

## Analisis Butir Soal Model Timss dengan Item Respon Theory

Muhammad Na'im Mahmud<sup>1\*</sup>, Afudin La Arua<sup>2</sup>, Samron<sup>3</sup>, Irawati Windani<sup>4</sup>

<sup>1</sup>Pendidikan Matematika, Universitas Muslim Buton, Kota Baubau, Indonesia;  
[naimmahmudmuhammad@gmail.com](mailto:naimmahmudmuhammad@gmail.com).

<sup>2</sup>Pendidikan Matematika, Universitas Muslim Buton, Kota Baubau, Indonesia;  
[afudinlaarua@umubuton.ac.id](mailto:afudinlaarua@umubuton.ac.id).

<sup>3</sup>Pendidikan Matematika, Universitas Muslim Buton, Kota Baubau, Indonesia;  
[samron2772@gmail.com](mailto:samron2772@gmail.com).

<sup>4</sup>Pendidikan Matematika, Universitas Muslim Buton, Kota Baubau, Indonesia;  
[irawatiwindani@gmail.com](mailto:irawatiwindani@gmail.com)

**Abstrak.** Penelitian ini bertujuan untuk menganalisis kualitas butir soal matematika model TIMSS yang digunakan pada siswa kelas VIII di Kota Baubau. Penelitian ini merupakan jenis ex-post facto yang memadukan dua pendekatan, yakni pendekatan rasional dan empiris. Sampel dalam penelitian ini berjumlah 400 siswa kelas VIII yang terdiri dari 5 SMP dari total keseluruhan 29 SMP yang ada di Kota Baubau. Adapun instrumen penelitian yang digunakan berupa tes pilihan ganda dengan soal-soal matematika yang mengadopsi model TIMSS. Analisis data dilakukan dengan bantuan perangkat lunak Bilog-MG dan SPSS. Berdasarkan hasil analisis kecocokan model menggunakan Teori Respon Butir, ditemukan bahwa model yang paling sesuai adalah model dua parameter logistik (2-PL). Hasil output dari Bilog-MG menunjukkan bahwa seluruh butir soal (sebanyak 30 butir) memiliki indeks daya beda yang tergolong baik. Sementara itu, untuk tingkat kesulitan butir, ditemukan bahwa 27 soal berada dalam kategori sedang, satu soal (butir 27) tergolong terlalu mudah, dan dua soal (butir 26 dan 30) termasuk dalam kategori terlalu sulit.

**Kata Kunci:** Analisis butir, Soal-soal model TIMSS, Bilog Mg

**Abstract.** This study aims to analyze the quality of TIMSS model mathematics questions used in grade VIII students in Baubau City. This study is an ex-post facto type that combines two approaches, namely rational and empirical approaches. The sample in this study consisted of 400 class VIII students consisting of 5 junior high schools from a total of 29 junior high schools in Baubau City. The research instrument used was a multiple-choice test with mathematics questions that

adopted the TIMSS model. Data analysis was carried out with the help of Bilog-MG and SPSS software. Based on the results of the model fit analysis using Item Response Theory, it was found that the most appropriate model was the two-parameter logistic model (2-PL). The output results from Bilog-MG showed that all questions (as many as 30 items) had a fairly good discrimination index. Meanwhile, for the level of item difficulty, it was found that 27 questions were in the moderate category, one question (item 27) was too easy, and two questions (items 26 and 30) were too difficult.

**Keywords.** Item analysis, TIMSS model questions, Bilog Mg

## **Pendahuluan**

Penilaian yang efektif dapat dilakukan dengan mengumpulkan bukti-bukti akurat mengenai capaian pembelajaran siswa dan memastikan bahwa proses dan hasil penilaian di kelas bermanfaat bagi siswa. Untuk memperoleh data yang akurat, penilaian harus didukung oleh proses pengukuran yang tepat. Dalam sistem penilaian hasil pembelajaran, penilaian merupakan tahap lanjutan setelah pengukuran. Sebagai seorang pendidik, salah satu metode untuk menilai pemahaman siswa terhadap materi adalah melalui tes.

