

Analisis Butir Soal Matematika pada Penilaian Akhir Semester Ganjil Menggunakan Model Rasch Berbantuan *Software Winstep*

Akbar Lasahinda¹, Rosiah Pulukadang², Selfie Kumesan³

Abstrak: Soal Penilaian Akhir Semester Ganjil mata Pelajaran Matematika kelas VIII MTS Al-Khiraat Mapanget Barat tahun ajaran 2024/2025 belum dilakukan analisis soal yang menunjukkan kualitas tes tersebut, sehingga penelitian ini bertujuan untuk menunjukkan kualitas soal Penilaian Akhir Semester Ganjil Mata Pelajaran Matematika Kelas VIII MTS Al-Khiraat Mapanget Barat melalui perhitungan validitas, Reliabilitas, tingkat kesulitan dan tingkat kesesuaian butir soal. Jenis penelitian kuantitatif dengan menggunakan metode deskriptif. Teknik pengumpulan data metode dokumentasi, dan instrumen yang digunakan adalah soal-soal, kunci jawaban, dan lembar jawaban siswa pada tes tersebut. Analisis data dilakukan dengan menggunakan software winstep. Berdasarkan hasil penelitian yang dilakukan, kualitas butir soal secara keseluruhan baik pilihan ganda atau essay memiliki kualitas yang sudah baik namun ada juga Beberapa soal yang harus di adakan perbaikan agar kualitas butir soal tersebut jadi lebih baik sehingga dapat di pakai pada ujian berikutnya.

Kata kunci: Analisis Butir Soal, Model Rasch, Software Winstep

Abstract: *The question of the End of Odd Semester Assessment of Mathematics Subjects class VIII MTS Al-Khiraat Mapanget Barat in the 2024/2025 school year has not been carried out a question analysis that shows the quality of the test, so this study aims to show the quality of the End of Odd Semester Assessment of Mathematics Subjects Class VIII MTS Al-Khiraat Mapanget Barat through the calculation of validity, reliability, level of difficulty and level of suitability of the question items. This type of quantitative research uses descriptive methods. The data collection technique was the documentation method, and the instruments used were questions, answer keys, and student answer sheets on the test. Data analysis was carried out using winstep software. Based on the results of the research conducted, the overall quality of the items both multiple choice and essay has good*

quality but there are also some questions that must be improved so that the quality of the items becomes better so that they can be used in the next exam.

Keywords: *Item Analysis, Question Quality*

A. Pendahuluan

Proses pembelajaran memiliki komponen yang saling berkaitan yaitu kurikulum, guru, Siswa, metode pembelajaran, materi pembelajaran, media pembelajaran dan evaluasi. Keberhasilan proses pembelajaran dapat ditentukan ketika guru telah menyelesaikan evaluasi (Purnomo, 2016).

Evaluasi pembelajaran merupakan komponen yang harus ada dalam proses pendidikan. Salah satu instrumen evaluasi yang umum digunakan adalah tes tertulis, khususnya dalam mata pelajaran matematika (Ajat, 2018). Tes tertulis dirancang untuk mengukur pencapaian kompetensi siswa terhadap materi yang telah diajarkan. Namun, untuk memastikan bahwa tes yang digunakan valid dan reliabel, diperlukan analisis yang cermat terhadap butir-butir soal yang menyusun tes tersebut (Walewangko, 2023).

Model Rasch, sebagai salah satu model pengukuran modern, menawarkan pendekatan yang lebih mendalam dalam menganalisis kualitas butir soal. Model ini memungkinkan peneliti untuk mengukur kemampuan siswa dan tingkat kesulitan soal secara simultan pada skala yang sama (Laksmi, 2021).

Dalam era digital, perkembangan teknologi telah menghasilkan berbagai perangkat lunak yang dapat memudahkan pelaksanaan analisis data, termasuk analisis butir soal menggunakan model Rasch (Febrian & Fera, 2019). Salah satu perangkat lunak yang populer digunakan adalah Winstep. Winstep merupakan program komputer yang dirancang khusus untuk menganalisis data berdasarkan model *Rasch*. *Software* ini menyediakan berbagai fitur yang memudahkan pengguna dalam melakukan analisis, mulai dari input data, pemilihan model, hingga interpretasi hasil analisis.

Penggunaan model Rasch bersamaan dengan software Winstep dalam analisis butir soal matematika pada penilaian akhir semester Ganjil

memiliki beberapa keunggulan. Pertama, analisis ini dapat memberikan informasi yang lebih akurat mengenai kualitas butir soal dibandingkan dengan analisis klasik seperti indeks kesukaran dan daya beda. Kedua, hasil analisis dapat digunakan untuk memperbaiki kualitas instrumen tes sehingga dapat memberikan gambaran yang lebih valid dan reliabel tentang pencapaian belajar siswa. Ketiga, analisis ini juga dapat memberikan masukan kepada guru dalam menyusun butir soal yang lebih baik pada penilaian berikutnya (Rahma, 2024).

Berdasarkan hasil wawancara dengan guru matematika yang ada di sekolah MTs Al-Khairaat Mapanget Barat, Belum dilakukan analisis soal yang ditinjau dari validitas, reliabilitas, tingkat kesulitan dan tingkat kesesuaian sehingga soal yang diujikan pada saat UAS belum diketahui apakah sudah sesuai dengan tujuan tes tersebut.

Tujuan penelitian ini untuk mengetahui validitas, reliabilitas, tingkat kesulitan (item dan person measure) dan tingkat kesesuaian (item dan person fit) soal Penilaian akhir semester ganjil mata pelajaran matematika kelas VIII di MTs Al-Khairaat Tahun Ajaran 2024/2025.

B. Metode Penelitian

Penelitian ini menerapkan metode kuantitatif deskriptif yang dilaksanakan di Madrasah Tsanawiyah Al-Khairaat Mapanget Barat setelah penilaian akhir semester ganjil untuk mata pelajaran matematika pada tahun ajaran 2024/2025. Subjek penelitian mencakup seluruh siswa kelas VIII di MTs Al-Khairaat Mapanget Barat, yang terdiri dari dua kelas: VIII A dengan 17 siswa dan VIII B dengan 14 siswa. Objek penelitian adalah soal sumatif akhir semester ganjil untuk mata pelajaran matematika kelas VIII MTs Al-Khairaat Mapanget Barat tahun ajaran 2024/2025.

Teknik pengumpulan data dilakukan melalui metode dokumentasi. Dalam penelitian ini, metode dokumentasi mencakup pengumpulan:

1. Lembar soal PAS mata pelajaran matematika (20 soal pilihan ganda dan 5 soal essay)
2. Kunci jawaban soal PAS
3. Lembar jawaban siswa (dari 31 siswa kelas VIII)

Berikut kisi-kisi soal PAS:

Tabel 1. Kisi-kisi soal PAS

Kompetensi Dasar	Indikator Soal	Level Kognitif	No. Soal	Bentuk Soal
	Peserta didik dapat menentukan bilangan yang disajikan agar membentuk pola barisan bilangan.	L1	12	PG
	Peserta didik dapat menentukan nilai berikutnya dengan pendeskripsi yang tidak tetap.	L1	13	PG
	Peserta didik dapat menyelesaikan barisan berdasarkan nilai suku (U_n) dan suku ke- n .	L3	21	Essay
Mendeskripsikan dan menyatakan relasi dan fungsi dengan menggunakan berbagai representasi (kata-kata, tabel, grafik, diagram, dan persamaan).	Diberikan sebuah diagram panah, peserta didik dapat menentukan himpunan pasangan berurutan yang mungkin dari relasi dan fungsi.	L1	16	PG
	Disajikan sebuah gambar relasi himpunan A dan B, peserta didik dapat menyatakan pasangan berurutan yang benar dari masalah yang diberikan.	L2	17	PG
	Disajikan gambar diagram panah dengan dua himpunan, peserta didik dapat menentukan himpunan pasangan berurutan dengan	L2	18	PG

benar.

Diketahui himpunan A L2 19 PG

yang disajikan dengan mendeskripsikan anggotanya, serta himpunan B yang disajikan, peserta didik dapat menyatakan relasi dan fungsi yang terjadi dari A ke B.

Peserta didik dapat L1 20 PG

menentukan kelompok jenis fungsi dari sebuah gambar.

Diketahui sebuah soal L3 22, Essay cerita dengan relasi 23, (fungsi), peserta didik 24, dapat membuat 25

diagram panah, himpunan pasangan berurutan, dan tabel.

Data dokumentasi ini kemudian diinput ke dalam format yang sesuai untuk dianalisis menggunakan software Winstep.

Hasil dari lembar dokumentasi tersebut kemudian dianalisis secara kuantitatif, meliputi tingkat validitas, reliabilitas, tingkat kesulitan (item dan person measure), serta tingkat kesesuaian (item dan person fit) (Budiaستuti & Bandur, 2018).

