



**ANALISIS KUALITAS BUTIR SOAL PADA KOMPETISI MATEMATIKA  
MENGUNAKAN TEORI RESPON ITEM**

Andi Saparuddin Nur<sup>1)\*</sup>, Dessy Rizki Suryani<sup>2)</sup>, Dian Mayasari<sup>3)</sup>

<sup>1,2,3</sup>Jurusan Pendidikan Matematika, FKIP, Universitas Musamus, Jalan Kamizaun Mopah Lama, Merauke, 99611, Indonesia

✉ [andisaparuddin@unmus.ac.id](mailto:andisaparuddin@unmus.ac.id)

ARTICLE INFO	ABSTRAK
<p><b>Article History:</b> Received: 25/10/2024 Revised: 22/11/2024 Accepted: 04/12/2024</p>	<p>Kualitas butir soal pada kompetisi matematika merupakan faktor yang wajib terpenuhi agar mampu menyeleksi peserta dengan kemampuan terbaik. Adakalanya peserta yang lolos pada babak penyisihan memanfaatkan kelemahan butir soal. Namun demikian, pendekatan klasik tidak mampu menganalisis kelemahan butir soal sehingga diperlukan pendekatan teori respon item untuk memperoleh gambaran kemampuan peserta. Penelitian ini bertujuan menganalisis kualitas butir soal yang digunakan pada babak penyisihan competition of mathematics (Comath) Universitas Musamus tahun 2023 menggunakan teori respon item. Terdapat 25 butir soal pilihan ganda dengan jumlah peserta sebanyak 24 tim. Teknik analisis menggunakan pendekatan respon item model rasch dengan bantuan software Ministep. Hasil penelitian menunjukkan skor reliabilitas butir item dan peserta pada kategori baik dengan koefisien alpha cronbach sebesar 0,91. Rata-rata kemampuan peserta berada di bawah item dengan -0,35 logit. Nilai Outfit MNSQ dan Outfit ZSTD berada pada kategori fit yang berarti bahwa seluruh item mampu mendeksripsikan kemampuan peserta secara rasional. Instrumen soal yang digunakan fit (cocok) dengan Model Rasch dalam menyeleksi kemampuan peserta dengan baik.</p>

**Kata kunci:** Butir soal, kompetisi matematika, model rasch, teori respon item

ABSTRACT
<p><i>The quality of the questions in the mathematics competition is a factor that must be met in order to be able to select participants with the best abilities. Sometimes participants who pass the preliminary round take advantage of the weaknesses of the questions. However, the classical approach is unable to analyze the weaknesses of the questions so that an item response theory approach is needed to obtain an overview of the participants' abilities. This study aims to analyze the quality of the questions used in the preliminary round of the 2023 Musamus University Competition of Mathematics (Comath) using item response theory. There are 25 multiple choice questions with 24 teams participating. The analysis technique uses the Rasch model item response approach with the help of Ministep software. The results of the study showed that the reliability scores of the items and participants were in the good category with a Cronbach alpha coefficient of 0.91. The average ability of the participants was below the item with -0.35 logit. The Outfit MNSQ and Outfit ZSTD values were in the fit category, meaning that all items were able to describe the participants' abilities rationally. The question instrument used fit the Rasch Model in selecting participants' abilities well.</i></p> <p><b>Keywords:</b> Item test, mathematics competition, rasch model, item response theory</p>

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



**Cara Menulis Sitasi:** Nur, A. S., Suryani, D. R., & Mayasari, D. (2024). Analisis Kualitas Butir Soal pada Kompetisi Matematika Menggunakan Teori Respon Item. *SIGMA: Jurnal Pendidikan Matematika*, 16 (2), 371-384. <https://doi.org/10.26618/sigma.v16i2.16307>

## Pendahuluan

Kompetisi matematika atau *competition of mathematics* (Comath) merupakan lomba cerdas cermat yang menjadi agenda tahunan di Universitas Musamus Merauke sejak tahun 2015. Comath menjadi ajang kompetisi matematika yang diikuti oleh siswa SMA di Kabupaten Merauke. Terdapat dua babak dalam perlombaan Comath, yaitu babak penyisihan dan babak final. Pada babak penyisihan disediakan soal pilihan ganda dan soal isian singkat dengan sistem penskoran 1 jika menjawab benar dan 0 jika menjawab salah. Peserta yang mengumpulkan skor terbanyak pada babak penyisihan berhak melaju ke babak final. Namun demikian, peserta yang mengumpulkan skor terbanyak pada babak penyisihan belum tentu mampu memenangkan babak final. Oleh karena itu, analisis terkait kualitas butir soal pada babak penyisihan perlu dilakukan untuk mengantisipasi adanya kemungkinan peserta memperoleh keberuntungan dari tebakan jawaban. Kualitas butir soal sangat diperlukan untuk mengetahui layak tidaknya soal digunakan dalam suatu penilaian (Hanna & Retnawati, 2022; Sumintono, 2016). Analisis butir soal dapat memberikan informasi terkait kemampuan peserta tes, demikian halnya soal yang digunakan pada kompetisi matematika (Apriyani et al., 2023).

Secara garis besar, terdapat dua pendekatan analisis butir soal yaitu; pendekatan klasik dan pendekatan item respon. Pendekatan klasik masih berpatokan pada jawaban peserta di setiap butir soal dengan menganalisis daya beda, tingkat kesulitan, reliabilitas, fungsi distraktor (pengecoh), dan kesalahan baku pengukuran (Sumintono, 2016). Pada umumnya, soal berbentuk objektif seperti pilihan ganda dan isian singkat lebih dipilih untuk digunakan pada lomba kompetisi matematika karena mudah diperiksa. Namun demikian, kelemahan soal objektif kurang memberikan gambaran utuh terkait kemampuan peserta yang sebenarnya. Hal ini disebabkan pada soal objektif jawaban telah tersedia pada pilihan sehingga peserta yang memiliki keberuntungan dalam menebak jawaban memperoleh banyak keuntungan (Hutabarat, 2009). Kadangkala, peserta mampu menjawab benar soal yang sangat sulit namun salah dalam menjawab soal paling mudah. Pendekatan klasik tidak memiliki acuan untuk menjelaskan anomali tersebut sehingga diperlukan pendekatan lain untuk menganalisis kualitas butir soal, yaitu dengan menggunakan pendekatan item respon (Marjiastuti & Wahyuni, 2014).