Tes merupakan suatu instrumen yang digunakan untuk mengumpulkan informasi, baik berupa serangkaian pertanyaan maupun soal yang secara khusus dirancang untuk diberikan kepada siswa dengan kriteria tertentu (Susilawati, 2018). Sejalan dengan pendapat (Kartowagiran, 2013), tes merupakan sebuah serangkaian kegiatan yang dilakukan secara sistematis dengan menyajikan sejumlah pertanyaan yang mempunyai jawaban benar atau salah, yang bertujuan untuk mengukur kemampuan siswa secara kuantitatif. Output yang dihasilkan oleh tes tidak hanya dijadikan sebagai alat evaluasi, tetapi juga dapat digunakan secara strategis untuk meningkatkan kualitas pembelajaran (El Hasbi et al., 2024). Selain itu, hasilnya dapat digunakan untuk mengidentifikasi area dimana siswa mengalami kesulitan atau masih rendah pemahamannya. Informasi yang diperoleh melalui analisis ini sangat membantu guru dalam merancang pembelajaran yang sesuai dengan kebutuhan siswa, baik secara individu maupun kelompok.

Tes yang baik terdiri dari butir-butir soal yang berkualitas. Instrumen yang berkualitas mampu memberikan informasi secara tepat sesuai dengan tujuan yang diinginkan (Safitri et al., 2024). Ada berbagai jenis butir soal, seperti pilihan ganda, benar-salah, menjodohkan, dan sebagainya. Untuk menilai kualitas suatu butir soal, diperlukan evaluasi melalui proses analisis. Tes yang baik dapat mengukur kemampuan siswa secara akurat. Oleh karena itu, butir-butir soal harus disusun dengan baik dan sesuai dengan tujuan tes yang akan dilaksanakan (Elvira & Samsul, 2016).

Copyright © 2025

*Buana Matematika* :

Jurnal Ilmiah Matematika dan Pendidikan Matematika

p-ISSN : 2088-3021

e-ISSN : 2598-8077

Untuk meningkatkan mutu soal yang telah disusun, analisis butir soal perlu dilakukan oleh guru atau pihak yang terlibat dalam tim penyusun soal (Mania et al., 2020). Soal yang kurang bermutu dapat mempengaruhi hasil evaluasi, sehingga diperlukan upaya untuk meningkatkan mutu soal yang telah dibuat. Analisis butir soal merupakan proses evaluasi yang dilakukan untuk menilai mutu soal dalam ujian atau tes. Dengan menganalisis butir soal merupakan sebuah langkah penting bagi guru untuk menilai kualitas soal yang digunakan, sehingga dapat menentukan mana soal yang layak digunakan kembali dan mana yang harus diperbaiki atau dihapus (Ida & Musyarofah, 2021). Sehingga, dengan melakukan analisis butir soal, pendidik dapat menentukan keputusan dalam menilai proses pengumpulan, penyimpulan, dan pemanfaatan informasi yang diperoleh dari respon peserta didik.

Tujuan dari analisis butir soal adalah untuk memastikan kualitas soal dengan cara meninjau setiap butir soal sebelum digunakan dalam melakukan pengukuran. Analisis butir soal dapat membantu mengidentifikasi kelemahan-kelemahan dalam tes (Urbina, 2014) dan menentukan apakah peserta didik telah memahami atau masih mengalami kesulitan dalam memahami materi yang telah diajarkan. Analisis butir soal dapat dilakukan dengan menggunakan pendekatan klasik maupun modern. Hambleton & Jones, 1993; Hambleton & Swaminathan, 1985 mengemukakan beberapa kelemahan dalam teori tes klasik, antara lain: Tingkat kesulitan dan daya pembeda butir soal tergantung pada kelompok peserta yang mengerjakannya, Analisis tes dilakukan dengan cara membandingkan kemampuan peserta didik berdasarkan kelompok atas, tengah, dan bawah, Reliabilitas skor didasarkan pada konsep tes paralel, Tidak ada dasar teori yang jelas untuk menentukan kesesuaian tes dengan kemampuan peserta didik, Standard Error of Measurement (SEM) berlaku secara umum untuk semua peserta tes. Hasil estimasi parameter sangat dipengaruhi oleh karakteristik peserta tes (bersifat *group dependent*). Konsekuensinya, tingkat kesukaran soal akan tampak lebih rendah apabila tes diberikan kepada kelompok peserta yang memiliki kemampuan tinggi. Sebaliknya, soal yang sama akan terlihat lebih sukar apabila diberikan kepada peserta dengan kemampuan rendah (Sarea & Ruslan, 2019). Keterbatasan ini menyebabkan estimasi kemampuan peserta menjadi rendah jika soal yang dikerjakan berada di atas kemampuan mereka. Sebaliknya, estimasi kemampuan akan tampak tinggi apabila soal yang diberikan berada di bawah tingkat kemampuan peserta

Penelitian ini menggunakan teori tes modern atau yang biasa dikenal dengan teori respons butir yang dikembangkan oleh para ahli di bidang pengukuran psikologi dan pendidikan. Teori Respon Butir (IRT) merupakan pendekatan analisis butir yang modern. Teori ini dikembangkan oleh para ahli untuk

mengatasi keterbatasan teori analisis butir klasik. Dengan demikian, IRT berfungsi sebagai penyempurnaan konsep dalam Teori Tes Klasik (CTT) (Mulyani et al., 2021). Teori respons butir hadir untuk menjawab berbagai kekurangan dalam teori tes klasik. Keterbatasan pendekatan klasik tersebut kerap kali menghambat penerapannya dalam proses seperti penyusunan butir tes, penyetaraan antar butir, atau perbandingan karakteristik peserta. Esensi utama teori respons butir adalah memisahkan pengaruh antara butir tes dan peserta tes secara independen.

