

**PERBANDINGAN AKURASI PENILAIAN KETERAMPILAN  
BERPIKIR KRITIS MATEMATIS SISWA SMP DENGAN  
PENDEKATAN *ITEM RESPONSE THEORY* (IRT)**

**TESIS**



**Disusun Oleh:**

**MUZIYATI HANIM**

**NIM. P2A923004**

**PROGRAM MAGISTER PENDIDIKAN MATEMATIKA  
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN  
UNIVERSITAS JAMBI  
JULI 2025**

**PERBANDINGAN AKURASI PENILAIAN KETERAMPILAN  
BERPIKIR KRITIS MATEMATIS SISWA SMP DENGAN  
PENDEKATAN *ITEM RESPONSE THEORY* (IRT)**

**TESIS**

**Diajukan sebagai salah satu syarat untuk memperoleh  
gelar Magister Pendidikan Matematika**



**Disusun Oleh:**

**MUZIYATI HANIM**

**NIM. P2A923004**

**PROGRAM MAGISTER PENDIDIKAN MATEMATIKA  
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN  
UNIVERSITAS JAMBI  
JULI 2025**

## HALAMAN PERSETUJUAN

### HALAMAN PERSETUJUAN

Tesis yang berjudul “Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)” merupakan Tesis Program Studi Magister Pendidikan Matematika Universitas Jambi yang disusun oleh Muziyati Hanim, Nomor Induk Mahasiswa P2A923004 telah diperiksa oleh pembimbing dan disetujui untuk diuji.

Jambi, Juli 2025

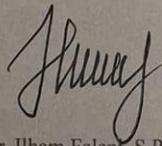
Pembimbing I



Dr. Dra. Mujahidawati, M.Si.  
NIP. 196411201990012001

Jambi, Juli 2025

Pembimbing II



Dr. Ilham Falani, S.Pd., M.Si.  
NIP. 198905182022031010

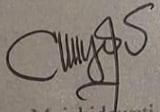
## HALAMAN PENGESAHAN

### HALAMAN PENGESAHAN

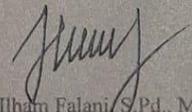
Tesis yang berjudul "Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory (IRT)*" Tesis Program Studi Magister Pendidikan Matematika yang disusun oleh **Muziyati Hanim, Nomor Induk Mahasiswa P2A923004** telah dipertahankan di depan Dewan Penguji pada tanggal 07 Juli 2025.

Tim Penguji	
Ketua	Dr. Dra. Mujahidawati, M.Si. NIP. 196411201990012001
Sekretaris	Dr. Ilham Falani, S.Pd., M.Si. NIP. 198905182022031010
Anggota	1. Drs. Jefri Marzal, M.Sc., D.I.T. NIP. 196806021993031004 2. Prof. Dr. Drs. Haryanto, M.Kes. NIP. 196803131993031003 3. Dr. Duano Sapta Nusantara, S.Pd. NIP. 199611022024061001

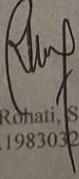
Mengetahui,  
Ketua Tim Penguji

  
Dr. Dra. Mujahidawati, M.Si.  
NIP. 196411201990012001

Jambi, Juli 2025  
Mengetahui,  
Sekretaris Tim Penguji

  
Dr. Ilham Falani, S.Pd., M.Si.  
NIP. 198905182022031010

Mengetahui,  
Ketua Program Studi  
Magister Pendidikan Matematika

  
Dr. Rohati, S.Pd., M.Pd.  
NIP. 198303242006042003

## HALAMAN PERNYATAAN

### HALAMAN PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Muziyati Hanim  
Tempat, Tanggal Lahir : Balai Jaya, 16 Mei 2000  
NIM : P2A923004  
Program Studi : Magister Pendidikan Matematika

Menyatakan dengan sesungguhnya bahwa:

1. Seluruh data, informasi, interpretasi serta pernyataan dalam pembahasan dan kesimpulan yang disajikan dalam tesis ini, kecuali yang disebutkan sumbernya adalah merupakan hasil pengamatan, pengolahan, serta pemikiran saya melalui pengarahan dari pembimbing I dan pembimbing II yang ditetapkan.
2. Tesis yang saya tulis ini adalah asli dan belum pernah disajikan untuk mendapatkan gelar akademik baik di Universitas Jambi maupun di perguruan tinggi lainnya.

Pernyataan ini dibuat dengan sebenar-benarnya dan apabila dikemudian hari ditemukan adanya bukti ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pembatalan gelar yang saya peroleh melalui tesis ini.

Demikian pernyataan ini saya buat dengan penuh kesadaran dan tanggung jawab serta ditandatangani diatas materai.

Jambi, Juli 2025  
Yang Membuat Pernyataan



Muziyati Hanim  
NIM. P2A923004

## ABSTRAK

Hanim, Muziyati. 2025. Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT): Tesis, Magister Pendidikan Matematika, FKIP Universitas Jambi, Pembimbing: (I) Dr. Dra. Mujahidawati (II) Dr. Ilham Falani, S.Pd., M.Si.

**Kata Kunci:** *Item Response Theory*, Keterampilan Berpikir Kritis Matematis, Penilaian

Keterampilan berpikir kritis matematis merupakan kemampuan yang penting dalam menghadapi tantangan abad ke-21. Namun, pengukuran keterampilan ini di sekolah masih belum optimal karena keterbatasan instrumen yang valid dan reliabel, sehingga hasil penilaian belum sepenuhnya akurat. Penelitian ini bertujuan membandingkan tingkat akurasi tiga model *Item Response Theory* (IRT) dalam mengestimasi parameter keterampilan berpikir kritis matematis siswa. Model yang digunakan meliputi *Partial Credit Model* (PCM), *Generalized Partial Credit Model* (GPCM), dan *Graded Response Model* (GRM).

Penelitian ini menggunakan metode kuantitatif dengan desain *ex post facto*. Data dikumpulkan dari 106 respons siswa SMP, yaitu 57 siswa dari SMPN 14 Muaro Jambi, 15 siswa dari SMPN 45 Muaro Jambi, 13 siswa dari SMPN 49 Muaro Jambi, dan 21 siswa dari SMPN 53 Muaro Jambi. Instrumen yang digunakan berupa tes uraian dengan dua butir soal, disusun berdasarkan indikator FRISCO (*Focus, Reason, Inference, Situation, Clarity, and Overview*). Analisis data dilakukan menggunakan perangkat lunak PARSCALE 4.1. Tingkat presisi diukur melalui varians hasil estimasi, dengan uji hipotesis menggunakan uji F.

Hasil penelitian menunjukkan bahwa GPCM memiliki tingkat presisi dan akurasi lebih tinggi dibandingkan PCM dan GRM. GPCM menghasilkan varians estimasi terkecil serta fungsi informasi tes yang paling optimal, dengan nilai *standard error of measurement* (SEM) sebesar 0,354. Adapun SEM pada PCM sebesar 0,417 dan GRM sebesar 0,406. GPCM direkomendasikan sebagai model paling efektif untuk mengestimasi parameter keterampilan berpikir kritis matematis, terutama pada tes esai. Saran penelitian berikutnya adalah mengembangkan sistem penilaian berbasis teknologi yang menerapkan GPCM.

## ABSTRACT

Hanim, Muziyati. 2025. *Comparison of Accuracy of Assessment of Mathematical Critical Thinking Skills of Junior High School Students with Item Response Theory (IRT) Approach*: Thesis, Master of Mathematics Education, FKIP Jambi University, Supervisor (I) Dr. Dra. Mujahidawati (II) Dr. Ilham Falani, S.Pd., M.Si.

**Keywords:** *Assessment, Item Response Theory, Mathematical Critical Thinking Skills*

*Mathematical critical thinking skills are an important ability in facing the challenges of the 21st century. However, the measurement of these skills in schools is still not optimal due to the limitations of valid and reliable instruments, so the assessment results are not fully accurate. This study aims to compare the accuracy of three Item Response Theory (IRT) models in estimating the parameters of students' mathematical critical thinking skills. The models used include Partial Credit Model (PCM), Generalized Partial Credit Model (GPCM), and Graded Response Model (GRM).*

*This research used quantitative method with ex post facto design. Data were collected from 106 junior high school student responses, namely 57 students from SMPN 14 Muaro Jambi, 15 students from SMPN 45 Muaro Jambi, 13 students from SMPN 49 Muaro Jambi, and 21 students from SMPN 53 Muaro Jambi. The instrument used was a description test with two items, prepared based on the FRISCO (Focus, Reason, Inference, Situation, Clarity, and Overview) indicators. Data analysis was carried out using PARSCALE 4.1 software. The level of precision was measured through the variance of the estimation results, with hypothesis testing using the F test.*

*The results showed that GPCM had higher precision and accuracy than PCM and GRM. GPCM produces the smallest estimation variance and the most optimal test information function, with a standard error of measurement (SEM) value of 0.354. The SEM for PCM was 0.417 and GRM was 0.406. GPCM is recommended as the most effective model for estimating mathematical critical thinking skill parameters, especially on essay tests. The next research suggestion is to develop a technology-based assessment system that applies GPCM.*

## KATA PENGANTAR

Alhamdulillah, puji Syukur kepada Allah ﷻ, atas limpahan Rahmat dan karunia-Nya sehingga penulis diberi kesempatan, Kesehatan, kekuatan, dan kesabaran untuk menyelesaikan tesis yang berjudul **“Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)”**. Selanjutnya, shalawat dan salam kepada Nabi Muhammad ﷺ yang telah memberikan teladan terbaik bagi umat manusia, serta senantiasa diharapkan syafa’atnya di dunia dan akhirat kelak. Penulisan Tesis ini dimaksudkan untuk memenuhi salah satu syarat akademik guna mendapatkan gelar Magister Pendidikan pada Pascasarjana Universitas Jambi. Penulis menyadari sepenuhnya bahwa penyelesaian tesis ini banyak melibatkan pihak yang telah memberikan motivasi, baik moril maupun materil, untuk itu pada bagian ini penulis menyampaikan terima kasih dan penghargaan kepada:

1. Bapak Prof. Dr. Helmi, S.H.,M.H., selaku Rektor Universitas Jambi.
2. Bapak Prof. Dr. Supian, S.Ag., M.Ag, selaku Dekan Fakultas Keguruan dan Ilmu Pendidikan.
3. Ibu Dr. Rohati, S.Pd., M.Pd, selaku ketua Program Studi Magister Pendidikan Matematika dan dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.
4. Ibu Dr. Dra. Mujahidawati, M.Si., selaku dosen pembimbing I yang banyak memberikan arahan dan bimbingan serta masukan yang sangat berharga dalam penyusunan Tesis ini.
5. Bapak Dr. Ilham Falani, S.Pd., M.Si., selaku dosen pembimbing II yang banyak memberikan arahan dan bimbingan serta masukan yang sangat berharga dalam penyusunan Tesis ini.
6. Bapak Dr. Jefri Marzal, M.Sc. D.I.T., selaku dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.
7. Bapak Prof. Dr. Drs. Haryanto, M.Kes. selaku dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.
8. Bapak Dr. Duano Sapta Nusantara, S.Pd., selaku dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.
9. Dengan penuh rasa syukur dan hormat, saya mengucapkan terima kasih yang

sebesar-besarnya kepada Beasiswa Unggulan atas dukungan pendanaan yang telah diberikan dalam menyelesaikan studi Magister saya di bidang Pendidikan Matematika.

10. Bapak dan Ibu dosen, khususnya dosen Program Studi Magister Pendidikan Matematika Universitas Jambi yang telah memberikan ilmu kepada penulis selama perkuliahan, dan staf Magister Pendidikan Matematika yang telah memberikan bantuan selama ini. Semoga ilmu dan bantuan yang diberikan menjadi amal ibadah yang baik.
  11. Dengan penuh rasa syukur, saya mengucapkan terima kasih yang sebesar-besarnya kepada pihak SMPN 53 Muaro Jambi, SMPN 45 Muaro Jambi, SMPN 49 Muaro Jambi dan SMPN 14 Muaro Jambi yang telah memberikan izin dan kesempatan kepada saya untuk melaksanakan penelitian di lingkungan sekolah tersebut. Dukungan dan kerja sama yang diberikan oleh kepala sekolah, para guru, serta seluruh siswa sangat berperan penting dalam kelancaran proses pengumpulan data penelitian ini.
  12. Kedua orang tua tercinta, Ayah Hamdan, Ibu Halimah Tusaqdhah, Adik Aminah Tuzariyah dan keluarga besar yang telah memberikan motivasi tiada henti dalam menyelesaikan tesis ini.
  13. Seluruh rekan Magister Pendidikan Matematika Angkatan 2023 dan semua pihak yang telah membantu dalam menyelesaikan tesis ini yang tidak dapat penulis sebutkan namanya satu persatu.
  14. Seluruh adik-adik kos yang telah menemani dalam penyusunan ini dan memberikan semangat terutama Meliyani Rumiyanthi yang selalu diminta bantuan untuk menemani bimbingan.
- Akhirnya, semoga Allah ﷻ berkenan membalas segala kebaikan dan amal semua pihak yang telah membantu penuliah dalam menyelesaikan tugas akhir tesis di Magister Pendidikan Matematika Universitas Jambi. Semoga tesis ini bermanfaat bagi pengembangan ilmu pengetahuan dan teknologi.

Jambi, 11 Juli 2025

Penulis

## DAFTAR ISI

	<b>Halaman</b>
<b>HALAMAN SAMPUL</b>	
<b>HALAMAN JUDUL</b>	
<b>HALAMAN PERSETUJUAN</b>	
<b>HALAMAN PENGESAHAN</b>	
<b>HALAMAN PERNYATAAN</b>	
<b>ABSTRAK .....</b>	<b>i</b>
<b>ABSTRACT .....</b>	<b>ii</b>
<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>DAFTAR ISI.....</b>	<b>v</b>
<b>DAFTAR TABEL .....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>viii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>ix</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1    Latar Belakang Masalah.....	1
1.2    Identifikasi Masalah .....	9
1.3    Rumusan Masalah .....	9
1.4    Tujuan Penelitian .....	10
1.5    Manfaat Hasil Penelitian .....	10
1.6    Batasan Penelitian .....	11
1.7    Definisi Operasional.....	11
<b>BAB II KAJIAN TEORI .....</b>	<b>13</b>
2.1    Deskripsi Teoritik.....	13
2.1.1    Penilaian Dalam Pendidikan .....	13
2.1.2    Penyekoran Politomi .....	15
2.1.3 <i>Item Response Theory</i> (IRT) .....	16
2.1.4    Model Politomi IRT .....	20
2.1.5    Keterampilan Abad 21 .....	23
2.1.6    Keterampilan Berpikir Kritis Matematis.....	24
2.1.7    Keterkaitan IRT terhadap Peningkatan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis .....	29
2.1.8    Tinjauan Materi Rasio.....	32
2.2    Penelitian yang Relevan .....	33
2.3    Kerangka Berpikir.....	35
2.4    Hipotesis Penelitian.....	37
<b>BAB III METODE PENELITIAN .....</b>	<b>38</b>
3.1    Waktu dan Tempat Penelitian .....	38
3.2    Metode Penelitian.....	38
3.3    Data Penelitian .....	39
3.4    Populasi dan Sampel .....	39
3.5    Prosedur Penelitian.....	40
3.6    Pengolahan Data.....	43
3.7    Teknik Analisis Data.....	45
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....</b>	<b>54</b>
4.1    Hasil Validasi Tes oleh Ahli/Pakar .....	54

4.2	Hasil Uji Coba Tes .....	55
4.3	Hasil Uji Asumsi IRT .....	56
4.4	Deskripsi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis .....	59
4.5	Deskripsi Hasil Estimasi Parameter Butir.....	64
4.6	Fungsi Informasi Tes .....	70
4.7	Pengujian Hipotesis Penelitian.....	72
4.8	Pembahasan Hasil Penelitian .....	77
4.8.1	Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan PCM, GPCM dan GRM .....	78
4.8.2	Perbandingan Akurasi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis antara Model PCM dan GPCM .....	83
4.8.3	Perbandingan Akurasi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis antara model GPCM dan GRM .....	85
4.8.4	Perbandingan Akurasi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis antara Model PCM dan GRM.....	88
<b>BAB V PENUTUP .....</b>		<b>90</b>
5.1	Kesimpulan .....	90
5.2	Implikasi.....	91
5.3	Saran.....	91
<b>DAFTAR PUSTAKA .....</b>		<b>93</b>
<b>LAMPIRAN.....</b>		<b>102</b>
<b>RIWAYAT HIDUP .....</b>		<b>136</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1.1 Rekap Survei Guru .....	2
2.1 Aspek Kemampuan Berpikir .....	27
2.3 Indikator Keterampilan Berpikir Kritis Matematis .....	28
2.4 Keterkaitan Keterampilan Berpikir Kritis Matematis Indikator FRISCO dengan Model IRT .....	31
3.1 Rancangan perbandingan Variansi Hasil Estimasi .....	38
3.2 Jumlah Peserta Tes .....	39
3.3 Biodata Ahli untuk Validasi Tes .....	41
3.4 Kriteria Keterampilan Berpikir Kritis Siswa Berdasarkan Nilai Logit ( $\theta$ ) ....	52
4.1 Rekapitulasi Penilaian Butir Soal oleh Ahli.....	55
4.2 Uji Reliabilitas Tes Keterampilan Berpikir Kritis Matematis .....	55
4.3 Varians Total Tes Keterampilan Berpikir Kritis Matematis .....	57
4.4 Uji Normalitas Keterampilan Berpikir Kritis Matematis Siswa .....	59
4.5 Distribusi Frekuensi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan PCM.....	60
4.6 Distribusi Frekuensi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GPCM.....	61
4.7 Distribusi Frekuensi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GRM .....	62
4.8 Hasil Estimasi Parameter Butir Tes Keterampilan Berpikir Kritis Matematis dengan PCM.....	64
4.9 Hasil Estimasi Parameter Butir Tes Keterampilan Berpikir Kritis Matematis dengan GPCM.....	66
4.10 Hasil Estimasi Parameter Butir Tes Keterampilan Berpikir Kritis Matematis dengan GRM .....	68
4.11 Ringkasan Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis .	72
4.12 Ringkasan Deskripsi Hasil Estimasi Keterampilan Berpikir Kritis Matematis .....	73
4.13 Pengujian Hipotesis Pertama antara PCM dan GPCM .....	73
4.14 Pengujian Hipotesis Kedua antara GPCM dan GRM .....	74
4.15 Pengujian Hipotesis Ketiga antara PCM dan GRM.....	75
4.16 Rangkuman Hasil Uji Hipotesis.....	76

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
2.1 Hubungan antara Taksonomi Bloom dan Indikator Berpikir Kritis (FRISCO) .....	28
2.2 Kerangka Berpikir.....	36
4.1 <i>Scree Plot</i> Unidimensi Keterampilan Berpikir Kritis Matematis .....	58
4.2 Histogram Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan PCM.....	63
4.3 Histogram Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GPCM.....	64
4.4 Histogram Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GRM.....	64
4.5 Kurva Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis PCM.	70
4.6 Kurva Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis GPCM .....	71
4.7 Kurva Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis GRM	71

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. Syntax PARSCALE PCM.....	102
2. Syntax PARSCALE GPCM.....	103
3. Syntax PARSCALE GRM .....	104
4. Form Validasi Instrumen Tes Keterampilan Berpikir Kritis Matematis.....	105
5. Jawaban Siswa .....	118
6. Uji Kecocokan Model Luaran PARSCALE ph.2 .....	120
7. Estimasi Keterampilan Berpikir Kritis Matematis Siswa Luaran PARSCALE ph. 3.....	121
8. Penyekoran Tes Keterampilan Berpikir Kritis Matematis .....	124
9. Surat Selesai Bimbingan Validasi Instrumen Keterampilan Berpikir Kritis Matematis.....	125
10. Surat Izin Selesai Penelitian.....	128
11. Dokumentasi Penelitian .....	132
12. Link Drive Langkah-langkah GPCM.....	135

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Penilaian merupakan proses penting dalam dunia pendidikan, yaitu mengumpulkan dan mengolah informasi guna mengukur pencapaian hasil belajar siswa. Penilaian tidak hanya berfungsi untuk mengetahui tingkat penguasaan materi, tetapi juga menjadi alat untuk memberikan umpan balik yang konstruktif bagi siswa dan guru, sehingga pembelajaran dapat terus ditingkatkan (Septiani et al., 2023). Oleh karena itu, guru dituntut memiliki keterampilan yang baik dalam merancang alat penilaian yang efektif.

Dalam menyusun alat penilaian, guru harus memperhatikan aspek materi, konstruksi, dan bahasa secara cermat. Hal ini penting untuk memastikan hasil penilaian yang akurat. Tanpa ketelitian tersebut, penilaian bisa menghasilkan informasi yang keliru atau tidak valid, sehingga tidak mencerminkan kemampuan siswa yang sebenarnya. Oleh karena itu, penilaian harus dirancang melalui analisis butir soal yang mempertimbangkan validitas dan reliabilitas, serta memperhatikan tingkat kesukaran soal secara proporsional (Saputri et al., 2023). Penilaian yang akurat akan membantu guru mengidentifikasi kebutuhan belajar siswa secara tepat dan memberikan intervensi yang sesuai, sehingga setiap siswa memiliki kesempatan yang optimal untuk berkembang.

Salah satu keterampilan penting yang perlu dinilai secara akurat dalam pembelajaran matematika adalah keterampilan berpikir kritis. Keterampilan ini merupakan bagian dari 6C Pendidikan Abad 21, yaitu *Critical Thinking*, *Creativity*, *Collaboration*, *Communication*, *Computational*, dan *Compassion* (Inganah et al., 2023; Mujahidawati et al., 2023; Partono et al., 2021; Sugianto et al., 2022). Critical

thinking atau keterampilan berpikir kritis menjadi salah satu kompetensi utama yang harus dikembangkan untuk menghadapi tantangan zaman.