Sebaliknya, pendekatan teori respon item mampu mengonfirmasi keterkaitan respon peserta terhadap butir soal dengan kemampuan yang diukur pada tes (Fajrianthi et al., 2016). Pendekatan teori respon item atau biasa dikenal sebagai Model Rasch didasarkan pada asumsi parameter logistik yaitu suatu kriteria yang digunakan untuk menentukan bobot jawaban peserta (Hutabarat, 2009). Misalnya, pada model rasch 1PL (satu parameter logistik) hanya mempertimbangkan tingkat kesulitan butir. Selanjutnya, terdapat Model Rasch 2PL yaitu melibatkan tingkat kesulitan dan daya beda butir serta Model Rasch 3PL yaitu mempertimbangkan tingkat kesulitan, daya beda, dan faktor tebakan semu. Semakin banyak kriteria yang digunakan untuk menjelaskan respon peserta maka parameter logistik juga semakin bertambah sehingga analisisnya menjadi lebih mendalam.

Model Rasch mampu menganalisis kualitas soal dengan tepat, baik secara kemampuan individu maupun identifikasi kebutuhan bantuan belajar (Sumintono, 2016). Model Rasch menjadi teknik analisis butir soal yang mampu mengungkap ketepatan alat tes dalam mendeskripsikan kemampuan peserta tes (Widhiarso, 2016). Model Rasch tidak hanya menggunakan fitur analisis yang telah ada pada pendekatan klasik, namun juga sejumlah komponen lainnya telah disediakan seperti; analisis fit item, analisis unidimensi respon peserta,

reliabilitas instrumen, peta respon-item, dan analisis fungsi daya beda (Asrijanty, 2014; Darmana et al., 2021). Model Rasch mampu mengungkap kemampuan sebenarnya yang dimiliki peserta melalui grafik *item characteristic curve* (ICC), dan kemungkinan peserta menebak (*guessing*) (Sumintono & Widhiarso, 2013). Model Rasch juga memiliki beberapa kelebihan yaitu (Kumalasari & Mahmudi, 2024; Marjiastuti & Wahyuni, 2014): (1) mampu memberikan skala liner dengan interval yang sama; (2) mampu melakukan prediksi terhadap data yang hilang; (3) mampu memberikan estimasi yang lebih tepat; (4) mampu mendeteksi ketidaktepatan model; dan (5) mampu menghasilkan pengukuran yang *replicable*.

Model Rasch telah digunakan secara luas dalam evaluasi dan pengembangan soal matematika untuk mengetahui kualitas dan reliabilitas soal (Azizah & Wahyuningsih, 2020; Hutabarat, 2009; Kurniawan & Andriyani, 2018). Lebih lanjut, Model Rasch juga telah digunakan untuk menganalisis kualitas butir soal kompetisi matematika (Apriyani et al., 2023). Namun demikian, analisis kualitas butir soal belum pernah dilakukan pada kompetisi Comath Universitas Musamus Merauke. Oleh karena itu, pendekatan Model Rasch sangat diperlukan untuk menganalisis kualitas butir soal yang digunakan pada babak penyisihan Comath Universitas Musamus Merauke. Adapun pertanyaan penelitian ini, yaitu; Bagaimana kualitas butir soal Comath menggunakan teori respon item?

### Metodologi Penelitian

Penelitian ini merupakan penelitian deskriptif yang bersifat kuantitatif dalam memetakan kualitas butir soal pada *competition of mathematics* (Comath) Universitas Musamus Tahun 2023 menggunakan teori respon item Model Rasch. Terdapat 24 tim yang berpartisipasi dengan masing-masing tim terdiri atas tiga siswa yang berasal dari SMA se Kabupaten Merauke. Sebanyak 13 tim berasal dari SMA di Kota Merauke, dan 11 tim berasal dari SMA pedesaan di Kabupaten Merauke. Jumlah soal sebanyak 25 butir dan telah divalidasi isi oleh para juri yang terdiri atas tiga orang. Adapun topik pada masing-masing item ditunjukkan pada Tabel 1.

Tabel 1. Topik pada setiap butir item

Item	Topik	Item	Topik
1	Eksponen	14	Persamaan garis
2	Persamaan kuadrat	15	Fungsi
3	Trigonometri	16	Bentuk akar
4	Fungsi komposisi	17	Program linear
5	Persamaan lingkaran	18	Integral
6	Barisan aritmatika	19	Eksponen
7	Integral	20	Persamaan garis
8	Polinomial	21	Fungsi
9	Matriks	22	Fungsi invers
10	Trigonometri	23	Persamaan lingkaran
11	Limit	24	Barisan aritmatika
12	Sistem persamaan linear	25	Integral substitusi
13	Program linear		

Analisis butir soal menggunakan Model Rasch berbentuk dikotomi. Menurut Bond & Fox (Sumintono & Widhiarso, 2013) Model Rasch menyatakan probabilitas yang mungkin dijawab benar oleh peserta terhadap item dengan mempertimbangkan kemampuan peserta dan tingkat kesulitan item.

$$P_{ni} \left( X_{ni} = \frac{1}{\beta_n}, \delta_i \right) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \quad (1)$$

Keterangan:

$P_{ni} \left( X_{ni} = \frac{1}{\beta_n}, \delta_i \right)$  =kemungkinan peserta  $n$  menjawab benar item  $i$

$\beta_n$  = kemampuan peserta

$\delta_i$  = tingkat kesulitan item  $i$

Apabila persamaan (1) disederhanakan melalui fungsi logaritma pada kedua ruas maka diperoleh:

$$\log \left( P_{ni} \left( X_{ni} = \frac{1}{\beta_n}, \delta_i \right) \right) = \beta_n - \delta_i \quad (2)$$

Melalui persamaan (2) disimpulkan bahwa kemungkinan peserta berhasil menjawab benar ditentukan oleh selisih antara kemampuan peserta dengan tingkat kesulitan item. Pada model Rasch, pola jawaban peserta menjadi variabel penting yang menunjukkan konsistensi ketepatan respon terhadap setiap item. Oleh karena itu, model Rasch mampu mendeteksi bias respon dan kecenderungan peserta menebak jawaban yang benar.