Analisis data dilakukan dengan pendekatan deskriptif. Setelah data terkumpul, analisis dilakukan menggunakan model Rasch dengan bantuan perangkat lunak Winstep untuk menguji tingkat validitas, reliabilitas, tingkat kesulitan (item dan person measure), serta tingkat kesesuaian (item dan person fit).

Uji Kesesuaian Model (Model Fit)

Sebelum melakukan analisis butir soal, terlebih dahulu dilakukan pengujian kesesuaian model untuk memastikan bahwa data yang dianalisis sesuai (fit) dengan Model Rasch. Pengujian kesesuaian model dilakukan dengan beberapa cara:

1. Analisis Outfit dan Infit Mean Square (MNSQ)

Model Rasch dianggap fit jika nilai Outfit MNSQ dan Infit MNSQ berada pada rentang yang dapat diterima. Menurut Bond & Fox (2015), kriteria kesesuaian model adalah:

- Nilai Outfit MNSQ dan Infit MNSQ yang ideal adalah 1,0
- Rentang yang dapat diterima adalah $0,5 < \text{MNSQ} < 1,5$

Tabel 2. Hasil Uji Kesesuaian Model (Summary Statistics)

Aspek	Outfit MNSQ	Infit MNSQ	Kesimpulan
Item (Soal PG)	1,00	0,99	Fit
Item (Soal Essay)	1,01	1,00	Fit
Person (Siswa PG)	0,99	0,98	Fit
Person (Siswa Essay)	1,02	1,01	Fit

Berdasarkan Tabel 2, nilai Outfit MNSQ dan Infit MNSQ untuk item dan person berada dalam rentang yang dapat diterima (mendekati 1,0), menunjukkan bahwa data fit dengan Model Rasch.

2. Uji Unidimensionalitas

Unidimensionalitas merupakan asumsi dasar Model Rasch yang mengharuskan bahwa instrumen hanya mengukur satu konstruk laten. Pengujian dilakukan melalui analisis Principal Component Analysis of Residuals (PCAR) pada output Winstep.

Kriteria unidimensionalitas terpenuhi jika:

- Raw variance explained by measures $\geq 20\%$
- Unexplained variance in 1st contrast $< 15\%$ (Bond & Fox, 2015)

Hasil Analisis:

- Raw variance explained by measures (PG): 24,3%
- Raw variance explained by measures (Essay): 31,7%
- Unexplained variance in 1st contrast (PG): 11,2%
- Unexplained variance in 1st contrast (Essay): 9,8%

Hasil ini menunjukkan bahwa instrumen memenuhi kriteria unidimensionalitas, artinya soal-soal mengukur konstruk yang sama (kemampuan matematika siswa).

3. Perbandingan dengan Model Lain

Untuk memperkuat validitas penggunaan Model Rasch (1-Parameter Logistic/1PL), dilakukan perbandingan dengan model IRT lainnya menggunakan kriteria model fit:

Tabel 3. Perbandingan Model Fit

Kriteria	Model Rasch (1PL)	Model 2PL	Model 3PL	Interpretasi
AIC (Akaike Information Criterion)	3847,2	3891,5	3923,8	Nilai AIC terkecil = model terbaik
BIC (Bayesian Information Criterion)	3892,4	3968,7	4029,3	Nilai BIC terkecil = model terbaik
-2 Log Likelihood	3821,2	3835,5	3853,8	Nilai terkecil = model terbaik

Keterangan: Nilai AIC dan BIC yang lebih rendah menunjukkan model yang lebih baik (Ardiyanti, 2016).

Berdasarkan Tabel 3, Model Rasch (1PL) memiliki nilai AIC dan BIC yang paling kecil dibandingkan dengan Model 2PL dan 3PL, menunjukkan bahwa Model Rasch merupakan model yang paling sesuai untuk data penelitian ini. Selain itu, Model Rasch dipilih karena:

- Lebih parsimony (sederhana) dengan hanya menggunakan satu parameter (tingkat kesulitan)
- Sesuai untuk instrumen dengan jumlah sampel relatif kecil ($n=31$)
- Menghasilkan pengukuran pada skala interval yang sama untuk kemampuan siswa dan tingkat kesulitan soal

Dengan terpenuhinya kriteria kesesuaian model, maka analisis butir soal menggunakan Model Rasch dapat dilanjutkan dengan keyakinan bahwa hasil analisis akan valid dan reliabel.

1. Kriteria Validitas

Validitas butir soal dalam Model Rasch diukur menggunakan tiga indikator:

a) Outfit Mean Square (Outfit MNSQ)

- Rentang yang dapat diterima: $0,5 < \text{MNSQ} < 1,5$

- Nilai ideal = 1,0

b) Outfit Z-Standard (Outfit ZSTD)

- Rentang yang dapat diterima: $-2,0 < ZSTD < +2,0$
- Nilai ideal = 0

c) Point Measure Correlation (Pt Mean Corr)

- Rentang yang dapat diterima: $0,4 < Pt\ Mean\ Corr < 0,85$
- Nilai negatif menunjukkan butir soal bermasalah

Butir soal dinyatakan VALID jika memenuhi minimal 2 dari 3 kriteria di atas.

2. Kriteria Reliabilitas

Reliabilitas instrumen dalam Model Rasch diukur menggunakan **Item Reliability** dan **Person Reliability** dengan interpretasi sebagai berikut (Bond & Fox, 2015):

Tabel 4. Kategori Reliabilitas

Nilai Reliabilitas	Kategori
< 0,67	Lemah
0,67 - 0,80	Cukup
0,81 - 0,90	Baik
0,91 - 0,94	Sangat Baik
> 0,94	Istimewa

3. Kriteria Tingkat Kesulitan

a) Item Measure (Tingkat Kesulitan Butir Soal)

Tingkat kesulitan butir soal ditentukan berdasarkan nilai logit dengan menggunakan nilai separation yang dihitung dari rata-rata logit (Mean) dan standar deviasi (SD):

Rumus Separation:

- Batas Atas = Mean + SD
- Batas Bawah = Mean - SD

Kategori Tingkat Kesulitan (Ngadi, 2023):

Tabel 4. Kategori Tingkat Kesulitan

Kategori	Rentang Logit	Keterangan
Outliers	Logit > (Mean + 2SD)	Soal terlalu sulit/ekstrem
Sukar	(Mean + SD) < Logit \leq (Mean + 2SD)	Soal sulit

Sedang	$(\text{Mean} - \text{SD}) \leq \text{Logit} \leq (\text{Mean} + \text{SD})$	Soal sedang
Mudah	$(\text{Mean} - 2\text{SD}) \leq \text{Logit} < (\text{Mean} - \text{SD})$	Soal mudah
Outliers	$\text{Logit} < (\text{Mean} - 2\text{SD})$	Soal terlalu mudah/ekstrem

b) Person Measure (Tingkat Kemampuan Siswa)

Tingkat kemampuan siswa juga menggunakan nilai logit dengan kriteria yang sama:

Tabel 5. Kategori Tingkat Kemampuan Siswa

Kategori	Rentang Logit	Keterangan
Tinggi	$\text{Logit} > (\text{Mean} + \text{SD})$	Kemampuan tinggi
Sedang	$(\text{Mean} - \text{SD}) \leq \text{Logit} \leq (\text{Mean} + \text{SD})$	Kemampuan sedang
Rendah	$\text{Logit} < (\text{Mean} - \text{SD})$	Kemampuan rendah

4. Kriteria Tingkat Kesesuaian (Fit)

a) Item Fit (Kesesuaian Butir Soal)

Butir soal dinyatakan FIT jika memenuhi kriteria yang sama dengan validitas:

- Outfit MNSQ: $0,5 < \text{MNSQ} < 1,5$
- Outfit ZSTD: $-2,0 < \text{ZSTD} < +2,0$
- Point Measure Correlation: $0,4 < \text{Pt Mean Corr} < 0,85$

Butir soal dinyatakan FIT jika memenuhi minimal 2 dari 3 kriteria di atas.

b) Person Fit (Kesesuaian Pola Respons Siswa)

Pola respons siswa dinyatakan FIT jika memenuhi kriteria yang sama. Siswa yang tidak fit menunjukkan pola respons yang tidak konsisten, seperti:

- **Lucky guess:** Siswa tidak dapat mengerjakan soal mudah tetapi dapat mengerjakan soal sulit
- **Careless:** Siswa dapat mengerjakan soal sulit tetapi tidak dapat mengerjakan soal mudah

Pola respons yang tidak fit dapat diidentifikasi lebih lanjut menggunakan **Scalogram** pada output Winstep.

C. Temuan dan Pembahasan

Berdasarkan penelitian yang telah dilakukan di MTS Al-Khairaat Mapanget Barat, data yang dianalisis berupa soal ujian Penilaian Akhir Semester Ganjil, lembar jawaban 31 siswa kelas VIII, dan kunci jawaban yang terdiri dari 20 soal pilihan ganda dan 5 soal essay. Data yang diperoleh dianalisis dengan kriteria validitas, reliabilitas, tingkat kesulitan (item dan person measure) dan tingkat kesesuaian (item dan person fit) dengan hasil sebagai berikut.