Keterampilan berpikir kritis dalam matematika mencakup kemampuan untuk menganalisis, mengevaluasi, dan menyelesaikan masalah secara logis dan sistematis. Siswa yang memiliki keterampilan ini mampu mengidentifikasi asumsi, menemukan pola, membuat generalisasi, serta menguji hipotesis berdasarkan data yang tersedia. Mereka juga mampu mempertanyakan solusi yang diberikan, mencari alternatif lain, dan membuat keputusan yang tepat melalui analisis mendalam (Nuryadi et al., 2024). Keterampilan ini tidak hanya penting untuk memahami konsep matematika secara lebih bermakna, tetapi juga sangat relevan untuk memecahkan masalah dalam kehidupan nyata secara analitis dan terstruktur. Oleh karena itu, penilaian yang akurat terhadap keterampilan berpikir kritis sangat penting untuk membantu siswa berkembang menjadi individu yang mampu berpikir secara reflektif, kritis, dan kreatif. Dengan demikian, keterampilan berpikir kritis yang kuat akan tercapai jika didukung oleh penilaian yang tepat dan akurat.

Namun, berdasarkan hasil survei terhadap guru-guru Sekolah Menengah Pertama (SMP) di wilayah Kecamatan Bahar Utara Kabupaten Muaro Jambi, ditemukan beberapa temuan yang menunjukkan masih lemahnya implementasi penilaian keterampilan berpikir kritis (KBK) matematis secara sistematis.

**Tabel 1.1 Rekap Survei Guru**

No	Aspek yang Disurvei	Persentase Responden (%)	Keterangan
1	Guru pernah mengukur KBK	80	Sebagian besar guru pernah melakukan pengukuran
2	Guru menggunakan rubrik/indikator penilaian KBK	40	Sebagian kecil guru menggunakan indikator yang jelas
3	Guru mengetahui pendekatan pengukuran modern (IRT)	20	Sebagian besar belum mengetahui pendekatan IRT

No	Aspek yang Disurvei	Persentase Responden (%)	Keterangan
4	Guru tertarik mendapatkan pelatihan tentang IRT	100	Seluruh responden menunjukkan ketertarikan
5	Tes sudah dirancang berdasarkan indikator kemampuan	30	Mayoritas soal belum berbasis indikator
6	Validasi soal dilakukan secara formal	20	Validasi masih dilakukan secara informal

Dari total responden survei, hanya sekitar 80% guru yang menyatakan pernah mengukur keterampilan berpikir kritis matematis siswa. Akan tetapi, saat ditelusuri lebih lanjut, hanya sebagian kecil dari mereka yang telah menggunakan rubrik atau indikator penilaian yang spesifik. Sebanyak 60% guru tidak menggunakan rubrik, dan beberapa guru yang menyatakan menggunakan indikator sederhana seperti "*siswa mampu mengerjakan soal cerita*".

Dalam hal teknik penilaian, metode yang paling dominan digunakan adalah diskusi kelompok dan observasi langsung, namun masih belum sepenuhnya terintegrasi dengan standar penilaian yang objektif. Sementara itu, lebih dari 80% guru tidak mengetahui adanya pendekatan modern dalam pengukuran kemampuan, seperti *Item Response Theory* (IRT), meskipun semua responden menyatakan tertarik untuk mendapatkan pelatihan mengenai pendekatan tersebut.

Hanya sebagian kecil guru yang berusaha memastikan soal yang mereka buat sudah valid, biasanya hanya melalui diskusi sederhana dengan rekan sejawat, tanpa melalui proses yang resmi dan terstandar. Selain itu, sekitar 70% guru mengakui bahwa soal tes tertulis yang mereka gunakan belum disusun berdasarkan indikator keterampilan berpikir kritis matematis. Kondisi ini tentu memengaruhi ketepatan hasil penilaian, karena soal yang tidak dirancang dengan baik bisa menghasilkan data yang tidak akurat. Oleh karena itu, sebelum melakukan penilaian, penting dilakukan pengukuran terlebih dahulu sebagai langkah awal, yaitu proses untuk

memperoleh data atau angka dari hasil belajar siswa. Pengukuran ini menjadi dasar dalam penilaian, karena penilaian tidak bisa dilakukan tanpa adanya data hasil pengukuran yang jelas.

Pengukuran dan penilaian merupakan dua proses yang berkesinambungan. Pengukuran adalah kegiatan pemberian atau penetapan angka pada objek yang diukur, disesuaikan dengan kriteria objek tersebut, sedangkan penilaian merupakan kegiatan membandingkan hasil pengukuran sifat atau objek dengan suatu acuan yang relevan sehingga diperoleh kuantitas suatu objek yang bersifat kualitatif (Teluma & Rivaie, 2019). Pengukuran dilaksanakan terlebih dahulu, yang menghasilkan skor dan dari hasil pengukuran dapat dilaksanakan penilaian. Seorang pendidik pada saat melaksanakan pengukuran kemampuan berpikir kritis perlu adanya alat ukur yang meliputi tes atau non tes.

Tes merupakan suatu teknik atau cara yang digunakan dalam rangka melaksanakan kegiatan pengukuran, yang didalamnya terdapat berbagai pertanyaan, pernyataan, atau serangkaian tugas yang harus dikerjakan atau dijawab oleh siswa untuk mengukur aspek perilaku atau pengetahuan siswa (Faiz et al., 2022; Zainal, 2020). Ada dua bentuk tes jika dilihat dari sistem penskorannya yaitu tes objektif (pilihan ganda) dan tes subjektif (uraian). Menurut Arini et al., (2024) tes subjektif seperti uraian dan penugasan proyek lebih sesuai untuk mengevaluasi keterampilan berpikir matematis tingkat tinggi seperti penalaran, pemecahan masalah kompleks, dan komunikasi matematis. Sejalan dengan hal tersebut Yokia Marsita & Ishaq Nuriadin (2024) menyatakan bahwa tes subjektif lebih baik dalam mengukur kemampuan berpikir kritis matematis siswa. Namun, tes kemampuan berpikir kritis matematis yang biasanya diterapkan di sekolah-sekolah umumnya

terbatas pada tugas-tugas atau soal pilihan ganda, sehingga tidak cukup efektif dalam mengukur aspek-aspek berpikir kritis yang lebih kompleks seperti interpretasi, analisis, dan evaluasi (Amanda & Nusantara, 2021; Habibah & Marlina, 2023). Tes uraian memiliki karakteristik yang berbeda dengan tes objektif. Skor pada tes uraian umumnya menggunakan skor politomi dengan beberapa kategori berdasarkan kriteria tertentu. Selain itu, diperlukan analisis butir soal tes dalam pembuatan soal untuk memastikan kualitasnya, sehingga dapat memberikan informasi yang berguna mengenai karakteristik dan analisis empiris dari butir tes tersebut.

Analisis butir tes memiliki beberapa manfaat, di antaranya: 1) membantu dalam mengidentifikasi butir-butir soal yang baik dan kurang baik; 2) dapat memperoleh informasi untuk memperbaiki tes yang digunakan baik dari segi isi dan konstruksinya; 3) memperoleh gambaran tentang keadaan tes yang disusun. Analisis butir soal ini dapat ditinjau dari segi kualitatif dan kuantitatif (Arikunto, 2013). Tinjauan dari segi kuantitatif yang populer adalah dengan menggunakan *Classical Test Theory* (CTT) atau Teori Tes Klasik. Sampai saat ini, CTT masih menjadi pendekatan yang paling banyak digunakan dalam menganalisis butir soal.

Penelitian yang dilakukan oleh Ciptari et al., (2024), Fernanda & Hidayah (2020), dan Hayat (2021) menunjukkan adanya kelemahan yang dimiliki oleh CTT yaitu kurang efektif dalam mengukur tingkat kesukaran butir soal dan daya pembeda dibandingkan dengan *modern test*. Kelemahan tersebut memicu teori baru yang lebih memadai, yaitu *modern test* (tes modern) yang dikenal juga sebagai teori respons butir (TRB) atau IRT. Meskipun CTT telah menjadi dasar, saling ketergantungan empiris dan asumsi keandalan yang seragam mengharuskan

pergeseran ke arah model yang lebih kompleks seperti IRT untuk akurasi pengukuran yang lebih baik (Frey, 2020).

IRT adalah analisis butir berdasarkan model parameter linguistik. Dalam IRT terdapat 3 model utama, yaitu model dikotomi, politomi, dan multidimensi. Pada setiap model terdiri dari beberapa jenis. IRT Dikotomi terdiri dari tiga model, yaitu model: satu parameter linguistik (1 PL), dua parameter linguistik (2 PL), dan tiga parameter linguistik (3PL). IRT politomi terdiri dari beberapa model, diantaranya: satu parameter (b) yang disebut *Partial Credit Model* (PCM), dua parameter (a dan b) yang disebut *Generalized Partial Credit Model* (GPCM), dan dua parameter (a dan b) yang disebut *Graded Response Model* (GRM) (Isgiyanto, 2013).

PCM dikembangkan oleh Masters pada tahun 1982, memungkinkan butir soal memiliki beberapa kategori skor, bukan hanya benar atau salah, sehingga lebih fleksibel dalam mengukur berbagai tingkat kinerja siswa. Setiap langkah atau bagian yang benar mendapat poin tertentu. Misalnya, sebuah soal dapat memiliki skor 0 (salah), 1 (sebagian benar), dan 2 (benar). Kemudian, PCM merupakan model satu parameter dalam IRT yang hanya mempertimbangkan tingkat kesukaran butir soal, tanpa memasukkan daya beda sebagai parameter dalam estimasi kemampuan peserta tes (Zain et al., 2022).

GPCM yang diperkenalkan oleh Muraki pada tahun 1992, memperluas PCM dengan mengizinkan diskriminasi butir yang berbeda-beda, sehingga dapat mengakomodasi variasi dalam seberapa baik suatu butir soal membedakan antara siswa dengan kemampuan yang berbeda. Misalnya, sebuah soal matematika dapat memiliki skor 0 (salah), 1 (sebagian benar), 2 (hampir benar), dan 3 (benar). Kedua model ini memberikan alat yang lebih canggih untuk mengukur kemampuan siswa

secara lebih tepat dan memberikan wawasan yang lebih mendalam tentang karakteristik butir soal tes. GPCM salah satu pendekatan dalam IRT yang digunakan untuk menganalisis data ujian dengan skala skor bertingkat. Penelitian oleh (Dewanti et al., 2021) menunjukkan bahwa GPCM sangat berguna dalam tes literasi matematika yang memiliki beberapa tingkat kesulitan, sehingga memberikan hasil yang lebih representatif. Menurut Harsana & Lumenyela (2023) berbeda dengan model klasik yang hanya mempertimbangkan jawaban benar atau salah, GPCM dalam IRT memberikan estimasi yang lebih baik terhadap kemampuan siswa dalam soal berbasis *constructed-response* dibandingkan model klasik. GPCM dapat memberikan gambaran lebih akurat mengenai kemampuan seseorang berdasarkan respons yang diberikan, sehingga hasil analisis lebih representatif dan informatif (Luo, 2018).

Model IRT lainnya yang digunakan adalah GRM. GRM, yang juga dikenal sebagai model respons kategoris terurut, berfokus pada kategori politomus yang terurut, di mana responden diharapkan mendapatkan berbagai tingkat skor, seperti 0 hingga 4 poin. Kategori yang digunakan dalam model ini adalah 0, 1, 2, 3, dan 4, dan kategori-kategori tersebut terurut GRM dan GPCM sama-sama merupakan model dua parameter dalam Item Response Theory yang memperhitungkan tingkat kesukaran dan daya beda butir soal. Namun, GPCM memiliki fleksibilitas lebih tinggi karena mampu mengakomodasi respons tidak lengkap atau nilai kosong tanpa mengganggu proses estimasi parameter. Fleksibilitas ini menjadikan GPCM lebih adaptif terhadap data politomus yang kompleks, terutama dalam asesmen keterampilan berpikir kritis matematis, di mana variasi respons dan kategori skor sering kali tidak seragam. Sementara GRM lebih cocok untuk data dengan struktur

kategori yang teratur dan lengkap, GPCM memberikan keunggulan dalam menangani ketidakteraturan data dan tetap menghasilkan estimasi yang presisi (Aji & Retnawati, 2024). Data yang diperoleh berbentuk data politomi yaitu data yang memiliki interval tertentu, sehingga analisis yang digunakan pada penelitian ini adalah analisis IRT politomi dengan model PCM, GPCM dan GRM.

IRT memiliki beberapa kelebihan dibandingkan dengan CTT. Penelitian yang dilakukan oleh Aini et al., (2024) menunjukkan bahwa IRT memiliki keunggulan dibandingkan dengan CTT, terutama dalam hal akurasi pengukuran kemampuan. Dinyatakan bahwa hubungan antara kemampuan individu dan respons mereka terhadap butir soal dipengaruhi oleh karakteristik tiap butir, seperti tingkat kesulitan dan daya beda. Ini memperkuat ide bahwa probabilitas menjawab benar bergantung pada keterampilan subjek. Kelebihan yang terdapat dalam analisis butir soal menggunakan IRT menjadi salah satu alasan peneliti untuk melakukan penelitian ini, selain dari keadaan lapangan yang menunjukkan belum adanya alat untuk mengukur keterampilan berpikir kritis matematis secara khusus.

Penelitian ini berfokus pada akurasi penilaian keterampilan berpikir kritis matematika dengan menerapkan pendekatan IRT untuk analisis data politomi. Penerapan pendekatan IRT dapat meningkatkan akurasi penilaian variabel laten yang dalam penelitian ini adalah keterampilan berpikir kritis siswa. Model ini cocok untuk analisis data yang dikumpulkan menggunakan skala penilaian, skala respon tipe Likert, atau data respons lainnya dengan kategori yang berurutan. Selain itu, model ini mempertahankan karakteristik utama analisis IRT untuk respons politomi (Mertasari & Candiasa, 2023). Model politomi memberikan fleksibilitas dan keakuratan yang tinggi dalam menganalisis data dengan berbagai format respons.

Dengan demikian, penerapan pendekatan IRT dapat menjadi langkah penting dalam meningkatkan akurasi penilaian keterampilan berpikir kritis siswa. Berdasarkan latar belakang yang telah dipaparkan di atas, penelitian ini berfokus pada penggunaan pendekatan IRT dalam meningkatkan akurasi penilaian keterampilan berpikir kritis matematis siswa.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang yang telah dipaparkan, dapat diidentifikasi permasalahan sebagai berikut:

1. Penilaian keterampilan berpikir kritis siswa dalam bidang matematika mungkin belum akurat atau belum optimal. Terdapat kebutuhan untuk alat atau metode penilaian yang lebih presisi untuk mengukur keterampilan ini.
2. Implementasi IRT dalam konteks penilaian pendidikan, khususnya untuk keterampilan berpikir kritis matematis. IRT adalah model yang digunakan untuk menganalisis data tes berdasarkan kinerja individu pada butir soal tertentu, tetapi mungkin belum diterapkan secara luas atau efektif dalam penilaian keterampilan berpikir kritis matematis di tingkat SMP.
3. Mencari cara untuk meningkatkan akurasi penilaian yang ada. Ini mencakup tantangan dalam menerapkan IRT untuk memastikan bahwa penilaian benar-benar mencerminkan keterampilan berpikir kritis matematis siswa dengan lebih akurat.

## **1.3 Rumusan Masalah**

Berdasarkan uraian pada latar belakang masalah, yang menjadi rumusan masalah pada penelitian adalah sebagai berikut:

1. Bagaimana keakuratan penilaian keterampilan berpikir kritis matematis siswa menggunakan model PCM dibandingkan dengan model GPCM?
2. Bagaimana keakuratan penilaian keterampilan berpikir kritis matematis siswa menggunakan model GPCM dibandingkan dengan model GRM?
3. Bagaimana keakuratan penilaian keterampilan berpikir kritis matematis siswa menggunakan model PCM dibandingkan dengan model GRM?

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah yang dikemukakan, maka tujuan penelitian yang akan dilakukan adalah:

1. Untuk mengetahui keakuratan penilaian keterampilan berpikir kritis matematis siswa menggunakan model PCM dibandingkan dengan model GPCM.
2. Untuk mengetahui keakuratan penilaian keterampilan berpikir kritis matematis siswa menggunakan model GPCM dibandingkan dengan model GRM.
3. Untuk mengetahui keakuratan penilaian keterampilan berpikir kritis matematis siswa menggunakan model PCM dibandingkan dengan model GRM.

#### **1.5 Manfaat Hasil Penelitian**

Diharapkan setelah melakukan penelitian ini dapat memberikan manfaat sebagai berikut:

1. Hasil dari penelitian ini diharapkan dapat memberikan manfaat bagi siswa. Dengan menyadari tingkat keterampilan berpikir kritis mereka, siswa dapat lebih termotivasi untuk terus belajar dan mencari cara untuk mengembangkan kemampuan analitis dan evaluatif mereka.
2. Hasil dari penelitian ini dapat memberikan manfaat bagi guru. Dengan pemahaman yang mendalam tentang keterampilan berpikir matematis siswa,

guru dapat merancang materi pelajaran dan kegiatan yang sesuai dengan tingkat kemampuan siswa, memastikan semua siswa mendapatkan tantangan yang tepat dan dukungan yang mereka butuhkan.

3. Hasil dari penelitian ini dapat memberikan sumbangan berupa kajian teori tentang implementasi IRT dalam peningkatan akurasi penilaian keterampilan berpikir kritis matematis siswa SMP.

### **1.6 Batasan Penelitian**

Adapun yang menjadi batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan merupakan data hasil tes keterampilan berpikir kritis siswa SMP. Tes yang digunakan berupa tes format esai pada materi rasio.
2. Data hasil tes dianalisis dengan 3 model IRT yaitu PCM, GPCM dan GRM.
3. Penelitian ini hanya berfokus pada hasil estimasi parameter keterampilan berpikir kritis siswa yang dihasilkan oleh ketiga model IRT tersebut.

### **1.7 Definisi Operasional**

1. Akurasi

Akurasi adalah ukuran seberapa dekat suatu pengukuran dengan nilai yang sebenarnya dari kuantitas yang diukur.

2. Penilaian

Penilaian diartikan sebagai proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar siswa.

3. Keterampilan Berpikir Kritis Matematis

Keterampilan berpikir kritis matematis merupakan kemampuan intelektual yang terstruktur untuk menganalisis, mengevaluasi, dan mensintesis informasi matematis secara logis dan reflektif.

#### 4. *Item Response Theory*

*Item Response Theory* (IRT) adalah pendekatan dalam psikometri yang memodelkan hubungan antara kemampuan laten responden dan respons terhadap subitem, yang dapat dianalisis melalui tiga model utama yaitu PCM, GPCM, dan GRM.

#### 5. *Item*

*Item*, butir soal, dan tes merujuk pada satuan soal yang dikembangkan untuk mengukur keterampilan berpikir kritis matematis. Setiap item terdiri atas satu atau lebih subitem yang merepresentasikan indikator spesifik dari keterampilan yang diukur, sehingga subitem berfungsi sebagai komponen analitik yang merefleksikan dimensi kognitif atau prosedural tertentu yang terdapat dalam item utama.

## **BAB II KAJIAN TEORI**

### **2.1 Deskripsi Teoritik**

#### **2.1.1 Penilaian Dalam Pendidikan**

Penilaian adalah alat penting yang digunakan untuk mengukur kompetensi siswa dalam berbagai bidang studi. Istilah penilaian dalam bahasa Indonesia dapat bersinonim dengan evaluasi (*evaluation*) dan kini juga populer istilah asesmen (*assessment*). Penilaian diartikan sebagai proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar siswa (PP No.19 Th 2005). Melalui penilaian, pendidik dapat menilai sejauh mana siswa memahami materi, menguasai keterampilan, dan mampu menerapkan pengetahuan mereka dalam situasi nyata.

Menurut Zainal (2020) penilaian adalah proses yang digunakan untuk menilai hasil belajar siswa berdasarkan data yang telah dikumpulkan melalui pengukuran. Pengukuran adalah langkah awal yang melibatkan pengumpulan data objektif menggunakan berbagai instrumen seperti tes dan non tes. Hasil pengukuran kemudian dianalisis dan diinterpretasikan dalam proses penilaian untuk memberikan gambaran yang lebih komprehensif tentang kemampuan dan perkembangan siswa. Penilaian ini bertujuan untuk memberikan umpan balik yang konstruktif dan membantu dalam perencanaan strategi pengajaran yang lebih efektif. Selanjutnya menurut Satria (2024) penilaian yang berfungsi untuk mengoptimalkan proses pembelajaran, penilaian bukan lagi hanya untuk mengukur hasil belajar, namun juga untuk memahami kebutuhan belajar siswa dan pemberian ruang pada optimalisasi pelaksanaan asesmen autentik.

Penilaian hasil belajar siswa dilakukan dengan berbagai teknik sesuai dengan kompetensi yang hendak dinilai. Penilaian kompetensi sikap dilakukan melalui observasi, penilaian diri, penilaian “teman sejawat” (*peer assessment*) oleh siswa, dan jurnal. Penilaian kompetensi pengetahuan dilakukan melalui tes tulis, tes lisan, dan penugasan. Penilaian kompetensi keterampilan dilakukan melalui penilaian kinerja berupa kinerja praktik, proyek, dan penilaian portofolio (Farman et al., 2021).

Pendekatan penilaian hasil belajar menekankan pada pengukuran tingkat berpikir siswa dari yang rendah sampai dengan yang tinggi; menggunakan pertanyaan mendalam, bukan sekadar hafalan; mengukur proses kerja siswa, bukan hanya hasil kerja siswa; dan menggunakan portofolio pembelajaran siswa. Tingkat berpikir siswa dikembangkan mulai dari tingkat berpikir mengingat, memahami, menerapkan, menganalisis, menilai, dan berkreasi. Pertanyaan dalam pembelajaran ditekankan pada jenis pertanyaan bagaimana dan mengapa yang bersifat rasional, bukan pada pertanyaan apa, dimana, siapa, dan kapan yang bersifat faktual.

