

## ANALISIS KUALITAS SOAL KEMAMPUAN MEMBEDAKAN RANGKAIAN SERI DAN PARALEL MELALUI TEORI TES KLASIK DAN MODEL RASCH

Muhammad Erfan<sup>1</sup>, Mohammad Archi Maulyda<sup>2</sup>, Vivi Rachmatul Hidayati<sup>3</sup>, Fitri Puji Astria<sup>4</sup>, Tursina Ratu<sup>5</sup>

<sup>1,2,3,4</sup>Program Studi Pendidikan Guru Sekolah Dasar, Universitas Mataram, Mataram, Indonesia

<sup>5</sup>Program Studi Pendidikan Fisika, Universitas Ssamawa, Sumbawa Besar, Indonesia

e-mail: [muhammaderfan@unram.ac.id](mailto:muhammaderfan@unram.ac.id), [archimaulyda@unram.ac.id](mailto:archimaulyda@unram.ac.id), [vivirachma@unram.ac.id](mailto:vivirachma@unram.ac.id), [fitripujia@unram.ac.id](mailto:fitripujia@unram.ac.id), [tursinaratu@universitas-samawa.ac.id](mailto:tursinaratu@universitas-samawa.ac.id)

### Abstrak

Penelitian ini bertujuan untuk mendeskripsikan hasil perbandingan analisis kualitas soal kemampuan membedakan rangkaian seri dan rangkaian paralel melalui pendekatan teori tes klasik dan model Rasch dilihat dari aspek validitas, reliabilitas, tingkat kesukaran soal, serta daya pembeda soal. Penelitian ini merupakan penelitian deskriptif kualitatif dan proses pengambilan data dilakukan pada bulan Februari 2020. Data diperoleh melalui metode dokumentasi berupa 40 butir soal pilihan ganda dengan dua opsi jawaban dan diujikan kepada 195 orang mahasiswa Program Studi Pendidikan Guru Sekolah Dasar FKIP Universitas Mataram. Data dianalisis dengan pendekatan teori tes klasik menggunakan SPSS 18.0 dan model Rasch dengan *software* Winstep. Kualitas soal dari aspek validitas dengan pendekatan teori tes klasik diperoleh 30 soal yang dikategorikan valid dan 10 soal dianggap tidak valid sedangkan dengan model Rasch diperoleh 5 soal kategorikan valid dan 35 soal dikategorikan tidak valid. Nilai reliabilitas *Cronbach Alpha* (KR-20) soal tes melalui pendekatan teori tes klasik adalah 0,470 (buruk) dan nilai reliabilitas *Cronbach Alpha* (KR-20) pada model Rasch adalah 0,640 (cukup) dengan nilai *Person Reliability* sebesar 0,60 (lemah) dan *Item Reliability* sebesar 0,97 (istimewa). Rata-rata indeks tingkat kesukaran soal pada pendekatan teori tes klasik adalah 0,631 dengan kategori sedang, sedangkan hasil analisis tingkat kesukaran berdasarkan model Rasch menunjukkan terdapat empat kategori tingkat kesulitan butir soal yaitu 7 soal yang termasuk sangat mudah, 13 soal mudah, 15 soal sulit, serta 5 soal sangat sulit. Daya pembeda butir soal dengan pendekatan teori tes klasik berada pada kategori buruk (20 item), kategori cukup 19 butir soal, serta 1 (satu) item dengan kategori baik. Daya pembeda butir soal dari pemodelan Rasch diperoleh 8 (delapan) kelompok butir soal yang didasarkan pada indeks separasi butir soal ( $H=8,28$ ).