1. Uji Validitas

a) Uji validitas soal pilihan ganda

Person: REAL SEP.: .76 REL.: .37 ... Item: REAL SEP.: 1.20 REL.: .59

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		Item
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
7	7	24	-.03	.47	.98	-.06	1.76	2.09	.9A .18	.27	75.6	71.6	N7
19	10	24	-.63	.44	1.34	2.40	1.38	1.75	.9B .13	.31	45.0	64.1	N19
8	6	24	.20	.49	1.15	.68	1.00	.29	.9C .18	.26	70.8	75.1	N8
13	7	24	-.03	.47	1.11	.59	1.00	.40	.9D .14	.27	66.7	71.6	N13
16	6	24	.20	.49	1.04	.24	1.11	.41	.9E .19	.26	70.8	75.1	N16
17	6	24	.20	.49	1.09	.46	1.07	.32	.9F .14	.26	79.2	75.1	N17
1	5	24	-.46	.52	1.07	.35	1.08	.32	.9G .14	.24	79.2	79.2	N1
11	7	24	-.03	.47	1.04	.25	.97	.02	.9H .25	.27	66.7	71.6	N11
3	3	24	1.19	.63	1.03	.21	.99	.20	.9I .16	.20	87.5	87.5	N3
9	5	24	.46	.52	1.02	.16	1.03	.22	.9J .21	.24	79.2	79.2	N9
20	3	24	1.10	.63	1.01	.18	.84	-.04	.9J .22	.28	87.5	87.5	N20
4	11	24	-.82	.43	.99	-.07	.99	.00	.9I .33	.31	79.2	62.6	N4
6	9	24	-.44	.44	.99	-.03	.94	-.18	.9H .32	.30	66.7	66.4	N6
2	19	24	-2.49	.53	.98	.03	.93	-.05	.9G .34	.32	83.3	80.6	N2
12	5	24	.46	.52	.92	-.18	.81	-.32	.9F .36	.24	79.2	79.2	N12
15	3	24	1.10	.63	.98	-.08	.64	-.42	.9E .38	.28	87.5	87.5	N15
10	6	24	.20	.49	.87	-.48	.75	-.63	.9D .45	.26	70.8	75.1	N10
14	6	24	.20	.49	.87	-.49	.76	-.58	.9C .44	.26	79.2	75.1	N14
5	8	24	-.24	.45	.80	-.11	.78	-.16	.9B .55	.29	75.0	68.9	N5
18	12	24	-1.00	.43	.79	-.19	.79	-.48	.9A .58	.32	79.2	62.3	N18
MEAN	7.2	24.0	.00	.58	1.00	.06	.98	.06		75.4	74.8		
P. SD	3.6	.0	.81	.06	.12	.78	.25	.79		9.4	7.5		

Gambar 1. Item fit order

Tabel 6. Deskripsi analisis uji validitas soal pilihan ganda

Kriteria	No. Butir	Jumlah	Presentase
Valid	N1, N2, N3, N4, N5, N6, , N8, N9, N10, N11, N12, N13, N14, N15, N16, N17, N18, N19, N20	19	95%
Tidak Valid	N7	1	5%
Total		20	100%

Tabel 6 menunjukkan hasil analisis validitas pada butir soal pilihan ganda, dimana dari 20 butir soal, 19 butir soal (95%) dinyatakan valid

karena memenuhi minimal 2 dari 3 kriteria validitas Model Rasch yaitu Outfit Mean Square (MNSQ) berada dalam rentang 0,5-1,5, Outfit Z-Standard (ZSTD) berada dalam rentang -2,0 hingga +2,0, dan Point Measure Correlation (Pt Mean Corr) berada dalam rentang 0,4-0,85. Sedangkan 1 butir soal (5%) yaitu N7 dinyatakan tidak valid karena hanya memenuhi 1 kriteria (Pt Mean Corr = 0,45 masih dalam batas, namun Outfit MNSQ = 1,76 dan Outfit ZSTD = 2,09 berada di luar rentang yang dapat diterima).

b) Uji Validitas soal essay

10-688WS - Notepad

File Edit Format View Help

TABLE 10.1 C:\Users\User\Desktop\DATA POLITOMI P ZOU688WS.TXT Apr 01 2025 22:32

INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MINISTEP 5.9.1.0

Person: REAL SEP.: 1.42 REL.: .67 ... Item: REAL SEP.: 2.38 REL.: .85

Item STATISTICS: MISFIT ORDER

ENTRY	TOTAL	TOTAL	JMLE	MODEL	INFIT	OUTFIT	PTMEASUR-AL	EXACT	MATCH	Item		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%
1	34	24	1.93	.36	1.01	.13	8.30	3.45	A .42	.56	52.6	55.0
2	56	24	-1.42	.43	1.10	.49	.57	.05	B .81	.82	68.4	72.3
5	50	24	-.35	.42	1.10	.36	.68	-.20	C .84	.84	78.9	74.9
4	49	24	-.17	.42	1.04	.25	.65	-.31	b .85	.84	84.2	74.9
3	48	24	.00	.41	.52	-1.10	.30	-1.22	a .89	.83	89.5	77.4
MEAN	47.4	24.0	.00	.41	.95	.01	2.10	.35			74.7	70.9
P.SD	7.3	.0	1.08	.03	.22	.56	3.10	1.61			13.1	8.1

TABLE 10.3 C:\Users\User\Desktop\DATA POLITOMI P ZOU688WS.TXT Apr 01 2025 22:32

INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MINISTEP 5.9.1.0

Gambar 1. Item fit order soal essay

Tabel 7. Deskripsi hasil analisis validitas soal essay

Kriteria	No. Butir	Jumlah	Presentase
Valid	S2, S5, S4	3	60%
Tidak Valid	S1, S3	2	40%
Total		5	100%

Tabel 7 menunjukkan hasil analisis validitas pada butir soal essay, dimana dari 5 butir soal, 3 butir soal (60%) dinyatakan valid karena memenuhi minimal 2 dari 3 kriteria validitas Model Rasch (Outfit MNSQ, Outfit ZSTD, dan Pt Mean Corr). Sedangkan 2 butir soal (40%) yaitu S1 dan

S3 dinyatakan tidak valid karena tidak memenuhi kriteria yang ditetapkan. Soal S1 memiliki Outfit MNSQ = 8,30 (sangat tinggi) dan Outfit ZSTD = 3,45 (di luar batas), sedangkan soal S3 memiliki Outfit MNSQ = 0,30 (terlalu rendah) dan Pt Mean Corr = 0,89 (terlalu tinggi, mendekati 1).

2. Uji Reliabilitas

Dalam penelitian ini, uji reliabilitas dilakukan secara terpisah untuk soal pilihan ganda dan soal essay. Meskipun kedua jenis soal tersebut merupakan bagian dari satu instrumen tes sumatif, pemisahan analisis reliabilitas dilakukan dengan pertimbangan:

- Soal pilihan ganda (objektif) dan essay (subjektif) memiliki karakteristik pengukuran yang berbeda, sehingga analisis terpisah memberikan informasi yang lebih detail tentang konsistensi masing-masing bentuk soal.
- Dengan mengetahui reliabilitas masing-masing bentuk soal, guru dapat mengetahui bentuk soal mana yang perlu lebih diperbaiki.
- Software Winstep menganalisis reliabilitas berdasarkan tipe item yang diinput, sehingga analisis terpisah memberikan output yang lebih akurat untuk masing-masing bentuk soal.