Berdasarkan beberapa uraian diatas peneliti menyimpulkan penilaian hasil belajar siswa adalah proses sistematis untuk mengukur dan mengevaluasi kemampuan, keterampilan, pengetahuan, dan sikap siswa setelah menyelesaikan suatu kegiatan pembelajaran. Proses ini melibatkan pengumpulan data melalui berbagai instrumen seperti tes, kuis, proyek, presentasi, dan observasi. Hasil dari penilaian ini digunakan untuk menentukan sejauh mana siswa telah mencapai tujuan pembelajaran yang telah ditetapkan, memberikan umpan balik yang konstruktif kepada siswa dan guru, serta untuk membuat keputusan terkait perbaikan dan pengembangan strategi pembelajaran di masa mendatang. Penilaian

hasil belajar siswa memberikan informasi yang akurat dan komprehensif mengenai kemajuan akademik dan keterampilan siswa, yang selanjutnya digunakan untuk meningkatkan kualitas pendidikan secara keseluruhan.

### **2.1.2 Penyekoran Politomi**

Pemberian skor (*scoring*) merupakan langkah pertama dalam proses pengolahan hasil tes, yaitu proses pengubahan jawaban soal tes menjadi menjadi angka-angka dengan kata lain pemberian skor itu merupakan tindakan kuantifikasi terhadap jawaban-jawaban yang diberikan oleh *testee* dalam suatu tes hasil belajar.

Cara pemberian skor terhadap hasil tes belajar pada umumnya disesuaikan dengan bentuk soal-soal yang dikeluarkan dalam tes tersebut, apakah tes uraian (*essay*) ataukah tes objektif (*objective*). Untuk soal-soal objektif biasanya setiap jawaban benar diberi skor (satu) dan setiap jawaban yang salah diberi skor (nol) total skor diperoleh dengan menjumlahkan skor yang diperoleh dari semua soal. Untuk soal-soal *essay* dalam penskorannya biasanya digunakan cara memberi bobot (*weighting*) kepada setiap soal menurut tingkat kesukarannya atau banyak sedikitnya unsur yang harus terdapat dalam jawaban yang dianggap paling baik. Misalnya: untuk soal no. 1 diberi skor maksimal 4, untuk soal no. 3 diberi skor maksimum 6, untuk skor no. 3 skor maksimum 4, untuk skor no. 5 skor maksimum 10 dan seterusnya, pemberian skor seperti ini juga dapat disebut penskoran politomi.

Menurut Osterlind (2002) kategori penyekoran politomi ditentukan berdasarkan rubrik penilaian yang memiliki skala kebenaran yang hierarki. Jenjang kategori penyekoran ini ditentukan berdasarkan kriteria yang telah ditetapkan oleh

penyusun tes. Kategori respon yang dapat digunakan, contohnya kategori “2” pada skala 1 sampai dengan 4. Semakin banyak gradasi dalam kategori, memungkinkan penyekoran ini untuk dapat mengukur kemampuan peserta tes dalam mendemonstrasikan pengetahuannya secara parsial pada tiap butir. Berbeda dengan penskoran dikotomi yang hanya mengklasifikasikan jawaban sebagai benar atau salah, penskoran politomi memungkinkan penilaian yang lebih halus dengan mempertimbangkan berbagai tingkat kesulitan dan kualitas jawaban peserta. IRT merupakan pendekatan utama dalam penskoran politomi. IRT mengkaji hubungan antara kemampuan peserta tes dan respons mereka terhadap butir-butir soal, memungkinkan estimasi yang lebih akurat terhadap kemampuan individu.

Dalam konteks ini, model PCM, GPCM, GRM menjadi salah satu model yang paling umum digunakan untuk penskoran politomi. Ketiga model ini memperkenalkan kategori skor yang mencerminkan langkah-langkah dalam menjawab soal secara bertahap.

### **2.1.3 *Item Response Theory (IRT)***

Teori Respons Butir (*Item Response Theory-IRT*) dikenal juga sebagai Teori Ciri Laten (*Latent Trait Theory-LTT*) atau lengkungan karakteristik butir (*Item Characteristic Curve-ICC*) atau Fungsi Karakteristik Butir (*Items Characteristic Function-ICF*) (Naga, 2012). Pada dasarnya, teori ini ingin memperbaiki kelemahan yang terdapat pada teori tes klasik yakni adanya sifat *group dependent* dan *item dependent*. Hal ini berarti indeks daya pembeda, tingkat kesulitan, dan koefisien reliabilitas tes tergantung kepada yang mengerjakan tes tersebut, selain dipengaruhi oleh soal atau butir yang ada (Fatkhudin et al., 2016).

IRT adalah sebuah model probabilitas yang berusaha menjelaskan hubungan antara respons seseorang terhadap sebuah butir dengan variabel laten (kemampuan/*ability* atau sifat/*trait*) yang diukur oleh tes tersebut (Fajrianti et al., 2016). Dalam hal ini, respons atau kinerja peserta tes merupakan hal yang dapat diamati (*observable*) sedangkan sifat atau kemampuan merupakan sesuatu yang tidak tampak (*unobservable*) yang mendasari kinerja pada tes tersebut. Variabel laten (misalnya kemampuan) dalam IRT disebut sebagai theta ( $\theta$ ) dimana semakin tinggi  $\theta$  yang dimiliki seseorang maka semakin tinggi pula probabilitasnya untuk menjawab soal dengan benar (Baker & Kim, 2017). Probabilitas seorang peserta tes berpikir kritis untuk menjawab benar sebuah butir merupakan fungsi dari tingkat kemampuan berpikir kritis yang dimiliki dengan tingkat kesulitan dari butir tersebut.

IRT merupakan kerangka kerja statistik yang digunakan untuk memodelkan hubungan antara kemampuan laten individu dan respons mereka terhadap butir tes. Salah satu model IRT yang paling dikenal adalah Rasch model, yang mengasumsikan bahwa probabilitas jawaban benar pada suatu butir bergantung pada perbedaan antara kemampuan individu dan kesulitan butir tersebut (Keliat et al., 2023; Nasution & Falani, 2024). Rasch model sering dikenal untuk data dikotomi, dimana respons dari suatu butir adalah benar atau salah (Falani et al., 2022; Ramalisa et al., 2023). Rasch model menyediakan dasar yang kuat untuk analisis data tes yang dapat diperluas ke berbagai model IRT lainnya. Dengan kemajuan dalam IRT, muncul model penskoran politomi seperti PCM yang menangani *item* dengan beberapa kategori jawaban. PCM memungkinkan pengukuran yang lebih terperinci dan akurat dengan memberikan kredit parsial

untuk respons yang tidak sepenuhnya benar, mencerminkan pemahaman yang lebih mendalam tentang kemampuan individu dibandingkan hanya menggunakan jawaban benar atau salah.

Model politomus dari IRT, digunakan untuk menganalisis skor *item* yang memiliki respons multi kategori. Contoh skor politomi adalah skala Likert untuk skala sikap atau juga respons tes *essay* (uraian). Banyak model yang menggunakan penskoran politomus, diantaranya adalah: GRM, PCM, GPCM, dan *Rating Scale Model* (RSM) (Embretson & Reise, 2000). Semua model tersebut mendasarkan pada asumsi bahwa respons (jawaban) peserta tes pada suatu *item* tes tergantung kemampuan peserta.

Penerapan analisis menggunakan IRT mensyaratkan terpenuhinya tiga asumsi dasar agar hasil yang diperoleh valid dan dapat diinterpretasikan secara akurat.

Pertama, asumsi unidimensi. Asumsi yang utama adalah unidimensi yang berarti variable yang diukur hanya memiliki satu dimensi atribut atau dikenal sebagai kemampuan ( $\theta$ ). Dalam kenyataan tidak mudah memperoleh atribut variable yang unidimensi. Sehingga dalam prakteknya, unidimensi dicapai melalui adanya satu dimensi yang dominan. Unidimensi terjadi pada setiap satu butir (bukan kelompok butir). Adapun beberapa cara yang sering digunakan untuk menguji syarat unidimensi yakni indeks berdasarkan residual pada penyelesaian unidimensi, analisis faktor melalui eigen value matriks korelasi interbutir, dan uji *Stout*. Pada penelitian ini uji asumsi dilakukan dengan analisis faktor, uji asumsi ini memerlukan uji kecukupan sampel. Selanjutnya analisis faktor dilakukan dengan menggunakan eigen value varians kovarians inter-butir (Kogar & Kogar, 2018).

Kedua asumsi independensi lokal. Asumsi IRT yang kedua adalah independensi lokal (*local independence*). Naga (2012) menyebutkan independensi lokal berarti hasil ukur pada satu responden tidak bergantung kepada hasil ukur pada responden lain dan hasil ukur pada satu butir tidak bergantung kepada hasil ukur pada butir lain. Uji independensi lokal digunakan untuk memastikan bahwa jawaban peserta didik terhadap satu soal tidak dipengaruhi oleh jawaban mereka terhadap soal lain. Ketiga, uji invariansi parameter bertujuan untuk melihat apakah karakteristik soal seperti tingkat kesulitan atau kemampuan membedakan peserta tetap konsisten di berbagai kelompok peserta yang memiliki kemampuan serupa. Jika parameter soal berubah antar kelompok, bisa jadi soal tersebut tidak adil atau cenderung menguntungkan kelompok tertentu. Selanjutnya, menurut Bahar & Retnawati (2022) jika uji unidimensi terpenuhi, maka uji independensi lokal dan invariansi parameter juga dianggap terpenuhi.

Setelah ketiga uji diatas terdapat uji kecocokan model yang harus terpenuhi sebelum dilakukan estimasi parameter kemampuan. Pemilihan model yang digunakan untuk kalibrasi tergantung pada butir-butir dalam tes. Setelah menentukan pilihan model yang digunakan untuk menganalisis butir, selanjutnya perlu dilakukan uji kecocokan model terhadap data. Menurut Alagoz (2000) dan Baker & Kim (2017) salah satu uji statistik yang dapat digunakan adalah uji *statistic Chi-Square*, yakni dengan menghitung  $-2\log L$ , di mana L adalah likelihood. Indikator kecocokan butir dengan uji *Chi-Square* adalah jika nilai *Chi-Square* hitung yang dihasilkan oleh butir melebihi nilai *Chi-Square* tabel. Menurut Hambleton et al., (1991) menggunakan lebih dari satu model untuk menganalisis

data tes lalu membandingkan hasil uji kecocokan model-model tersebut merupakan langkah yang sangat membantu dalam pemilihan model yang paling tepat.

Setelah asumsi-asumsi terpenuhi, prosedur analisis dengan model-model IRT dapat ditetapkan. Ada beberapa kategori model analisis dalam pendekatan IRT yang dapat digunakan berdasarkan spesifikasi penyekoran yang digunakan. Pada penelitian ini pembahasan difokuskan pada 3 model yaitu PCM, GPCM, GRM.

## 2.1.4 Model Politomi IRT

### 2.1.4.1 *Partial Credit Model*

PCM merupakan pengembangan dari model rasch butir dikotomi yang diterapkan pada butir politomi, jika  $i$  adalah butir politomi dalam kategori skor , 0, 1, 2, ...,  $m$ , probabilitas dari individu  $n$  skor  $x$  pada butir  $i$  yang diwujudkan dalam persamaan berikut.

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^x(\theta_n - \delta_{ij})\right]}{\sum_{r=0}^{mi} \left[\exp\left[\sum_{j=0}^r(\theta_n - \delta_{ij})\right]\right]} \quad (1)$$

Keterangan:

$P_{ix}$  = probabilitas peserta dengan kemampuan  $\theta$  yang memperoleh skor kategori  $x$  pada butir  $i$

$\theta_n$  = level trait individu (lokasi trait individu pada kontinum trait laten)

$\delta_{ij}$  = persimpangan garis antar kategori ( $j$ ) pada Butir ( $i$ ).

(Sumintono & Widhiarso, 2015)

### 2.1.4.2 *Generalized Partial Credit Model*

Selain model PCM model lainnya dalam pendekatan IRT untuk data politomi adalah GPCM. Model ini disebut GPCM karena Muraki (1992) mengembangkan model ini sebagai bentuk umum dari PCM yang dikembangkan oleh Masters (1982)

yang merupakan pengembangan dari Model Rasch (DeMars, 2010). GPCM merupakan Model IRT yang dikembangkan untuk menganalisis data politomus, di mana respons diskor 0, 1, ..., k, dengan k adalah kategori skor tertinggi untuk butir.

GPCM mempunyai probabilitas penskoran kategori k pada butir. Model GPCM dalam menghitung estimasi kemampuan peserta memperhitungkan tingkat kesulitan pada setiap langkah-langkahnya. Model GPCM sendiri mirip dengan PCM. Tetapi pada model GPCM terdapat parameter daya pembeda ( $a$ ) dan faktor skala ( $D$ ) yang merupakan faktor skala yang telah ditetapkan sebesar 1,7. Model matematisnya sebagai berikut.

$$P_{jk}(\theta) = \frac{\exp \sum_{v=0}^k Z_{jh}(\theta)}{\sum_h^m \exp \sum_{v=0}^k Z_{jh}(\theta)} \quad (2)$$

$$k = 0, 1, 2, 3, \dots, m$$

Dengan  $Z_{jh}(\theta)$  persamaan sebagai berikut:

$$Z_{jh}(\theta) = D_{aj}(\theta - b_{jh}) = D_{aj}(\theta - b_{jh} + d_h) \quad (3)$$

$$d_{j0} = 0$$

Dengan,

- $P_{jk}(\theta)$  = probabilitas peserta tes dengan kemampuan  $\theta$  yang memperoleh nilai skor kategori k pada butir j.
- $a_j$  = merupakan indeks daya beda pada butir ke j.
- $\theta$  = Kemampuan peserta tes
- $D$  = Faktor skala sebesar 1,7.
- $b_{jh}$  = Indeks kesukaran pada kategori k butir ke j.
- $b_j$  = Indeks kesukaran pada lokasi butir j.
- $(d_h)$  = Parameter pada kategori k. (Aji & Retnawati, 2024)

### 2.1.4.3 Graded Response Model

GRM merupakan sebuah model politomi IRT yang dikembangkan untuk butir dengan karakteristik kategorinya terurut seperti skala Likert (Baker & Kim, 2017). Kategori-kategori ini mencakup *partial credit* yang diberikan sesuai dengan tingkat pencapaian peserta tes dalam menyelesaikan masalah.

Model ini tidak membutuhkan jumlah kategori yang sama untuk tiap butirnya. Misalkan tiap butir  $i$  memiliki alternatif jumlah respons  $k_i$ . Tiap butir dideskripsikan oleh parameter  $a$  dan parameter  $\beta$  ambang batas (*threshold*)  $m_i$ , di mana  $m_i = k - 1$  adalah jumlah kategori dalam butir dikurangi satu. Misalkan kemampuan  $\theta$ ,  $X$  adalah variabel bilangan acak yang menunjukkan respons bertingkat butir ke I dan  $x = 1, \dots, k_i$  dinotasikan sebagai aktual respons. Peluang  $P_{ix}(\theta) = P_i(X = x|\theta)$  dari peserta tes dengan kemampuan  $\theta$  mendapatkan score  $x$  untuk butir I, dengan  $x = 1, \dots, k_i$  dapat dirumuskan sebagai berikut,

$$P_{ix}(\theta) = P_{ix}^* - P_{i(x+1)}^* \quad (4)$$

Di mana  $P_{ix}^*(\theta) = P_i(X > x|\theta)$  merepresentasikan peluang peserta tes memperoleh skor  $X$  atau lebih dari skor  $x$  berdasarkan pada tingkat kemampuan.  $P_{ix}^*(\theta)$  ditulis (Nering & Ostini, 2010),

$$P_{ix}^* = \frac{e^{Da_i(\theta - b_i(x-1))}}{1 + e^{Da_i(\theta - b_i(x-1))}} \quad (5)$$

Untuk  $x = 2, \dots, k_i$ . Untuk melengkapi definisi model, ditambahkan bahwa peluang untuk respons skor sama dengan atau lebih dari nilai terendah  $P_{i1}^* = 1$ .

Sedangkan peluang untuk respons sama dengan atau lebih dari nilai kategori tertinggi  $P_{i(ki+1)}^* = 0$ .

### **2.1.5 Keterampilan Abad 21**

Keterampilan abad 21 mencakup berbagai kemampuan yang esensial untuk menghadapi tantangan modern dalam era globalisasi dan teknologi. Salah satu keterampilan utama adalah literasi digital, yang melibatkan kemampuan menggunakan teknologi informasi dan komunikasi secara efektif dan efisien. Literasi digital mencakup penguasaan perangkat lunak, pemahaman tentang keamanan siber, dan kemampuan untuk mengevaluasi serta menyaring informasi yang diperoleh dari internet. Keterampilan ini penting agar individu dapat beradaptasi dengan cepat terhadap perubahan teknologi dan memanfaatkan peluang yang ada.

Selain literasi digital, keterampilan kolaborasi dan komunikasi juga sangat vital. Di abad 21, kemampuan untuk bekerja sama dalam tim yang multikultural dan lintas disiplin menjadi sangat penting. Kolaborasi yang efektif membutuhkan kemampuan berkomunikasi dengan jelas, mendengarkan dengan aktif, dan menghargai perbedaan pendapat. Di dunia kerja yang semakin terhubung secara global, kemampuan ini membantu individu untuk berkontribusi secara produktif dalam tim yang beragam, menyelesaikan masalah secara kolektif, dan mencapai tujuan bersama.

Keterampilan berpikir kritis dan kritis juga menjadi bagian tak terpisahkan dari keterampilan abad 21. Berpikir kritis melibatkan kemampuan untuk menganalisis informasi secara mendalam, mengevaluasi argumen, dan membuat keputusan yang berlandaskan bukti. Sementara itu, berpikir kritis melibatkan

kemampuan untuk menghasilkan ide-ide inovatif dan solusi baru yang out-of-the-box. Keterampilan ini membantu individu untuk menghadapi situasi yang kompleks dan dinamis dengan cara yang fleksibel dan adaptif. Dengan menggabungkan keterampilan berpikir kritis dan kritis, individu dapat memberikan kontribusi yang bernilai tinggi di berbagai bidang, mulai dari bisnis hingga sains dan seni.

## **2.1.6 Keterampilan Berpikir Kritis Matematis**

### **2.1.6.1 Pengertian Keterampilan Berpikir Kritis Matematis**

Dalam Kamus Besar Bahasa Indonesia (KBBI) keterampilan berasal dari kata “terampil” yang berarti cakap dalam menyelesaikan tugas, mampu dan cekatan. Sedangkan keterampilan adalah kecakapan untuk menyelesaikan tugas. Menurut Zahri et al., (2017) berpendapat bahwa keterampilan merupakan kepandaian melakukan suatu pekerjaan dengan cepat dan benar, dalam hal ini ruang lingkup keterampilan sangat luas yang melingkupi berbagai kegiatan antara lain, perbuatan, berpikir, berbicara, melihat, mendengar, dan lain sebagainya. Keterampilan adalah ukuran kemampuan yang dimiliki seseorang (Nasihudin & Hariyadin, 2021).

Pada era Revolusi Industri 4.0, kompetensi yang mencakup kreativitas (*creativity*), berkomunikasi (*communication*), berpikir kritis (*critical thinking*), pemecahan masalah (*problem solving*), dan berkolaborasi (*collaboration*), merupakan kompetensi-kompetensi yang harus dimiliki siswa (Fitri et al., 2025). Secara eksplisit, kemampuan-kemampuan tersebut tertuang pada capaian pembelajaran pada pendidikan anak usia dini, pendidikan dasar, dan pendidikan menengah.

Kemampuan berpikir kritis berasal dari kata “pikir” dan “kritis”. Menurut Kamus Besar Bahasa Indonesia (KBBI), kata “pikir” memiliki arti “akal budi, ingatan, dan angan-angan”, sedangkan “berpikir” memiliki arti “menggunakan akal budi untuk mempertimbangkan dan memutuskan sesuatu, menimbang-nimbang dalam ingatan”. Kata “Kritis” berasal dari bahasa Yunani “*kritikos*” atau “*criterion*”. “*kritikos*” berarti “pertimbangan”, sedangkan “*criterion*” berarti “ukuran baku atau standar”. Dalam KBBI, “kritis” berarti “sifat tidak mudah percaya, berusaha menemukan kesalahan dan tajam dalam menganalisis”. Dapat disimpulkan bahwa berpikir kritis merupakan suatu tindakan atau sikap di mana seseorang tidak mudah mempercayai informasi yang diterimanya.

Kemampuan berpikir kritis sebagai salah satu kompetensi yang harus dimiliki siswa, dimaknai sebagai suatu keterampilan yang seharusnya dimiliki oleh manusia modern agar bisa berkomunikasi dan bertahan hidup di era global (Susanti & Hernawati, 2022). Definisi lain menyatakan bahwa “*Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions, when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task*” (Halpern & Dunn, 2022). Definisi ini memiliki arti bahwa berpikir kritis adalah penggunaan keterampilan atau strategi kognitif yang meningkatkan kemungkinan hasil yang diinginkan. Ini digunakan untuk menggambarkan pemikiran yang bertujuan, beralasan, dan terarah pada tujuan, jenis pemikiran yang terlibat dalam memecahkan masalah, merumuskan kesimpulan, menghitung kemungkinan, dan

membuat keputusan, ketika pemikir menggunakan keterampilan yang cermat dan efektif untuk konteks tertentu dan jenis tugas berpikir tersebut.

Menurut Nuryadi et al., (2024) keterampilan berpikir kritis matematis merupakan kemampuan siswa untuk menganalisis, mengevaluasi, dan menyimpulkan informasi dalam konteks pemecahan masalah matematika yang melibatkan proses kognitif yang kompleks, termasuk interpretasi data, analisis situasi, dan pengambilan keputusan yang logis. Matematika adalah materi yang dapat dipahami melalui berpikir kritis, dan berpikir kritis dilatih melalui serangkaian proses pembelajaran matematika (D. Fitri et al., 2025; Leniati & Indarini, 2021). Sedangkan menurut Angraini & Wahyuni, (2021) keterampilan berpikir matematis merupakan kemampuan untuk mengidentifikasi pola dan hubungan dalam berbagai konsep matematika, mengevaluasi informasi dan membandingkan antara konsep-konsep yang ada, baik dalam bentuk pernyataan maupun representasi matematika lainnya, menyusun alasan yang jelas dan logis untuk memecahkan masalah matematika serta mengkomunikasikan pemecahan masalah secara sistematis.

