki kemungkinan lebih benar karena item pertanyaan harus memenuhi tiga kriteria yaitu nilai *Outfit* MNSQ, nilai *Outfit* ZSTD, dan nilai *Point Measure Correlation* maka item pertanyaan tersebut akan dinyatakan valid.

### Reliabilitas

Reliabilitas atau keandalan adalah konsistensi dari serangkaian alat ukur (Arikunto, 2017). Reliabilitas ini menunjukkan kondisi instrumen penelitian dalam berbagai bentuk, yaitu jika hasil pengujian yang sama dilakukan oleh orang yang berbeda (inter-penilai), dan hasil pengujian yang sama dilakukan oleh orang yang berbeda dalam waktu yang sama dengan tes yang berbeda (parallel), dan hasil pengujian yang sama dengan menggunakan berbagai pertanyaan-pertanyaan membangun (konsistensi internal). Kualitas butir soal dari aspek reliabilitas butir soal pada pada pendekatan teori tes klasik menggunakan rumus Alpha Cronbach (KR-20) (Sugiyono,2021). Butir soal akan dianggap reliable apabila memenuhi kriteria koefisien korelasi reliabilitas instrumen yang ditunjukkan pada tabel berikut.

Tabel 3. Kriteria Koefisien Realiabilitas

Koefisien	Kriteria	Reliabilitas
$0,90 \leq r \leq 1,00$	Sangat Tinggi	Sangat baik
$0,70 \leq r \leq 0,90$	Tinggi	Baik
$0,40 \leq r \leq 0,70$	Sedang	Cukup baik
$0,20 \leq r \leq 0,40$	Rendah	Buruk
$r < 0,20$	Sangat Rendah	Sangat Buruk

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

Tabel 4. Kriteria Reliabilitas Rasch Model

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 pada reliabilitas butir soal dengan cara klasik dan modern ditunjukkan pada table 4 berikut

Tabel 5. Perbandingan Hasil Analisis Reliabilitas Butir Soal Melalui Pendekatan Teori Tes Klasik dan Model Rasch

Classical Theory Test		Model Rasch					
Alpha Cronbach	Kategori	Person Reliability	Kategori	Item Reliability	Kategori	Alpha Cronbach	Kategori
0,5551	Sedang	0,50	Lemah	0,90	Bagus	0,56	Lemah

Dari tabel berikut terlihat bahwa hasil analisis berdasarkan teori tes klasik kualitas butir soal pada aspek reliabilitas secara keseluruhan memiliki nilai yang berkategori sedang. Pada Rasch Model memiliki nilai person reliability 0,50 dalam kategori lemah, nilai ini menunjukkan konsistensi siswa saat menjawab soal dalam kondisi lemah. Sedangkan nilai item reliability 0,90 dalam kategori bagus, nilai ini menunjukkan instrument kualitas butir soal dalam memiliki aspek reliabilitas yang bagus.

### Tingkat Kesukaran

Tingkat Kesukaran/kesulitan butir soal akan menunjukkan besarnya kemungkinan banyaknya responden yang dapat menjawab suatu butir soal dengan benar (Arikunto, 2017). Tingkat kesukaran butir soal akan menunjukkan apakah butir soal dalam kategori sukar, sedang, atau mudah. Soal yang baik adalah soal yang tidak terlalu susah ataupun tidak terlalu mudah. Pada teori tes klasik tingkat kesukaran butir soal ditentukan berdasarkan kriteria pada tabel 5 sebagai berikut.