Item Response Theory (IRT) merupakan penyempurnaan dari kelemahan-kelemahan yang terdapat pada teori klasik. Salah satu kelebihan IRT adalah Probabilitas seorang individu menjawab benar suatu butir soal bergantung pada kemampuan subjek dan sifat karakteristik dari butir tersebut (Sarea & Ruslan, 2019). Teori respons butir soal memiliki sejumlah karakteristik penting, yaitu: (1) karakteristik butir soal tidak bergantung pada peserta tes, (2) skor yang diperoleh peserta tidak dipengaruhi oleh tes yang digunakan, (3) pendekatan ini berfokus pada analisis tiap butir soal, bukan keseluruhan soal, (4) tidak mensyaratkan penggunaan tes paralel yang ketat untuk menaksir reliabilitas, dan (5) memungkinkan adanya dasar pengambilan keputusan melalui fungsi hubungan antara kemampuan peserta dengan respons yang diberikan (Hambleton et al., 1991). Selain itu, Hambleton dan Swaminathan (1985) menambahkan bahwa secara matematis, peluang seseorang menjawab suatu butir soal dengan benar ditentukan oleh kemampuan peserta dan karakteristik butir soal.

Mengacu pada latar belakang tersebut, penelitian ini bertujuan untuk menganalisis kualitas butir soal dan mengkarakterisasikan tes yang dikembangkan untuk mengukur kemampuan siswa SMP se-Kota Baubau dalam mata pelajaran matematika dengan menggunakan soal-soal model TIMSS.

### Metode

Penelitian ini dikategorikan sebagai penelitian *ex-post facto* dengan menggabungkan dua pendekatan, yaitu pendekatan rasional dan pendekatan empiris. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi butir-butir soal matematika berbasis model TIMSS yang digunakan pada siswa SMP dengan pendekatan Teori Respon Butir (Item Response Theory). Teknik pengambilan sampel yang digunakan adalah *proportionately stratified random sampling*, yakni metode yang memperhitungkan keberadaan subkelompok (strata) dengan ukuran berbeda dalam populasi (Mertens, 2019). Metode ini bertujuan untuk memastikan bahwa jumlah sampel dari setiap strata proporsional terhadap ukuran strata dalam populasi. Populasi dalam penelitian ini mencakup seluruh siswa kelas VIII SMP yang berada di wilayah Kota Baubau. Sebanyak lima sekolah (SMPN 1 Baubau, SMPN 3 Baubau, SMPN 4 Baubau, SMPN 6 Baubau, dan SMP Swasta Darussalam) dipilih sebagai sampel dengan total responden sebanyak 400 siswa kelas VIII. Instrumen tes yang digunakan berupa soal-soal matematika

Copyright © 2025

*Buana Matematika* :

Jurnal Ilmiah Matematika dan Pendidikan Matematika

p-ISSN : 2088-3021

e-ISSN : 2598-8077

berformat model TIMSS, terdiri atas 30 soal pilihan ganda dengan empat opsi jawaban (A, B, C, dan D). Soal TIMSS yang digunakan sebagai instrumen penelitian terdiri dari materi bilangan, aljabar, geometri, dan statistik & peluang yang telah disesuaikan dengan kurikulum yang berlaku. Analisis empiris terhadap butir soal dilakukan menggunakan pendekatan Teori Respon Butir (IRT) dengan bantuan perangkat lunak Bilog-MG.