Terdapat beberapa analisis yang digunakan pada model Rasch, yaitu (Kumalasari & Mahmudi, 2024); (1) analisis peta person-item (Wright) yang memetakan antara kekuatan setiap butir soal dalam menjelaskan kemampuan peserta. Sebaran kemampuan peserta dan tingkat kesulitan item dapat dijelaskan melalui peta person-item. (2) Analisis item atau biasa disebut item fit yaitu digunakan untuk menjelaskan kesesuaian butir dalam pengukuran. Informasi yang diberikan pada item fit menunjukkan validitas butir yang ditentukan oleh tiga kriteria, yaitu (Boone et al., 2014); (a) nilai outfit mean square (MNSQ) diantara  $0,5 < MNSQ < 1,5$ . (b) nilai outfit Z-standard diantara  $-2 < ZSTD < 2$ , dan (c) nilai point measure correlation diantara 0,4 dan 0,85. Selanjutnya, analisis bias terkait kemampuan item dalam membedakan kemampuan peserta berdasarkan atribut tertentu atau biasa disebut DIF (*Differential Item Functioning*). Analisis DIF dalam penelitian ini berdasarkan asal sekolah peserta yaitu; Kota (K) dan Desa (D). Validitas item sangat bergantung pada tidak adanya bias dengan kriteria probabilitas ( $p \geq 0,05$ ). (3) Analisis person berkaitan dengan konsistensi jawaban peserta sehingga peluang tebakan atau kecurangan dapat diketahui. Kriteria yang digunakan untuk menentukan kesesuaian jawaban peserta sama dengan kriteria pada item fit. Tingkat kemampuan peserta dikelompokkan berdasarkan indikasi jawaban yang diperoleh dengan cara menyontek, atau menebak sehingga berbeda dengan prediksi model dan ditunjukkan pada nilai person measure. (4) Analisis instrumen meliputi ringkasan statistik dari keseluruhan hasil penilaian seperti; (a) reliabilitas yaitu gambar interaksi antara peserta dengan item. Jika skor alpha cronbach  $> 0,8$  maka kriteria sangat baik, (b) reliabilitas peserta dan reliabilitas item menunjukkan kekonsistenan dengan skor minimal 0,8.

## Hasil Penelitian

Hasil analisis menggunakan Model Rasch dikelompokkan ke dalam tiga bagian utama, yaitu; ringkasan statistik, analisis person-item, dan peta Wright. Analisis dalam penelitian ini menggunakan bantuan aplikasi software Ministep.

### Ringkasan Statistik

Bagian ringkasan statistik menunjukkan rata-rata kemampuan seluruh peserta terhadap item (logit), reliabilitas peserta, reliabilitas item, dan reliabilitas instrumen. Hasil analisis ringkasan statistik ditunjukkan pada tabel 2

Tabel 2. Hasil ringkasan statistik

Analisis	Informasi	Nilai	Kategori
Logit	Peserta	-0,35	Rata-rata kemampuan peserta di bawah item
	Item	0	
Reliabilitas	Reliabilitas Peserta	0,88	Baik
	Reliabilitas Item	0,89	Baik
	Alpha cronbach	0,91	Sangat baik
Outfit MNSQ	Peserta	1,37	Fit
	Item	1,38	Fit
Outfit ZSTD	Peserta	0,30	Fit
	Item	0,24	Fit

Informasi pada tabel 2 menunjukkan bahwa logit person (peserta) 0,35 lebih rendah dari logit item. Hal tersebut berarti rata-rata kemampuan peserta di bawah tingkat kesulitan item. Peserta kesulitan bahkan tidak mungkin dapat menjawab benar seluruh butir soal. Sementara itu, nilai reliabilitas peserta sebesar 0,88 dan reliabilitas item sebesar 0,89 menunjukkan peserta konsisten dalam menjawab butir soal sedangkan item berfungsi baik dalam menyeleksi kemampuan peserta. Koefisien alpha cronbach sebesar 0,91 mengonfirmasi interaksi antara kemampuan peserta dengan item secara keseluruhan sangat baik. Pada nilai outfit MNSQ, tampak peserta memiliki skor 1,37 dan item memiliki skor 1,38 termasuk ke dalam kategori fit ( $0,5 < MNSQ < 1,5$ ). Hal ini menunjukkan instrumen yang digunakan sudah layak untuk menyeleksi peserta di babak penyisihan kompetisi matematika. Lebih lanjut, nilai outfit ZSTD peserta dan item masing-masing sebesar 0,30 dan 0,24 juga termasuk ke dalam kategori fit ( $-2,0 < ZSTD < 2,0$ ) yang bermakna seluruh item mampu mendeskripsikan kemampuan peserta secara rasional. Dengan demikian, hasil ringkasan statistik menunjukkan instrumen soal babak penyisihan kompetisi matematika Comath 2023 telah sesuai dengan Model Rasch dalam menyeleksi kemampuan peserta.

### Analisis person-item

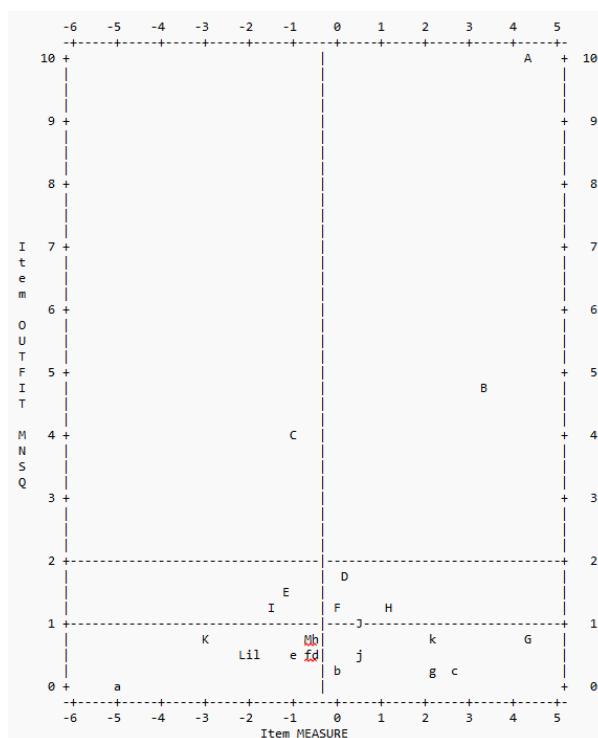
Meskipun secara keseluruhan butir item dinyatakan sesuai dengan Model Rasch, namun analisis lanjutan terkait item yang memenuhi kriteria fit atau misfit perlu dilakukan. Untuk menyatakan butir item fit maka terdapat 3 syarat yang diperlukan, yaitu (Boone et al., 2014); (a) nilai outfit MNSQ ( $0,5 < MNSQ < 1,5$ ), (b) nilai outfit ZSTD ( $-2,0 < ZSTD < 2,0$ ), dan (c) skor korelasi item dengan skor keseluruhan ( $0,4 < point\ measure\ correlation < 0,85$ ). Namun demikian, menurut Sumintono & Widhiarso

(Kumalasari & Mahmudi, 2024) minimal dua kriteria terpenuhi telah cukup untuk menyatakan item fit. Hasil analisis item menggunakan ministep ditunjukkan pada tabel 3.