Dari beberapa pengertian peneliti menyimpulkan bahwa keterampilan berpikir kritis matematis mengacu pada kemampuan individu untuk menganalisis dan mengevaluasi masalah matematis secara mendalam dan komprehensif. Keterampilan ini melibatkan serangkaian proses intelektual seperti mengidentifikasi pola, menginterpretasi data, menghubungkan konsep-konsep yang relevan, dan membuat keputusan yang didasarkan pada penalaran logis dan bukti yang valid.

### 2.1.6.2 Indikator Keterampilan Berpikir Kritis Matematis

Adapun tingkatan berpikir manusia menurut taksonomi Bloom yang telah direvisi, terdiri (Anderson et al., 2001) dari enam tingkatan berpikir sebagai berikut:

**Tabel 2.1 Aspek Kemampuan Berpikir**

<b>Aspek Berpikir</b>	<b>Definisi</b>
Mengingat	Pengambilan memori pengetahuan yang relevan dari memori jangka panjang
Memahami	Kemampuan merumuskan makna pesan pembelajaran dan mengkomunikasikannya secara efektif dalam bentuk lisan, tulisan, dan bentuk grafis
Menerapkan	Kemampuan untuk menggunakan prosedur untuk memecahkan masalah
Menganalisis	Kemampuan untuk memecah sebuah unit menjadi bagian-bagiannya dan menentukan bagaimana bagian-bagian ini terhubung ke keseluruhan
Mengevaluasi	Kemampuan untuk membuat penilaian berdasarkan kriteria dan standar tertentu
Membuat	Menghasilkan ide, produk, atau perspektif baru dari acara tertentu

Ennis (2018) menjelaskan bahwa berpikir kritis adalah proses yang memungkinkan kita untuk membuat keputusan yang masuk akal sehingga sehingga apa yang kita yakini sebagai kebenaran yang terbaik dapat dilakukan dengan benar. Walaupun Robert H. Ennis sebagai salah satu pelopor dalam kajian berpikir kritis tidak secara langsung menyebutkan indikator FRISCO (*Focus, Reason, Inference, Situation, Clarity, Overview*) dalam karya-karya konseptualnya, namun dalam praktiknya, indikator ini telah banyak diadopsi dalam berbagai penelitian pendidikan. FRISCO dianggap sebagai representasi praktis dari keterampilan berpikir kritis karena mencakup aspek fokus terhadap masalah, kemampuan penalaran logis, penyimpulan, kejelasan dalam mengemukakan gagasan, serta kemampuan menilai secara menyeluruh.

Peneliti-peneliti kontemporer dalam bidang pendidikan, khususnya pada penelitian pengembangan instrumen dan asesmen berpikir kritis, memanfaatkan indikator FRISCO karena lebih mudah diimplementasikan dalam perumusan indikator soal dan analisis hasil belajar siswa. Penggunaan FRISCO seringkali

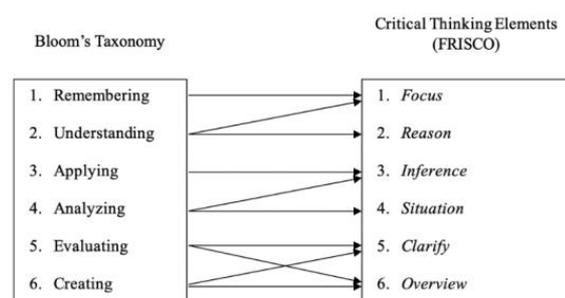
dinilai sebagai bentuk pengadaptasian dari prinsip-prinsip berpikir kritis yang digagas oleh Ennis, namun dalam versi yang lebih aplikatif. Hal ini menunjukkan bahwa meskipun tidak berasal langsung dari teori Ennis, indikator FRISCO mendapatkan legitimasi akademik melalui penerapan empiris yang luas dalam penelitian-penelitian di jurnal nasional dan internasional (Ambarwati & Alim, 2023; Indrawatiningsih et al., 2019; Nufus & Kusaeri, 2020; Nuryadi et al., 2024b; Setiana & Purwoko, 2020; Siburian et al., 2023). Berikut indikator keterampilan berpikir kritis matematis yang digunakan dalam penelitian ini.

**Tabel 2.2 Indikator Keterampilan Berpikir Kritis Matematis**

<b>Aspek</b>	<b>Deskriptor</b>
F ( <i>Focus</i> )	Siswa dapat memahami masalah dan dapat menuliskan apa yang diketahui dari masalah
R ( <i>Reason</i> )	Siswa dapat memahami masalah dan dapat menuliskan apa yang ditanya dari masalah
I ( <i>Inference</i> )	Siswa membuat kesimpulan terkait langkah yang tepat untuk penyelesaian masalah seperti rumus yang digunakan atau cara tertentu
S ( <i>Situation</i> )	Siswa menyelesaikan permasalahan dengan mengumpulkan semua informasi yang relevan untuk digunakan dalam penyelesaian
C ( <i>Clarity</i> )	Siswa dapat memberikan penjelasan (membenarkan atau mengklarifikasi) lebih lanjut tentang kesimpulan akhir
O ( <i>Overview</i> )	Siswa memeriksa kembali langkah-langkah yang dikerjakan atau menguji jawaban

(Modifikasi Ennis, 1996)

Berdasarkan penjelasan mengenai taksonomi Bloom dan indikator berpikir kritis Ennis, yang diwakili oleh FRISCO, hubungan antara keduanya dapat divisualisasikan sebagai berikut:



Sumber: 10.53555/kuvey.v30i4.1228

**Gambar 2.1 Hubungan antara Taksonomi Bloom dan Indikator Berpikir Kritis (FRISCO)**

Dalam konteks pengembangan keterampilan berpikir kritis matematis, indikator FRISCO dapat dipetakan secara sistematis terhadap level kognitif yang terdapat dalam Taksonomi Bloom. Indikator *Focus*, yang menekankan kemampuan siswa dalam mengidentifikasi inti masalah, bersesuaian dengan tahap memahami dan menganalisis, karena siswa dituntut untuk menyaring informasi yang relevan dalam suatu situasi. *Reason* dan *Inference* berkaitan erat dengan tahapan menganalisis dan mengevaluasi, di mana siswa menyusun alasan logis dan menarik kesimpulan yang valid berdasarkan bukti. *Situation* mendorong siswa untuk mempertimbangkan konteks atau kondisi penyelesaian masalah, yang memerlukan kemampuan sintesis dan aplikasi pengetahuan ke dalam situasi yang kompleks. Sementara itu, *Clarity* dan *Overview* mengarah pada tahapan mengevaluasi dan mencipta, karena siswa harus menyampaikan argumen secara jelas serta meninjau ulang seluruh proses berpikir untuk membangun keputusan akhir yang kokoh dan reflektif (Nuryadi et al., 2024). Dengan demikian, FRISCO tidak hanya relevan sebagai indikator operasional berpikir kritis, tetapi juga mendukung pengembangan keterampilan kognitif tingkat tinggi sebagaimana diuraikan dalam Taksonomi Bloom.

### **2.1.7 Keterkaitan IRT terhadap Peningkatan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis**

Berpikir kritis dianggap sebagai salah satu keterampilan utama yang mempersiapkan siswa menghadapi perubahan cepat dan berkesinambungan di abad ke-21. Kesuksesan hidup seseorang sangat ditentukan oleh kemampuannya dalam memecahkan masalah, baik skala besar maupun kecil. Dalam konteks ini, berpikir kritis menjadi syarat penting bagi setiap individu untuk menyelesaikan permasalahan.

Keterampilan berpikir kritis yang perlu ditanamkan pada siswa meliputi kemampuan merumuskan masalah, menganalisis beberapa argumen, serta mengambil keputusan yang tepat dan melaksanakan tindakan terhadap permasalahan yang dihadapi. Pengembangan keterampilan ini tidak hanya dilakukan dalam pembelajaran, tetapi juga perlu didukung oleh alat yang mencerminkan keterampilan berpikir kritis. Salah satu alat ukur yang dapat digunakan untuk menilai keterampilan berpikir kritis adalah tes uraian (Handayani et al., 2021). Banyak peneliti menggunakan bentuk tes ini untuk mengukur keterampilan berpikir kritis siswa, menunjukkan bahwa soal uraian mampu mengukur keterampilan tersebut.

Dalam pelaksanaan tes uraian, penskoran dilakukan secara bertahap, di mana skor diperoleh dengan menjumlahkan nilai dari setiap langkah yang dikerjakan siswa. Namun, model penskoran ini belum tentu tepat karena tidak mempertimbangkan tingkat kesulitan masing-masing langkah. Selain itu, peluang siswa untuk menjawab benar berdasarkan respons tertentu juga sulit diprediksi.

Pendekatan alternatif yang dapat diterapkan adalah teori respons butir untuk penskoran politomi. Penskoran politomi memberikan nilai pada hasil tes yang terdiri dari dua nilai atau lebih, dengan penskoran dilakukan langkah per langkah dalam suatu butir soal. Dalam analisis ini, tingkat kesulitan tiap langkah diperhitungkan (Falani et al., 2020a). Skor tertinggi diperoleh jika peserta mampu menyelesaikan soal hingga langkah akhir dengan benar.

Salah satu model yang dapat digunakan dalam analisis butir politomi adalah PCM, GPCM, dan GRM. PCM, GPCM dan GRM adalah teori IRT politomus yang berguna untuk mengevaluasi karakteristik tes serta kemampuan peserta tes.

Parameter yang diestimasi dalam PCM, GPCM, dan GRM adalah tingkat kesulitan butir tes, yang merupakan salah satu karakteristik tes. Tabel 2.3 menunjukkan keterkaitan antara ketiga model IRT dengan indikator keterampilan berpikir kritis FRISCO.

**Tabel 2.3 Keterkaitan Keterampilan Berpikir Kritis Matematis Indikator FRISCO dengan Model IRT**

<b>Indikator FRISCO</b>	<b>PCM</b>	<b>GPCM</b>	<b>GRM</b>
<i>Focus</i>	Dapat menilai fokus siswa dalam menyelesaikan langkah-langkah soal.	Mampu mengukur fokus dengan mempertimbangkan kesulitan setiap langkah.	Menyediakan penilaian yang lebih terperinci terkait fokus siswa.
<i>Reason</i>	Mengukur kemampuan siswa dalam memberikan alasan yang tepat.	Memperhitungkan variasi dalam alasan yang diberikan oleh siswa.	Menilai alasan dengan mempertimbangkan kategori jawaban yang berjenjang.
<i>Inference</i>	Dapat digunakan untuk menilai kemampuan menarik kesimpulan dari data.	Lebih fleksibel dalam menilai inferensi berdasarkan variasi item.	Mengukur inferensi melalui skala penilaian yang lebih halus.
<i>Situation</i>	Menilai kemampuan siswa dalam menerapkan konsep pada situasi tertentu.	Dapat memperhitungkan konteks situasi yang berbeda dalam penilaian.	Menyediakan pemahaman situasional melalui kategori respons yang beragam.
<i>Clarity</i>	Mengukur kejelasan dalam penyampaian jawaban siswa.	Memungkinkan penilaian kejelasan dengan mempertimbangkan kesulitan item.	Menilai kejelasan tingkatan kategori jawaban yang lebih terstruktur.
<i>Overview</i>	Dapat menilai kemampuan siswa dalam melakukan pemeriksaan kembali.	Memperhitungkan langkah-langkah pemeriksaan kembali dalam analisis.	Mampu menilai seberapa baik siswa melakukan overview terhadap jawaban mereka.

Hasil analisis PCM, GPCM, dan GRM dapat memberikan informasi mengenai sebaran kemampuan peserta tes, hubungan antara kemampuan peserta dan peluang menjawab benar, serta fungsi informasi dan SEM. Penggunaan PCM, GPCM, dan GRM diharapkan dapat memberikan informasi yang berguna dalam mengembangkan instrumen tes yang lebih akurat untuk mengukur kemampuan siswa.

### 2.1.8 Tinjauan Materi Rasio

Adapun materi rasio yang dijabarkan sebagai berikut (Marsigit & Susilo, 2006):

#### Konsep Rasio

Rasio merupakan perbandingan dua besaran, dapat berupa ukuran atau jumlah benda. Rasio dapat dituliskan dalam bentuk  $a : b$ .

#### Skala

Skala berkaitan dengan perbesaran atau pengecilan suatu ukuran tetapi tidak mengubah bentuknya. Skala ( $S$ ) adalah perbandingan antara jarak pada peta ( $JP$ ) dengan jarak sebenarnya ( $JS$ ).

$$\text{Rumus skala: } S = \frac{JP}{JS}$$

Keterangan:

$S$  = Skala

$JP$  = Jarak pada peta

$JS$  = Jarak sebenarnya

#### Perbandingan Senilai

Perbandingan senilai adalah perbandingan yang memiliki nilai pembanding yang sama. Secara matematis dapat ditulis dengan rumus:

$$\frac{x_1}{y_1} = \frac{x_2}{y_2}$$

#### Perbandingan Berbalik Nilai

Perbandingan berbalik nilai adalah perbandingan yang memiliki nilai pembanding yang berbalik.

$$\frac{x_1}{y_2} = \frac{x_2}{y_1}$$

## 2.2 Penelitian yang Relevan

Penggunaan tes format *essay* untuk mengukur kemampuan berpikir kritis siswa dengan menggunakan model IRT (PCM, GPCM dan GRM), beberapa penelitian terdahulu telah mencoba melakukan penelitian lebih lanjut tentang penggunaan model IRT untuk data politomi. Penelitian yang relevan diantaranya adalah:

Penelitian oleh Komala et al., (2020) tentang karakterisasi soal tes keterampilan berpikir kritis menggunakan analisis IRT pada materi fluida statis. Karakterisasi soal ini bertujuan untuk menggambarkan kualitas soal keterampilan berpikir kritis dengan kategori memberikan alasan (*verbal reasoning*), pengujian hipotesis (*hypothesis testing*), analisis argumen (*argument analysis*), kemungkinan dan analisis ketidakpastian (*likelihood and uncertainty analysis*), serta pemecahan masalah dan pembuatan keputusan (*problem-solving and decisioan making*). Versi pertama dari soal diujicobakan kepada 11 orang siswa kelas X MIPA (N=11), uji coba ini meliputi wawancara kognitif dan pelaksanaan uji coba skala kecil. Setelah itu dibuat modifikasi dan soal diujicobakan kepada subjek uji yang berbeda yaitu kelas X MIPA dan kelas XI IPA (N=50). Hasil dari uji coba ini menunjukkan 15 dari 17 butir soal ini adalah valid untuk subjek uji yang memiliki tingkat kemampuan rata-rata sampai kemampuan tinggi sementara reliabiitas soal tes menunjukkan bahwa 13 soal reliabel untuk subjek uji yang memiliki kemampuan rata-rata dan empat soal lainnya reliabel untuk subjek uji yang memiliki kemampuan tinggi (diatas rata-rata). Berdasarkan temuan tersebut dapat dinyatakan bahwa soal yang telah dibuat mampu untuk mengukur keterampilan berpikir kritis pada materi

fluida statis untuk subjek uji yang memiliki kemampuan rata-rata dan diatas rata-rata.

Kemudian berdasarkan penelitian yang dilakukan oleh Rosidah et al., (2018) yang berjudul Karakteristik Tes Keterampilan Berpikir Kritis (KBK) Berdasarkan Pendekatan Teori Respons Butir. Analisis data dalam penelitian ini didasarkan pada respons 108 siswa yang telah mendapatkan pembelajaran tentang suhu dan kalor kelas X SMA di Tasikmalaya dan Bandung. Hasil analisis dari kurva karakteristik tes menunjukkan bahwa indeks kesukaran (*bjk*) terentang antara -1,49 sampai 1,34 dengan indeks kesukaran yang berbeda pada tiap kategorinya. Perpotongan antara total fungsi informasi dan SEM berada pada rentang -1,70 sampai 1,20. Hal ini berarti, keseluruhan tes ini dapat digunakan untuk mengukur kemampuan berpikir kritis siswa dari tingkat kemampuan rendah (di bawah rata-rata) sampai kemampuan tinggi (di atas rata-rata).

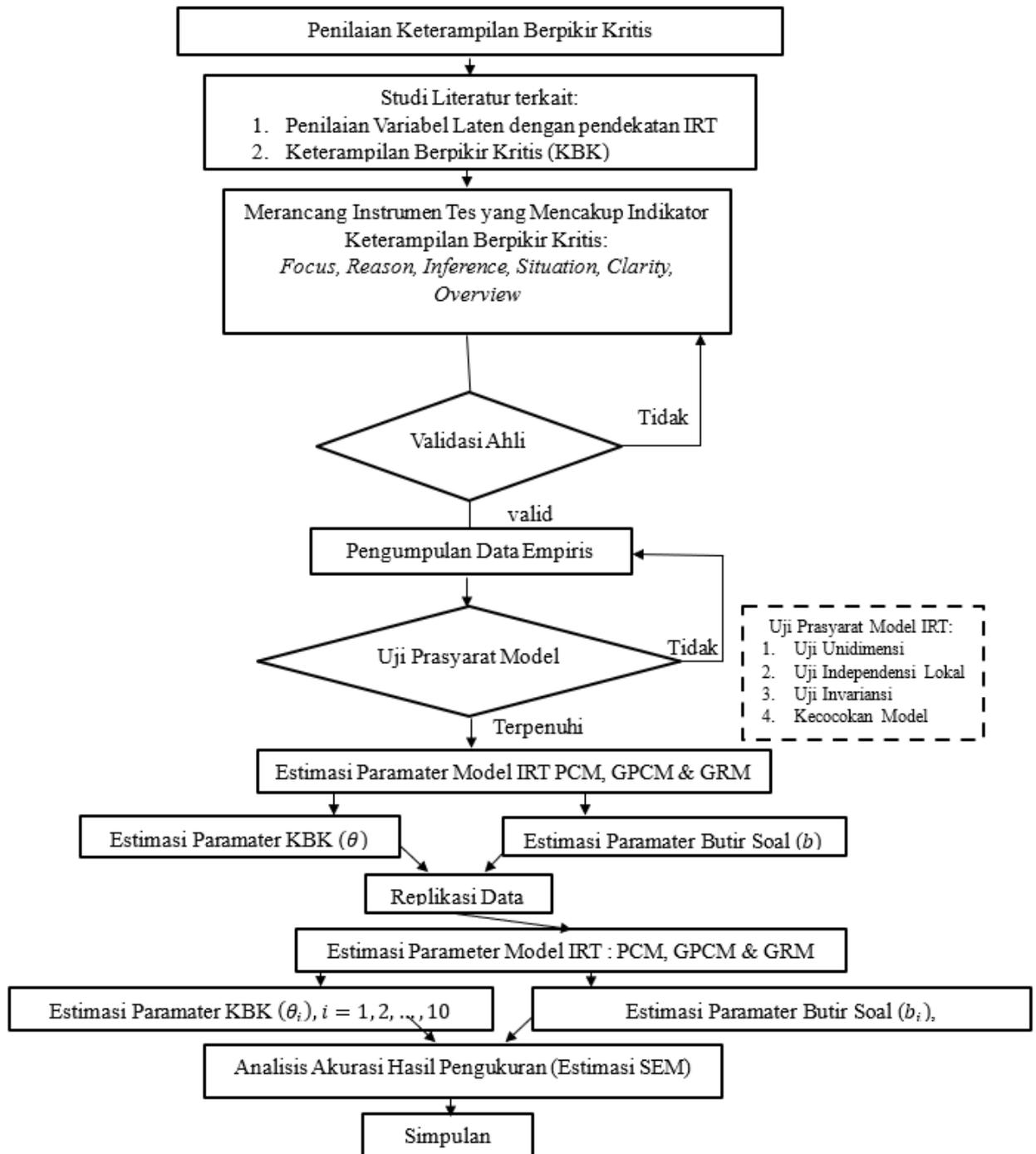
Penelitian oleh Falani, Akbar, dan Naga (2020) menunjukkan bahwa kombinasi model *Multiple-Choice Model* (MCM) untuk data dikotomis dan *Graded Response Model* (GRM) untuk data politomis menghasilkan estimasi kemampuan siswa yang lebih presisi dibandingkan dengan kombinasi model *3 Parameter Logistic Model* (3PLM) dan *Generalized Partial Credit Model* (GPCM). Dengan menggunakan data dari 1625 siswa SMP di Depok dan dianalisis melalui perangkat lunak PARSCALE, hasil estimasi dari kombinasi MCM+GRM menunjukkan nilai *Root Mean Squared Error* (RMSE) yang lebih rendah serta korelasi yang lebih tinggi dibandingkan 3PLM+GPCM. Temuan ini menegaskan bahwa MCM+GRM lebih direkomendasikan untuk mengestimasi kemampuan matematis siswa dalam

tes format campuran, karena mampu menangkap karakteristik respons secara lebih akurat dan efisien.

Berdasarkan penelitian-penelitian diatas, peneliti tertarik untuk mengukur kemampuan berpikir kritis siswa dengan menggunakan model IRT (PCM, GPCM dan GRM) . Perbedaan penelitian yang akan peneliti lakukan dengan penelitian sebelumnya yaitu peneliti akan mengukur keterampilan berpikir kritis matematis siswa dengan menggunakan tiga model IRT yaitu PCM, GPCM dan GRM.