**Kata kunci:** soal, seri, paralel, teori tes klasik, model rasch

### Abstract

*This study aims to describe the results of a comparative analysis of the quality of the ability to distinguish series and parallel series through the classical test theory approach and the Rasch model viewed from aspects of validity, reliability, level of difficulty of the questions, and the distinguishing power of items. This research is a qualitative descriptive study and the data collection process was carried out in February 2020. The data was obtained through a documentation method in the form of 40 multiple choice questions with two answer options and was tested on 195 students of Elementary School Teacher Education Study Program, Faculty of teacher Training and Education, Universitas Mataram. Data were analysed with the classical test theory approach using SPSS 18.0 and the Rasch model with Winstep. The quality of the questions from the aspect of validity with the classical test theory approach obtained 30 questions that were categorized as valid and 10 questions were considered invalid while with the Rasch model obtained 5 questions categorized as valid and 35 questions were categorized invalid. The reliability value of questions through the classical test theory approach is 0.470 (bad) and the reliability value on the Rasch model is 0.640 (enough) with a Person Reliability value of 0.60 (weak) and Item Reliability of 0.97 (special). The average index of the difficulty level of the questions on the classical test theory approach is 0.631 in the medium category, while the results of the difficulty level analysis based on the Rasch model show that there are four categories of difficulty items namely 7 items are very easy, 13 are easy, 15 are difficult, and 5 questions are very difficult. The distinguishing power items with the classical test theory approach is in the bad category (20 items), enough categories of 19 items, and 1 item in the good category. The distinguishing power of items from the Rasch model was obtained by 8 (eight) groups of items based on its separation index value ( $H=8,28$ ).*

**Keywords:** *questionnaire, series, parallel, Classical test theory, Rasch model*

## 1. Pendahuluan

Evaluasi merupakan salah satu komponen penting yang harus ada dalam proses pembelajaran sebagaimana evaluasi itu sendiri merupakan kegiatan untuk mengumpulkan data-data untuk mengukur sejauh mana tujuan pembelajaran telah tercapai. Oleh karena itu guru sebagai ujung tombak sistem pendidikan dituntut tidak hanya mampu mendidik tetapi juga harus dibekali kemampuan mengevaluasi proses pembelajaran karena keberhasilan kegiatan pembelajaran dapat dilihat dari hasil evaluasi yang dilakukan setelah kegiatan proses pembelajaran (Widyaningsih & Yusuf, 2018).

Pendidik atau guru selain harus mampu mengevaluasi proses pembelajaran, juga harus mampu mengembangkan alat-alat atau instrumen evaluasi yang digunakan dalam proses pembelajaran sesuai dengan jenis capaian pembelajaran yang diharapkan. Selain itu, alat-alat atau instrumen evaluasi juga harus disesuaikan dengan teknik dalam mengukur dan memperoleh data yang selanjutnya dapat berfungsi sebagai indikator ketercapaian tujuan pembelajaran.

Teknik yang biasa digunakan dalam mengukur ketercapaian tujuan pembelajaran pada kegiatan evaluasi adalah dengan teknik tes dan non tes. Tes merupakan suatu alat untuk mengumpulkan informasi tentang ketercapaian tujuan pendidikan atau tujuan pembelajaran (Wahyudi, 2010), selain itu tes juga merupakan cara tertentu yang dapat digunakan atau prosedur yang perlu ditempuh dalam rangka pengukuran dan penilaian di bidang pendidikan (Kadir, 2015). Teknik non tes merupakan teknik penilaian hasil belajar peserta didik yang dilakukan dengan tanpa “menguji” peserta didik, melainkan dengan melakukan pengamatan secara sistematis atau dikenal dengan observasi, wawancara, penyebaran angket (kuesioner), analisis dokumen skala (baik skala sikap maupun skala penilaian), studi kasus, serta sosiometri (Mania, 2008). Kemampuan pendidik dalam memilih dan mengatur teknik yang sesuai dalam kegiatan evaluasi merupakan suatu keterampilan mengajar yang wajib dikuasai oleh calon pendidik.

Instrumen penilaian yang baik harus memenuhi beberapa kriteria di antaranya validitas butir soal yang baik, reliabilitas butir soal yang bagus, memiliki tingkat kesulitan butir soal yang bervariasi, serta memiliki daya beda soal yang mampu membedakan peserta didik yang cerdas dan berkemampuan menjawab soal dengan peserta didik yang tidak mampu menjawab soal. Semakin tinggi nilai validitas dan reliabilitas suatu instrumen, maka akan semakin jitu data yang didapat dari suatu penelitian (Hayati & Lailatussaadah, 2016), validitas dan reliabilitas juga merupakan faktor penting dalam menentukan apakah suatu tes berkriteria baik atau tidak (Tri Wahyuningsih, 2015).