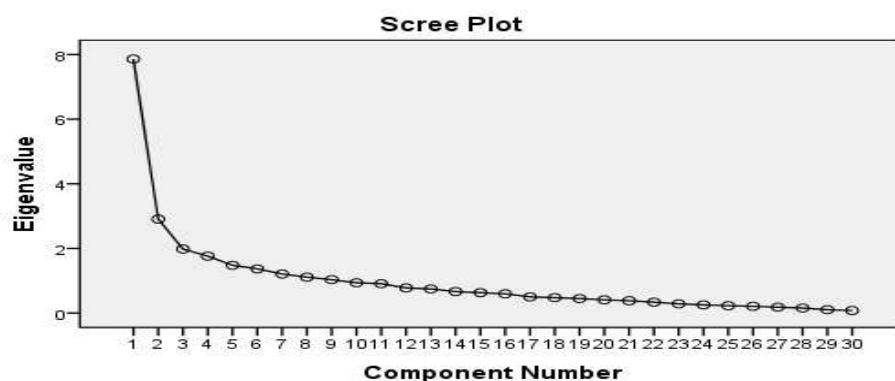
### Hasil dan Pembahasan

Secara umum, terdapat 3 model yang paling umum diaplikasikan dalam pendekatan teori respons butir yaitu model 1-PL, 2-PL, dan 3-PL. Untuk melakukan analisis lebih lanjut, diperlukan pemenuhan dua asumsi dasar terlebih dahulu: unidimensi dan independensi lokal. Uji asumsi dilakukan melalui analisis faktor eksploratori dengan bantuan SPSS, bertujuan mengevaluasi kelayakan sampel serta kevalidan instrumen dalam mengukur satu kemampuan, yaitu matematika.

Tabel 1. Bartlett KMO and Bartlett's Test

<b>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</b>		0,709
<b>Bartlett's Test of Sphericity</b>	<b>Approx. Chi-Square</b>	1183,433
	<b>df</b>	435
	<b>Sig.</b>	0,000

Dari tabel 1 di atas, diketahui bahwa nilai KMO yang didapatkan yaitu sebesar 0,709. Angka tersebut mengindikasikan bahwa sampel penelitian cukup memadai untuk digunakan, karena nilainya lebih besar dari 0,05. Oleh karena itu, analisis dapat diteruskan dengan menguji unidimensionalitas menggunakan scree plot sebagaimana ditampilkan pada gambar di bawah.



Gambar 1. Hasil output Scree Plot

Berdasarkan hasil output SPSS 22 pada Gambar 1, instrumen yang digunakan pada penelitian yaitu berupa soal-soal model TIMSS terdapat satu faktor yang terlihat dominan. Perbedaan begitu signifikan jika melihat pada hasil scree plot yang terjadi pada faktor pertama ke faktor kedua. Adapun untuk faktor yang kedua tidak cukup jauh berbeda dengan faktor lainnya yang mempunyai perubahan nilai eigen yang tidak jauh berbeda atau begitu besar. Berdasarkan hasil tersebut instrumen yang digunakan yaitu soal-soal model TIMSS uji asumsi unidimensinya sudah terpenuhi. Berikutnya yaitu akan disajikan nilai eigen berdasarkan hasil analisis dengan menggunakan bantuan program SPSS 22 pada tabel 2 di bawah ini .

Tabel 2. Faktor Nilai Eigen > 1 dan Komponen Variansi Hasil Analisis Faktor

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	7,860	26,201	26,201
2	2,908	9,693	35,894
3	1,981	6,603	42,497
4	1,760	5,868	48,364
5	1,475	4,918	53,282
6	1,366	4,553	57,835
7	1,208	4,027	61,862
8	1,111	3,703	65,565
9	1,033	3,443	69,008

Uji berikutnya yaitu uji asumsi independensi lokal. Uji tersebut otomatis dapat terpenuhi jika uji sebelumnya yaitu uji asumsi unidimensi terpenuhi. Langkah berikutnya yaitu melakukan uji kecocokan antara model 1-PL, 2-PL dan 3-PL. Hasil uji tersebut dilihat dengan membandingkan nilai signifikansi dari nilai  $\chi^2$  dengan nilai alfa yang digunakan ( $\alpha = 0.05$ ). Kriteria yang menunjukkan butir yang cocok yaitu dengan melihat apakah nilai  $X^2_{hitung}$  kurang dari nilai  $X^2_{tabel}$ . Berikut tabel yang menunjukkan kecocokan model antara 1 PL, 2 PL, dan 3 PL.

Tabel 3. Kecocokan Butir Soal Model TIMSS  
Berdasarkan IRT (*Item Respon Theory*)

Item (i)	Model		
	1-PL	2-PL	3-PL
1	Fit Model	Fit Model	Fit Model
2	Tidak Fit Model	Fit Model	Fit Model
3	Tidak Fit Model	Fit Model	Fit Model
4	Tidak Fit Model	Fit Model	Fit Model
5	Tidak Fit Model	Fit Model	Tidak Fit Model

Copyright © 2025

*Buana Matematika* :