**Tabel 3.** Hasil analisis item misfit menggunakan Model Rasch

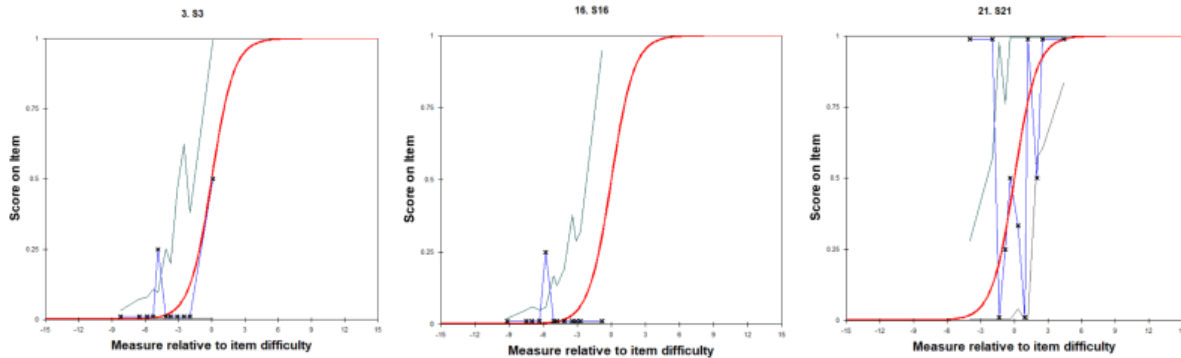
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL CORR.		EXACT OBS%	MATCH EXP%	Item	
					MNSQ	ZSTD	MNSQ	ZSTD	EXP.	EXP%				
16	1	24	4.28	1.10	1.45	.78	9.90	3.63	A	-.13	.30	95.8	95.7	S16
3	2	24	3.36	.85	1.14	.43	4.66	1.89	B	.21	.40	91.7	91.9	S3
21	14	24	-.94	.54	1.95	2.89	4.05	2.77	C	.19	.61	58.3	79.2	S21
11	10	24	.21	.54	1.12	.51	1.86	1.19	D	.53	.62	83.3	80.4	S11
13	15	24	-1.23	.54	1.34	1.28	1.54	.86	E	.46	.60	70.8	79.0	S13
14	11	24	-.08	.54	1.44	1.43	1.19	.49	F	.49	.63	66.7	80.1	S14
25	1	24	4.28	1.10	1.31	.63	.69	.22	G	.20	.30	95.8	95.7	S25
17	7	24	1.13	.57	.96	-.04	1.27	.58	H	.57	.58	87.5	80.5	S17
7	16	24	-1.52	.55	1.08	.40	1.15	.46	I	.55	.59	83.3	79.1	S7
12	9	24	.51	.55	1.08	.36	.88	.12	J	.60	.61	83.3	80.7	S12
18	20	24	-2.92	.66	.94	-.05	.73	.17	K	.51	.49	91.7	87.0	S18
10	18	24	-2.15	.58	.92	-.21	.52	-.10	L	.60	.55	83.3	81.9	S10
20	13	24	-.65	.54	.91	-.24	.76	-.17	M	.66	.62	83.3	79.2	S20
6	17	24	-1.83	.56	.89	-.34	.56	-.19	N	.62	.57	79.2	80.0	S6
19	4	24	2.24	.67	.89	-.16	.66	.10	O	.53	.50	91.7	87.6	S19
1	9	24	.51	.55	.86	-.40	.61	-.29	P	.67	.61	83.3	80.7	S1
8	18	24	-2.15	.58	.83	-.56	.50	-.13	Q	.62	.55	83.3	81.9	S8
15	13	24	-.65	.54	.73	-.98	.63	-.44	R	.72	.62	91.7	79.2	S15
24	4	24	2.24	.67	.68	-.82	.33	-.34	S	.61	.50	91.7	87.6	S24
4	13	24	-.65	.54	.65	-1.36	.45	-.84	T	.75	.62	83.3	79.2	S4
5	14	24	-.94	.54	.63	-1.53	.44	-.79	U	.75	.61	91.7	79.2	S5
23	13	24	-.65	.54	.57	-1.73	.41	-.94	V	.78	.62	91.7	79.2	S23
22	3	24	2.73	.74	.56	-1.00	.25	-.32	W	.60	.46	95.8	90.2	S22
2	11	24	-.08	.54	.50	-2.03	.35	-1.08	X	.80	.63	91.7	80.1	S2
9	23	24	-5.05	1.16	.43	-.61	.07	-.77	Y	.47	.32	95.8	95.8	S9
MEAN	11.2	24.0	.00	.65	.96	-.13	1.38	.24				85.8	83.7	
P.SD	6.0	.0	2.20	.19	.35	1.07	2.04	1.11				9.2	5.8	

Item pada tabel 3 disusun secara berurut berdasarkan kriteria misfit. Terdapat 5 butir soal (20%) yang misfit, yaitu; item 16, 3, 21, 11, dan 13. Butir soal yang misfit menunjukkan adanya penyimpangan prediksi model dalam menjelaskan item. Terdapat 7 butir soal dikategorikan soal sulit (nilai measure >1.0 logit), 12 butir soal kategori sedang (-1.0 logit < nilai measure < 1.0 logit), dan 6 butir soal kategori mudah (nilai measure < -1.0 logit). Untuk mengetahui penyebab item yang misfit maka perlu analisis lanjutan terkait sebaran nilai outfit MNSQ dan item seperti ditunjukkan pada gambar 1.



**Gambar 1.** Hubungan antara outfit MNSQ dan item

Pada gambar 1, tampak bahwa item nomor 16, 21, dan 3 memiliki nilai outfit di luar batas toleransi. Hal ini berarti, soal tersebut seharusnya dapat dijawab benar oleh peserta dengan kemampuan tinggi namun kenyataannya tidak. Sebaliknya, peserta dengan kemampuan sedang atau rendah justru mampu menjawab benar soal tersebut. Dengan kata lain, peserta memiliki kemungkinan menebak dengan benar pada item tersebut. Hubungan antara kesesuaian pola prediksi model dengan hasil observasi menggunakan model Rasch dapat dianalisis melalui kurva karakteristik item (ICC) seperti ditunjukkan pada gambar 1.