Soal dikatakan valid jika soal tersebut mengukur sesuatu yang harus diukur, artinya jika hasil pelajaran yang diinginkan meliputi perubahan dalam pengetahuan, keterampilan, dan sikap, maka soal-soal yang dibuat pun haruslah mencakup ketiga hal tersebut (pengetahuan, keterampilan, dan sikap) (Sumintono & Widhiarso, 2015). Reliabilitas suatu tes pada hakikatnya menguji keajegan pertanyaan tes yang di dalamnya berupa seperangkat butir soal apabila diberikan berulang kali pada objek yang sama (Nuswowati, Binadja, Soeprodjo, & Ifada, 2010). Misalnya, ujian yang diberikan hari ini kepada siswa oleh seorang guru, seharusnya memberikan nilai yang tidak jauh berbeda apabila diberikan esoknya (karena tidak ada aktivitas pembelajaran atau lupa pada jangka waktu yang hanya satu hari) (Sumintono & Widhiarso, 2015).

Aspek lain yang mendukung baiknya kualitas perangkat soal adalah aspek tingkat kesukaran butir soal dan daya pembeda soal (Tri Wahyuningsih, 2015). Tingkat kesukaran butir soal merupakan proporsi antara siswa yang dapat menjawab benar suatu butir soal dengan jumlah seluruh peserta tes (Rahayu & Djazari, 2016). Soal yang baik kualitasnya adalah soal yang tidak terlalu mudah dan juga tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang mahasiswa untuk mempertinggi usaha menjawab soal tersebut, dan soal yang terlalu sukar akan menyebabkan mahasiswa berputus asa dan tidak mempunyai

semangat untuk mencoba lagi karena berada di luar jangkauannya (Iskandar & Rizal, 2018). Daya pembeda butir soal adalah kemampuan suatu butir soal untuk membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada dalam kelompok itu (Bagiyono, 2017). Soal dikatakan memiliki daya pembeda yang baik jika soal tersebut mampu membedakan individu yang memiliki kemampuan yang tinggi dan rendah (Sumintono & Widhiarso, 2015). Berdasarkan uraian uraian di atas, pengujian validitas, reliabilitas, tingkat kesukaran serta daya pembeda soal mutlak perlu dilakukan supaya diperoleh instrumen tes yang berkualitas (Perdana, 2018).

Analisis instrumen tes dalam evaluasi di bidang pendidikan dapat dilakukan melalui dua pendekatan. Pendekatan pertama yaitu pendekatan yang paling banyak dan umum diterapkan dalam bidang pendidikan hingga saat ini adalah teori tes klasik atau *classical test theory* (CTT). Aspek yang sangat menentukan kualitas butir soal dalam pendekatan teori tes klasik adalah tingkat kesukaran dan daya pembeda soal. Namun, karakteristik butir soal yang dihasilkan oleh teori tes klasik inkonsisten (berubah) bergantung pada kemampuan peserta tes.

Pendekatan yang kedua adalah pendekatan modern dengan permodelan Rasch. Permodelan Rasch memberikan pendekatan yang berbeda dalam penggunaan skor atau data mentah ujian pada konteks penilaian pendidikan. Pengaplikasian pengukuran permodelan Rasch pada data mentah hasil ujian bertujuan untuk menghasilkan suatu skala pengukuran dengan interval yang sama yang pada akhirnya dapat memberikan informasi secara akurat tentang kemampuan peserta tes maupun kualitas soal yang dikerjakan oleh peserta didik. Analisis butir soal dengan pemodelan Rasch akan menghasilkan informasi mengenai karakteristik butir dan siswa yang telah dibentuk menjadi metrik yang sama (Sumintono & Widhiarso, 2015).

Penelitian ini mencoba untuk menganalisis perbandingan kualitas instrumen dilihat dari beberapa aspek seperti validitas, reliabilitas, serta tingkat kesukaran soal melalui kedua pendekatan yaitu pendekatan teori tes klasik dan model Rasch. Item tes yang dianalisis dalam penelitian ini adalah instrumen pengukuran berupa item soal kemampuan membedakan rangkaian seri dan rangkaian paralel. Analisis butir soal dilakukan sebagai salah satu upaya dalam meningkatkan kualitas alat evaluasi, mengembangkan alat evaluasi yang dapat mengukur kemampuan peserta didik dalam membedakan rangkaian seri dan rangkaian paralel sebagai umpan balik terhadap kegiatan pembelajaran khususnya pembelajaran IPA pada materi kelistrikan pada tingkat mahasiswa calon guru sekolah dasar. Analisis butir soal juga penting dilakukan untuk mempertahankan dan meningkatkan kualitas instrumen penelitian khususnya pada ranah kemampuan membedakan rangkaian seri dan rangkaian paralel.