### **2.3 Kerangka Berpikir**

Kerangka berpikir untuk penelitian ini sebagai berikut:



Gambar 2.2 Kerangka Berpikir

## 2.4 Hipotesis Penelitian

Hipotesis penelitian disusun sebagai berikut:

1. Terdapat perbedaan nilai varians hasil estimasi parameter keterampilan berpikir kritis matematis antara model PCM dan GPCM, dimana model GPCM memiliki varians estimasi yang lebih kecil dibandingkan model PCM.
2. Terdapat perbedaan nilai varians hasil estimasi parameter keterampilan berpikir kritis matematis antara model GPCM dan GRM, dimana model GPCM memiliki varians estimasi yang lebih kecil dibandingkan model GRM.
3. Terdapat perbedaan nilai varians hasil estimasi parameter keterampilan berpikir kritis matematis antara model PCM dan GRM, dimana model PCM memiliki varians estimasi yang lebih kecil dibandingkan model GRM.

## BAB III METODE PENELITIAN

### 3.1 Waktu dan Tempat Penelitian

Penelitian ini menggunakan respons siswa SMP Kelas VII semester genap tahun ajaran 2024/2025 di wilayah Kecamatan Bahar Utara Kabupaten Muaro Jambi. Penelitian ini dilaksanakan pada bulan April hingga Mei.

### 3.2 Metode Penelitian

#### 3.2.1 Jenis Penelitian

Penelitian yang dilakukan berupa penelitian kuantitatif dengan metode *ex post facto*. Penelitian *ex post facto* merupakan penelitian yang meneliti hubungan sebab akibat yang tidak dimanipulasi atau diberi perlakuan oleh peneliti (Sappaile, 2010). Variabel terikat dalam penelitian ini adalah varians hasil estimasi parameter kemampuan berpikir kreatif matematis, sedangkan variabel bebasnya adalah model IRT yang digunakan yaitu: PCM, GPCM, dan GRM.

#### 3.2.2 Desain penelitian

Penelitian ini didesain dengan membandingkan tingkat presisi parameter keterampilan berpikir kritis. Adapun rancangan yang digunakan dalam penelitian ini yaitu dengan membandingkan varians hasil estimasi parameter keterampilan berpikir kritis dengan ketiga model tersebut.

**Tabel 3.1 Rancangan perbandingan Variansi Hasil Estimasi**

Model	PCM	GPCM	GRM
PCM	-	$\frac{\sigma^2 PCM}{\sigma^2 GPCM}$	$\frac{\sigma^2 GRM}{\sigma^2 GPCM}$
GPCM	-	-	$\frac{\sigma^2 PCM}{\sigma^2 GRM}$
GRM	-	-	-

Keterangan:

$\sigma^2 PCM$  = nilai varians hasil estimasi parameter kemampuan dengan PCM

$\sigma^2 GPCM$  = nilai varians hasil estimasi parameter kemampuan dengan GPCM

$\sigma^2 GRM$  = nilai varians hasil estimasi parameter kemampuan dengan GRM

### 3.3 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data hasil respons siswa dari tes soal matematika yang diberikan. Data yang diperoleh bersifat politomi dengan format penskoran (0,1,2,3).

### 3.4 Populasi dan Sampel

Populasi dalam penelitian ini adalah peserta tes, dimana peserta tes tersebut adalah siswa kelas VII SMP semester genap tahun 2024/2025 di wilayah Kecamatan Bahar Utara Kabupaten Muaro Jambi yang terdiri dari 4 SMP Negeri. Pengambilan sampel menggunakan teknik total sampling yang artinya seluruh populasi dijadikan sampel (Sugiyono, 2019a). Untuk sampel menggunakan teknik total sampling yang artinya seluruh populasi dijadikan sampel. Total sampling memiliki kelemahan seperti kebutuhan sumber daya yang besar, serta ketidakefisienan jika populasi sangat besar (Sugiyono, 2015). Namun kelebihanannya cukup signifikan karena menghilangkan kesalahan sampling, meningkatkan validitas eksternal serta memberikan gambaran menyeluruh yang sangat berguna terutama untuk populasi kecil (Sugiyono, 2019b).

**Tabel 3.2 Jumlah Peserta Tes**

No	Nama Sekolah	Jumlah Siswa
1.	SMPN 14 Muaro Jambi	57
2.	SMPN 45 Muaro Jambi	15
3.	SMPN 49 Muaro Jambi	13
4.	SMPN 53 Muaro Jambi	21
	Jumlah	106

### 3.5 Prosedur Penelitian

#### 3.5.1 Penyusunan Instrumen Penelitian

Instrumen penelitian ini berupa tes *essay* untuk mata pelajaran matematika yang disusun berdasarkan capaian pembelajaran (CP) kurikulum Merdeka kelas VII. Matematika adalah bidang studi yang selalu hadir di setiap tingkat pendidikan. Pelajaran matematika mencakup berbagai topik, mulai dari aritmatika dasar seperti penjumlahan dan pengurangan, hingga konsep yang lebih kompleks seperti aljabar, geometri, trigonometri, dan kalkulus.

Setiap topik dalam matematika bersifat saling berhubungan dan mengikuti struktur yang hierarkis, sehingga pemahaman yang kuat pada konsep dasar sangat penting sebelum melanjutkan ke materi yang lebih lanjut, sehingga untuk memahami materi secara keseluruhan, siswa perlu memahami konsep-konsep prasyarat secara bertahap. Sifat ini memungkinkan penyusunan soal yang langkah penyelesaiannya bersifat bertahap.

Penyelesaian bertahap ini membantu mengukur sejauh mana siswa menguasai konsep-konsep yang telah dipelajari. Proses pengukuran ini terlihat dari prosedur penilaian yang memberikan tingkat-tingkat berdasarkan pemahaman konsep yang tidak sepenuhnya lengkap. Selain itu, matematika tidak hanya melatih kemampuan berhitung, tetapi juga melibatkan pemecahan masalah, berpikir kritis, berpikir kreatif dan pengambilan keputusan yang rasional. Dalam proses pembelajarannya, siswa diajak untuk mengidentifikasi masalah, menganalisis informasi yang ada, dan merumuskan solusi yang tepat berdasarkan data dan fakta yang relevan.

Tes *essay* disusun berdasarkan capaian pembelajaran (CP) pada fase D kemudian dijabarkan menjadi alur tujuan pembelajaran (ATP) sebanyak 3 soal yang akan mengukur kemampuan berpikir kritis siswa, penyusunan butir tes merupakan hasil diskusi serta saran dari para ahli dan guru matematika yang ada disekolah tempat penelitian berlangsung. Soal-soal tes yang telah dirancang tidak langsung digunakan untuk pengumpulan data penelitian. Sebelum itu, beberapa langkah dilakukan terlebih dahulu untuk memastikan bahwa instrumen tes tersebut valid dan reliabel dalam mengukur kemampuan berpikir kritis matematika siswa.

Tahap awal melibatkan penilaian butir tes oleh para ahli atau pakar. Beberapa ahli yang memiliki keahlian di bidang Matematika diminta untuk mengevaluasi instrumen tes yang telah disusun oleh peneliti. Berikut adalah deskripsi mengenai para penilai tes.

**Tabel 3.3 Biodata Ahli untuk Validasi Tes**

No	Nama	Bidang Keahlian	Institusi	Jabatan
1.	Dr. Syamsir Sainuddin, S.Pd. M.Pd.	Matematika	Universitas Jambi	Dosen
2.	Aliffia Teja Prasasty, M.Pd	Matematika	Universitas Indraprasta PGRI	Dosen
3.	Yuni Astarina, S.Pd	Matematika	SMPN 14 Muaro Jambi	Guru
4.	Nanang Sriadi, S.Pd	Matematika	SMPN 54 Muaro Jambi	Guru
5.	Sahnar, S.Pd	Matematika	SMPN 12 Muaro Jambi	Guru

Penilaian yang dilakukan oleh para ahli tersebut meliputi: kesesuaian Capain Pembelajaran (CP), Alur Tujuan Pembelajaran (ATP) dan Tujuan Pembelajaran (TP), kesesuaian indikator dan materi, kesesuaian indikator dan soal, kejelasan perumusan soal, penggunaan bahasa pada butir tes, dan usulan penyekoran. Penilaian ini dilakukan secara kuantitatif dengan menggunakan metode Aiken. Metode menggunakan indeks kesepakatan para ahli mengenai validitas butir, indeks ini merupakan kesepakatan para ahli terhadap kesesuaian butir dengan

indikator yang ingin diukur dengan menggunakan butir tersebut (Retnawati, 2015).

Perhitungan indeks Aiken dengan menggunakan rumus,

$$V = \frac{\sum s}{n(c-1)}, \text{ dengan } s = r - l_0 \quad (10)$$

$V$  adalah indeks kesepakatan ahli atau pakar mengenai validitas butir,  $s$  adalah skor yang ditetapkan setiap penilai  $r$  dikurangi skor terendah dalam kategori penilaian  $l_0$  yang digunakan.  $n$  adalah banyaknya penilai, dan  $c$  adalah banyaknya kategori penilaian. Instrumen asesmen dikatakan valid apabila memiliki indeks Aiken  $\geq 0,75$ . Item dengan indeks Aiken  $< 0,75$  dianggap tidak valid (Nabil et al., 2022).

Tahap kedua adalah uji coba tes. Tes yang telah disesuaikan berdasarkan penilaian indeks Aiken diujikan kepada 35 siswa kelas 8 SMP di Kecamatan Sungai Bahar. Uji reliabilitas tes dilakukan menggunakan metode *Cronbach Alpha* pada data tersebut untuk memastikan bahwa butir-butir tes memenuhi standar keandalan (reliabilitas) dalam mengukur kemampuan matematika siswa. Hal ini penting agar pengambilan data penelitian dapat dilakukan dengan baik. Setelah tahap-tahap ini selesai, tes kemudian diujikan kepada sampel penelitian untuk mengumpulkan data.

### 3.5.2 Penyekoran Data Hasil Tes

Setelah tes diberikan diujikan kepada sampel penelitian, kemudian selanjutnya hasil tes diskor secara politomi berdasarkan rubrik penilaian kemampuan berpikir kritis siswa yang disusun oleh peneliti. Penyekoran hasil tes dapat dilihat di Lampiran 4. Butir soal diskor secara politomi dengan empat kategori skor yaitu 0, 1, 2, 3 dan 4. Pemberian skor dilakukan berdasarkan tahapan respons jawaban benar pada penyelesaian butir soal.

### 3.6 Pengolahan Data

Setelah penyekoran selesai dilakukan terhadap data hasil respon ujites. Selanjutnya dilakukan pengolahan data. Pengolahan data yang dilakukan meliputi: uji asumsi prasyarat Model IRT, tingkat kecocokan kombinasi model, estimasi parameter kemampuan, dan nilai fungsi informasi tes. Pengolahan data dimulai dengan melakukan uji asumsi prasyarat pada model IRT, di mana asumsi utama yang diuji adalah asumsi unidimensi. Asumsi unidimensi diperiksa menggunakan metode analisis faktor.

Asumsi berikutnya adalah tingkat kecocokan model IRT terhadap data lapangan. Tingkat kecocokan model terhadap data. Tingkat kecocokan untuk masing-masing model dievaluasi dengan menggunakan teknik pemeriksaan prediksi model. Teknik pemeriksaan prediksi model dilakukan pada hasil ujites empiris dan pada hasil ujites harapan matematika dari model. Pencocokan dengan model ini diawali dengan mencari residu mentah yakni selisih antara hasil ujites empiris dan pada hasil ujites harapan matematika dari model. Selanjutnya residu mentah di transformasi menjadi residu baku ( $z$ ). Dengan asumsi bahwa jumlah kuadrat dari residu baku berdistribusi khi-khuadrat ( $\chi^2$ ). Selanjutnya dengan menggunakan taraf signifikansi  $\alpha$  dilakukan pengujian kecocokan model dengan membandingkan nilai khi kuadrat dari residu dengan nilai kritis distribusi probabilitas khi kuadrat. Nilai khi kuadrat ( $\chi^2$ ) di atas nilai kritis ini mengindikasikan bahwa butir tidak cocok dengan model yang digunakan, dan sebaliknya.

Pada penelitian ini kalibrasi dilakukan dengan menggunakan PARSCALE 4.1. Luaran PARSCALE 4.1 terdiri dari empat phase (ph) bagian, yakni ph.0, ph.1,

ph.2, dan ph.3. ph.0 berisikan informasi tentang spesifikasi analisis serta inputan dari data peserta tes. Ph.1 berisi beberapa tabel informasi penting terkait: Butir summary statistiks yakni proporsi dari peserta tes pada tiap kategori response, rata-rata dan *standard error* tiap butir, korelasi pearson dan polyserial antar masing-masing butir dan sekor total. Ph.2 berisi informasi tentang estimasi parameter, konvergensi algoritma yang digunakan, serta estimasi akhir dari butir parameter, serta dilengkapi dengan statistik kecocokan butir untuk setiap butirnya. Ph.3 berisi informasi sekor untuk tes, Estimasi parameter kemampuan dan standar errorsnya.

Pemilihan model yang tepat diharapkan dapat menghasilkan estimasi parameter kemampuan yang memiliki tingkat presisi tinggi. Hal ini sangat mempengaruhi hasil penilaian dan evaluasi yang dilakukan dalam rangka memperbaiki kualitas pembelajaran. Indikator tingkat presisi yang digunakan dalam penelitian ini adalah varians hasil estimasi. Semakin kecil varians estimasi yang dihasilkan oleh suatu kombinasi model (Dali, 2012; Si *et al.*, 2002).

Pengolahan data selanjutnya adalah penentuan nilai fungsi informasi tes untuk masing-masing model. Nilai fungsi informasi tes merupakan akumulasi dari nilai fungsi informasi untuk tiap butir pada tes (Supriyati *et al.*, 2021). Oleh karena itu, terlebih dahulu dihitung nilai fungsi informasi untuk setiap butir tes pada masing-masing kombinasi model. Selanjutnya nilai fungsi informasi ujites dihitung dengan menjumlahkan seluruh nilai fungsi informasi butir tes. Nilai fungsi informasi ujites hasil analisis untuk kedua kombinasi model akan dibandingkan untuk mengetahui kombinasi model yang menghasilkan nilai fungsi informasi tes yang lebih tinggi. Menurut Naga (2012) dan Retnawati (2014) Penentuan fungsi informasi dilakukan dengan menggunakan varians yakni kuadrat simpangan baku estimasi parameter

kemampuan . Makin tinggi fungsi informasi makin kecil kekeliruan baku estimasi parameter kemampuan . Oleh karena itu, rumusan masalah pada penelitian ini hanya menguji tingkat presisi estimasi parameter kemampuan.

Selain menggunakan data empiris, peneliti ini juga menggunakan data replikasi yang dibangun berdasarkan data empiris. Data replikasi di generate dengan menggunakan bantuan perangkat lunak WINGEN. Data simulasi ini di generate berdasarkan hasil estimasi parameter kemampuan dan parameter butir dari data hasil ujites empiris. Data replikasi ini selanjutnya diolah dengan prosedur yang sama dengan data hasil ujites empiris sebelumnya, hasil pengolahan data replikasi ini selanjutnya dikorelasikan dengan hasil pengolahan data ujites empiris. Hal ini dilakukan untuk mendapatkan hasil analisis yang lebih mendalam terutama terkait tingkat presisi estimasi parameter kemampuan dan parameter butir yang dihasilkan oleh kedua kombinasi model. Sehingga pengambilan keputusan dalam pemilihan kombinasi model terbaik untuk menganalisis tes format campuran dapat semakin akurat

### **3.7 Teknik Analisis Data**

Analisis data dimaksud untuk melakukan pengujian dan menjawab masalah yang telah diajukan. Sebelum melakukan uji hipotesis, perlu dilakukan uji prasyarat.

#### **3.7.1 Uji Prasyarat Model**

Terdapat tiga uji prasyarat model yang digunakan dalam IRT diantaranya yaitu uji unidimensi, uji independensi lokal, dan uji invariansi parameter.

### 1. Uji Unidimensi

Uji unidimensi dalam IRT menghendaki kurva yang sesuai dengan karakteristik butir. Kurva itu harus terus naik mengikuti kenaikan pada parameter kemampuan. Agar kurva jawaban betul terus naik menurut kenaikan parameter kemampuan, maka kemampuan yang diukur dibatasi hanya pada satu macam atau satu dimensi kemampuan saja. Apabila pengukuran dilakukan dengan lebih dari satu dimensi, maka kurva karakteristik butir akan menjadi lain dan tidak meningkat (naik) dengan meningkatnya kemampuan yang akan diukur. Maka sebab itulah pada IRT disyaratkan unidimensi. Pemeriksaan unidimensi dapat dilakukan melalui pencocokan model karakteristik butir yang dipilih dengan data dari lapangan. Jika kurva karakteristik butir yang dipilih sesuai dengan data dari lapangan, maka tujuan unidimensi sudah tercapai.

### 2. Uji Independensi Lokal

Uji independensi lokal menunjukkan bahwa setiap butir tes dan responden (peserta tes) itu bersifat independen yang artinya tidak ada keterkaitan antara satu butir dengan butir lainnya dan jawaban peserta tes terhadap suatu butir tes tidak mempengaruhi jawaban peserta tes terhadap butir tes lainnya. Pengujian independensi dapat dilakukan tidak hanya untuk mengetahui apakah butir soal bersifat independen, tetapi juga untuk mengetahui apakah responden bersifat independen.

### 3. Uji Invariansi Parameter

Uji invariansi parameter merupakan karakteristik yang paling utama yang sangat berpengaruh dalam IRT. Didalam pengujian instrumen tes, tentunya terdapat berbagai macam kemampuan yang dimiliki oleh responden (peserta tes) mulai dari

yang rendah, sedang, sampai yang tinggi. Invariansi parameter dapat diperiksa dengan cara melihat kecocokan model. Jika data dari lapangan sudah cocok dengan karakteristik butir yang dipakai, maka syarat invariansi parameter sudah terpenuhi.

Dalam pengujian prasyarat model IRT dan replikasi data, digunakan dua *software* berikut:

#### 1. Parscale

Parscale digunakan untuk pengujian prasyarat model IRT. Parscale merupakan salah satu *software* analisis data berdasarkan model IRT, yaitu metode psikometri untuk mengukur validitas dan reliabilitas suatu instrumen pengukuran. Penggunaan Parscale berguna untuk memberikan perkiraan yang lebih akurat dan pengukuran yang dapat direplikasi, serta mampu menangani data yang tidak lengkap atau hilang. Salah satu fitur utama Parscale yaitu pemberian skala linier dengan interval yang sama, yang membantu mempermudah interpretasi hasil analisis. Dalam konteks penelitian, Parscale sering digunakan untuk menguji kemampuan atau sikap kognitif responden, atau untuk memvalidasi instrumen pengukuran seperti angket atau tes. Parscale menjadi pilihan populer dalam penelitian karena mudah digunakan dan memberikan hasil analisis yang valid.

#### 2. Wingen

Wingen digunakan untuk melakukan replikasi atau tiruan data dari hasil pengujian prasyarat IRT. Tujuan dari dilakukannya replikasi data yaitu agar memperoleh lebih dari satu kali pengujian data agar didapatkan hasil yang lebih meyakinkan atau mendekati kebenaran. Wingen membantu pengguna dalam menjalankan simulasi pengujian untuk memperoleh data baru yang diperlukan yang memiliki karakteristik hampir mirip dengan data awal tanpa harus turun lapangan

untuk melakukan pengujian secara langsung, sehingga analisis IRT dapat dilakukan dengan lebih efisien. Salah satu fitur utama Wingen yaitu mampu menghasilkan data berdasarkan parameter yang telah ditentukan, seperti jumlah *item* dan karakteristik *item* tes, sehingga memungkinkan peneliti untuk mempertimbangkan berbagai skenario tes dan memperoleh hasil yang lebih akurat dalam mengestimasi keterampilan peserta yang mengikuti tes. Dengan demikian, Wingen mampu memberikan kontribusi yang signifikan terhadap pengembangan metode pengukuran yang lebih baik untuk berbagai aplikasi penelitian.

### 3.7.2 Estimasi Parameter Butir

Estimasi parameter butir merupakan proses untuk memperoleh informasi karakteristik butir soal dalam suatu tes. Dalam teori respon butir (IRT), terdapat beberapa parameter yang diestimasi untuk menggambarkan karakteristik butir, seperti parameter tingkat kesulitan (*difficulty*) dan parameter daya beda (*discrimination*), yang disesuaikan dengan model yang digunakan.

Pada penelitian ini, estimasi parameter butir dilakukan menggunakan tiga model politomus, yaitu PCM, GPCM, dan GRM. PCM digunakan untuk menganalisis butir soal politomus dengan asumsi bahwa setiap butir memiliki tingkat kesulitan (*b*) pada setiap kategori skor, namun semua butir dianggap memiliki daya beda (*discrimination*) yang sama atau tetap. PCM fokus pada estimasi parameter kesulitan butir untuk setiap kategori (Falani et al., 2017).

Sementara itu, GPCM merupakan pengembangan dari PCM, yang memungkinkan adanya variasi daya beda antar butir. Dengan kata lain, GPCM mengestimasi parameter diskriminasi (*a*) untuk setiap butir, selain juga parameter

kesulitan (b) pada setiap kategori skor, sehingga model ini lebih fleksibel dibanding PCM (Falani et al., 2017).

Adapun GRM juga digunakan untuk menganalisis data politomus dengan urutan kategori yang bergradasi. Model ini mengestimasi parameter diskriminasi (a) serta sejumlah parameter kesulitan (b) berupa *threshold* (ambang batas) yang menunjukkan titik perubahan peluang respon antar kategori skor.

Proses estimasi parameter butir pada ketiga model ini dilakukan menggunakan perangkat lunak analisis IRT, seperti PARSCALE atau program sejenisnya. Hasil estimasi ini digunakan untuk menilai kualitas butir soal, apakah butir tersebut efektif dalam membedakan tingkat kemampuan responden serta sesuai dengan tujuan pengukuran yang diharapkan.