Gambar 2. Kurva ICC item nomor 3, 16, dan 21

Berdasarkan Gambar 2. Model Rasch dapat mendeskripsikan pola jawaban peserta di setiap butir soal. Kurva merah menunjukkan proporsi ideal dari tiap butir soal sedangkan garis biru menunjukkan respon peserta. Beberapa butir soal terlihat memiliki pola kurva yang berbeda dengan penaksiran Model Rasch. Hal ini menunjukkan pada butir soal tersebut, Model Rasch memprediksi respon peserta lebih rendah atau lebih tinggi dari kemampuan sebenarnya. Sebagai contoh, pada butir soal nomor 3 terlihat kurva merah dibatasi oleh ruang kepercayaan (garis biru sisi bawah dan atas), namun terdapat respon yang tidak sesuai (misfit) ditandai titik yang berada di luar ruang kepercayaan. Hal ini menunjukkan banyak peserta yang seharusnya mampu menjawab benar, namun kenyataannya salah. Kurva serupa juga teramati pada butir 16. Selanjutnya, pada soal nomor 21 garis biru tampak acak-acakan yang menunjukkan ruang kepercayaan dibatasi pada kurva ideal. Akan tetapi, terdapat beberapa garis yang keluar dari ruang kepercayaan sebagai respon misfit. Hal ini menunjukkan Model Rasch memprediksi lebih banyak peserta yang menjawab salah, namun kenyataannya menjawab benar atau beberapa peserta kemungkinan menebak dengan benar soal nomor 21.

Selanjutnya, analisis person digunakan untuk mengonfirmasi kecocokan peserta dengan Model Rasch. Peserta dikatakan cocok (fit) jika konsisten dalam menjawab item dengan prediksi model. Kriteria person fit sama dengan item fit yang telah diuraikan sebelumnya. Hasil analisis person fit ditunjukkan pada Tabel 4.

**Tabel 4.** Hasil analisis person misfit menggunakan Model Rasch

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Person	
18	7	25	-1.62	.54	1.16	.67	9.90	4.08	A	.34	.53	84.0	80.3	11K
16	8	25	-1.33	.53	1.32	1.33	5.15	2.27	B	.35	.55	76.0	78.7	06D
24	1	25	-4.91	1.20	1.84	1.08	2.19	1.10	C	.09	.33	92.0	96.3	13K
21	6	25	-1.93	.57	1.43	1.43	1.71	.89	D	.34	.51	76.0	82.2	09D
6	18	25	1.55	.62	1.40	1.05	1.29	.60	E	.54	.65	80.0	86.6	05K
14	10	25	-.80	.51	1.23	1.06	.95	.26	F	.51	.58	72.0	76.3	05D
3	17	25	1.19	.59	1.21	.68	1.02	.30	G	.59	.65	84.0	84.9	02K
2	22	25	3.46	.78	.96	.06	1.16	.58	H	.49	.53	96.0	90.9	01D
12	11	25	-.53	.51	1.13	.65	1.08	.37	I	.54	.59	76.0	76.4	03D
22	3	25	-3.16	.74	1.13	.41	.76	.29	J	.40	.44	92.0	90.8	10D
8	17	25	1.19	.59	1.08	.33	.83	.06	K	.64	.65	84.0	84.9	02D
11	13	25	-.01	.52	1.06	.31	.97	.23	L	.60	.62	84.0	78.1	09K
10	17	25	1.19	.59	.88	-.23	1.05	.34	l	.66	.65	92.0	84.9	08K
17	7	25	-1.62	.54	.99	.04	.63	.05	k	.55	.53	76.0	80.3	07D
5	18	25	1.55	.62	.89	-.17	.92	.19	j	.67	.65	88.0	86.6	04K
13	11	25	-.53	.51	.89	-.45	.64	-.20	i	.64	.59	76.0	76.4	04D
15	8	25	-1.33	.53	.82	-.75	.57	-.06	h	.61	.55	84.0	78.7	10K
23	4	25	-2.67	.66	.75	-.56	.43	-.11	g	.55	.47	92.0	87.7	11D
20	5	25	-2.27	.60	.64	-1.13	.34	-.30	f	.61	.49	88.0	84.6	08D
9	14	25	.27	.53	.55	-1.95	.37	-.78	e	.77	.63	92.0	79.4	07K
19	6	25	-1.93	.57	.52	-1.93	.31	-.41	d	.66	.51	92.0	82.2	12K
7	16	25	.86	.57	.51	-1.73	.32	-.80	c	.79	.64	92.0	83.1	06K
1	22	25	3.46	.78	.42	-1.54	.14	-.55	b	.67	.53	96.0	90.9	01K
4	18	25	1.55	.62	.37	-2.06	.21	-1.14	a	.82	.65	96.0	86.6	03K
MEAN	11.6	25.0	-.35	.62	.97	-.14	1.37	.30				85.8	83.7	
P.SD	6.0	.0	2.01	.14	.35	1.08	2.03	1.04				7.4	5.2	

Peserta pada tabel 4 disusun secara berurut dari misfit terbesar sampai terkecil. Terdapat 4 peserta (16%) yang misfit, yaitu; peserta 11K, 06D, 13K, 09D. Hal ini menunjukkan kemungkinan peserta tersebut menebak jawaban pada soal yang diberikan atau menjawab salah pada soal yang seharusnya mampu dijawab benar. Untuk mengetahui faktor penyebab peserta tersebut dinyatakan misfit maka perlu dilakukan analisis lanjutan melalui scalogram seperti ditunjukkan pada gambar 3.