## 2. Metode

Penelitian ini merupakan penelitian deskriptif kualitatif dengan menggunakan data sekunder instrumen soal pengukuran kemampuan membedakan rangkaian seri dan rangkaian paralel. Item soal berupa pertanyaan dengan opsi pilihan ganda dengan dua pilihan jawaban dan berjumlah 40 butir soal. Pertanyaan diujikan pada mahasiswa program studi pendidikan guru sekolah dasar FKIP Universitas Mataram serta soal diberikan melalui *Google Form*.

Penelitian ini bermaksud untuk mencari informasi dan data yang dapat digunakan untuk mendeskripsikan kualitas instrumen tes secara empiris ditinjau dari aspek validitas, reliabilitas, serta tingkat kesulitan soal melalui pendekatan teori tes klasik yang diolah dengan bantuan program SPSS versi 18.0 dan pendekatan model Rasch dengan bantuan *software* Winstep.

### 3. Hasil dan Pembahasan

#### Hasil Uji Validitas

Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel yang dilihat dari aspek validitas, reliabilitas, serta tingkat kesulitan soal. Validitas butir soal diuji dengan metode korelasi *product moment*. Soal dikatakan valid jika koefisien korelasi  $r_{xy} > r_{tabel}$  dan jika  $r_{xy} \leq r_{tabel}$  maka soal dapat dianggap tidak valid (Syofian, 2015). 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*):  $0,4 < \text{Point Measure Corr} < 0,85$

Setelah dilakukan pengujian aspek validitas pada tiap butir soal baik secara teori tes klasik dan model Rasch, diperoleh hasil perbandingan hasil analisis validitas butir soal kemampuan membedakan rangkaian seri dan rangkaian paralel sebagaimana dapat dilihat pada Tabel 1.

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

No.	Hasil	Nomor Soal	
		Teori Tes Klasik	Permodelan Rasch
1.	Valid	1, 2, 3, 4, 5, 6, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 29, 30, 31, 32, 33, 35, 36, 37, 39, 40	9, 11, 16, 23, 39
2.	Tidak Valid	7, 8, 10, 13, 25, 26, 27, 28, 34, 38	1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40

Berdasarkan hasil analisis kualitas soal kemampuan membedakan rangkaian seri dan rangkaian paralel dilihat dari segi validitas butir soal, diperoleh informasi bahwa secara teori tes klasik terdapat 30 soal yang dinyatakan valid dan 10 soal termasuk dalam kategori tidak valid. Sedangkan pada analisis kualitas soal dengan permodelan Rasch diperoleh 5 butir soal yang dikatakan valid dan 35 soal dikatakan tidak valid karena tidak memenuhi syarat outfit MNSQ, Outfit ZSTD, serta *Point Measure Correlation* (Pt Measure Corr).

Berdasarkan hasil analisis kualitas butir soal kemampuan membedakan rangkaian seri dan rangkaian paralel pada aspek validitas, diketahui bahwa terdapat perbedaan hasil analisis yang dianggap valid oleh kedua pendekatan baik itu secara teori tes klasik dan pemodelan Rasch. Secara teori tes klasik diperoleh butir soal valid yang lebih banyak dari pada butir soal valid pada pemodelan Rasch. Lima soal valid dari pemodelan Rasch juga valid menurut pendekatan teori tes klasik. Kualitas butir soal lebih banyak melalui pengujian kualitas pada pemodelan Rasch dari pada dengan pendekatan teori tes klasik karena pada pemodelan Rasch, analisis butir soal dengan pendekatan pemodelan Rasch dirasa lebih akurat karena ketika item soal memenuhi 3 (tiga) kriteria yaitu nilai *Outfit MNSQ*, nilai *Outfit ZSTD*, serta nilai *Point Measure Correlation* maka soal tersebut dianggap valid.

#### Reliabilitas Butir Soal

Kualitas butir soal dari aspek reliabilitas butir soal pada pendekatan teori tes klasik menggunakan rumus Alpha Cronbach (KR-20). Butir soal dianggap reliabel jika memenuhi kriteria koefisien korelasi reliabilitas instrumen yang ditunjukkan pada Tabel 2.