Jurnal Ilmiah Matematika dan Pendidikan Matematika

p-ISSN : 2088-3021

e-ISSN : 2598-8077

Item (i)	Model		
	1-PL	2-PL	3-PL
6	Fit Model	Fit Model	Fit Model
7	Tidak Fit Model	Fit Model	Fit Model
8	Fit Model	Fit Model	Tidak Fit Model
9	Fit Model	Fit Model	Fit Model
10	Fit Model	Tidak Fit Model	Tidak Fit Model
11	Tidak Fit Model	Fit Model	Fit Model
12	Fit Model	Fit Model	Tidak Fit Model
13	Tidak Fit Model	Fit Model	Tidak Fit Model
14	Tidak Fit Model	Fit Model	Tidak Fit Model
15	Fit Model	Fit Model	Fit Model
16	Fit Model	Fit Model	Tidak Fit Model
17	Tidak Fit Model	Tidak Fit Model	Fit Model
18	Fit Model	Fit Model	Fit Model
19	Fit Model	Fit Model	Fit Model
20	Tidak Fit Model	Tidak Fit Model	Fit Model
21	Tidak Fit Model	Fit Model	Fit Model
22	Fit Model	Fit Model	Fit Model
23	Tidak Fit Model	Fit Model	Fit Model
24	Fit Model	Tidak Fit Model	Fit Model
25	Tidak Fit Model	Tidak Fit Model	Tidak Fit Model
26	Tidak Fit Model	Fit Model	Fit Model
27	Tidak Fit Model	Fit Model	Tidak Fit Model
28	Tidak Fit Model	Fit Model	Fit Model
29	Tidak Fit Model	Fit Model	Tidak Fit Model
30	Fit Model	Fit Model	Fit Model
<b>Fit Model</b>	<b>13</b>	<b>25</b>	<b>20</b>
<b>Tidak Fit Model</b>	<b>17</b>	<b>5</b>	<b>10</b>

Tabel 3 di atas menunjukkan bahwa model 2 PL merupakan model yang paling banyak memiliki kecocokan dengan butir soal. Sehingga, penentuan parameter butir yang berupa soal-soal model TIMSS menggunakan model 2 PL. Adapun dua parameter penting yang dihasilkan pada model 2 PL yaitu parameter diskriminasi (a) dan parameter kesulitan (b). Berikut rincian karakteristik butir hasil analisis dengan model 2-PL melalui program Bilog-MG ditampilkan pada tabel 4 di bawah ini.

*Tabel 4. Hasil Analisis Model 2-PL menggunakan Bilog-MG*

Copyright © 2025

*Buana Matematika* :

Jurnal Ilmiah Matematika dan Pendidikan Matematika

p-ISSN : 2088-3021

e-ISSN : 2598-8077

Item	Daya Beda (Slope)		Tingkat Kesukaran (Threshold)		Chi-Square	
	<i>a</i>	Keterangan	<i>b</i>	Keterangan	Prob	Keterangan
Item 1	0.923	Baik	-0.734	Baik	0.5260	Fit model
Item 2	1.503	Baik	-0.546	Baik	0.4630	Fit model
Item 3	1.958	Baik	-0.369	Baik	0.8837	Fit model
Item 4	1.029	Baik	-0.800	Baik	0.0753	Fit model
Item 5	1.394	Baik	-1.088	Baik	0.5245	Fit model
Item 6	1.073	Baik	-1.547	Baik	0.3564	Fit model
Item 7	0.821	Baik	-0.444	Baik	0.7345	Fit model
Item 8	0.722	Baik	-0.706	Baik	0.0546	Fit model
Item 9	0.731	Baik	-0.273	baik	0.8779	Fit model
Item 10	0.699	Baik	-0.029	Baik	0.0290	Tidak Fit model
Item 11	1.383	Baik	-0.657	Baik	0.8347	Fit model
Item 12	1.150	Baik	-1.559	Baik	0.4828	Fit model
Item 13	1.163	Baik	-0.453	Baik	0.2479	Fit model
Item 14	1.228	Baik	-0.459	Baik	0.1536	Fit model
Item 15	0.549	Baik	-0.603	Baik	0.6294	Fit model
Item 16	0.819	Baik	-0.533	Baik	0.1352	Fit model
Item 17	1.129	Baik	0.150	Baik	0.0327	Tidak Fit model
Item 18	0.795	Baik	-0.388	baik	0.1861	Fit model
Item 19	0.787	Baik	-0.243	baik	0.8490	Fit model
Item 20	1.149	Baik	0.284	Baik	0.0236	Tidak Fit model
Item 21	1.499	Baik	-0.559	baik	0.5448	Fit model
Item 22	1.141	Baik	-0.399	Baik	0.0676	Fit model
Item 23	1.451	Baik	-0.872	Baik	0.4232	Fit model
Item 24	0.824	Baik	-0.685	Baik	0.0203	Tidak Fit model
Item 25	0.221	Baik	0.598	Baik	0.0441	Tidak Fit model
Item 26	0.352	Baik	2.514	Tidak Baik	0.3694	Fit model
Item 27	0.236	Baik	-5.057	Tidak Baik	0.1537	Fit model
Item 28	0.352	Baik	0.352	Baik	0.0605	Fit model
Item 29	0.285	Baik	0.215	baik	0.9030	Fit model
Item 30	0.318	Baik	2.612	Tidak Baik	0.6686	Fit model
<b>Total</b>	<b>Baik</b>	<b>30</b>		<b>27</b>		
	<b>Tidak Baik</b>	<b>-</b>		<b>3</b>		
<b>Jumlah Item Fit</b>						<b>20</b>