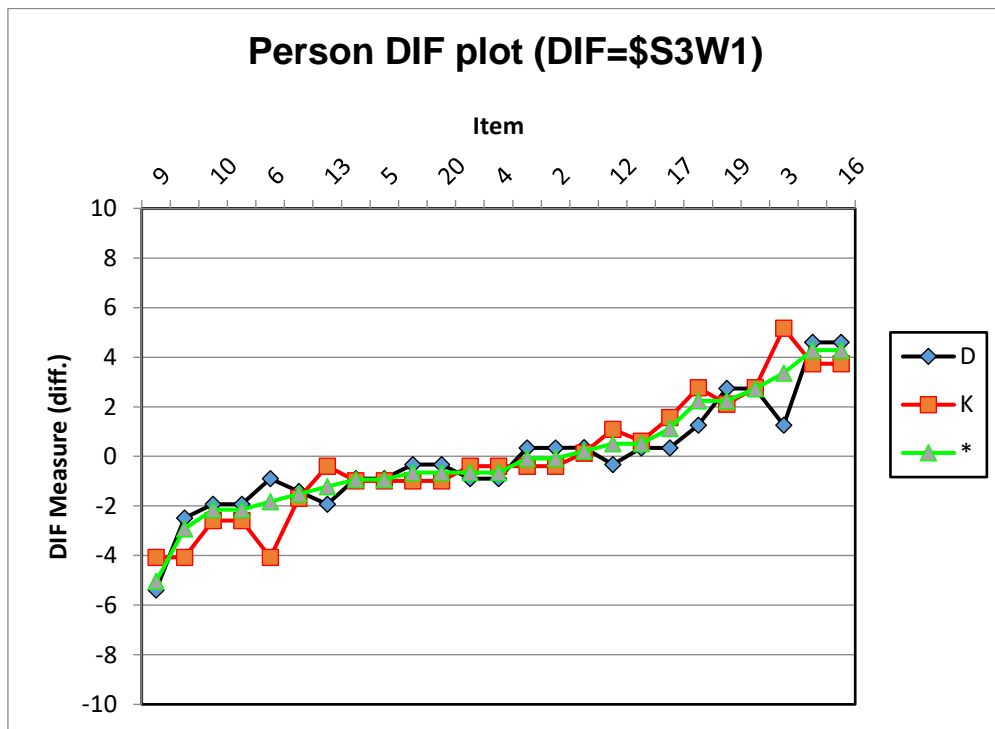




**Tabel 5.** Hasil analisis DIF

Person CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	Person CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Rasch-Welch t	d.f.	Prob.	Mantel-Haenszel Chi-squ	Size Prob.	Active Slices	Item Number	Name	
D	.02	.35	.87	K	-.02	.62	.69	-.27	1.11	-.24	19	.8111	.1250	.7237	5	1	S1	
D	-.06	.35	.87	K	.05	-.40	.75	.74	1.15	.65	20	.5245			5	2	S2	
D	.12	1.26	1.06	K	-.10	5.16	1.91	-3.90	2.18	-1.78	18	.0911	.5000	.4795	5	3	S3	
D	.04	-.89	.73	K	-.03	-.40	.75	-.50	1.04	-.48	21	.6388	.0000	1.000	5	4	S4	
D	-.01	-.89	.73	K	.01	-.99	.80	.09	1.08	.09	21	.9315	.0000	1.000	5	5	S5	
D	-.16	-.89	.73	K	.14	-4.07	1.42	3.18	1.60	1.99	17	.0635	1.3333	.2482	5	6	S6	
D	-.02	-1.42	.72	K	.02	-1.68	.88	.27	1.13	.24	21	.8150	.2353	.6276	.41	5	7	S7
D	-.04	-1.93	.72	K	.03	-2.59	1.04	.65	1.27	.51	20	.6123	.5000	.4795	.00	5	8	S8
D	-.04	-5.39	1.88	K	-.03	-4.07	1.42	-1.31	2.36	-.56	19	.5843			5	9	S9	
D	-.04	-1.93	.72	K	.03	-2.59	1.04	.65	1.27	.51	20	.6123	.0000	1.000	5	10	S10	
D	-.02	.35	.87	K	.01	.13	.71	.22	1.13	.19	20	.8488	.5000	.4795	.00	5	11	S11
D	.11	-.33	.78	K	-.10	1.09	.69	-1.42	1.04	-1.36	21	.1871	.0000	1.000	5	12	S12	
D	.12	-1.93	.72	K	-.10	-.40	.75	-1.53	1.04	-1.48	21	.1547	.0000	1.000	-.69	5	13	S13
D	-.06	.35	.87	K	.05	-.40	.75	.74	1.15	.65	20	.5245	.9615	.3268	5	14	S14	
D	.04	-.89	.73	K	-.03	-.40	.75	-.50	1.04	-.48	21	.6388	.2353	.6276	5	15	S15	
D	-.03	4.60	2.04	K	.03	3.74	1.14	.86	2.34	.37	15	.7194	.0000	1.000	5	16	S16	
D	.08	.35	.87	K	-.07	1.57	.71	-1.23	1.13	-1.09	20	.2887	.0000	1.000	5	17	S17	
D	-.06	-2.48	.76	K	.06	-4.07	1.42	1.59	1.62	.99	18	.3371	.0000	1.000	5	18	S18	
D	-.03	2.74	1.40	K	.02	2.11	.76	.63	1.59	.40	15	.6952	.1250	.7237	5	19	S19	
D	-.05	-.33	.78	K	.04	-.99	.80	.66	1.11	.59	21	.5594	.1250	.7237	5	20	S20	
D	-.01	-.89	.73	K	.01	-.99	.80	.09	1.08	.09	21	.9315	.0385	.8445	.51	5	21	S21
D	.00	2.73	1.40	K	.00	2.77	.87	-.03	1.65	-.02	17	.9833			5	22	S22	
D	-.05	-.33	.78	K	.04	-.99	.80	.66	1.11	.59	21	.5594	.0000	1.000	5	23	S23	
D	.07	1.26	1.06	K	-.06	2.77	.87	-1.50	1.37	-1.10	20	.2860	.1250	.7237	5	24	S24	
D	-.03	4.60	2.04	K	.03	3.74	1.14	.86	2.34	.37	15	.7194			5	25	S25	

Pada tabel 5 terlihat nilai Prob. untuk keseluruhan item ( $P \geq 0,05$ ) yang berarti bahwa seluruh item dapat dikerjakan oleh peserta baik yang berasal dari kota maupun desa. Dengan demikian, tidak terdapat bias antara item dengan sekolah asal yang menyebabkan kerugian peserta. Namun demikian, analisis lebih lanjut terkait item mana saja yang menguntungkan peserta berdasarkan sekolah asal dapat ditelusuri melalui DIF plot seperti terlihat pada gambar 4.