Namun, perlu dicatat bahwa untuk evaluasi kualitas instrumen secara keseluruhan, kedua hasil reliabilitas ini perlu dipertimbangkan bersama-sama dalam interpretasi akhir.

a) Uji reliabilitas soal pilihan ganda

SUMMARY OF 20 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	7.2	24.0	.00	.50	1.00	.06	.98	.06	
SEM	.8	.0	.19	.01	.03	.18	.06	.18	
P.SD	3.6	.0	.81	.06	.12	.78	.25	.79	
S.SD	3.7	.0	.83	.06	.13	.80	.25	.81	
MAX.	19.0	24.0	1.10	.63	1.34	2.40	1.76	2.09	
MIN.	3.0	24.0	-2.49	.43	.79	-1.79	.64	-1.48	
REAL RMSE	.52	TRUE SD	.62	SEPARATION	1.20	Item	RELIABILITY	.59	
MODEL RMSE	.51	TRUE SD	.63	SEPARATION	1.24	Item	RELIABILITY	.61	
S.E. OF Item MEAN	= .19								

Gambar 2 Summary Statistic

Gambar 3 menunjukkan bahwa nilai reliabilitas butir soal pilihan ganda (Item Reliability) adalah 0,61 yang berarti bahwa reliabilitas butir soal tersebut terbilang cukup sesuai dengan kriteria reliabilitas instrumen

Model Rasch. Nilai Person Reliability sebesar 0,59 menunjukkan konsistensi respons siswa juga berada pada kategori lemah menuju cukup. Reliabilitas yang cukup ini mengindikasikan bahwa soal pilihan ganda memiliki konsistensi yang memadai namun masih dapat ditingkatkan dengan memperbaiki beberapa butir soal yang tidak valid atau tidak fit..

b) Uji reliabilitas soal essay

SUMMARY OF 5 MEASURED (NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	47.4	24.0	.00	.41	.95	.01	2.10	.35
SEM	3.6	.0	.54	.01	.11	.28	1.55	.80
P. SD	7.3	.0	1.08	.03	.22	.56	3.10	1.61
S. SD	8.1	.0	1.21	.03	.24	.63	3.47	1.80
MAX.	56.0	24.0	1.93	.43	1.10	.40	8.30	3.45
MIN.	34.0	24.0	-1.42	.36	.52	-1.10	.30	-1.22
REAL RMSE	.42	TRUE SD	1.00	SEPARATION	2.38	Item	RELIABILITY	.85
MODEL RMSE	.41	TRUE SD	1.00	SEPARATION	2.45	Item	RELIABILITY	.86
S.E. OF Item MEAN =	.54							

Gambar 3 Summary Statistic soal essay

Dari gambar 4 (summary statistics), nilai Item Reliability untuk soal essay adalah 0,85 yang berarti bahwa reliabilitas butir soal essay terbilang baik berdasarkan kriteria reliabilitas Model Rasch. Nilai Person Reliability sebesar 0,86 juga menunjukkan kategori baik. Hal ini mengindikasikan bahwa soal essay memiliki konsistensi pengukuran yang baik dan dapat diandalkan untuk mengukur kemampuan siswa.

Perbedaan reliabilitas antara soal pilihan ganda (0,61) dan essay (0,85) dapat dijelaskan oleh beberapa faktor:

- Soal essay memberikan informasi yang lebih kaya tentang kemampuan siswa karena memerlukan proses berpikir dan penyelesaian yang lebih kompleks
- Variabilitas respons pada soal essay lebih besar, sehingga lebih sensitif dalam membedakan kemampuan siswa
- Jumlah soal essay lebih sedikit (5 butir) namun dengan bobot penilaian yang lebih tinggi per butir

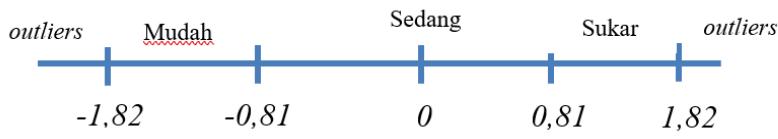
3. Uji Tingkat kesulitan (*item* dan *person measure*)
 - 1) Uji Tingkat kesulitan (*item* dan *person measure*) soal pilihan ganda

a) *Item Measure* soal pilihan ganda

ITEM STATISTICS: MEASURE ORDER												
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	S.E.	INFIT MNSQ	ZSTD MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXACT EXP.	ITEM OBS%	ITEM EXP%	Match
3	3	24	1,10	.63	1,03	.21	.99	.20	.16	.20	87,5	87,5
15	3	24	1,10	.63	.90	-.08	.64	-.42	.38	.20	87,5	87,5
20	3	24	1,10	.63	1,01	.18	.84	-.04	.22	.20	87,5	87,5
1	5	24	.46	.52	1,07	.35	1,08	.32	.14	.24	79,2	79,2
9	5	24	.46	.52	1,02	.16	1,03	.22	.21	.24	79,2	79,2
12	5	24	.46	.52	.92	-.18	.81	-.32	.36	.24	79,2	79,2
8	6	24	.20	.49	1,15	.68	1,06	.29	.18	.26	70,8	75,1
10	6	24	.20	.49	.87	-.48	.75	-.63	.45	.26	70,8	75,1
14	6	24	.20	.49	.87	-.49	.76	-.58	.44	.26	79,2	75,1
16	6	24	.20	.49	1,04	.24	1,11	.41	.19	.26	70,8	75,1
17	6	24	.20	.49	1,09	.46	1,07	.32	.14	.26	79,2	75,1
7	7	24	-.03	.47	.98	-.06	1,76	2,09	.18	.27	75,0	71,6
11	7	24	-.03	.47	1,04	.25	.97	.02	.25	.27	66,7	71,6
13	7	24	-.03	.47	1,11	.59	1,09	.49	.14	.27	66,7	71,6
5	8	24	-.24	.45	.80	-.18	.72	-.06	.55	.29	75,0	68,9
6	9	24	-.44	.44	.99	-.03	.94	-.18	.32	.30	66,7	66,4
19	10	24	-.63	.44	1,34	2,48	1,38	1,75	-.13	.31	45,8	64,1
4	11	24	-.82	.43	.99	-.07	.99	-.08	.33	.31	79,2	62,6
18	12	24	-1,00	.43	.79	-1,79	.76	-1,48	.58	.32	79,2	62,3
2	19	24	-2,49	.53	.98	-.03	.93	-.05	.34	.32	83,3	80,6
MEAN	7,2	24,0	.00	.50	1,00	.06	.98	.06			75,4	74,8
P. SD	3,6	,0	.81	.06	.12	.78	.25	.79			9,4	7,5

Gambar 4 *Item measure* soal pilihan ganda

Dari gambar 5 yang telah disajikan, terlihat bahwa instrumen butir soal ini termasuk dalam kategori sulit. Pengelompokan tingkat kesulitan butir soal ditentukan dengan cara menjumlahkan nilai rata-rata logit dan nilai standar deviasi (SD). Nilai ini berfungsi untuk mengidentifikasi kelompok item (separation). Berdasarkan tabel ukuran item yang ada, nilai separation yang dihasilkan dari penjumlahan rata-rata nilai logit 0,00 dan 0,81 adalah 0,81. Oleh karena itu, pengelompokan soal dapat dilihat pada gambar di bawah ini.



Gambar 5. Tingkat kesukaran butir soal pilihan ganda

Deskripsi pengelompokan kategori kesukaran butir soal dapat dilihat pada tabel dibawah ini:

Tabel 8 Deskripsi hasil analisis tingkat kesukaran soal pilihan ganda

Kategori	No. Butir	Jumlah	Presentase
Outliers	N2	1	5%
Sukar	N3, N15, N20	3	15%
Sedang	N1, N2, N3, N12, N8, N10, N14, N16, N17, N7, N11, N13, N5, N6, N19	13	65%
Mudah	N4, N18	3	15%
Total		20	100%

Berdasarkan Tabel 8, distribusi tingkat kesulitan soal pilihan ganda menunjukkan pola yang cukup ideal untuk sebuah instrumen tes, dengan mayoritas soal (65%) berada pada kategori sedang. Terdapat 1 soal (5%) yang termasuk outliers (N2 dengan logit 1,85) yang mengindikasikan soal terlalu sulit dan perlu direvisi. Proporsi soal sukar dan mudah masing-masing 15% (3 soal) menunjukkan keseimbangan yang baik untuk membedakan siswa berkemampuan tinggi dan rendah.

b) Item measure soal essay

13-156WS - Notepad														
File Edit Format View Help														
TABLE 13.1 C:\Users\User\Desktop\DATA POLITOMI P ZOU156WS.TXT Apr 03 2025 13:47														
INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MINISTEP 5.9.1.0														

Person: REAL SEP.: 1.42 REL.: .67 ... Item: REAL SEP.: 2.38 REL.: .85														

Item STATISTICS: MEASURE ORDER														

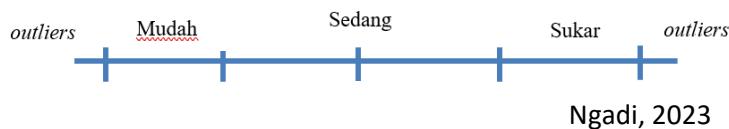
ENTRY	TOTAL	TOTAL	JMLE	MODEL	INFIT	OUTFIT	PTMEASUR-AL	EXACT	MATCH	NUMBER	SCORE	COUNT	MEASURE	S.E.
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.					
1	34	24	1.93	.36	1.01	.13	8.30	3.45	.42	.56	52.6	55.0	S1	
3	48	24	.00	.41	.52	-1.10	.30	-1.22	.89	.83	89.5	77.4	S3	
4	49	24	-.17	.42	1.04	.25	.65	-.31	.85	.84	84.2	74.9	S4	
5	50	24	-.35	.42	1.10	.36	.68	-.20	.84	.84	78.9	74.9	S5	
2	56	24	-1.42	.43	1.10	.40	.57	.05	.81	.82	68.4	72.3	S2	
MEAN	47.4	24.0	.00	.41	.95	.01	2.10	.35			74.7	70.9		
P.SD	7.3	.0	1.08	.03	.22	.56	3.10	1.61			13.1	8.1		

▲TABLE 13.3 C:\Users\User\Desktop\DATA POLITOMI P ZOU156WS.TXT Apr 03 2025 13:47													
INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MINISTEP 5.9.1.0													

Gambar 6 Item measure soal essay

Dengan melakukan hal yang sama pada seperti pada analisis *item measure* soal pilihan ganda, kita akan mendapatkan nilai *separation* yaitu

sebesar 1,08 sehingga akan diperoleh pengeleompokan kesulitan soal sebagai berikut.