### 3.7.3 Estimasi Parameter Kemampuan

Pada tahap awal analisis menggunakan model IRT, nilai parameter kemampuan peserta tes dan karakteristik butir soal masih belum diketahui, sehingga diperlukan estimasi terlebih dahulu. Proses estimasi ini disebut dengan kalibrasi, di mana parameter kemampuan peserta dan karakteristik butir soal harus dihitung. Tantangan dalam kalibrasi adalah menentukan nilai parameter kemampuan untuk setiap peserta tes, serta parameter butir soal, berdasarkan respons yang telah diberikan pada setiap butir tes.

Demars (2010) dan Retnawati (2015) menyebutkan bahwa metode sering digunakan dalam kalibrasi parameter butir adalah Metode *Marginal Maximum A Posteriori* (MMAP) dan Metode *Marginal Maximum Likelihood Estimation* (MMLE). Sedangkan kalibrasi parameter kemampuan dapat dilakukan dengan tiga metode, yakni Metode *Bayes Modal* atau *Maximum a posteriori* (MAP), Metode

*Bayes* atau *Expected a Posteriori* (EAP), dan Metode *Maksimum Likelihood Estimation* (MLE). Pembahasan hanya difokuskan pada kalibrasi parameter kemampuan, karena seperti yang telah dijelaskan sebelumnya, parameter butir tidak menjadi objek perbandingan dalam penelitian ini.

Untuk kalibrasi parameter kemampuan peneliti menggunakan metode MLE, hal ini didukung oleh penelitian yang dilakukan oleh Retnawati (2015) yang membandingkan akurasi hasil estimasi parameter kemampuan dengan menggunakan metode MLE dan Bayes. Hasil penelitian tersebut mengemukakan bahwa pada ukuran sampel minimal 1500 responden dengan panjang butir 25 dan 30 diperoleh hasil metode MLE lebih akurat dibandingkan dengan Metode Bayes, akan tetapi pada panjang tes 15 dan 20 diperoleh hasil bahwa kedua metode menghasilkan tingkat akurasi yang tidak jauh berbeda. Penelitian Retnawati (2015) tersebut menjadi salah satu pembuktian empiris bahwa tingkat akurasi estimasi kedua metode relatif sama.

Seorang peserta tes dengan kemampuan  $\theta$  mengerjakan soal tes yang terdiri dari  $n$  butir politomi. Misalkan  $u_i$  adalah skor untuk butir ke  $i$   $u_i = 0, \dots, m_i$ , di mana  $m_i$  merupakan maksimum kemungkinan skor untuk butir  $i$ , dan misalkan  $U = u_1, u_2, \dots, u_n$  adalah vector respons untuk semua butir. Jika asumsi independensi lokal terpenuhi, Likelihood dari dapat dihitung sebagai berikut (Tao et al., 2012).

$$L = L(U|\theta) = \prod_{i=1}^n P_{iu_i} \quad (6)$$

Di mana  $P_{iu_i}$  peluang peningkatan skor  $u_i$  pada butir politomus ke  $i$ . nilai dari  $\theta$  yang memaksimalkan fungsi likelihood dari  $\theta$  dinotasikan dengan  $\hat{\theta}_{ML}$ . *Standart error* dari estimasi  $\hat{\theta}_{ML}$  dinotasikan dengan  $SE(\hat{\theta}_{ML})$  merupakan akar dari variansi  $\hat{\theta}_{ML}$  dan dituliskan dengan persamaan,

$$SE(\hat{\theta}_{ML}) = \sqrt{Var_{ML}} = \sqrt{\frac{1}{I(\theta)}}, \quad (7)$$

Di mana  $I(\theta)$  adalah informasi dari tes, yang dihitung dengan Retnawati (2015),

$$I(\theta) = \sum_{i=1}^n \sum_{k=0}^{m_i} \frac{(P_{ik})^2}{P_{ik}} \quad (8)$$

Untuk menghitung SEM ditulis dengan persamaan,

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} \quad (9)$$

Berdasarkan persamaan (7) dilihat bahwa kesalahan baku pengukuran *standard error measurement* (SEM) memiliki hubungan kuadratik yang berbanding terbalik dengan fungsi informasi tes. Hal ini berarti bahwa semakin besar nilai informasi tes yang dihasilkan oleh suatu model, maka semakin tinggi tingkat presisi estimasi model tersebut, dan sebaliknya. Perlu dipahami terlebih dahulu mengenai konsep Fungsi Informasi. Istilah “informasi” adalah sebuah indikator statistika terkait kualitas estimasi parameter, dalam penelitian ini adalah parameter kemampuan (Lee et al., 2020). Istilah ini digunakan untuk menggambarkan seberapa yakin kita terhadap lokasi ( $\theta$ ) peserta tes. Untuk menjelaskan ketidakpastian tentang estimasi, dapat digunakan konsep kesalahan baku (*standard error*). Dalam konteks pengukuran kemampuan, SEM yang lebih kecil mencerminkan tingkat presisi estimasi yang lebih tinggi. Beberapa studi merekomendasikan nilai SEM di bawah 0,5 sebagai indikator kualitas pengukuran yang dapat diterima, terutama dalam pendekatan berbasis IRT (Huebner & Skar, 2021).

Rentang skor kemampuan responden pada analisis IRT umumnya berada antara  $-3$  sampai  $+3$  logit, mengikuti distribusi normal untuk interpretasi kemampuan (Falani et al., 2020). Tabel 2.1 merupakan kriteria logit keterampilan berpikir kritis matematis siswa.

**Tabel 3. 4 Kriteria Keterampilan Berpikir Kritis Siswa Berdasarkan Nilai Logit ( $\theta$ )**

<b>Logit</b>	<b>Kategori Kemampuan</b>	<b>Deskripsi</b>
$\leq -2.00$	Sangat Rendah	Siswa memiliki kemampuan berpikir kritis matematis yang sangat terbatas
$-1.99$ s.d. $-1.00$	Rendah	Siswa cenderung kesulitan dalam menyelesaikan soal dengan tuntutan analitis
$-0.99$ s.d. $0.00$	Cukup	Siswa mulai menunjukkan kemampuan berpikir kritis, namun belum konsisten
$0.01$ s.d. $1.00$	Sedang	Siswa mampu menyelesaikan soal dengan tuntutan berpikir kritis secara umum
$1.01$ s.d. $2.00$	Tinggi	Siswa menunjukkan kemampuan berpikir kritis yang kuat dan stabil
$> 2.00$	Sangat Tinggi	Siswa sangat mahir dalam berpikir kritis matematis dan mampu menyelesaikan soal kompleks

### 3.7.4 Uji Prasyarat Pengujian Hipotesis

Sebelum melakukan pengujian hipotesis penelitian, diperlukan uji prasyarat, yaitu uji normalitas data. Uji ini bertujuan untuk memastikan bahwa data hasil estimasi parameter kemampuan memiliki distribusi normal. Metode yang digunakan untuk uji normalitas adalah uji Kolmogorov-Smirnov, dengan hipotesis yang menyertainya.

$H_0$  = data berdistribusi normal

$H_1$  = data tidak berdistribusi normal

Kriteria uji yang digunakan adalah nilai  $\text{sig} > 0.05$  maka  $H_0$  diterima, yang berarti bahwa data tersebut berdistribusi normal. Uji normalitas pada penelitian ini dilakukan dengan menggunakan SPSS 26.

### 3.7.5 Pengujian Hipotesis Penelitian

Hipotesis pada penelitian ini di uji dengan uji beda varians, dengan membandingkan varians dari dua kelompok, dengan persamaan sebagai berikut:

$$F = \frac{S_1^2}{S_2^2}$$

Keterangan:

$S_1^2$  = varian terbesar dari kedua kelompok uji

$S_2^2$  = Varian terkecil dari kedua kelompok uji

Dengan kriteria jika  $F_{hitung} > F_{tabel}$  maka  $H_0$  ditolak (Akdon & Riduwan, 2009)

Selain melakukan uji beda varians pada data empiris, analisis inferensi tambahan juga dilakukan pada data replikasi hasil simulasi (*generated*). Analisis ini menggunakan uji beda rerata RMSE dari dua kelompok data dengan Uji Independent T-test. Uji tersebut bertujuan untuk menguji perbedaan rerata RMSE hasil estimasi dari dua kelompok data berdasarkan 10 replikasi yang telah dilakukan. Persamaan yang digunakan untuk uji ini adalah sebagai berikut.

Jika  $\sigma_x^2 = \sigma_y^2 = \sigma$

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{S_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_g = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Jika  $\sigma_x^2 \neq \sigma_y^2$

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Selanjutnya nilai  $t_{hitung}$  tersebut kemudian dibandingkan dengan  $t_{tabel}$ . Kriteria uji yang digunakan adalah tolak  $H_0$  jika  $t_{hitung} > t_{tabel}$  (Akdon & Riduwan, 2009)

## **BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Pada bab ini disajikan hasil penelitian yang meliputi hasil validasi tes oleh ahli/pakar, hasil uji coba tes, deskripsi data, pengujian prasyarat analisis, pengujian hipotesis, dan pembahasan hasil penelitian.

### **4.1 Hasil Validasi Tes oleh Ahli/Pakar**

Validasi tes dilakukan secara kuantitatif dan kualitatif. Pada tahap awal, peneliti memberikan instrumen tes matematika dengan format butir campuran yang telah disusun kepada para ahli di bidang Matematika, Deskripsi profil Ahli/Pakar dapat dilihat pada Tabel 3.3.

Para ahli memberikan masukan dan saran untuk perbaikan tes, meliputi beberapa aspek: 1) kesesuaian antara kompetensi dasar dan indikator, 2) kesesuaian antara indikator dan soal, 3) kejelasan dalam perumusan soal, 4) penggunaan bahasa baku, dan 5) kesesuaian antara soal dan penyekoran. Peneliti melakukan perbaikan instrumen tes berdasarkan masukan dan saran dari para ahli.

Tahap selanjutnya, peneliti memberikan tes yang telah direvisi berdasarkan masukan dan saran sebelumnya kepada para ahli. Para ahli diminta untuk melakukan penilaian secara objektif untuk setiap butir dengan meninjau lima aspek yang telah disebutkan sebelumnya. Pemberian skor dilakukan dengan ketentuan: skor 1 jika hanya satu kriteria yang muncul, skor 2 jika hanya ada dua kriteria yang muncul, skor 3 jika hanya ada tiga kriteria yang muncul, skor 4 jika hanya ada empat kriteria yang muncul, skor 5 jika semua kriteria muncul. Rekapitulasi penilaian oleh para ahli dapat dilihat pada Tabel 4.1.

Tabel 4.1 Rekapitulasi Penilaian Butir Soal oleh Ahli

Butir	Pakar					Nilai V	Nilai V Tabel	Keterangan
	1	2	3	4	5			
1	4	4	5	5	5	0.9	0.8	Valid
2	4	5	5	4	5	0.9	0.8	Valid
3	4	4	5	4	5	0.85	0.8	Valid

Rangkuman hasil penilaian dari para ahli untuk tiap butir tes dapat dilihat pada tabel di atas. Perhitungan validasi isi oleh para ahli dilakukan dengan menggunakan metode Aiken. Perhitungan nilai V dengan menggunakan persamaan (10). Nilai V untuk masing-masing butir soal ditunjukkan pada kolom delapan tabel di atas. Hasil perhitungan nilai V untuk tiap butir dibandingkan dengan kriteria minimal sesuai nilai  $V_{tabel}$  yakni sebesar 0.75. Berdasarkan hasil perbandingan pada Tabel 4.1 diketahui bahwa butir soal memiliki validitas yang baik.

#### 4.2 Hasil Uji Coba Tes

Instrumen tes yang telah divalidasi oleh pakar atau ahli berjumlah tiga butir. Selanjutnya tes tersebut diujicobakan kepada 35 siswa kelas VIII SMPN 53 Muaro Jambi. Uji reliabilitas dilakukan terhadap data respons siswa menggunakan Metode Cronbach's Alpha. Berikut adalah hasil uji reliabilitas berdasarkan *output* SPSS 26.

Tabel 4.2 Uji Reliabilitas Tes Keterampilan Berpikir Kritis Matematis

<b>Case Processing Summary</b>			
		N	%
Cases	Valid	35	100.0
	Excluded <sup>a</sup>	0	.0
	Total	35	100.0
a. Listwise deletion based on all variables in the procedure.			
<b>Reliability Statistics</b>			
	Cronbach's Alpha	N of Items	
	.981	3	

Berdasarkan tabel di atas diperoleh Nilai Cronbach's Alpha sebesar 0,981 lebih dari kriteria nilai minimal Cronbach's Alpha yakni 0,6 Maulana (Maulana, 2022). Dapat disimpulkan tes yang digunakan dalam pengumpulan dalam

penelitian ini handal atau reliabel. Berdasarkan hasil uji-uji statistik pada data respon siswa yang telah dilakukan, dapat disimpulkan bahwa instrumen tes pada penelitian ini telah siap digunakan pengumpulan data penelitian karena telah memenuhi kriteria validitas dan reliabilitas.

Dalam penelitian hanya diambil dua butir soal untuk pengumpulan data keterampilan berpikir kritis matematis siswa yaitu butir soal satu dan butir soal dua dengan nilai  $V(0,9)$ . Masing-masing butir soal telah memuat keseluruhan indikator yang ingin dinilai. Dalam soal *esai representative*, setiap butir soal disusun untuk mencakup keseluruhan indikator kompetensi, menjadikan alat evaluasi yang holistik dalam menilai kedalaman dan keterampilan berpikir tingkat tinggi siswa (Arifin, 2009). Selanjutnya, dalam penelitian terdahulu mengenai asesmen *higher order thinking skills*, seperti yang dikemukakan oleh Brookhart (2010) dan Zohar & Dori (2003), lebih disarankan untuk mengutamakan soal yang mampu mengeksplorasi kemampuan bernalar secara mendalam dibanding sekadar memperbanyak jumlah soal. Dengan demikian, dua soal yang telah dikembangkan dan divalidasi telah memadai untuk mencapai tujuan penelitian ini.

#### **4.3 Hasil Uji Asumsi IRT**

Sebelum dilakukan analisis data lebih lanjut menggunakan PARSCALE 4.1. Perlu adanya pengujian asumsi prasyarat IRT. Analisis prasyarat model meliputi uji unidimensi, uji independensi lokal, dan uji invariansi parameter yang saling terkait. Jika uji unidimensi terpenuhi, maka uji independensi lokal dan invariansi parameter juga dianggap terpenuhi (Bahar & Retnawati, 2022). Menurut Alagoz (2000), asumsi unidimensi dapat diuji menggunakan analisis faktor, dengan bantuan *software Statistical Package for Social Science (SPSS) 26*. Uji unidimensi

dilakukan terhadap data-data sebelum digunakan untuk estimasi parameter kemampuan peserta tes. Data-data yang diuji tersebut adalah data politomi untuk PCM, GPCM, dan GRM.

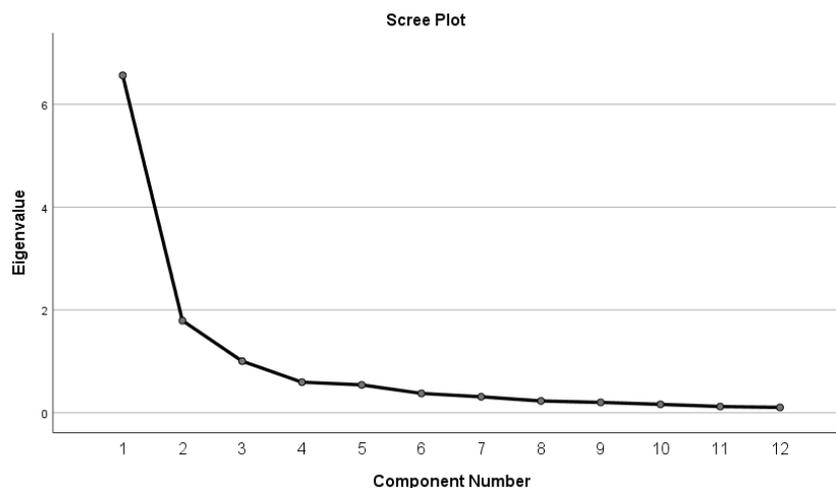
Uji unidimensi bertujuan memastikan instrumen hanya mengukur satu aspek, yaitu keterampilan berpikir kritis matematis siswa. Berdasarkan luaran SPSS 26, hasil dari *dimension reduction* untuk data-data yang diuji menunjukkan bahwa *Principal Component Analysis* mengekstraksi data menjadi sejumlah faktor dengan nilai eigen lebih dari satu.

**Tabel 4.3 Varians Total Tes Keterampilan Berpikir Kritis Matematis**

Component	Total Variance Explained	
	Total	% of Variance
1	6.563	54.693
2	1.793	14.939
3	1.004	8.363
4	.594	4.954
5	.542	4.514
6	.376	3.133
7	.310	2.584
8	.229	1.912
9	.201	1.678
10	.163	1.360
11	.121	1.006
12	.104	.865

Extraction Method: Principal Component Analysis.

Berdasarkan Tabel 4.3 data menghasilkan faktor utama dengan total *variance explained* sebesar 54,6%. Nilai eigen terbesar dimiliki oleh komponen pertama, yang menunjukkan bahwa komponen ini merupakan dimensi utama yang paling dominan dalam menjelaskan keragaman data. Data dikatakan Unidimensi apabila komponen pertama dalam analisis PCA menjelaskan lebih dari 50% varians total (Friyatmi, 2018). Hasil ini menunjukkan bahwa data yang dianalisis memiliki struktur faktor yang baik dan unidimensionalitas dengan faktor utama yang mendominasi. Berikut Gambar 4.1 merupakan grafik *scree plot* data.



**Gambar 4.1 Scree Plot Unidimensi Keterampilan Berpikir Kritis Matematis**

Hasil perhitungan analisis faktor terhadap data yang diuji serta *Scree Plot* pada Gambar 4.1 tampak bahwa grafik *scree plot* memiliki penurunan yang sangat curam pada komponen pertama ke komponen kedua, yang kemudian diikuti oleh garis yang cenderung mendatar. *scree plot* yang curam di awal menunjukkan dominasi satu faktor utama, yang merupakan indikator kuat dari struktur unidimensi. Penurunan tajam ini mencerminkan bahwa komponen pertama menjelaskan proporsi varians yang jauh lebih besar dibandingkan komponen berikutnya (Gustiawan et al., 2019). Hal ini mengindikasikan bahwa instrumen yang digunakan memiliki kecenderungan unidimensional, yakni mengukur satu konstruk utama secara dominan. Dalam konteks penelitian ini, konstruk tersebut adalah keterampilan berpikir kritis matematis.

Setelah dilakukan uji prasyarat unidimensi, selanjutnya dilakukan uji kecocokan model. Berdasarkan hasil uji kecocokan pada butir tes dengan bantuan PARSCALE 4.1 yang dapat dilihat pada Lampiran 6. diperoleh *item fit statistics* nilai  $\chi^2$  PCM sebesar 106. ( $p$ -value= 0,264), nilai  $\chi^2$  GPCM sebesar 26. ( $p$ -value= 0,987), nilai  $\chi^2$  GRM sebesar 44. ( $p$ -value= 0,819). Nilai  $p$ -value (PROB.) harus

diatas batas signifikasi umum 0,05 yang mengindikasikan kecocokan model baik secara keseluruhan (Niken et al., 2024). Selain itu, nilai-nilai  $\chi^2$  yang dihasilkan kurang dari nilai kritis distribusi  $\chi^2$  nya, hal ini menunjukkan bahwa model yang digunakan memiliki kecocokan yang tinggi dengan data empiris (Naga, 2012; Yilmaz, 2019). Selanjutnya model-model tersebut dapat digunakan untuk mengestimasi parameter kemampuan peserta tes.

#### 4.4 Deskripsi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis

Estimasi parameter keterampilan dilakukan pada hasil respon tes dari 108 siswa. Estimasi parameter keterampilan dilakukan secara simultan dan terpisah dengan menggunakan bantuan PARSCALE 4.1. Telah dijelaskan sebelumnya bahwa hasil estimasi parameter keterampilan peserta tes terdapat pada ph.3 luaran PARSCALE 4.1.

Pengecekan normalitas distribusi dilakukan terhadap hasil estimasi parameter baik secara simultan maupun terpisah yang telah dilakukan. Pengecekan distribusi keterampilan ini dilakukan dengan menggunakan program SSPS 26, normalitas Kolmogorov-Smirnov.

**Tabel 4.4 Uji Normalitas Keterampilan Berpikir Kritis Matematis Siswa**

	Tests of Normality					
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PCM	.069	106	.200*	.987	106	.399
GPCM	.084	106	.064	.984	106	.227
GRM	.071	106	.200*	.979	106	.093

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Berdasarkan table 4.4 dapat dilihat dari ketiga data yang dianalisis menghasilkan nilai sig > 0.05. Menurut kriteria uji Kolmogorov-Smirnov, jika nilai

sig > 0.05 maka data-data tersebut berdistribusi normal (Kadir, 2017). Data ini selanjutnya digunakan untuk keperluan analisis dan pengujian hipotesis pada penelitian ini. Berdasarkan Lampiran 7. berikut deskripsi statistik hasil estimasi parameter keterampilan yang telah dilakukan.