Tabel 6. Kriteria Tingkat Kesukaran Butir Soal

No	Klasifikasi	Kriteria
1	0,00-0,29	Sukar
2	0,30-0,69	Sedang
3	0,70-1,00	Mudah

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

Tabel 7. Kriteria Tingkat Kesukaran Butir Soal dengan rasch Model

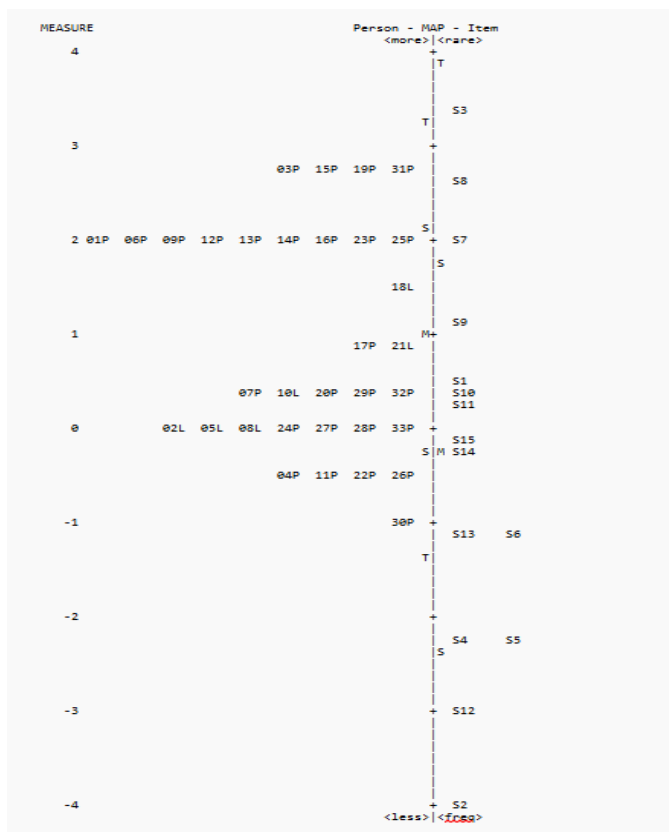
Nilai <i>Measure</i> (logit)	Interpretasi Kesulitan Butir Soal
$Measure \text{ logit} > SD \text{ logit}$	Sangat sukar
$0 \leq Measure \text{ logit} \leq SD \text{ logit}$	Sukar
$- SD \text{ logit} \leq Measure \text{ logit} \leq 0$	Sedang
$Measure \text{ logit} < - SD \text{ logit}$	Mudah

Nilai standar deviasi pada butir soal bernilai 2,07. Hasil analisis tingkat kesukaran pada butir soal baik dengan pendekatan klasik dan modern ditunjukkan pada tabel 7 berikut

Tabel 8. Hasil Analisis Butir Soal menggunakan Pendekatan Klasik dan Modern

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

Berdasarkan data pada tabel tersebut bahwa pada teori tes klasik butir soal terbagi dalam tiga kelompok yaitu mudah, sedang, dan sukar. Sedangkan pada Rasch Model butir soal terbagi menjadi empat kelompok yaitu mudah, sedang, sukar, dan sangat sukar. Pada Rasch Model tingkat kesukaran butir soal bisa dikonfirmasi melalui item map yang ditunjukkan pada **Gambar 2**.



Gambar 2. Wright Map siswa dan Butir Soal

Dari item map ini terlihat sebaran kemampuan siswa dalam menjawab butir soal dan tingkat kesulitan soal. *Wright map* sebelah kiri menggambarkan kemampuan siswa bisa kita lihat bahwa siswa dengan kode 03P, 15P, 19P, dan 31P (4 orang) memiliki kemampuan yang sangat tinggi dalam mengerjakan soal pada instrumen tes dibandingkan dengan siswa yang lain. Sedangkan untuk siswa yang memiliki kemampuan paling rendah dalam mengerjakan soal pada instrumen tes adalah siswa dengan kode 30P (1 orang). Pada *Wright map* sebelah kanan menggambarkan tingkat kesulitan butir soal, butir soal dengan kode S3 merupakan butir soal yang sangat sulit dan butir soal dengan kode S2 merupakan butir soal yang sangat mudah.

**Daya beda**

Daya beda/pembeda pada butir soal merupakan kemampuan suatu soal untuk dapat membedakan para peserta didik yang mampu menjawab soal atau para peserta didik yang memiliki tingkat kemampuan menjawab soal yang tinggi dengan peserta didik yang memiliki kemampuan yang rendah. Menurut (Novytasari 2014) indeks daya pembeda dapat dikatakan layak apabila hasilnya lebih dari 0,20 (dengan kriteria cukup). Pada analisis butir soal dengan teori klasik memiliki kriteria sebagai berikut.

Tabel 9. Hasil analisis Daya Beda Butir Soal menggunakan Teori Tes Klasik

Daya Pembeda (DP)	Interpretasi	Jumlah Butir soal
0,70 – 1,00	Baik sekali	0
0,40 - 0,69	Baik	0
0,20 - 0,39	Cukup	4
0,00 – 0,19	Jelek	11
DP < 0,00	Buruk	0



Hasil analisis dengan menggunakan teroi tes klasik sebagian besar butir soal menunjukkan dengan daya pembeda yang buruk.