**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 teta/buruk
$r < 0,20$	Sangat Rendah	Sangat tidak tetap/sangat buruk

Sedangkan dalam model Rasch, kriteria dalam menentukan nilai *Item Reliability* dan *Person Reliability* berdasarkan kriteria 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

Perbandingan hasil analisis reliabilitas soal kemampuan membedakan rangkaian seri dan rangkaian paralel melalui kedua pendekatan baik secara teori tes klasik dan pemodelan Rasch diuraikan pada Tabel 4.

**Tabel 4.** Perbandingan hasil analisis reliabilitas soal dengan pendekatan teori tes klasik dan pemodelan Rasch

Teori Tes Klasik		Pemodelan Rasch			
Alpha Cronbach	Kategori	Person Reliability	Kategori	Item Reliability	Kategori
0,470	Sedang	0,60	Lemah	0,97	Istimewa

Berdasarkan Tabel 4 dapat diketahui bahwa nilai terdapat perbedaan nilai *Person Reliability* yang bernilai 0,60 yang termasuk dalam kategori lemah dengan nilai *Item Reliability* sebesar 0,97 dengan kategori istimewa. Perbedaan nilai yang sangat signifikan ini menunjukkan bahwa konsistensi dari jawaban dari mahasiswa pada saat menjawab soal kemampuan membedakan rangkaian seri dan rangkaian paralel termasuk sangat lemah atau tidak konsisten. Berdasarkan nilai *Item Reliability* yang sebesar 0,97 dengan kategori istimewa dapat disimpulkan bahwa kualitas butir-butir soal dalam instrumen memiliki aspek reliabilitas yang bagus.

#### *Tingkat Kesulitan Butir Soal*

Tingkat kesulitan butir soal menunjukkan besarnya kemungkinan seberapa banyak responden yang dapat menjawab suatu butir soal dengan benar. Pada teori pengukuran butir soal secara klasik, indeks kesukaran butir soal diinterpretasikan pada kriteria sebagai berikut (Lestari & Yudhanegara, 2017):

**Tabel 5.** Perbandingan hasil analisis reliabilitas soal dengan pendekatan teori tes klasik dan pemodelan Rasch

Nilai Indeks Kesukaran	Interpretasi Indeks Kesukaran
IK = 0,00	Sangat Sukar
0,00 < IK ≤ 0,30	Sukar
0,30 < IK ≤ 0,70	Sedang
0,70 < IK ≤ 1,00	Mudah
IK = 1	Terlalu Mudah

Pada pemodelan Rasch, tingkat kesulitan 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 <i>Measure (logit)</i>	Interpretasi Kesulitan Butir Soal
$Measure\ logit < -SD\ logit$	Item sangat mudah
$-SD\ logit \leq Measure\ logit \leq 0$	Item mudah
$0 \leq Measure\ logit \leq SD\ logit$	Item sulit
$Measure\ logit > SD\ logit$	Item sangat sulit

**Tabel 7.** Hasil Distribusi Tingkat Kesukaran Butir Soal dengan Pendekatan Teori Tes Klasik

Nilai Indeks Kesukaran Soal	Interpretasi Indeks Kesukaran	Jumlah Butir Soal
IK = 0,00	Sangat Sukar	0
0,00 < IK ≤ 0,30	Sukar	2
0,30 < IK ≤ 0,70	Sedang	20
0,70 < IK ≤ 1,00	Mudah	18
IK = 1	Terlalu Mudah	0

**Tabel 8.** Hasil Distribusi Tingkat Kesukaran Butir Soal dengan Pemodelan Rasch

Nilai <i>Measure (logit)</i>	Interpretasi Kesulitan Butir Soal	Jumlah Butir Soal
$Measure\ logit < -1,12$	Item sangat mudah	7
$-1,12 \leq Measure\ logit \leq 0,00$	Item mudah	13
$0,00 \leq Measure\ logit \leq 1,12$	Item sulit	15
$Measure\ logit > 1,12$	Item sangat sulit	5

Berdasarkan Tabel 7 dan Tabel 8 dapat diketahui perbandingan antara hasil analisis soal pada aspek tingkat kesukaran melalui pendekatan teori tes klasik dan pemodelan Rasch. Berdasarkan kedua tabel tersebut dapat diketahui bahwa sebagian besar tingkat kesukaran butir soal yang dianalisis dengan pendekatan teori tes klasik berada pada kategori sedang dan mudah, hanya 2 (dua) soal yang termasuk dalam kategori soal sukar.