Berdasarkan kriteria teori respon butir bahwa Indeks daya beda yang dikatakan baik yaitu apabila memiliki nilai rentang antara 0 – 2. Tabel diatas

Copyright © 2025

*Buana Matematika* :

Jurnal Ilmiah Matematika dan Pendidikan Matematika

p-ISSN : 2088-3021

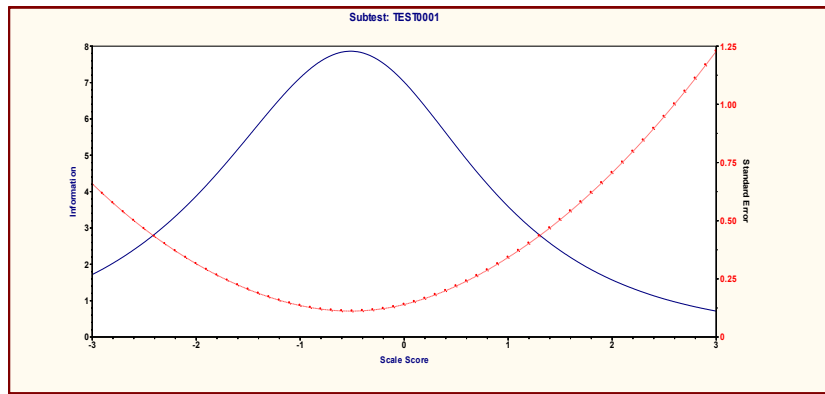
e-ISSN : 2598-8077

menunjukkan bahwa analisis model 2-PL seluruh itemnya yang terdiri dari 30 butir soal model TIMSS memiliki daya beda yang baik. Berikutnya, kriteria tingkat kesulitan dikatakan baik jika nilai interval yang didapatkan yaitu berkisar antara  $-2 \leq \theta \leq 2$ . Berdasarkan penjelasan diatas dapat disimpulkan, bahwa untuk tingkat kesulitan butir didapatkan 27 butir tergolong kategori baik sedangkan butir lainnya yaitu sebanyak 3 butir tergolong tidak baik. Berikut akan disajikan tingkat kesulitan butir dari butir yang termudah ke butir yang sulit.

Tabel 5. Daya Beda dan Tingkat Kesulitan Butir Matematika Model TIMSS

Item (i)	a	b	Item (i)	a	b
27	0.236	-5.057	13	0,753	-0,453
12	1.150	-1.559	7	1,346	-0,444
6	1.073	-1.547	22	1,194	-0,399
5	1.394	-1.088	18	1,150	-0,388
23	1,297	-0,872	3	1,052	-0,369
4	1,240	-0,8	9	0,899	-0,273
1	2,191	-0,734	19	0,662	-0,243
8	1,141	-0,706	10	1,067	-0,029
24	2,052	-0,685	17	1,231	0,15
11	1,280	-0,657	29	0,895	0,215
15	1,471	-0,603	20	0,859	0,284
21	1,388	-0,559	28	0,951	0,352
2	0,507	-0,546	25	0,452	0,598
16	1,445	-0,533	26	0,544	2,514
14	1,046	-0,459	30	0,391	2,612

Tabel 5 di atas menunjukkan bahwa indeks daya beda pada butir soal-soal model TIMSS yang digunakan sebagai penelitian dengan nilai parameter a berada antara 0 sampai 2. Adapun parameter lainnya yaitu tingkat kesulitan (b), yang dikatakan baik yaitu jika nilai yang didapatkan berada pada -2 hingga +2. Output yang didapatkan berdasarkan hasil analisis menggunakan bilog menunjukkan bahwa dari 30 butir soal terdapat satu butir yang terlalu mudah, dua butir tergolong sulit, sebagaimana tercantum dalam tabel.