Pada analisis Rasch Model menggunakan analisis pada tingkat abilitis individu sebagai alat untuk dapat membedakan kemampuan peserta didik yang mampu menjawab dan yang tidak mampu menjawab soal. Nilai daya beda dapat dilihat dari nilai standar *error*. Nilai Model SE apabila kurang dari 0,5 mengindikasikan bahwa daya beda item tersebut bagus, sementara jika nilainya antara 0,5 – 1 maka daya bedanya dikategorikan cukup mampu untuk membedakan dan jika model Standar *Error* lebih besar dari 1 daya bedanya jelek atau tidak mampu membedakan (Purniasari et al., 2021). Selain itu kita dapat juga menggunakan cara untuk mengidentifikasi kelompok responden berdasarkan indeks separasi responden. Apabila semakin besar nilai separasi itemnya maka kualitas instrumennya secara keseluruhan responden dan butir soal akan semakin bagus, akrena mampu mengidentifikasi antara kelompok responden dan kelompok butir (Sumintono & Widhiarso, 2015). Persamaan untuk mengetahui pengelompokkan secara lebih teliti menggunakan strata (H):

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

SUMMARY OF 14 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	20.9	33.0	.00	.53	.99	.01	.98	.08
SEM	2.4	.0	.51	.05	.04	.21	.12	.25
P. SD	8.8	.0	1.83	.18	.16	.75	.44	.89
S. SD	9.1	.0	1.90	.18	.16	.77	.46	.93
MAX.	32.0	33.0	3.41	1.03	1.39	1.57	1.83	1.54
MIN.	4.0	33.0	-3.03	.40	.72	-1.72	.41	-1.77
REAL RMSE	.57	TRUE SD	1.74	SEPARATION	3.03	Item	RELIABILITY	.90
MODEL RMSE	.56	TRUE SD	1.75	SEPARATION	3.10	Item	RELIABILITY	.91
S.E. OF Item MEAN = .51								

SUMMARY OF 33 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	9.9	15.0	.96	.75	.96	-.15	.98	.15
SEM	.4	.0	.21	.01	.09	.22	.15	.14
P. SD	2.2	.0	1.16	.08	.52	1.25	.86	.82
S. SD	2.2	.0	1.18	.08	.53	1.27	.87	.83
MAX.	13.0	15.0	2.77	.91	2.38	2.42	3.25	2.01
MIN.	6.0	15.0	-.98	.68	.40	-1.80	.20	-1.07
REAL RMSE	.83	TRUE SD	.82	SEPARATION	.99	Person	RELIABILITY	.50
MODEL RMSE	.75	TRUE SD	.88	SEPARATION	1.17	Person	RELIABILITY	.58
S.E. OF Person MEAN = .21								

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00  
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .56 SEM = 1.44

Gambar 3. Output Statistik Analisis

Berdasarkan hasil analisis pada daya beda butir soal ditunjukkan pada gambar berikut. Berdasarkan pada gambar berikut diketahui bahwa nilai separasi butir soal sebesar 3,03 maka nilai H = 4,373 dibulatkan menjadi 4 , sehingga terdapat 4 kelompok butir soal yang dapat diidentifikasi, dan untuk responden diperoleh nilai separasi sebesar 0,99 dengan H = 1,653 bulatkan menjadi 2, ini menunjukkan bahwa pada kelompok responden berdasarkan nilai tersebut menyatakan kemampuan pada peserta didik.

Pada hasil analisis daya beda butir soal menggunakan pendekatan klasik dan modern mendapatkan hasil yang hampir sama berdasarkan nilai separasi dikelompokkan menjadi dua, daya beda buruk 11 soal dan cukup 4 soal.

## SIMPULAN

Hasil dari penelitian ini menunjukkan kualitas instrumen tes materi dinamika rotasi dan kesetimbangan benda tegar pada aspek validitas menggunakan teori tes klasik terdapat 5 soal yang valid dan 10 soal yang tidak valid. Sedangkan pada analisis instrumen tes menggunakan Model Rasch diperoleh 10 soal dinyatakan valid dan 5 soal dinyatakan tidak valid. Analisis reliabilitas instrumen soal melalui pendekatan teori tes klasik diperoleh nilai Alpha Cronbach secara keseluruhan berkategori sedang. Sedangkan Pada Rasch Model nilai *person reliability* 0,50 dalam kategori lemah, nilai ini menunjukkan konsistensi siswa saat menjawab soal dalam kondisi lemah. Sedangkan nilai item reliability 0,90 dalam kategori bagus, nilai ini menunjukkan instrumen kualitas butir soal dalam memiliki aspek reliabilitas yang bagus. Tingkat kesukaran pada pendekatan teori tes klasik butir soal terbagi dalam tiga kelompok yaitu mudah 8 soal, sedang 5 soal, dan sukar 2 soal. Sedangkan pada Rasch Model butir soal terbagi menjadi empat kelompok yaitu mudah 5 soal, sedang 4 soal, sukar 4 soal, dan sangat sukar 2 soal. Pada aspek daya beda dengan menggunakan pendekatan teori uji klasik dan Rasch Model, nilai  $H = 1,653$  dibulatkan menjadi 2 yaitu menunjukkan bahwa 11 soal berkategori kurang baik dan 4 soal cukup. Hasil perbandingan ini menunjukkan bahwa dengan kedua pendekatan ini bisa memberikan informasi terkait kualitas instrumen, namun pada Model Rasch mampu memberikan informasi/gambaran yang lebih detail mengenai butir soal. Analisis perbedaan ini bahwa Model Rasch mampu menganalisis secara baik daripada CTT dalam aspek validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Selain itu model rasch juga bisa mengukur kemampuan peserta didik. Maka model rasch ini bisa menjadi pelengkap untuk analisis CTT, khususnya dalam pengembangan instrumen tes yang lebih akurat.