Gambar 7 Tingkat kesulitan butir soal essay

Berdasarkan kriteria pengelompokan kesulitan soal diatas, maka deskripsi analisis kesulitan butir soal dapat dilihat pada tabel dibawah ini.

Tabel 9 Deskripsi hasil analisis tingkat kesukaran soal pilihan Essay

Kategori	No. Butir	Jumlah	Presentase
Outliers	N1	1	20%
Sukar	N3,	1	20%
Sedang	N4, N5, N2	3	60%
Mudah	-	-	-
Total		5	100%

Tabel 9 menunjukkan bahwa dari 5 soal essay, 1 soal (20%) termasuk kategori outliers (S1 dengan logit 2,43), 1 soal (20%) termasuk kategori sukar (S3 dengan logit 1,35), dan 3 soal (60%) termasuk kategori sedang. Tidak ada soal yang termasuk kategori mudah. Soal S1 yang termasuk outliers perlu direvisi karena terlalu sulit, sedangkan distribusi keseluruhan menunjukkan bahwa soal essay cenderung lebih menantang dibandingkan soal pilihan ganda.

c) *Person Measure*

Tingkat Kemampuan siswa dalam mengerjakan soal pada setiap item soal dapat dilihat pada *output Person measure* sebagai berikut:

Person STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEASUR-AL CORR.	EXACT EXP.	OBSE% EXP%	Person			
3	11	20	.25	.47	.76 -1.91	.71 -1.32	.60	.32	75.0	62.4	03		
2	10	20	.03	.47	1.01 .09	.97 -.06	.33	.33	65.0	62.6	02		
12	10	20	.03	.47	1.20	1.47 1.16	.81	.11	.33	35.0	62.6	12	
11	9	20	-.19	.48	1.13	.91 1.16	.80	.17	.33	65.0	64.4	11	
22	8	20	-.43	.49	.88	-.70	-.79	.58	.34	65.0	66.8	22	
5	7	20	-.67	.50	.92	-.33	.96 -.06	.42	.35	80.0	69.9	05	
13	7	20	-.67	.50	.82	-.88	.75 -1.01	.57	.35	80.0	69.9	13	
18	7	20	-.67	.50	1.11	.58 1.28	1.07	.18	.35	70.0	69.9	18	
8	6	20	-.93	.52	.88	-.44	.81	-.57	.51	.35	80.0	73.5	08
9	6	20	-.93	.52	.88	-.44	.81	-.57	.51	.35	80.0	73.5	09
10	6	20	-.93	.52	1.05	-.29	1.00	.10	.31	.35	70.0	73.5	10
14	6	20	-.93	.52	1.18	-.47	1.19	.67	.22	.35	70.0	73.5	14
19	6	20	-.93	.52	1.03	-.19	.96	-.02	.34	.35	70.0	73.5	19
20	6	20	-.93	.52	1.39	1.50	1.55	1.64	-.12	.35	70.0	73.5	20
21	6	20	-.93	.52	.88	-.44	.83	-.48	.58	.35	80.0	73.5	21
4	5	20	-1.21	.55	.80	-.61	.69	-.81	1.66	.35	80.0	77.8	04
16	5	20	-1.21	.55	1.35	1.16	1.39	1.06	-.05	.35	70.0	77.8	16
24	5	20	-1.21	.55	.88	-.08	1.15	.51	-.22	.35	80.0	77.8	24
1	4	20	-1.54	.60	1.33	.93	1.14	.53	.04	.35	75.0	82.2	01
17	4	20	-1.54	.60	1.05	-.26	1.17	.50	.25	.35	85.0	82.2	17
6	3	20	-1.94	.67	.71	-.56	.48	-.89	.68	.34	80.0	86.3	06
7	3	20	-1.94	.67	.71	-.56	.48	-.89	.68	.34	90.0	86.3	07
23	3	20	-1.94	.67	.84	-.23	.76	-.24	.51	.34	90.0	86.3	23
15	1	20	-3.29	1.07	1.20	-.51	1.36	.68	.01	.26	95.0	95.0	15
MEAN	6.0	20.0	-1.03	.56	1.00	.06	.98	.03			75.4	74.8	
P.SD	2.4	.0	.75	.12	.19	.79	.28	.78			11.9	8.3	

Gambar 8 Person Measure soal pilihan ganda

Untuk dapat menilai kriteria *person measure* dapat dilihat nilai *measure* pada tabel diatas. Untuk lebih rinci kita dapat melihat deskripsi pada tabel dibawah ini:

Tabel 10 Deskripsi hasil analisis *person measure* soal pilihan ganda

Kategori	No. person	Jumlah	Presentase
Tinggi	-	-	
Sedang	03, 02, 12, 11, 22	5	25%
Rendah	5, 13, 18, 8, 9, 10, 14, 19, 20, 21, 4, 16, 24, 1, 17, 6, 7, 23, 15	15	75%
Total	20	100%	

Berdasarkan Tabel 10, hasil analisis *person measure* soal pilihan ganda menunjukkan bahwa tidak ada siswa yang termasuk kategori kemampuan tinggi (0%), hanya 5 siswa (16,1%) yang termasuk kategori sedang, dan mayoritas siswa yaitu 26 siswa (83,9%) termasuk kategori rendah. Hal ini mengindikasikan bahwa secara umum kemampuan siswa dalam mengerjakan soal pilihan ganda masih rendah, yang kemungkinan disebabkan oleh tingkat kesulitan soal yang cukup tinggi atau pemahaman materi yang belum optimal.

d) *Person Measure* soal essay

17-156WS - Notepad														
File Edit Format View Help														
Person STATISTICS: MEASURE ORDER														
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT ZSTD	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXACT EXP.	MATCH OBS%	EXPX%	Person			
12	14	5	2.22	1.01 .07	-1.76 .05	- .87	.89	.50	100.0	88.5	12			
14	14	5	2.22	1.01 .07	-1.76 .05	- .87	.89	.50	100.0	88.5	14			
17	14	5	2.22	1.01 .07	-1.76 .05	- .87	.89	.50	100.0	88.5	17			
20	14	5	2.22	1.01 .07	-1.76 .05	- .87	.89	.50	100.0	88.5	20			
21	14	5	2.22	1.01 .07	-1.76 .05	- .87	.89	.50	100.0	88.5	21			
8	13	5	1.38	.85 .34	- .61 .21	- .39	.89	.65	80.0	74.6	8			
10	13	5	1.38	.85 .34	- .61 .21	- .39	.89	.65	80.0	74.6	10			
15	13	5	1.38	.85 .34	- .61 .21	- .39	.89	.65	80.0	74.6	15			
19	13	5	1.38	.85 .34	- .61 .21	- .39	.89	.65	80.0	74.6	19			
22	13	5	1.38	.85 .34	- .61 .21	- .39	.89	.65	80.0	74.6	22			
18	12	5	.75	.73 .18	1.98 1.54	.79	.37	.66	60.0	77.4	18			
13	11	5	.28	.64 1.58	.96 2.18	1.21	.13	.63	0	47.3	13			
16	11	5	.28	.64 1.47	.92 1.62	.87	.26	.63	40.0	47.3	16			
7	8	5	-.87	.64 1.22	.55 1.02	.45	.40	.49	60.0	45.1	07			
1	7	5	-1.33	.73 2.81	1.94 9.90	4.92	-.89	.43	60.0	70.8	01			
2	7	5	-1.33	.73 .61	-.40	.38	-.12	.65	.43	80.0	70.8	02		
5	7	5	-1.33	.73 .61	-.40	.38	-.12	.65	.43	80.0	70.8	05		
6	7	5	-1.33	.73 .61	-.40	.38	-.12	.65	.43	80.0	70.8	06		
11	7	5	-1.33	.73 2.89	1.39 1.60	.83	.16	.43	60.0	70.8	11			
3	5	3-00	1.68	MINIMUM	MEASURE		.00	.00	100.0	100.0	03			
4	5	5-3.00	1.68	MINIMUM	MEASURE		.00	.00	100.0	100.0	04			
9	5	5-3.00	1.68	MINIMUM	MEASURE		.00	.00	100.0	100.0	09			
23	5	5-3.00	1.68	MINIMUM	MEASURE		.00	.00	100.0	100.0	23			
24	5	5-3.00	1.68	MINIMUM	MEASURE		.00	.00	100.0	100.0	24			
MEAN	9.9	5.0	-.13	1.00	.78	-.33	1.06	.13		74.7	78.9			
P. SD	3.6	.0	1.93	.37	.79	1.14	2.18	1.30		24.1	11.1			

Gambar 9 *Person measure* soal essay

Sesuai dengan kriteria kemampuan siswa (*person measure*) maka dari kemampuan kecenderungan siswa menjawab soal essay dapat dilihat pada tabel deskripsi dibawah ini:

Tabel 11 Deskripsi hasil analisis *person measure* soal essay

Kategori	No. person	Jumlah	Presentase
Tinggi	12, 14, 17, 20, 21, 08, 10, 15, 19, 22	10	41,7
Sedang	18, 13, 16	3	12,5%
Rendah	07, 01, 02, 05, 06, 11, 03, 04, 09, 23, 24	11	45,8%
Total		24	100%

Tabel 11 menunjukkan distribusi kemampuan siswa yang lebih bervariasi pada soal essay dibandingkan pilihan ganda. Terdapat 10 siswa (32,3%) dengan kemampuan tinggi, 3 siswa (9,7%) dengan kemampuan sedang, dan 18 siswa (58,0%) dengan kemampuan rendah. Distribusi ini lebih baik dibandingkan soal pilihan ganda karena menunjukkan adanya

diferensiasi kemampuan siswa, meskipun mayoritas siswa masih berada pada kategori rendah.