Hasil yang berbeda diperoleh setelah butir soal dianalisis dengan pemodelan Rasch. Tingkat kesulitan atau kesukaran butir soal pada pemodelan Rasch dilihat dari nilai *measure* pada satuan *logit* tiap butir soal. Berdasarkan kriteria pada tabel 7 dapat diketahui bahwa dari nilai *Measure (logit)* item dapat dikelompokkan menjadi 4 (empat) kategori tingkat kesukaran butir soal yaitu item sangat mudah, item mudah, item sulit, serta item sangat sulit dengan distribusi yang hampir merata.

SUMMARY OF 195 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	25.2	40.0	.71	.38	1.00	.0	.98	.0
S.D.	4.5	.0	.63	.03	.23	1.4	.35	1.3
MAX.	34.0	40.0	2.12	.48	1.68	3.7	2.36	4.3
MIN.	15.0	40.0	-.65	.36	.58	-3.0	.48	-2.5
REAL RMSE	.40	TRUE SD	.49	SEPARATION	1.24	Person RELIABILITY	.60	
MODEL RMSE	.38	TRUE SD	.50	SEPARATION	1.32	Person RELIABILITY	.63	
S.E. OF Person MEAN = .05								

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00  
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .64

SUMMARY OF 40 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	123.0	195.0	.00	.18	1.00	.1	.98	.0
S.D.	40.4	.0	1.12	.04	.13	2.2	.22	2.4
MAX.	182.0	195.0	2.86	.29	1.35	7.1	1.48	7.4
MIN.	23.0	195.0	-2.10	.15	.79	-4.4	.66	-4.5
REAL RMSE	.19	TRUE SD	1.10	SEPARATION	5.96	Item RELIABILITY	.97	
MODEL RMSE	.18	TRUE SD	1.11	SEPARATION	6.08	Item RELIABILITY	.97	
S.E. OF Item MEAN = .18								

Gambar 1. Hasil *Summary Statistic* untuk 40 butir soal dengan pemodelan Rasch

### Daya Pembeda Soal

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.

Hasil perhitungan indeks daya pembeda butir soal berdasarkan teori tes klasik secara umum dapat kategorikan menjadi tiga kategori sebagaimana terdapat pada Tabel 9.

**Tabel 9.** Indeks Daya Bada Butir Soal Melalui Teori Tes Klasik

Daya Pembeda (DP)	Interpretasi	Jumlah butir soal
$DP \geq 0,70$	Baik Sekali	0
$0,40 \leq DP < 0,70$	Baik	1
$0,20 \leq DP < 0,40$	Cukup	19
$DP < 0,20$	Buruk	20

Berdasarkan Tabel 8 dapat diketahui bahwa hasil analisis daya pembeda soal melalui pendekatan teori tes klasik menunjukkan bahwa sebagian besar daya beda soal berada pada kategori buruk. Daya pembeda soal berada pada kategori cukup sebanyak 19 soal dan hanya satu soal yang memiliki daya pembeda yang baik yaitu item nomor 3.

Berbeda dengan pendekatan teori tes klasik, 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. 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 ( $H$ ):

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

Berdasarkan Gambar 1 diketahui bahwa nilai separasi butir soal sebesar 5,96 maka nilai  $H = 8,28$  dibulatkan menjadi 8, sehingga terdapat delapan kelompok butir soal dan untuk responden diperoleh nilai separasi sebesar 1,24 dengan  $H = 1,98$  dibulatkan menjadi 2, menunjukkan bahwa kelompok responden dapat dibedakan menjadi dua kelompok berdasarkan nilai separasi responden.

#### 4. Simpulan

Hasil penelitian menunjukkan kualitas butir soal kemampuan membedakan rangkaian seri dan rangkaian paralel melalui pendekatan teori tes klasik secara umum dikategorikan sebagai soal dengan kualitas baik dari segi validitas butir soal (30 butir soal valid dari keseluruhan 40 butir soal) sedangkan melalui analisis pemodelan Rasch belum diperoleh kualitas soal yang baik (hanya 5 butir soal yang dikategorikan valid dari keseluruhan 40 butir soal). Analisis reliabilitas butir soal melalui pendekatan teori tes klasik diperoleh nilai Alpha Cronbach untuk nilai reliabilitas sebesar 0,470 dengan kategori sedang dan dari *Item Reliability* diperoleh nilai reliabilitas sebesar 0,97 dengan kategori istimewa. Rata-rata indeks tingkat kesukaran soal pada pendekatan teori tes klasik adalah 0,631 dengan kategori sedang, sedangkan hasil analisis tingkat kesukaran berdasarkan model Rasch menunjukkan terdapat empat kategori tingkat kesulitan butir soal yaitu 7 soal yang termasuk sangat mudah, 13 soal mudah, 15 soal sulit, serta 5 soal sangat sulit. Setelah dianalisis dengan pendekatan teori tes klasik, instrumen pengukuran memiliki daya pembeda yang berada pada kategori buruk (20 item), kategori cukup 19 butir soal, serta hanya 1 item yang dianggap memiliki daya pembeda yang baik. Daya pembeda butir soal dari pemodelan Rasch diperoleh delapan kelompok butir soal yang didasarkan pada indeks separasi butir soal  $H = 8,28$ .