**Tabel 4.5 Distribusi Frekuensi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan PCM**

<b>No</b>	<b>Kelompok</b>		<b>Frekuensi</b>	<b>Median</b>	<b>Fk</b>	<b>Prob</b>	<b>Prob Komul</b>
1	-2.59	-1.89	4	-2.24	4	4%	4%
2	-1.88	-1.24	9	-1.56	13	8%	12%
3	-1.23	-0.53	19	-0.88	32	18%	30%
4	-0.52	0.18	29	-0.17	61	27%	58%
5	0.19	0.89	24	0.54	85	23%	80%
6	0.90	1.60	18	1.25	103	17%	97%
7	1.61	2.31	2	1.96	105	2%	99%
8	2.32	3.02	1	2.67	106	1%	100%

Berdasarkan Tabel 4.5, diperoleh distribusi kemampuan berpikir kritis matematis siswa yang bervariasi. Sebagian besar siswa berada pada kategori cukup hingga sedang, yaitu sebanyak 48 siswa (45%) pada kategori cukup dan 24 siswa (23%) pada kategori sedang. Hal ini menunjukkan bahwa mayoritas siswa mulai mampu menunjukkan kemampuan berpikir kritis, meskipun sebagian di antaranya masih belum konsisten. Sementara itu, terdapat 20 siswa (19%) yang berada pada kategori tinggi, menunjukkan bahwa mereka memiliki kemampuan berpikir kritis yang kuat dan stabil. Pada kategori sangat tinggi, hanya terdapat 1 siswa (1%) yang menunjukkan kemampuan berpikir kritis sangat baik dalam menyelesaikan soal-soal kompleks. Di sisi lain, masih terdapat 13 siswa (12%) yang berada pada kategori rendah hingga sangat rendah, yang mengindikasikan bahwa mereka mengalami kesulitan dalam berpikir kritis dan memerlukan pendampingan lebih

lanjut. Secara keseluruhan, hasil ini menunjukkan bahwa kemampuan berpikir kritis matematis siswa cenderung berada pada tingkat cukup hingga sedang.

**Tabel 4.6 Distribusi Frekuensi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GPCM**

No	Kelompok		Frekuensi	Median	Fk	Prob	Prob Kumul
1	-2.91	-2.21	3	-2.56	3	3%	3%
2	-2.20	-1.46	4	-1.83	7	4%	7%
3	-1.45	-0.75	15	-1.10	22	14%	21%
4	-0.74	-0.04	24	-0.39	46	23%	43%
5	-0.03	0.67	29	0.32	75	27%	71%
6	0.68	1.38	25	1.03	100	24%	94%
7	1.39	2.09	5	1.74	105	5%	99%
8	2.10	2.80	1	2.45	106	1%	100%

Tabel 4.6 menunjukkan bahwa sebanyak 3 siswa (3%) berada pada kategori sangat rendah dengan median -2.56. Artinya, siswa pada kategori ini memiliki kemampuan berpikir kritis matematis yang sangat terbatas dan kemungkinan besar mengalami kesulitan dalam menyelesaikan soal yang menuntut analisis mendalam. Selanjutnya, terdapat 4 siswa (4%) yang masuk dalam kategori rendah dengan median -1.83. Siswa dalam kategori ini masih menunjukkan kesulitan dalam menyelesaikan soal-soal yang membutuhkan kemampuan berpikir analitis. Selain itu, terdapat 15 siswa (14%) yang juga tergolong pada kategori rendah dengan median -1.10. Sehingga secara keseluruhan terdapat 19 siswa (18%) yang masih berada pada kategori kemampuan berpikir kritis rendah hingga sangat rendah. Pada kategori cukup, terdapat 24 siswa (23%) dengan median -0.39. Siswa pada kategori ini sudah mulai menunjukkan adanya kemampuan berpikir kritis matematis, meskipun masih belum konsisten dalam menyelesaikan soal-soal yang diberikan.

Sementara itu, kategori sedang diisi oleh 29 siswa (27%) dengan median 0.32. Artinya, sebagian besar siswa dalam kategori ini sudah mampu menyelesaikan soal dengan tuntutan berpikir kritis secara umum. Adapun pada kategori tinggi, terdapat

25 siswa (24%) dengan median 1.03 dan 5 siswa (5%) dengan median 1.74. Hal ini menunjukkan bahwa sebanyak 30 siswa (28%) memiliki kemampuan berpikir kritis matematis yang kuat dan stabil. Terakhir, terdapat 1 siswa (1%) yang berada pada kategori sangat tinggi dengan median 2.45. Siswa ini sangat mahir dalam berpikir kritis matematis serta mampu menyelesaikan soal-soal yang bersifat kompleks dengan baik. Jika dilihat secara keseluruhan, mayoritas siswa berada pada kategori cukup hingga sedang, yaitu sebanyak 53 siswa (50%).

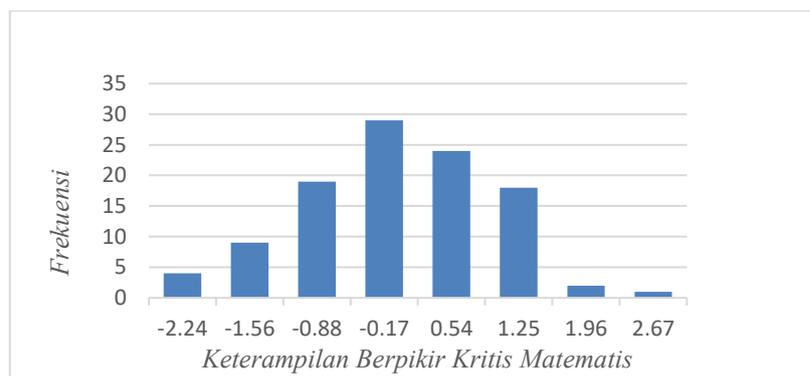
**Tabel 4.7 Distribusi Frekuensi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GRM**

No	Kelompok		Frekuensi	Median	Fk	Prob	Prob Kumul
1	-2.77	-2.07	3	-2.42	3	3%	3%
2	-2.06	-1.37	9	-1.72	12	8%	11%
3	-1.36	-0.66	10	-1.01	22	9%	21%
4	-0.65	0.05	31	-0.30	53	29%	50%
5	0.06	0.76	25	0.41	78	24%	74%
6	0.77	1.47	24	1.12	102	23%	96%
7	1.48	2.18	3	1.83	105	3%	99%
8	2.19	2.89	1	2.54	106	1%	100%

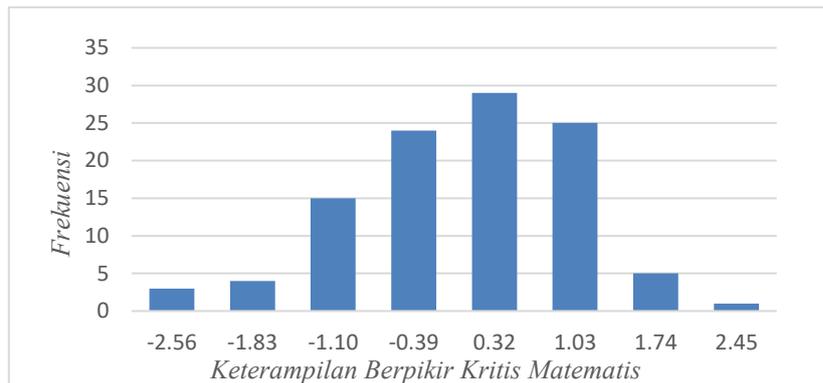
Tabel 4.7 menggambarkan distribusi frekuensi estimasi parameter keterampilan berpikir kritis matematis siswa berdasarkan pendekatan GRM. Berdasarkan hasil analisis distribusi kemampuan berpikir kritis matematis siswa, diperoleh hasil yang bervariasi pada tiap kategori. Sebanyak 3 siswa (3%) berada pada kategori sangat rendah dengan rentang skor  $\leq -2.00$ , yang menunjukkan bahwa siswa dalam kategori ini memiliki kemampuan berpikir kritis matematis yang sangat terbatas dan mengalami kesulitan besar dalam menyelesaikan soal-soal dengan tuntutan analitis. Selanjutnya, terdapat 19 siswa (17%) yang berada pada kategori rendah (-1.99 s.d. -1.00), yang mengindikasikan bahwa siswa dalam kategori ini masih cenderung mengalami kesulitan dalam menyelesaikan soal yang menuntut analisis lebih mendalam. Pada kategori cukup (-0.99 s.d. 0.00), sebanyak

31 siswa (29%) mulai menunjukkan kemampuan berpikir kritis, meskipun belum konsisten dalam menerapkannya secara menyeluruh. Sementara itu, terdapat 25 siswa (24%) yang berada pada kategori sedang (0.01 s.d. 1.00), yang berarti mereka umumnya sudah mampu menyelesaikan soal yang memerlukan kemampuan berpikir kritis secara umum. Pada kategori tinggi (1.01 s.d. 2.00), terdapat 24 siswa (23%) yang telah menunjukkan kemampuan berpikir kritis yang kuat dan stabil. Terakhir, sebanyak 4 siswa (4%) berada pada kategori sangat tinggi ( $> 2.00$ ), yang menunjukkan bahwa siswa tersebut sangat mahir dalam berpikir kritis matematis dan mampu menyelesaikan soal-soal kompleks dengan baik.

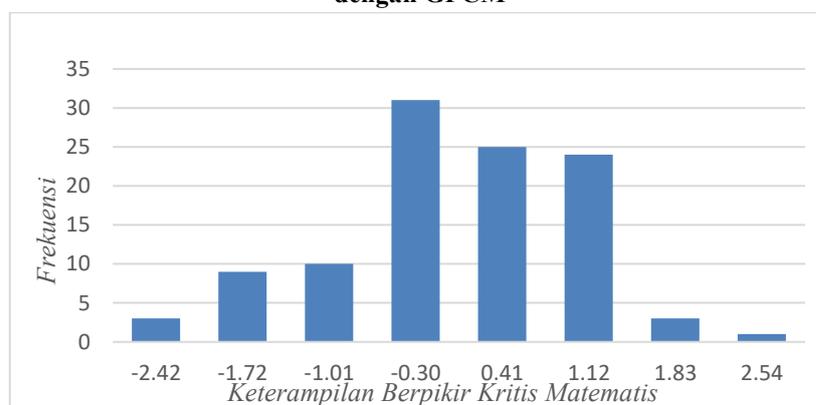
Secara keseluruhan, mayoritas siswa berada pada kategori cukup hingga tinggi, yaitu sebesar 76%, yang menunjukkan bahwa sebagian besar siswa sudah memiliki kemampuan berpikir kritis yang cukup baik, meskipun masih terdapat sebagian siswa yang perlu mendapatkan perhatian khusus karena berada pada kategori rendah hingga sangat rendah. Selain itu, distribusi data juga divisualisasikan melalui histogram berikut.



**Gambar 4.2 Histogram Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan PCM**



**Gambar 4.3** Histogram Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GPCM



**Gambar 4.4** Histogram Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan GRM

#### 4.5 Deskripsi Hasil Estimasi Parameter Butir

Estimasi parameter butir soal merupakan suatu parameter yang menggambarkan daya beda (a) dan tingkat kesulitan suatu butir soal (b). Hasil analisis terdapat pada luaran *Parscale 4.1* pada Ph.2.

**Tabel 4.8** Hasil Estimasi Parameter Butir Tes Keterampilan Berpikir Kritis Matematis dengan PCM

SUBITEM	BLOCK	SLOPE (a)	S.E.	LOCATION (b)	S.E.	GUESSING	S.E.
01	1	1.144	0.017	-0.874	0.028	0.000	0.000
02	2	1.144	0.017	-0.917	0.037	0.000	0.000
03	3	1.144	0.017	-0.749	0.033	0.000	0.000
04	4	1.144	0.017	-0.018	0.028	0.000	0.000
05	5	1.144	0.017	0.520	0.030	0.000	0.000
06	6	1.144	0.017	1.017	0.036	0.000	0.000
07	7	1.144	0.017	-1.239	0.047	0.000	0.000
08	8	1.144	0.017	-0.724	0.040	0.000	0.000
09	9	1.144	0.017	-0.554	0.041	0.000	0.000
10	10	1.144	0.017	-0.049	0.029	0.000	0.000

SUBITEM	BLOCK	SLOPE (a)	S.E.	LOCATION (b)	S.E.	GUESSING	S.E.
11	11	1.144	0.017	0.316	0.026	0.000	0.000
12	12	1.144	0.017	1.223	0.038	0.000	0.000

Pada PCM mengasumsikan bahwa semua subitem (indikator) memiliki kemampuan yang sama dalam membedakan siswa dengan tingkat kemampuan berbeda. Fokus utama PCM hanya pada tingkat kesukaran tiap kategori dalam satu butir, tanpa memperhitungkan seberapa tajam subitem tersebut membedakan siswa berkemampuan tinggi dan rendah. Dari Tabel 4.8 hasil analisis dengan menggunakan PCM dari output PARSCALE Phase 2, diketahui bahwa tingkat kesukaran dari 6 indikator keterampilan berpikir kritis matematis tersebar dalam beberapa kategori. Tingkat kesukaran ini ditentukan berdasarkan nilai parameter lokasi (*step difficulty*) yang diperoleh dari masing-masing indikator dalam satuan logit. Interpretasi dilakukan dengan merujuk pada rentang nilai logit, di mana butir dengan nilai  $\leq -2.00$  dikategorikan sangat mudah,  $-2.00$  s.d.  $-0.50$  sebagai mudah,  $-0.49$  s.d.  $+0.49$  sebagai sedang,  $+0.50$  s.d.  $+2.00$  sebagai sulit, dan  $\geq +2.01$  sebagai sangat sulit (Amelia & Kriswantoro, 2017).

Dari hasil analisis, sebanyak enam indikator tergolong dalam kategori mudah, yaitu *focus* (01) ( $-0.874$ ), *reason* (02) ( $-0.917$ ), *inference* (03) ( $-0.749$ ), *focus* (07) ( $-1.239$ ), *reason* (08) ( $-0.724$ ), dan *inference* (09) ( $-0.554$ ). Hal ini menunjukkan bahwa sebagian besar siswa mampu menjawab butir-butir ini dengan benar karena tidak membutuhkan kemampuan berpikir yang terlalu kompleks. Selanjutnya, terdapat tiga butir yang berada dalam kategori sedang, yaitu *situation* (04) ( $-0.018$ ), *clarity* (10) ( $-0.049$ ), dan *clarity* (11) ( $0.316$ ). Indikator tersebut cocok untuk siswa dengan kemampuan rata-rata dan dapat digunakan untuk menilai kemampuan dasar hingga menengah. Sementara itu, tiga indikator lainnya tergolong sulit, yakni

*clarity* (05) (0.520), *overview* (06) (1.017), dan *overview* (12) (1.223). indikator-indikator ini menuntut siswa untuk memiliki kemampuan berpikir kritis di atas rata-rata agar dapat menjawab dengan benar.

**Tabel 4.9 Hasil Estimasi Parameter Butir Tes Keterampilan Berpikir Kritis Matematis dengan GPCM**

SUBITEM	BLOCK	SLOPE (a)	S.E.	LOCATION (b)	S.E.	GUESSING	S.E.
01	1	0.348	0.017	-2.426	0.092	0.000	0.000
02	2	0.650	0.027	-1.731	0.062	0.000	0.000
03	3	0.705	0.044	-1.427	0.051	0.000	0.000
04	4	0.940	0.065	-0.297	0.036	0.000	0.000
05	5	1.361	0.085	0.432	0.032	0.000	0.000
06	6	1.361	0.106	0.990	0.035	0.000	0.000
07	7	0.614	0.035	-2.306	0.087	0.000	0.000
08	8	1.252	0.086	-1.208	0.041	0.000	0.000
09	9	1.724	0.142	-0.951	0.036	0.000	0.000
10	10	0.832	0.067	-0.357	0.042	0.000	0.000
11	11	0.774	0.048	0.173	0.037	0.000	0.000
12	12	1.063	0.080	1.276	0.041	0.000	0.000

Berdasarkan Tabel 4.9 menggunakan model IRT GPCM pada *software* PARSCALE Phase 2, diperoleh parameter lokasi (*difficulty parameter*) dari masing-masing indikator merepresentasikan tingkat kesukaran. Berdasarkan kategori rentang logit, diketahui bahwa dari total 6 indikator, terdapat indikator *frisco* (01 dan 07) yang tergolong sangat mudah, dengan nilai logit di bawah -2.00. Selanjutnya, indikator *reason* dan *inference* (subitem 02, 03, 08, dan 09) termasuk dalam kategori mudah, karena memiliki nilai logit di antara -2.00 hingga -0.50. Kemudian indikator *situation* dan *clarity* (item 04, 05, 10, dan 11) berada pada rentang logit -0.49 hingga +0.49, sehingga diklasifikasikan dalam kategori sedang, atau cocok untuk siswa dengan kemampuan rata-rata. Sementara itu, indikator *overview* (06 dan 12) memiliki nilai logit antara +0.50 hingga +2.00 dan tergolong dalam kategori sulit, yang berarti indikator-indikator ini menuntut kemampuan berpikir

kritis siswa di atas rata-rata. Tidak terdapat butir soal yang masuk dalam kategori sangat sulit ( $\logit \geq +2.01$ ). Distribusi ini menunjukkan bahwa instrumen tes memiliki penyebaran tingkat kesukaran yang cukup seimbang, dengan dominasi soal-soal pada kategori mudah hingga sedang. Keberagaman tingkat kesukaran ini penting dalam mengukur kemampuan siswa secara proporsional, karena memungkinkan identifikasi siswa dengan kemampuan rendah, sedang, maupun tinggi secara lebih adil dan akurat.

Selanjutnya hasil analisis parameter *slope* (a) atau daya beda subitem. Daya beda item ditentukan oleh nilai slope (a). Nilai  $< 0.25$  menunjukkan item sangat lemah dalam membedakan peserta,  $0.25-0.64$  rendah,  $0.65-1.34$  sedang,  $1.35-1.69$  tinggi, dan  $\geq 1.70$  sangat tinggi, yang berarti item sangat efektif mengidentifikasi perbedaan kemampuan peserta (Amelia & Kriswantoro, 2017). diperoleh bahwa sebagian besar subitem berada pada kategori sedang, yaitu sebanyak 7 subitem dari 12 subitem yang dianalisis. Hal ini menunjukkan bahwa mayoritas subitem memiliki kemampuan diskriminasi yang cukup baik dalam membedakan peserta dengan tingkat kemampuan berpikir kritis matematis yang berbeda. Dua subitem, yaitu subitem 05 (*clarity*) dan subitem 06 (*overview*), masuk dalam kategori daya beda tinggi, sehingga mampu membedakan peserta dengan baik. Selain itu, terdapat satu subitem, yaitu subitem 09 (*inference*), yang memiliki daya beda sangat tinggi, yang berarti subitem tersebut sangat efektif dalam membedakan peserta dengan kemampuan tinggi dan rendah. Namun demikian, terdapat dua subitem yang memiliki daya beda rendah, yaitu subitem 01 (*frisco*) dan subitem 07 (*frisco*). Subitem-subitem ini kurang efektif dalam membedakan peserta, karena memiliki

daya beda sangat rendah. Secara keseluruhan, instrumen ini sudah cukup baik dalam mengukur kemampuan berpikir kritis matematis peserta.

**Tabel 4.10 Hasil Estimasi Parameter Butir Tes Keterampilan Berpikir Kritis Matematis dengan GRM**

SUBITEM	BLOCK	SLOPE (a)	S.E.	LOCATION (b)	S.E.	GUESSING	S.E.
01	1	0.492	0.082	-2.669	0.435	0.000	0.000
02	2	0.952	0.161	-2.084	0.318	0.000	0.000
03	3	1.017	0.180	-1.537	0.284	0.000	0.000
04	4	1.342	0.251	-0.341	0.156	0.000	0.000
05	5	1.691	0.463	0.541	0.161	0.000	0.000
06	6	1.283	0.277	1.251	0.172	0.000	0.000
07	7	0.945	0.194	-2.433	0.276	0.000	0.000
08	8	1.656	0.393	-1.391	0.199	0.000	0.000
09	9	1.812	0.554	-1.074	0.156	0.000	0.000
10	10	1.175	0.215	-0.368	0.166	0.000	0.000
11	11	1.295	0.367	0.199	0.145	0.000	0.000
12	12	1.346	0.261	1.435	0.1670	0.000	0.000

Berdasarkan Tabel 4.10 *output* PARSCALE Phase 2, diketahui bahwa tingkat kesukaran dari enam indikator keterampilan berpikir kritis matematis yang dianalisis memiliki distribusi yang cukup beragam. Berdasarkan kategori tingkat kesukaran menurut rentang logit, ditemukan bahwa terdapat dua indikator yang tergolong sangat mudah, yaitu *focus* dan *reason* 01, 02, dan 07, karena memiliki nilai parameter lokasi  $\leq -2.00$ . Hal ini menunjukkan bahwa soal-soal tersebut dapat dijawab dengan benar oleh siswa yang memiliki kemampuan sangat rendah. Selain itu, terdapat tiga indikator yang tergolong mudah, yakni *inference* dan *reason*, 0003, 0008, dan 0009, dengan nilai parameter lokasi antara  $-2.00$  hingga  $-0.50$ , yang berarti soal ini cenderung dijawab benar oleh sebagian besar siswa.

Selanjutnya, terdapat dua indikator *situation* dan *clarity* yang berada pada kategori sedang, yaitu subitem 04, 10, dan 11, dengan nilai logit berkisar antara  $-0.49$  hingga  $+0.49$ . Soal-soal dalam kategori ini cocok digunakan untuk mengukur

kemampuan siswa dengan tingkat rata-rata. Sementara itu, indikator lainnya tergolong sulit, yaitu *clarity* dan *overview* subitem 05, 06, dan 12, karena nilai parameter lokasi berada dalam rentang +0.50 hingga +2.00. Soal-soal dalam kategori ini menuntut kemampuan berpikir kritis matematis yang lebih tinggi dan hanya dapat dijawab dengan baik oleh siswa dengan kemampuan di atas rata-rata.

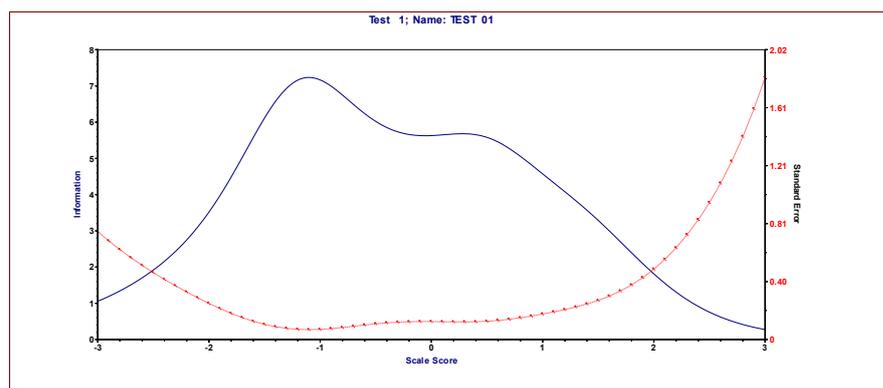
Menariknya, dari seluruh butir soal yang dianalisis, tidak terdapat soal yang tergolong sangat sulit (nilai logit  $\geq +2.01$ ), yang berarti seluruh butir soal masih berada dalam cakupan kemampuan siswa, meskipun dengan variasi tingkat kesulitan yang baik. Distribusi ini menunjukkan bahwa secara umum, perangkat tes telah mencakup berbagai level kesulitan.