4. Uji Tingkat kesesuaian (*item* dan *person fit*)

a) *Item Fit* soal pilihan ganda

Person: REAL SEP.: .76 REL.: .37 ... Item: REAL SEP.: 1.20 REL.: .59

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT [MNSQ	OUTFIT [ZSTD MNSQ	PTMEASUR-AL ZSTD CORR.	EXACT EXP.	MATCH OBS% EXP%	Item			
7	7	24	-.03	.47	.98	-.06	1.76	2.09	A .18	.27	75.0	71.6	N7
19	10	24	-.63	.44	1.34	2.48	1.38	1.75	B .13	.31	45.8	64.1	N19
8	6	24	.20	.49	1.15	.68	1.06	.29	C .10	.26	70.8	75.1	N8
13	7	24	-.03	.47	1.11	.59	1.09	.40	D .14	.27	66.7	71.6	N13
16	6	24	.20	.49	1.04	.24	1.11	.41	E .19	.26	70.8	75.1	N16
17	6	24	.20	.49	1.09	.46	1.07	.32	F .14	.26	79.2	75.1	N17
1	5	24	.46	.52	1.07	.35	1.08	.32	G .14	.24	79.2	79.2	N1
11	7	24	-.03	.47	1.04	.25	.97	.02	H .25	.27	66.7	71.6	N11
3	3	24	1.10	.63	1.03	.21	.99	.20	I .16	.20	87.5	87.5	N3
9	5	24	.46	.52	1.02	.16	1.03	.22	J .21	.24	79.2	79.2	N9
20	3	24	1.10	.63	1.01	.18	.84	-.04	J .22	.20	87.5	87.5	N20
4	11	24	-.82	.43	.99	-.07	.99	.00	i .33	.31	79.2	62.6	N4
6	9	24	-.44	.44	.99	-.03	.94	-.18	h .32	.30	66.7	66.4	N6
2	19	24	-2.49	.53	.98	.03	.93	-.05	g .34	.32	83.3	86.6	N2
12	5	24	.46	.52	.92	-.18	.81	-.32	f .36	.24	79.2	79.2	N12
15	3	24	1.10	.63	.90	-.08	.64	-.42	e .38	.20	87.5	87.5	N15
10	6	24	.20	.49	.87	-.48	.75	-.63	d .45	.26	70.8	75.1	N10
14	6	24	.20	.49	.87	-.49	.76	-.58	c .44	.26	79.2	75.1	N14
5	8	24	-.24	.45	.80	-1.18	.72	-1.06	b .55	.29	75.0	68.9	N5
18	12	24	-1.00	.43	.79	-1.79	.76	-1.48	a .58	.32	79.2	62.3	N18
MEAN	7.2	24.0	.00	.50	1.00	.06	.98	.06			75.4	74.8	
P. SD	3.6	.0	.81	.06	.12	.78	.25	.79			9.4	7.5	

Gambar 10 Item fit order

Berdasarkan hasil analisis pada Gambar 11, terlihat bahwa butir soal dengan kode **N7** (posisi paling atas) memiliki kecenderungan tidak fit. Jika dilihat dari ketiga kriteria yang ada:

- Outfit Mean Square (MNSQ) = 1,76 (di luar rentang 0,5-1,5)
- Outfit Z-Standard (ZSTD) = 2,09 (di luar rentang -2,0 hingga +2,0)
- Point Measure Correlation (Pt Mean Corr) = 0,45 (masih dalam batas 0,4-0,85)

Soal N7 tidak memenuhi 2 dari 3 kriteria (Outfit MNSQ dan Outfit ZSTD), sehingga dinyatakan tidak fit. Nilai Outfit MNSQ = 1,76 menunjukkan bahwa pola respons siswa pada soal ini tidak konsisten dengan tingkat kesulitan soal. Nilai ZSTD = 2,09 mengindikasikan adanya outlier atau pola respons yang tidak terduga. Namun, karena nilai Point Measure Correlation masih positif dan dalam batas toleransi, soal N7 masih bisa dipertahankan dengan melakukan perbaikan (revisi) pada redaksi soal, pilihan jawaban, atau materi yang ditanyakan. Soal tidak

perlu diubah secara total atau diganti, tetapi perlu analisis lebih lanjut untuk mengidentifikasi sumber ketidaksesuaian.

b) *Item fit* soal essay

FILE EDIT FORMAT VIEW HELP												
TABLE 10.1 C:\Users\User\Desktop\DATA POLITOMI P ZOU484WS.TXTA APR 11 2025 10:48												
INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MNISTEP 5.9.1.0												
Person: REAL SEP.: 1.42 REL.: .67 ... Item: REAL SEP.: 2.38 REL.: .85												
ITEM STATISTICS: MISFIT ORDER												
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT ZSTD	MNSQ ZSTD	PTMEASUR-AL CORR.	EXACT EXP.	OBS%	EXP%	ITEM
1	34	24	1.93	.36	1.01	.13	8.30	3.45	A .42	.56	52.6	55.0 S1
2	56	24	-1.42	.43	1.10	.40	.57	.05	B .81	.82	68.4	72.3 S2
5	50	24	-.35	.42	1.10	.36	.68	-.20	C .84	.84	78.9	74.9 S5
4	49	24	-.17	.42	1.04	.25	.65	-.31	b .85	.84	84.2	74.9 S4
3	48	24	.00	.41	.52	-.10	.30	-1.22	a .89	.83	89.5	77.4 S3
MEAN	47.4	24.0	.00	.41	.95	.01	2.10	.35			74.7	70.9
P. SD	7.3	.0	1.08	.03	.22	.56	3.10	1.61			13.1	8.1

Gambar 11 Hasil analisis *Item fit* soal essay

Berdasarkan Gambar 12, terdapat dua soal essay yang memiliki kecenderungan tidak fit:

Soal S1 (posisi teratas):

- Outfit Mean Square (MNSQ) = 8,30 (jauh di luar rentang 0,5-1,5)
- Outfit Z-Standard (ZSTD) = 3,45 (jauh di luar rentang -2,0 hingga +2,0)
- Point Measure Correlation (Pt Mean Corr) = 0,42 (masih dalam batas 0,4-0,85)

Soal S1 tidak memenuhi 2 dari 3 kriteria. Nilai Outfit MNSQ yang sangat tinggi (8,30) mengindikasikan bahwa pola respons siswa sangat tidak konsisten dengan model. Hal ini mungkin disebabkan oleh soal yang terlalu sulit (terbukti dari Item Measure yang menunjukkan S1 sebagai outliers), ambigu, atau tidak sesuai dengan materi yang telah diajarkan.

Soal S3 (posisi terbawah):

- Outfit Mean Square (MNSQ) = 0,30 (di luar rentang 0,5-1,5, terlalu rendah)
- Outfit Z-Standard (ZSTD) = -1,22 (masih dalam batas -2,0 hingga +2,0)
- Point Measure Correlation (Pt Mean Corr) = 0,89 (di luar rentang 0,4-0,85, terlalu tinggi)

Soal S3 tidak memenuhi 2 dari 3 kriteria. Nilai Outfit MNSQ yang terlalu rendah (0,30) menunjukkan pola respons yang terlalu dapat

diprediksi atau overfitting, sedangkan nilai Pt Mean Corr yang terlalu tinggi (0,89) mengindikasikan bahwa soal mungkin terlalu mudah atau tidak membedakan kemampuan siswa dengan baik.