#### Daftar Pustaka

- Bagiyono, B. (2017). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat 1. *WIDYANUKLIDA*, 16(1), 1–12.
- Hayati, S., & Lailatussaadah, L. (2016). VALIDITAS DAN RELIABILITAS INSTRUMEN PENGETAHUAN PEMBELAJARAN AKTIF, KREATIF DAN MENYENANGKAN (PAKEM) MENGGUNAKAN MODEL RASCH. *Jurnal Ilmiah Didaktika*, 16(2), 169. <https://doi.org/10.22373/jid.v16i2.593>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12. <https://doi.org/10.21831/pep.v22i1.15609>
- Kadir, A. (2015). MENYUSUN DAN MENGANALISIS TES HASIL BELAJAR. *AL-TA'DIB: Jurnal Kajian Ilmu Kependidikan*, 8(2), 70–81. <https://doi.org/10.31332/atdb.v8i2.411>
- Lestari, K., & Yudhanegara, M. (2017). *Penelitian Pendidikan Matematika*. Bandung: PT Refika Aditama.
- Mania, S. (2008). TEKNIK NON TES: TELAHAH ATAS FUNGSI WAWANCARA DAN KUESIONER DALAM EVALUASI PENDIDIKAN. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah Dan Keguruan*, 11(1), 45–54. <https://doi.org/10.24252/lp.2008v11n1a4>

- Nuswowati, M., Binadja, A., Soeprodjo, S., & Ifada, K. E. N. (2010). PENGARUH VALIDITAS DAN RELIABILITAS BUTIR SOAL ULANGAN AKHIR SEMESTER BIDANG STUDI KIMIA TERHADAP PENCAPAIAN KOMPETENSI. *Jurnal Inovasi Pendidikan Kimia*, 4(1), 566–573.
- Perdana, S. A. (2018). ANALISIS KUALITAS INSTRUMEN PENGUKURAN PEMAHAMAN KONSEP PERSAMAAN KUADRAT MELALUI TEORI TES KLASIK DAN RASCH MODEL. *Jurnal Kiprah*, 6(1), 41–48. <https://doi.org/10.31629/kiprah.v6i1.574>
- Rahayu, R., & Djazari, M. (2016). ANALISIS KUALITAS SOAL PRA UJIAN NASIONAL MATA PELAJARAN EKONOMI AKUNTANSI. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1). <https://doi.org/10.21831/jpai.v14i1.11370>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata.
- Syofian. (2015). *Metode Penelitian Kuantitatif dilengkapi Perbandingan Perhitungan Manual & SPSS*. Jakarta: PRENAMEDIA GROUP.
- Tri Wahyuningsih, E. (2015). ANALISIS BUTIR SOAL TES OBJEKTIF BUATAN GURU ULANGAN SEMESTER GANJIL MATA PELAJARAN EKONOMI KELAS X DI SMA NEGERI 1 MLATI TAHUN AJARAN 2013/2014 (Universitas Negeri Yogyakarta). Retrieved from <http://eprints.uny.ac.id/id/eprint/26627>
- Wahyudi, W. (2010). Assesment Pembelajaran Berbasis Portofolio di Sekolah. *Jurnal Visi Ilmu Pendidikan*, 2(1), 288–296. <https://doi.org/10.26418/jvip.v2i1.370>
- Widyaningsih, S. W., & Yusuf, I. (2018). ANALISIS SOAL MODUL LABORATORIUM FISIKA SEKOLAH I MENGGUNAKAN RACSH MODEL. *Gravity : Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, 4(1). <https://doi.org/10.30870/gravity.v4i1.3116>