Gambar 2. Grafik Plot

Berdasarkan grafik plot pada Gambar 2 di atas, garis biru menggambarkan kurva informasi tes, sementara garis merah menunjukkan tingkat kesalahan pengukuran. Kedua garis tersebut membentuk kurva fungsi informasi tes dan kesalahan pengukuran. Sumbu vertikal sebelah kiri merepresentasikan total informasi tes, sedangkan sumbu vertikal kanan menunjukkan besar kesalahan pengukuran. Dari grafik terlihat bahwa garis merah berpotongan dengan garis biru pada rentang kemampuan sekitar  $-2,4$  hingga  $+1,4$ . Dengan demikian, dapat disimpulkan bahwa instrumen tes yang digunakan paling optimal atau paling sesuai untuk mengukur kemampuan siswa yang berada dalam kisaran  $-2,4$  sampai  $+1,4$ .

### Simpulan

Berdasarkan hasil analisis data dengan menggunakan program Bilog-MG dapat disimpulkan bahwa instrumen tes matematika model TIMSS yang digunakan dalam penelitian ini telah memenuhi asumsi unidimensional yang berarti semua soal mengukur satu konstruk utama yaitu kemampuan matematika. Dengan terpenuhinya asumsi unidimensional tersebut, maka secara umum asumsi independensi lokal dianggap telah terpenuhi, karena jawaban peserta pada salah satu butir soal tidak dipengaruhi oleh jawaban butir soal yang lain. Model Item Response Theory yang paling tepat untuk menganalisis soal dalam penelitian ini adalah model logistik dua parameter (2-PL). Model ini menghasilkan dua parameter utama yaitu daya pembeda (parameter  $a$ ) dan tingkat kesukaran butir soal (parameter  $b$ ) yang menunjukkan kemampuan soal dalam membedakan siswa berdasarkan tingkat kemampuannya dan tingkat kesukaran relatif masing-masing butir soal. Dengan analisis model 2-PL seluruh itemnya yang terdiri dari 30 butir soal model TIMSS memiliki daya beda yang baik. Adapun untuk tingkat kesulitan butir didapatkan 27 butir tergolong kategori baik sedangkan butir lainnya yaitu sebanyak 3 butir tergolong tidak baik. Output software Bilog MG bahwa dari 30 butir soal ditemukan satu butir yang terlalu mudah, dua butir tergolong sulit dan lainnya tergolong sedang. Berdasarkan

Copyright © 2025

*Buana Matematika* :

Jurnal Ilmiah Matematika dan Pendidikan Matematika

p-ISSN : 2088-3021

e-ISSN : 2598-8077

kurva informasi tes, instrumen menunjukkan tingkat informasi tertinggi dan galat ukur rendah pada rentang kemampuan antara -2,4 sampai dengan +1,4. Hal ini menunjukkan bahwa instrumen ini paling efektif dan akurat untuk siswa dengan kemampuan pada rentang tersebut. Temuan penelitian ini mengindikasikan bahwa guru perlu melakukan analisis butir soal secara berkala guna memastikan kualitas instrumen evaluasi yang digunakan, baik dari segi validitas maupun tingkat kesukarannya

### Ucapan Terima Kasih

Peneliti menyampaikan rasa terima kasih yang setulus-tulusnya kepada seluruh pihak yang telah terlibat memberikan dukungan, bantuan, serta kontribusi selama penyusunan artikel ini. Ucapan terima kasih ditujukan kepada seluruh stakeholder yang terlibat dan pihak sekolah di SMP yang ada di kota Baubau khususnya yang menjadi tempat penelitian yang telah mengizinkan dan memudahkan dalam pengambilan data. Semoga karya tulis ini dapat memberikan sumbangsih dalam pengembangan ilmu pengetahuan, pengembangan butir-butir soal yang dijadikan alat tes dan menjadi rujukan yang bermanfaat bagi penelitian-penelitian sejenis di masa yang akan datang.

### Daftar Pustaka

- El Hasbi, A. Z., Huda, N., & Hermina, D. (2024). Teknik Pengolahan Tes Pada Bidang Pendidikan (Tes Tertulis, Tes Lisan, Tes Perbuatan). *Al Furqan : Jurnal Agama, Sosisal, Dan Budaya*, 3(3), 1428–1449.
- Elvira, M., & Samsul, H. (2016). Karakteristik butir soal ujian semester dan kemampuan siswa SMA di kabupaten Muaro Jambi. *Jurnal Evaluasi Pendidikan*, 4(1), 58–68.
- Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on. *Educational Measurement: Issues and Practice*, 12(3), 38–47.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals of Item Response Theory*.
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Muarrif Journal of Arabic Education*, 1(1), 34–44.
- Kartowagiran, B. (2013). Optimalisasi evaluasi pembelajaran teknik mesin melalui logic model untuk meningkatkan soft skills lulusan. *Yogyakarta: Pidato Pengukuhan Guru Besar Dalam Bidang Ilmu Evaluasi Pembelajaran Teknik Mesin Pada Fakultas Teknik Universitas Negeri Yogyakarta Tanggal, 10*.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis

- butir soal ujian akhir sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274–284.
- Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications.
- Mulyani, S., Efendi, R., & Ramalis, T. R. (2021). Karakterisasi tes keterampilan pemecahan masalah fisika berdasarkan teori respon butir. *Jurnal Pendidikan Dan Ilmu Fisika*, 1(1), 1–14.
- Safitri, I., Lestarani, D., Imtikhanah, R. D. N. W., Akbarini, N. R., Sari, M. W., Fitrah, M., & Hapsan, A. (2024). *Teori Pengukuran dan Evaluasi*. CV. Ruang Tentor.
- Sarea, M. S., & Ruslan, R. (2019). Karakteristik Butir Soal: Classical Test Theory vs Item Response Theory? *Didaktika: Jurnal Kependidikan*, 13(1), 1–16.
- Susilawati, D. (2018). *Tes dan pengukuran*. UPI Sumedang Press.
- Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.

### Riwayat Hidup Penulis



Muhammad Naim Mahmud, S.Pd., M.Pd. Lahir di Mola Utara, pada tanggal 10 April 1994. Pada Tahun 2012 melanjutkan Pendidikan Sarjana (S1) pada Program Studi Pendidikan Matematika di Universitas Dayanu Ikhsanuddin (UNIDAYAN) Baubau selesai tahun 2016, dan pada Tahun 2018 melanjutkan Program Magister (S2) jurusan Penelitian dan Evaluasi Pendidikan di Universitas Negeri Yogyakarta (UNY) dan selesai pada tahun 2021. Pada Tahun 2022 menjadi Dosen Luar Biasa (LB) di Universitas Muhammadiyah Buton (UMB). Tahun 2024-sekarang aktif sebagai Dosen Tetap di Universitas Muslim Buton pada Program Studi Pendidikan Matematika.



Afudin La Arua, S.Pd., M.Pd. Lahir di Nania pada tanggal 3 Februari 1989. Pendidikan sarjana di Universitas Dayanu Ikhsanuddin Baubau selesai pada tahun 2012, kemudian melanjutkan pendidikan magister di Pasca Sarjana Pendidikan Matematika Universitas Negeri Surabaya dan selesai pada tahun 2017. Mengawali karir menjadi Dosen di Universitas Muslim Buton dan berikan tanggung jawab menjadi Kaprodi pendidikan Matematika UMU Buton pada tahun 2019-2021 dan sebagai Direktur Akademik pada tahun 2021-2022 dan sekarang menjadi wakil rektor II di Universitas Muslim Buton (UMU)



Samron, S.Pd., M.Pd. Lahir disuatu daerah pesisir penghasil/pembudidaya Mutiara Mabe di Kota Baubau yaitu Palabusa pada tanggal 18 Desember 1987, memulai karir pendidikan SD Negeri 1 Bataraguru (Tamat 2000), SLTP Negeri 9 Baubau (Tamat 2003), dan SMK Negeri 2 Baubau (Tamat 2006). Pada Tahun 2008 melanjutkan Pendidikan Sarjana (S1) pada Program Studi Pendidikan Matematika Universitas Dayanu Ikhsanuddin (UNIDAYAN) Baubau selesai tahun 2012, dan pada Tahun 2014 melanjutkan Program Magister (S2) Pendidikan Matematika di Universitas Haluoleo (UHO) dan selesai pada tahun 2016. Pada Tahun 2016-2019 menjadi Dosen Luar Biasa (LB) di Universitas Dayanu Ikhsanuddin (Unidayan). Tahun 2019-sekarang aktif sebagai Dosen Tetap di Universitas Muslim Buton pada Program Studi Pendidikan Matematika.



Irawati Windani, S.Pd., M.Pd, Lahir di Ambon, 19 Desember 1991, Tahun 2014 LuLus Pendidikan S1 pada program Studi Pendidikan Matematika di Universitas Dayanu Ikhsanuddin Baubau dan pada tahun 2018 LuLus Pendidikan S2 pada program studi yg sama di Universitas Negeri Yogyakarta. Sekarang aktif sebagai Dosen Tetap di Universitas Muslim Buton pada Program Studi Pendidikan Matematika.