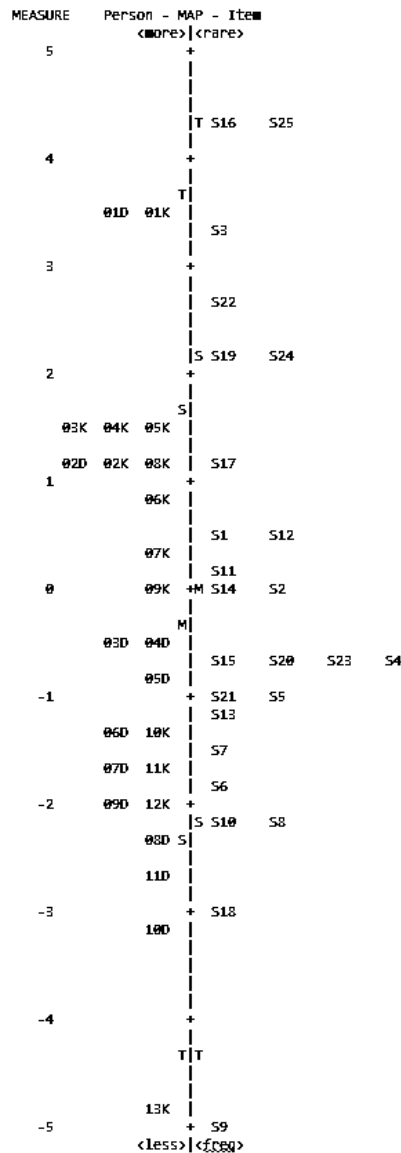


**Gambar 4.** Plot DIF peserta berdasarkan sekolah asal

Berdasarkan gambar 4, tampak pada item nomor 9 peserta dari kota lebih diuntungkan. Sedangkan, pada item nomor 6 peserta dari desa lebih diuntungkan. Peserta dari kota memiliki keuntungan dalam menjawab item nomor 13, dan 3. Akan tetapi, peserta dari desa juga memiliki keuntungan pada item nomor 2 dan 4. Secara umum, plot DIF menunjukkan setiap item tidak merugikan secara signifikan peserta berdasarkan sekolah asal sehingga dapat dikatakan instrumen telah memenuhi kriteria valid.

*Peta Wright*

Hubungan antara kemampuan peserta dengan tingkat kesulitan item dinyatakan melalui peta Wright menggunakan seperti ditunjukkan pada gambar 5.



**Gambar 5.** Peta Wright antara kemampuan peserta dengan tingkat kesulitan item

Berdasarkan gambar 5 tampak bahwa S9 merupakan item termudah yang mewakili peserta dengan kemampuan rendah, yaitu peserta 13 K. Sementara itu, S16, dan S25 merupakan item *outlier* yang tidak mewakili kemampuan peserta manapun. Selanjutnya, S3, S22, S19, dan S24 yang mewakili kemampuan peserta dengan kemampuan tinggi, yaitu peserta 01 K, dan 01D.

Item lainnya mampu menjelaskan kemampuan peserta pada kategori sedang ditunjukkan dengan sebaran data pada bagian tengah. Sebagian besar item dan peserta berkumpul di bagian tengah yang mengindikasikan tingkat kesulitan item berada pada kategori sedang berbanding lurus dengan kemampuan peserta. Oleh karena itu, karakteristik soal yang digunakan pada babak penyisihan Comath Universitas Musamus 2023 merupakan item dengan tingkat kesulitan sedang mewakili karakteristik peserta dengan kemampuan sedang.

### **Pembahasan**

Soal yang paling sedikit dijawab benar oleh peserta merupakan topik eksponen dan barisan aritmatika (16,67%), fungsi invers (12,5%), trigonometri (8,33%), bentuk akar dan integral substitusi (4,17%). Trigonometri merupakan topik yang sulit dipahami siswa karena berbagai jenis hambatan belajar (Jatisunda & Nahdi, 2019). Kesulitan peserta juga banyak dialami pada topik integral. Hal ini sejalan dengan pendapat Monariska (2019) yang menyebutkan bahwa kesalahan dalam menyelesaikan masalah integral terkait lemahnya pemahaman dan keterampilan proses. Selanjutnya, kesulitan peserta dalam menyelesaikan topik barisan disebabkan kelemahan dalam penerapan rumus. Hal ini sejalan dengan temuan Wulandari & Setiawan (2021) yang menyebutkan bahwa siswa mengalami kebingungan dalam menentukan rumus dan tidak memahami konsep barisan aritmatika dengan baik. peserta juga mengalami kesulitan dalam menyelesaikan soal bentuk akar, khususnya dalam memahami sifat-sifat serta konsep merasionalkan bentuk akar. Sejalan dengan itu, Hermawan & Andrianto (2018) mengidentifikasi kesulitan siswa disebabkan rendahnya kemampuan dalam memanipulasi bentuk akar dan rendahnya pemahaman terkait sifat-sifat bentuk akar.

Temuan dalam penelitian ini menunjukkan latar belakang demografis peserta tidak berbeda dalam menyelesaikan soal. Peserta dari sekolah perkotaan dan pedesaan memiliki kemampuan yang sama dalam menyelesaikan setiap soal. Hal ini menunjukkan distribusi materi pada soal secara umum telah diajarkan di setiap sekolah. Temuan ini juga mengonfirmasi hasil penelitian Sugiharto, et al (2020) yang menyebutkan bahwa tidak terdapat perbedaan signifikan kemampuan pengetahuan kognitif siswa di kota dan di desa. Namun demikian, beberapa item memberikan keuntungan baik siswa yang berasal dari kota maupun desa. Peserta dari kota diuntungkan pada topik trigonometri dan program linear yang ditandai dari jumlah peserta menjawab benar, sedangkan peserta dari desa diuntungkan pada topik persamaan kuadrat dan fungsi komposisi.

Topik pada materi matriks merupakan item yang mewakili karakteristik peserta kemampuan rendah. Hampir seluruh peserta dapat menjawab dengan benar item pada topik matriks. Berbeda dengan temuan Pratiwi & Adirakasiwi (2022) yang menyebutkan bahwa siswa dalam memecahkan masalah matriks berada pada kemampuan rendah. Sementara itu, item pada topik bentuk akar dan integral substitusi tidak mewakili karakteristik peserta manapun. Hal ini disebabkan karakteristik peserta kemampuan tinggi tidak mampu menjawab benar item tersebut sehingga perlu ada peninjauan ulang terkait topik bentuk akar dan integral substitusi untuk tidak digunakan dalam kompetisi berikutnya. Hal ini sejalan dengan pendapat Darmana et al (2021) yang menyebutkan bahwa kemungkinan peserta belum mempelajari topik tertentu sehingga gagal dalam memecahkan masalah. Sebagian besar item dapat menjelaskan karakteristik peserta dengan baik. Sejalan dengan itu, menurut Hutabarat (2009) penyusunan

instrumen untuk kompetisi harus mampu memetakan peserta berdasarkan tingkat kompetensinya.