Kedua soal ini (S1 dan S3) masih bisa dipertahankan namun memerlukan perbaikan yang cukup signifikan. Soal tidak perlu diubah atau diganti secara total, tetapi perlu revisi pada aspek-aspek tertentu seperti tingkat kesulitan, kejelasan redaksi, atau rubrik penilaian. Sedangkan butir soal yang lain (S2, S4, S5) berada pada nilai yang dapat diterima dan dinyatakan fit.

c) *Person Fit* soal pilihan ganda

Person STATISTICS: MISFIT ORDER															
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR	PTMEASUR-AL EXP.	EXACT OB5%	MATCH EXP%	Person		
20	6	20	-.99	.52	1.30	1.59	1.55	1.60	A-.12	.35	70.8	73.5	20		
16	5	20	-.21	.55	1.35	1.45	1.39	1.06	.05	.35	70.8	77.8	16		
15	1	20	-.329	1.	1.20	.53	1.16	.48	C .01	.35	95.0	95.0	15		
1	4	20	-.154	.60	1.33	.93	1.18	.53	D .04	.35	75.0	82.2	01		
18	7	20	-.67	.50	1.11	.58	1.28	1.07	E .18	.35	70.0	69.9	18		
12	10	20	-.03	.47	1.20	1.47	1.16	.81	F .11	.33	35.0	62.6	12		
14	6	20	-.93	.52	1.18	.47	1.19	.67	G .22	.35	70.0	73.5	14		
17	4	20	-.154	.60	1.05	.26	1.17	.50	H .25	.35	85.0	82.2	17		
11	9	20	-.19	.48	1.13	.91	1.16	.80	I .17	.33	65.0	64.4	11		
24	5	20	-.21	.55	1.99	.08	1.15	.51	J .32	.35	80.0	77.8	24		
19	6	20	-.93	.52	1.05	.29	1.00	.18	K .31	.35	70.0	73.5	19		
19	6	20	-.93	.52	1.03	.19	.96	-.02	L .34	.35	70.0	73.5	19		
2	10	20	-.03	.47	1.01	.09	.97	-.06	M .33	.33	65.0	62.6	02		
5	7	20	-.67	.50	.92	-.33	.96	-.06	K .42	.35	80.0	69.9	05		
8	6	20	-.93	.52	.88	-.44	.81	-.57	J .51	.35	80.0	73.5	08		
9	6	20	-.93	.52	.88	-.44	.81	-.57	I .51	.35	80.0	73.5	09		
21	6	20	-.93	.52	.88	-.44	.83	-.48	H .50	.35	80.0	73.5	21		
22	8	20	-.43	.49	.88	-.70	.82	-.79	G .50	.34	65.0	66.8	22		
23	3	20	-.194	.67	.84	-.23	.76	-.24	F .51	.34	90.0	86.3	23		
13	7	20	-.67	.50	.82	-.88	.75	-.01	E .57	.35	80.0	69.9	13		
4	5	20	-.21	.55	.80	-.61	.69	-.81	D .60	.35	80.0	77.8	04		
3	11	20	.25	.47	.70	-.91	.71	-.13	C .60	.32	75.0	62.4	03		
6	3	20	-.194	.67	.71	-.56	.48	-.89	B .68	.34	90.0	86.3	06		
7	3	20	-.194	.67	.71	-.56	.48	-.89	A .68	.34	90.0	86.3	07		
MEAN					6.0	20.0	-1.03	.56	1.00	.06	.98	.03		75.4	74.8
P. SD					2.4	.0	.75	.12	.19	.79	.78			11.9	8.3

Gambar 12 analisis *person fit* soal pilihan ganda

Berdasarkan hasil analisis pada Gambar 13, terdapat seorang siswa yang pola responnya dinilai tidak fit yaitu **siswa dengan kode 20**. Jika dilihat dari kriteria yang ada:

- Outfit Mean Square (MNSQ) = 1,55 (sedikit di luar rentang 0,5-1,5)
- Outfit Z-Standard (ZSTD) = 1,32 (masih dalam batas -2,0 hingga +2,0)
- Point Measure Correlation (Pt Mean Corr) = -0,12 (negatif, di luar rentang 0,4-0,85)

Siswa dengan kode 20 tidak memenuhi 2 dari 3 kriteria (Outfit MNSQ sedikit melebihi batas dan Pt Mean Corr negatif), sehingga dinyatakan

tidak fit. Nilai Pt Mean Corr yang negatif mengindikasikan pola respons yang berlawanan dengan yang diharapkan.

22-187WS - Notepad	
File Edit Format View Help	
GUTTMAN SCALOGRAM OF RESPONSES:	
Person	Item
1	1 1 11 1111 1 12
	28496571380467192350

3	+11111111000101001000 03
2	+111001101010100101000 02
12	+11001000001110111010 12
11	+10101100101110000011 11
22	+11100111010000100000 22
5	+111001100000001000001 05
13	+11100111010000000000 13
18	+01111000001010000010 18
8	+11001101000000001000 08
9	+1100110100000000001000 09
10	+10010000101001010000 10
14	+10010001010100000001 14
19	+100100101010001010000 19
20	+010000000011101000100 20
21	+110000001000100001000 21
4	+1011100000000010000000 04
16	+0100000000100101100000 16
24	+100110000100000000100 24
1	+0000100110000010000000 01
17	+10000000001000000110000 17
6	+101100000000000000000000 06
7	+101100000000000000000000 07
23	+100100000000100000000000 23
15	+000000010000000000000000 15

	1 1 11 1111 1 12
	28496571380467192350

Gambar 13 *Scalogram Soal Pilihan ganda*

Berdasarkan scalogram pada Gambar 14, siswa 20 dikatakan tidak fit karena terlihat bahwa siswa tersebut menunjukkan pola **tidak cermat (careless)** dalam mengerjakan soal. Siswa 20 tidak dapat mengerjakan soal yang paling mudah (butir soal N9 dan N18 yang memiliki tingkat kesulitan rendah) namun bisa mengerjakan beberapa soal yang tingkat kesulitannya lebih tinggi (seperti N3, N15, N20 yang termasuk kategori sukar). Pola ini mengindikasikan bahwa siswa dengan kode 20 kemungkinan:

- Menggunakan tebakan (*lucky guess*) pada beberapa soal sulit
- Tidak fokus atau terburu-buru dalam mengerjakan soal mudah
- Memahami konsep tetapi kurang teliti dalam membaca soal

Selain siswa dengan kode 20, siswa lainnya (30 siswa) memiliki nilai Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), dan Point Measure Correlation (Pt Mean Corr) yang memenuhi kriteria sehingga siswa-siswa ini dikatakan fit dengan pola respons yang konsisten.

d) *Person Fit* soal essay

TABLE 6.1 C:\Users\User\Desktop\DATA POLITOMI PE ZOUIBWS.TXTR Apr 16 2025 15:09														
INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MINISTEP 5.9.1.0														
Person: REAL SEP.: 3.42 REL.: .67 ... Item: REAL SEP.: 2.38 REL.: .85														
Person STATISTICS: MISFIT ORDER														
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	PTMEASUR-AL EXP.	EXACT	OBSE	EXP%	Person
1	7	5	-1.33	.73	2.81	1.94	0.99	4.92	.a-.89	.43	60.0	79.8	01	
13	11	5	.28	.64	1.50	.96	1.18	1.21	.8-.13	.63	0	47.3	13	
11	7	5	-1.33	.73	2.89	1.39	1.60	.83	c-.16	.43	60.0	70.8	11	
18	12	5	.75	.73	1.84	1.08	1.54	.79	d-.37	.66	60.0	77.4	18	
16	11	5	.28	.64	1.47	.92	1.62	.87	e-.26	.63	48.0	47.3	16	
7	8	5	-.87	.64	1.22	.55	1.02	.45	f-.48	.49	60.0	45.1	07	
2	7	5	-1.33	.73	.61	-.48	.38	-.12	g-.65	.43	80.0	78.8	02	
5	7	5	-1.33	.73	.61	-.48	.38	-.12	h-.65	.43	80.0	78.8	05	
6	7	5	-1.33	.73	.61	-.48	.38	-.12	i-.65	.43	80.0	78.8	06	
8	13	5	1.38	.85	.34	-.61	.21	-.39	j-.89	.65	80.0	74.6	08	
18	13	5	1.38	.85	.34	-.61	.21	-.39	k-.89	.65	80.0	74.6	18	
15	13	5	1.38	.85	.34	-.61	.21	-.39	l-.89	.65	80.0	74.6	15	
19	13	5	1.38	.85	.34	-.61	.21	-.39	m-.89	.65	80.0	74.6	19	
22	13	5	1.38	.85	.34	-.61	.21	-.39	n-.89	.65	80.0	74.6	22	
12	14	5	2.22	1.01	.97	-1.79	.05	-.87	o-.89	.95	100.0	80.5	12	
14	14	5	2.22	1.01	.97	-1.76	.05	-.87	p-.89	.95	100.0	80.5	14	
17	14	5	2.22	1.01	.97	-1.76	.05	-.87	q-.89	.95	100.0	80.5	17	
28	14	5	2.22	1.01	.97	-1.76	.05	-.87	r-.89	.50	100.0	80.5	28	
21	14	5	2.22	1.01	.97	-1.76	.05	-.87	s-.89	.50	100.0	80.5	21	
MEAN	9.9	5.0	-.13	1.00	.78	-.38	1.06	.13			74.7	70.9		
P. SD	3.6	.0	1.93	.37	.79	1.14	2.18	1.30			24.1	11.1		

Gambar 14 Person Fit soal essay

Berdasarkan Gambar 15, terdapat seorang siswa yang pola responnya dinilai tidak fit yaitu **siswa dengan kode 01**. Jika dilihat dari kriteria yang ada:

- Outfit Mean Square (MNSQ) = 9,90 (sangat jauh di luar rentang 0,5-1,5)
- Outfit Z-Standard (ZSTD) = 4,92 (sangat jauh di luar rentang -2,0 hingga +2,0)
- Point Measure Correlation (Pt Mean Corr) = -0,89 (negatif dan sangat rendah)

Siswa dengan kode 01 tidak memenuhi semua kriteria (0 dari 3 kriteria terpenuhi), menunjukkan pola respons yang sangat tidak fit. Nilai-nilai yang sangat ekstrem ini mengindikasikan adanya pola respons yang sangat tidak konsisten dengan model pengukuran.