## DAFTAR PUSTAKA

- Abi Hamid, M. (2016). Pengembangan instrumen penilaian hasil belajar siswa berbasis tik pada pembelajaran dasar listrik jielektronika. *VOLT: Jurnal Ilmiah Pendidikan Teknik Elektro*, 1(1), 37–46.
- Akhtar, H. (2023, Oktober 16). *Perbedaan Classical Test Theory (CTT), Item Response Theory (IRT), dan Rasch Model*. Semesta Psikometrika.
- Alfarisa, F., & Purnama, D. N. (2019). *Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Ekonomi SMA Menggunakan RASCH Model*. 11(2).
- Arikunto, Suharsimi. 2013. *Dasardasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Arikunto, S. (2017). *Pengembangan Instrumen Penelitian dan Penilaian Program*. Yogyakarta: Pustaka Pelajar.
- 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.
- 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.
- Fatimah, L. U., & Alfath, K. (2019). Analisis kesukaran soal, daya pembeda dan fungsi distraktor. *ALMANAR: Jurnal Komunikasi dan Pendidikan Islam*, 8(2), 37–64.
- Farida, dan Anna Musyarofah. 2021. Validitas dan Reliabilitas dalam Analisis Butir Soal. *Jurnal Pendidikan Bahasa Arab*. Vol. No. 1.

- Hanifah, N. (2017). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *Sosio eKONS*, 6(1), 41–55.
- 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>
- Hayati, S., & Lailatussadiyah, L. (2016). Validitas dan reliabilitas instrumen pengetahuan pembelajaran aktif, kreatif dan menyenangkan (pakem) menggunakan model Rasch. *Jurnal Ilmiah Didaktika: Media Ilmiah Pendidikan dan Pengajaran*, 16(2), 169–179.
- Linacre. (1994). *Sample Size and Item Calibration Stability*. *Rasch Measurement Transactions*. 7:4.328
- Magdalena, I., Fauzi, H. N., & Putri, R. (2020). Pentingnya evaluasi dalam pembelajaran dan akibat memanipulasinya. *Bintang*, 2(2), 244–257.
- Marjiastuti, K., & Wahyuni, S. (2015). *Analisis kemampuan peserta didik dengan model Rasch*. Seminar Nasional Evaluasi Pendidikan II.
- Novytasari, Y. P. (2014). *Analisis Butir Soal Pilihan Ganda Menggunakan Teori Pengukuran Klasik Pada Ulangan Umum Akhir Semester Genap Bahasa Prancis SMA Negeri 9 Yogyakarta Tahun Ajaran 2013/2014*. Universitas Negeri Yogyakarta.
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224–233.
- Pisca, C. C. (2014). *Analisis Perbandingan Kualitas Butir Soal Ujian Sekolah Bahasa Prancis SMAN 10 Yogyakarta Tahun Ajaran 2013/2014 Dilihat dari Paradigma Klasik dan Modern (PhD Thesis)*. Universitas Negeri Yogyakarta.
- 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. <https://jurnal.uns.ac.id/jpkim>
- 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.))*.
- Syahrul, S. (2010a). Pengembangan Model Asesmen Kompetensi Siswa SMK dalam Konteks Pembelajaran Berbasis Kerja di Industri. *Jurnal Penelitian dan Evaluasi Pendidikan*, 14(2), 246–268.
- Taramizi, Pebrian, Panut Setiono, Yuli Amaliah, Arief Agrian. 2020. Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V Sd Negeri 04 Kota Bengkulu. *Elementary School Education Journal*. Vol. 4. No. 2.
- Widoyoko, E. P. (2012). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.
- Widoyoko, S. E. P. (2017). *Evaluasi Program Pelatihan*. Yogyakarta: Pustaka Pelajar.