### Simpulan

Soal pada babak penyisihan kompetisi matematika Comath Universitas Musamus Tahun 2023 telah sesuai dengan Model Rasch. Nilai reliabilitas item sebesar 0,89 dan reliabilitas peserta sebesar 0,88 menunjukkan peserta konsisten dalam menjawab butir soal sedangkan item berfungsi baik dalam menyeleksi kemampuan peserta. Koefisien alpha cronbach sebesar 0,91 menunjukkan instrumen berfungsi dengan sangat baik dan didukung oleh nilai outfit MNSQ sebesar 1,37 untuk peserta dan 1,38 untuk item. Selanjutnya, nilai pada outfit ZSTD sebesar 0,30 untuk person dan 0,24 untuk item yang menunjukkan bahwa data dapat diprediksi secara logis oleh model. Terdapat 5 butir soal (20%) yang misfit, dan 4 peserta (16%) yang misfit.

Soal yang digunakan pada babak penyisihan kompetisi matematika Comath Universitas Musamus secara umum telah memenuhi kriteria dan memperhatikan tingkat pemahaman peserta. Namun, fakta bahwa terdapat peserta yang memperoleh skor tinggi pada babak penyisihan mengalami penurunan skor pada babak final memerlukan analisis lebih lanjut, khususnya prosedur atau tata tertib lomba. Hal tersebut dapat berkaitan dengan bentuk peraturan yang mungkin merugikan beberapa peserta saat babak final. Rekomendasi penelitian ini yaitu; panitia perlu mendiskusikan secara komprehensif tentang aturan lomba yang memberikan rasa keadilan bagi setiap peserta. Lebih lanjut, penelitian ini terbatas pada jumlah peserta yang tidak terlalu banyak sehingga penafsiran data dapat berbeda jika dilakukan pada peserta dengan jumlah yang lebih besar.

### Daftar Pustaka

- Apriyani, D. C. N., Susanto, H. P., & Hidayat, T. (2023). Analysis of Pre-Olympic Middle School Mathematics Test Instruments Based on Item Response Theory. *AlphaMath : Journal of Mathematics Education*, 9(2), 145. <https://doi.org/10.30595/alphamath.v9i2.18021>
- Asrijanty. (2014). Model Rasch sebagai Kerangka Acuan Penyusunan Alat Ukur. *Jurnal Pendidikan Dan Kebudayaan*, 20(1), 109–123. <https://doi.org/10.24832/jpnk.v20i1.130>
- 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>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.
- Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa\*, I., & Nurhaswinda, N. (2021). Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments. *Jurnal Pendidikan Sains Indonesia*, 9(3), 329–345. <https://doi.org/10.24815/jpsi.v9i3.19618>
- Fajrianti, F., Hendriani, W., & Septarini, B. G. (2016). Pengembangan Tes Berpikir Kritis Dengan Pendekatan Item Response Theory. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1), 45–55. <https://doi.org/10.21831/pep.v20i1.6304>
- Hanna, W. F., & Retnawati, H. (2022). Analisis Kualitas Butir Soal Matematika Menggunakan Model Rasch Dengan Bantuan Software Quest. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 11(4), 3695. <https://doi.org/10.24127/ajpm.v11i4.5908>
- Hermawan, V., & Andrianto, A. (2018). Analisis Kesulitan Siswa Terhadap Pokok Bahasan Pangkat Rasional Dan Bentuk Akar Di Kelas 1 Smu Bina Dharma 2 Bandung. *Symmetry*:

- Pasundan Journal of Research in Mathematics Learning and Education*, 3, 116–124.  
<https://doi.org/10.23969/symmetry.v3i2.1320>
- Hutabarat, I. M. (2009). Analisis Butir Soal dengan Teori Tes Klasik dan Teori Respons Butir. *Pythagoras: Jurnal Pendidikan Matematika*, 5(2), 1–13.
- Jatisunda, M. G., & Nahdi, D. S. (2019). Kesulitan Siswa Dalam Memahami Konsep Trigonometri Di Lihat Dari Learning Obstacles. *Didactical Mathematics*, 2(1), 9. <https://doi.org/10.31949/dmj.v2i1.1664>
- Kumalasari, E. D., & Mahmudi, I. (2024). *Analisis Pemodelan Rasch Pada Asesmen Pendidikan* (Issue February).
- Kurniawan, & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, 6(1), 34–39.
- Marjiastuti, K., & Wahyuni, S. (2014). Analisis Kemampuan Peserta Didik Dengan Model Rasch. *Seminar Nasional Evaluasi Pendidikan*, 121–128.
- Monariska, E. (2019). Analisis kesulitan belajar mahasiswa pada materi integral. *Jurnal Analisa*, 5(1), 9–19. <https://doi.org/10.15575/ja.v5i1.4181>
- Pratiwi, P., & Adirakasiwi, A. . (2022). Kemampuan Pemecahan Masalah Matematis Siswa SMA Dalam Menyelesaikan Soal Matriks. *Jurnal Pembelajaran Matematika Inovatif*, 5(5), 1419–1433. <https://doi.org/10.22460/jpmi.v5i5.1419-1434>
- Sugiharto, B., Malinda, E. R., Azizzah, H., Anugerah, J. F., Rani, M. J. M., Padmi, N. R. C., & Alifah, N. (2020). Perbedaan Kesadaran Metakognisi Siswa SMA di Desa dan di Kota. *Jurnal Pendidikan Sains Indonesia*, 8(1), 78–91. <https://doi.org/10.24815/jpsi.v8i1.15354>
- Sumintono, B. (2016). Penilaian keterampilan Berpikir Tingkat Tinggi: Aplikasi Pemodelan Rasch Pada Pendidikan. *Prosiding Seminar Nasional Pendidikan IPA*, 19–32.
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata Publishing House.
- Widhiarso, W. (2016). Penerapan model Rasch untuk mengevaluasi tes UKKS dan UKPS. *Tenaga Kependidikan*, 1(1), 50–51.
- Wulandari, M., & Setiawan, W. (2021). Analisis kesulitan dalam menyelesaikan soal materi barisan pada siswa SMA. *Jurnal Pembelajaran Matematika Inovatif*, 4(3), 571–578. <https://doi.org/10.22460/jpmi.v4i3.571-578>