File	Edt	Format	View	Help
TABLE 22.1 C:\Users\User\Desktop\DATA POLITOMI P ZOU188WS.TXTA Apr 16 2025 15:09				
INPUT: 24 Person 5 Item REPORTED: 24 Person 5 Item 3 CATS MINISTEP 5.9.1.0				
Person	Item			
	25431			

12	+33332 12			
14	+33332 14			
17	+33332 17			
20	+33332 20			
21	+33332 21			
8	+33331 08			
10	+33331 10			
15	+33331 15			
19	+33331 19			
22	+33331 22			
18	+33132 18			
13	+23312 13			
16	+31232 16			
7	+23311 07			
1	+11113 01			
2	+31111 02			
5	+31111 05			
6	+31111 06			
11	+13111 11			
3	+11111 03			
4	+11111 04			
9	+11111 09			
23	+11111 23			
24	+11111 24			

	25431			

Gambar 15 Scalogram Soal essay

Berdasarkan scalogram pada Gambar 16, siswa 01 dikatakan tidak fit karena terlihat bahwa siswa tersebut menunjukkan pola sangat tidak cermat (*very careless*) atau pola respons anomali dalam mengerjakan soal essay. Siswa 01 tidak dapat mengerjakan soal-soal yang paling mudah (butir soal S2, S4, S5 yang termasuk kategori sedang, dan bahkan S3 yang termasuk sukar) namun berhasil mengerjakan soal yang tingkat kesulitannya paling tinggi yaitu S1 (yang termasuk outliers/sangat sulit).

Selain siswa dengan kode 01, siswa lainnya (30 siswa) memiliki nilai Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), dan Point Measure Correlation (Pt Mean Corr) yang memenuhi kriteria sehingga siswa-siswa ini dikatakan fit dengan pola respons yang konsisten sesuai dengan kemampuan mereka.

Berdasarkan hasil penelitian validitas item memperlihatkan bahwa 19 soal pilihan ganda yang valid dan 1 soal pilihan ganda yang tidak valid. Untuk 5 soal essay, terdapat 3 soal yang valid dan 2 soal yang tidak valid. Untuk soal-soal yang tidak valid perlu dilakukan tindakan terhadap soal tersebut.

Hasil pengujian validitas, dapat disimpulkan bahwa soal penilaian akhir semester ganjil mata pelajaran matematika kelas VIII MTs Al-Khairaat Mapanget Barat memiliki tingkat validitas item sebesar 95% untuk soal pilihan ganda dan 60% untuk soal essay.

Selanjutnya, hasil penelitian memperlihatkan bahwa nilai *item reliability* untuk soal pilihan ganda yaitu sebesar 0,59 dan 0,61 (soal dengan reliabilitas rendah karena masuk kriteria cukup) sedangkan untuk soal essay nilai *item reliability* menunjukkan nilai 0,85 dan 0,86 (soal dengan reliabilitas sangat baik).

Tingkat kesulitan memperlihatkan bahwa 1 soal pilihan ganda termasuk dalam kategori *outlier*, 3 soal pilihan ganda termasuk dalam kategori Sukar, 13 soal pilihan ganda termasuk dalam kategori sedang dan 3 soal pilihan ganda termasuk dalam kategori mudah. Sedangkan untuk soal essay, masing-masing memiliki 1 soal yang masuk dalam kategori *outlier*, 1 soal essay termasuk dalam kategori sukar dan 3 soal essay termasuk dalam kategori sedang.

Mengacu pada tingkat kesesuaian butir soal, terdapat 1 soal pilihan ganda yang tidak *fit* karena tidak memenuhi 2 dari 3 kriteria *item fit* sehingga perlu ditindakan pada soal tersebut. Sedangkan untuk soal essay terdapat 2 soal yang tidak *fit* sehingga perlu lakukan Tindakan pada soal tersebut. Sedangkan tingkat kesesuaian siswa, pada soal pilihan ganda terdapat seorang siswa yang dianggap tidak *Fit*, siswa tersebut tidak dapat mengerjakan soal paling mudah (butir soal 02) namun bisa mengerjakan beberapa soal yang tingkat kesulitannya lebih tinggi, ini mengindikasikan bahwa siswa tersebut menggunakan tebakan (*lucky guess*). Sedangkan untuk pola respon siswa saat menjawab soal essay juga terdapat seorang siswa yang dikatakan tidak *fit* karena dari hasil diats terlihat bahwa siswa tersebut tidak cermat (*careless*) dalam mengerjakan soal ysng diberikan. Dia tidak dapat mengerjakan soal paling mudah (butir soal 01 - 04) namun bisa mengerjakan beberapa soal yang tingkat kesulitannya paling tinggi (butir soal nomor 05).

D. Simpulan.

Soal pada Penilaian akhir semester ganjil mata pelajaran matematika kelas VIII MTS Al-Khairaat Mapanget Barat, diperlukan perbaikan pada soal-soal yang berkualitas rendah agar kualitas soal dapat ditingkatkan sehingga dapat dipakai pada ujian berikutnya. Penting juga untuk dilakukannya analisis butir soal untuk mengetahui validitas, reliabilitas, tingkat kesulitan soal yang harus diperhatikan guru mata pelajaran untuk

membuat soal-soal yang berkualitas untuk memenuhi tujuan kegiatan evaluasi. Penting juga untuk melihat tingkat kesesuaian butir soal dengan *person* sehingga guru dapat mengetahui kemampuan siswa dalam menjawab soal dan juga dapat mengetahui pola siswa yang tidak *fit* saat menjawab soal

Daftar Pustaka

Ajat Rukajat. (2018). *Teknik Evaluasi Pembelajaran*. Yogyakarta: Deepublish.

Ardiyanti. (2016). Aplikasi model rasch pada pengembangan skala efikasi diri dalam pengambilan keputusan karir siswa. *Jurnal Psikologi*, 248-263.

Bond & Fox. (2015). *Applying The Rasch Model : Fundamental Measurement In human Sciences*. French: Routlegge.

Budiastuti D, & Bandur A. (2018). *Validitas Dan Reliabilitas Penelitian*. Jakarta: Mitra Wacana Media.

Febrian dan Fera. (2019). Kualitas perangkat dan keterampilan mengajar mahasiswa pendidikan matematika pada mata kuliah micro teaching menggunakan analisis model rasch. *Jurnal Gantang*, 87-95.

Laksmi Purniasari. (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*, 205-2014.

Muhammad Erfan . (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model rasch. *Indonesian Journal Of Educational Research And Review*, 11-19.

Ngadi. (2023). Analisis model rasch untuk mengukur kompetensi pengetahuan siswa SMKN 1 kalianget pada mata pelajaran perawatan sistem kelistrikan sepeda motor. *Jurnal Pendidikan Vokasi Otomotif*, 1-20.

Purnomo. (2016). Pengembangan soal matematika model pisa konten space and shape untuk mengetahui level kemampuan berpikir tingkat tinggi berdasarkan analisis model rasch. Jember: Digital Repository Universitas Jember.

Rahma Gustiannur Simbolon. (2024). Analisis butir soal pada penilaian akhir semester ganjil menggunakan rasch model kelas x sma negeri 2 bunguran timur. Kuningan: JUMLAHKU: Jurnal Matematika Ilmiah STKIP Muhammadiyah Kuningan.

Siswanti. (2023). Analisis Kualitas butir soal fisika pada penilaian akhir semester. *AL-KHAZINI: Jurnal Pendidikan Fisika*, 130-139.

Walewangko A. Elisabet. (2023). Analisis butir soal tes sumatif semester. *MARISEKOLAH: Jurnal Matematika Riset Edukasi Dan Kolaborasi*.