Berdasarkan Tabel 4.10 parameter daya beda (*slope*) terhadap enam indikator keterampilan berpikir kritis matematis, diketahui bahwa nilai *slope* ( $a$ ) berkisar antara 0.492 hingga 1.812. Mengacu pada kriteria klasifikasi daya beda, sebagian besar item menunjukkan kemampuan yang baik dalam membedakan peserta didik berdasarkan tingkat kemampuannya. *Focus* (01) memiliki nilai *slope* sebesar 0.492, yang termasuk dalam kategori rendah. Hal ini menunjukkan bahwa item ini kurang efektif dalam membedakan peserta didik dengan kemampuan rendah dan tinggi. Sebaliknya, indikator 02, 03, 06, 10, dan 11 memiliki nilai *slope* antara 0.945 hingga 1.295, yang tergolong dalam kategori sedang. Ini mengindikasikan bahwa item-item tersebut memiliki kemampuan diskriminatif yang cukup baik dan mampu membedakan peserta didik berdasarkan variasi kemampuan secara seimbang. Indikator *situation* (04) dan *overview* (12) menunjukkan nilai *slope* masing-masing 1.342 dan 1.346, yang berada tepat di batas bawah kategori tinggi, menandakan bahwa indikator-indikator ini memiliki kemampuan diskriminasi yang baik.

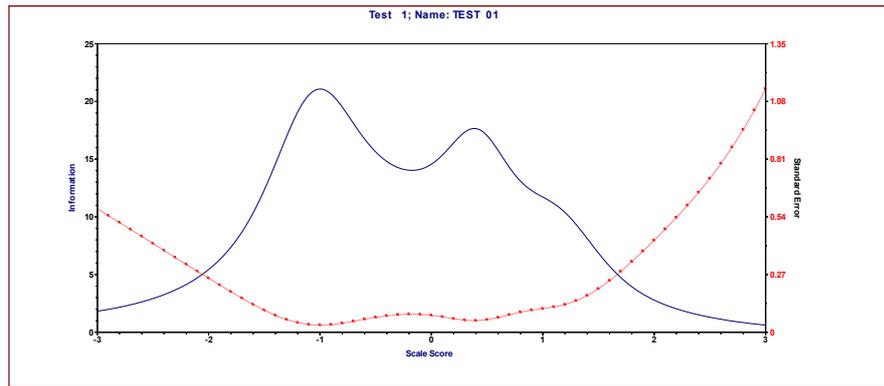
Indikator *clarity* (05) dan *reason* (08) memiliki nilai *slope* masing-masing 1.691 dan 1.656, yang termasuk dalam kategori tinggi, menunjukkan bahwa kedua item ini efektif dalam mengidentifikasi perbedaan kemampuan peserta didik. Sementara itu, indikator *inference* (09) memiliki nilai *slope* tertinggi sebesar 1.812, yang termasuk dalam kategori sangat tinggi, menunjukkan bahwa item ini sangat efektif membedakan peserta didik dengan kemampuan tinggi dan rendah.

#### 4.6 Fungsi Informasi Tes

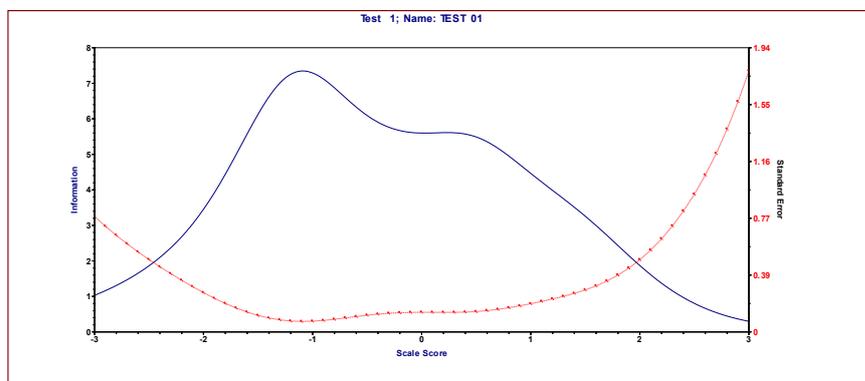
Fungsi informasi tes menyajikan data yang berkaitan dengan karakteristik keterampilan siswa. Terdapat hubungan terbalik antara fungsi informasi butir dan kesalahan baku pengukuran; artinya, semakin kecil tingkat kesalahan pengukuran, maka semakin tinggi informasi yang dapat diberikan oleh butir tes tersebut (Falani et al., 2020b). Selain itu, fungsi ini membantu menentukan bagian mana dari rentang kemampuan yang paling dapat diukur secara tepat oleh suatu tes, dan berguna untuk merancang tes yang sesuai dengan tujuan pengukuran, yaitu estimasi parameter keterampilan berpikir kritis matematis.



Gambar 4.5 Kurva Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis PCM



**Gambar 4.6 Kurva Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis GPCM**



**Gambar 4.7 Kurva Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis GRM**

Gambar 4.6 hingga 4.8 menyajikan informasi terkait seluruh butir dalam tes, yang digambarkan melalui garis biru solid, sedangkan kurva standar kesalahan pengukuran ditunjukkan dengan garis merah putus-putus. Nilai informasi total untuk setiap tingkat skor dapat diamati pada sumbu vertikal sebelah kiri, sementara standar kesalahan pengukuran ditampilkan pada sumbu vertikal di sisi kanan melalui garis merah putus-putus tersebut. Dari visualisasi ini, dapat diketahui bahwa standar kesalahan pengukuran meningkat secara signifikan pada area di mana informasi yang diberikan oleh butir soal relatif rendah. Berikut ringkasan perbedaan antara ke tiga model tersebut.

**Tabel 4.11 Ringkasan Fungsi Informasi Tes Keterampilan Berpikir Kritis Matematis**

<b>Aspek</b>	<b>PCM</b>	<b>GPCM</b>	<b>GRM</b>
Nilai Informasi	~7	~20	~7
Posisi Puncak	Skor -1	Skor -1 s/d 0,5	Skor -1
Lebar Kurva	-2 s.d. +2	-2 s.d. +2	-2 s.d. +2
Standar Error Minimum	~0,1	~0.1	~0.1
Standar Error Maksimum	~1.99 (ujung)	~1,09 (ujung)	~1,90 (ujung)
Akurasi Pengukuran	Tinggi di skor -1	Paling tinggi	Tinggi di skor -1

Berdasarkan kurva fungsi informasi dan standar kesalahan dari ketiga model, terlihat bahwa model GPCM memiliki puncak informasi tertinggi, yaitu mendekati 20, diikuti oleh PCM dan GRM dengan nilai sekitar 7. Ketiga model menunjukkan bahwa informasi tertinggi dan standar kesalahan terendah terjadi pada skor kemampuan sekitar -1 hingga 0,5, yang berarti ketepatan pengukuran tertinggi terdapat pada peserta dengan kemampuan rendah hingga sedang. Dari segi lebar kurva, PCM dan GPCM menunjukkan sebaran informasi yang cukup luas, mencakup rentang -2 hingga +2, sementara GRM memberikan informasi yang lebih terbatas pada rentang tersebut. Selain itu, model PCM menunjukkan standar kesalahan yang paling tinggi di area ekstrem, yang mengindikasikan penurunan akurasi pengukuran secara signifikan pada peserta dengan kemampuan sangat rendah maupun sangat tinggi. Dengan demikian, GPCM dapat dikatakan sebagai model yang paling efektif dalam hal ketepatan dan efisiensi pengukuran, diikuti oleh GRM dan PCM.

#### **4.7 Pengujian Hipotesis Penelitian**

Pengujian hipotesis dalam penelitian ini dilakukan dengan menguji hipotesis yang telah dirumuskan sebelumnya, sebelum dilakukan uji hipotesis data hasil estimasi keterampilan dari tiga model terlebih dahulu di replikasi menggunakan *software* WinGen. Setelah itu hasil varians dari ke tiga model yang telah di replikasi di uji menggunakan uji perbedaan varians terhadap hasil

estimasi parameter. Mengacu pada desain penelitian yang disajikan dalam Tabel 3.1, terdapat tiga hipotesis yang diuji.

**Tabel 4.12 Ringkasan Deskripsi Hasil Estimasi Keterampilan Berpikir Kritis Matematis**

<b>Statistik</b>	<b>PCM</b>	<b>GPCM</b>	<b>GRM</b>
Mean	0,0058	0,0098	0,0035
Median	0,0985	0,0340	0,0660
Standar Deviasi	1,11009	1,93903	1.10417
varians	1,232	0,882	1,219
Range	5,08	4,13	5,19
Minimum	-2,66	-1,84	-2,67
Maximum	2,42	2,29	2,52

Tabel 4.12 menyajikan ringkasan deskriptif dari data hasil estimasi parameter kemampuan peserta tes berdasarkan model-model yang digunakan dalam penelitian ini. Hasil estimasi menunjukkan rentang nilai yang relatif sama, rata-rata yang tertinggi diperoleh oleh model GPCM, sementara untuk hasil varians, GPCM memiliki hasil varians yang paling rendah diantara ketiganya.

#### 1. Pengujian Hipotesis Pertama

Hipotesis pertama dalam penelitian ini adalah membandingkan nilai variansi antara model PCM dan model GPCM, berikut hipotesis statistiknya.

$$H_0 = \sigma^2 GPCM = \sigma^2 PCM \quad H_1 = \sigma^2 GPCM < \sigma^2 PCM$$

Keterangan:

$\sigma^2 PCM$  = nilai varians hasil estimasi parameter keterampilan dengan PCM

$\sigma^2 GPCM$  = nilai varians hasil estimasi parameter keterampilan dengan GPCM

**Tabel 4.13 Pengujian Hipotesis Pertama antara PCM dan GPCM**

<b>Statistik</b>	<b>Hasil Estimasi Parameter</b>	
	<b>PCM</b>	<b>GPCM</b>
$n$	$n_1 = 106$	$n_2 = 106$
$df$	$df_1 = n_1 - 1 = 105$	$df_2 = n_2 - 1 = 105$
$\sigma^2$	1,232	0,882
$F_{hitung}$	$\frac{1,232}{0,882} = 1,396$	
$F_{tabel}$	$F_{(0,05;105;105)} = 1,381$	

Berdasarkan hasil pada tabel di atas, diperoleh bahwa nilai  $F_{hitung} > F_{tabel}$ , sehingga  $H_0$  ditolak. Ini menunjukkan bahwa varians estimasi parameter keterampilan peserta tes menggunakan GPCM lebih kecil dibandingkan dengan varians estimasi menggunakan PCM. Hal ini mengindikasikan bahwa estimasi parameter keterampilan berpikir kritis matematis peserta tes dengan menggunakan GPCM memiliki varians yang lebih kecil dibandingkan dengan PCM. Varians yang lebih kecil mencerminkan tingkat presisi yang lebih tinggi dalam pengukuran kemampuan berpikir kritis matematis.

## 2. Pengujian Hipotesis Kedua

Hipotesis kedua dalam penelitian ini adalah membandingkan nilai variansi antara model GPCM dan model GRM, berikut hipotesis statistiknya

$$H_0 = \sigma^2 GPCM = \sigma^2 GRM \quad H_1 = \sigma^2 GPCM < \sigma^2 GRM$$

Keterangan:

$\sigma^2 GPCM$  = nilai varians hasil estimasi parameter keterampilan dengan GPCM

$\sigma^2 GRM$  = nilai varians hasil estimasi parameter keterampilan dengan GRM

**Tabel 4.14 Pengujian Hipotesis Kedua antara GPCM dan GRM**

Statistik	Hasil Estimasi Parameter	
	GPCM	GRM
$n$	$n_1 = 106$	$n_2 = 106$
$df$	$df_1 = n_1 - 1 = 105$	$df_2 = n_2 - 1 = 105$
$\sigma^2$	0,882	1,219
$F_{hitung}$	$\frac{1,219}{0,882} = 1,382$	
$F_{tabel}$	$F_{(0,05;105;105)} = 1,381$	

Berdasarkan hasil pada tabel 4.14, diperoleh bahwa nilai  $F_{hitung} > F_{tabel}$ , sehingga  $H_0$  ditolak. Ini menunjukkan bahwa varians estimasi parameter keterampilan peserta tes menggunakan GPCM lebih kecil dibandingkan dengan varians estimasi menggunakan GRM. Temuan ini mengindikasikan bahwa estimasi parameter keterampilan berpikir kritis matematis peserta tes dengan menggunakan

GPCM memiliki varians yang lebih kecil dibandingkan dengan GRM. Varians yang lebih kecil mencerminkan tingkat presisi yang lebih tinggi dalam pengukuran kemampuan berpikir kritis matematis.

### 3. Pengujian Hipotesis Ketiga

Hipotesis ketiga dalam penelitian ini adalah membandingkan nilai variansi antara model PCM dan model GRM, berikut hipotesis statistiknya.

$$H_0 = \sigma^2 GPCM = \sigma^2 PCM \quad H_1 = \sigma^2 PCM < \sigma^2 GRM$$

Keterangan:

$\sigma^2 PCM$  = nilai varians hasil estimasi parameter keterampilan dengan PCM

$\sigma^2 GRM$  = nilai varians hasil estimasi parameter keterampilan dengan GRM

**Tabel 4.15 Pengujian Hipotesis Ketiga antara PCM dan GRM**

Statistik	Hasil Estimasi Parameter	
	GRM	PCM
$n$	$n_1 = 106$	$n_2 = 106$
$df$	$df_1 = n_1 - 1 = 105$	$df_2 = n_2 - 1 = 105$
$\sigma^2$	1,219	1,232
$F_{hitung}$	$\frac{1,232}{1,219} = 1,0106$	
$F_{tabel}$	$F_{(0,05;105;105)} = 1,381$	

Berdasarkan hasil Tabel 4.15, diperoleh bahwa nilai  $F_{hitung} < F_{tabel}$ , sehingga  $H_0$  diterima. Hal ini berarti bahwa varians hasil estimasi parameter kemampuan peserta tes dengan PCM tidak memiliki perbedaan yang signifikan dibandingkan varians hasil estimasi parameter kemampuan peserta tes dengan GRM. Artinya, varians hasil estimasi parameter kemampuan peserta tes dengan menggunakan PCM tidak berbeda secara signifikan dibandingkan dengan GRM. Dalam konteks pengukuran keterampilan berpikir kritis matematis, temuan ini menunjukkan bahwa kedua model memberikan tingkat presisi yang relatif setara dalam mengestimasi kemampuan siswa.

Berikut adalah rangkuman dari hasil pengujian hipotesis yang telah dilakukan.

**Tabel 4.16 Rangkuman Hasil Uji Hipotesis**

Hipotesis	Uji Beda Varians antar Model (Uji F)		Hasil Uji	Kesimpulan
	PCM	GPCM		
1	PCM	GPCM	Tolak $H_0$	GPCM lebih presisi
2	GPCM	GRM	Tolak $H_0$	GPCM lebih presisi
3	GRM	PCM	Terima $H_0$	Perbedaan Tidak Signifikan

Selanjutnya, dilakukan Uji SEM dalam konteks IRT untuk menilai tingkat presisi dari estimasi kemampuan peserta tes berdasarkan respons mereka terhadap subitem-subitem yang disusun menggunakan tiga model, yaitu PCM, GPCM, dan GRM. Hasil SEM ini bukan sekadar pelengkap analisis, melainkan berfungsi sebagai indikator penting dalam mendukung pengujian hipotesis. Ketika nilai SEM lebih kecil, hal ini menunjukkan bahwa model tersebut menghasilkan estimasi keterampilan berpikir kritis matematis yang lebih stabil dan akurat. Dengan demikian, perbandingan nilai SEM antar model memberikan dasar empiris yang kuat untuk menerima atau menolak hipotesis terkait efektivitas model dalam mengukur keterampilan berpikir kritis matematis. Presisi estimasi yang tinggi sangat krusial dalam penilaian keterampilan ini, karena berpikir kritis matematis melibatkan proses kognitif kompleks yang menuntut pengukuran yang sensitif terhadap variasi keterampilan berpikir kritis matematis siswa.

#### 1. Model GRM

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} = 0,406$$

## 2. Model GPCM

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} = 0,354$$

## 3. Model PCM

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} = 0,417$$

Berdasarkan hasil pengukuran SEM dari ketiga model, diketahui bahwa Model GPCM memiliki nilai SEM terkecil yaitu 0,354, diikuti oleh GRM sebesar 0,406, dan PCM sebesar 0,417. Semakin kecil nilai SEM menunjukkan bahwa estimasi kemampuan peserta tes semakin presisi, karena kesalahan pengukuran yang menyertai estimasi tersebut semakin rendah (Huebner & Skar, 2021b). Presisi tinggi sangat penting dalam pengukuran keterampilan berpikir kritis matematis, karena kemampuan ini melibatkan proses kognitif yang kompleks dan beragam.

### 4.8 Pembahasan Hasil Penelitian

Penelitian ini menggunakan data empiris dengan jumlah sampel sebanyak 106 siswa. Proses pengumpulan, penskoran, dan pengelolaan data memerlukan waktu yang cukup panjang. Pengujian hipotesis dalam penelitian ini berkaitan dengan seberapa presisi suatu model dalam mengestimasi parameter kemampuan peserta tes. Indikator presisi yang digunakan adalah besarnya varians dari hasil estimasi parameter kemampuan yang dihasilkan oleh masing-masing model. Semakin rendah nilai varians tersebut, maka semakin tinggi tingkat presisi estimasi model yang bersangkutan (Dali, 2012).

#### **4.8.1 Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis dengan PCM, GPCM dan GRM**

Dalam PCM, PARSCALE menghitung estimasi kemampuan ( $\theta$ ) setiap siswa berdasarkan seberapa jauh mereka berhasil mencapai tingkat skor tertentu dalam setiap butir soal politomus. Model ini mengasumsikan bahwa setiap butir memiliki beberapa langkah (step) kesulitan yang harus dilalui siswa untuk mencapai skor tertinggi. Estimasi  $\theta$  didasarkan pada peluang seorang siswa untuk melampaui ambang batas tersebut, dengan asumsi bahwa semua butir memiliki parameter diskriminasi yang sama (Falani et al., 2020a). Oleh karena itu, dalam PCM, perbedaan dalam respons siswa terhadap tingkatan skor di setiap soal digunakan untuk memperkirakan posisi mereka pada kontinum kemampuan laten, tanpa mempertimbangkan perbedaan tajam dalam kekuatan soal membedakan siswa.

Dalam GPCM dan GRM, PARSCALE menghasilkan estimasi kemampuan ( $\theta$ ) siswa dengan mempertimbangkan respons mereka terhadap soal-soal politomus sekaligus memperhitungkan parameter diskriminasi ( $a$ ) dari setiap butir soal. Tidak seperti PCM, GPCM dan GRM memungkinkan setiap indikator memiliki kemampuan yang berbeda dalam membedakan siswa dengan kemampuan rendah dan tinggi. Dengan demikian, GPCM memberikan estimasi  $\theta$  yang lebih fleksibel dan akurat karena mengintegrasikan informasi tentang tingkat kesulitan setiap level skor serta seberapa tajam soal dapat memisahkan siswa berdasarkan kemampuan berpikir kritis matematis mereka (Falani & Irdiyansyah, 2018). Model ini sangat cocok digunakan dalam pengukuran kompetensi kompleks yang tidak hanya dilihat dari benar atau salahnya jawaban.

Berdasarkan hasil analisis dengan model PCM, diketahui bahwa tingkat kesukaran butir soal tersebar pada beberapa kategori. Sesuai dengan teori PCM, model ini hanya memfokuskan pada tingkat kesukaran setiap kategori dalam satu butir soal, tanpa memperhatikan daya beda antar siswa (Sumintono & Widhiarso, 2015a). Hal ini menjadikan PCM sebagai model satu parameter (1-PL) yang kurang sensitif dalam membedakan siswa berkemampuan tinggi dan rendah. Sebaliknya, GPCM dan GRM merupakan model dua parameter (2-PL) yang memperhitungkan parameter diskriminasi butir, sehingga lebih mampu menangkap variasi kemampuan siswa secara tajam. Menurut Widhiarso (Dai et al., 2021), GPCM dan GRM mengestimasi parameter lereng ( $\alpha$ ) yang menunjukkan seberapa kuat suatu item membedakan individu dengan kemampuan berbeda.

Dari hasil analisis model PCM, GPCM, dan GRM berdasarkan Tabel 4,9 hingga Tabel 4.10, sebagian besar indikator keterampilan berpikir kritis matematis termasuk dalam kategori mudah, yaitu indikator *focus*, *reason*, dan *inference*. Ini menunjukkan bahwa soal-soal tersebut lebih mudah dijawab oleh siswa karena hanya menuntut kemampuan berpikir tingkat dasar, seperti mengingat dan memahami informasi, yang sesuai dengan level *remembering* dan *understanding* dalam Taksonomi Bloom (Anderson et al., 2001b). Selanjutnya, terdapat beberapa soal pada kategori sedang, seperti indikator *situation* dan *clarity*, yang cocok untuk siswa dengan kemampuan rata-rata karena menuntut keterampilan menganalisis dan menerapkan konsep, sesuai dengan level *applying* dan *analyzing* dalam Taksonomi Bloom. Sementara itu, soal yang tergolong sulit terdapat pada indikator *clarity* dan *overview*, yang memerlukan kemampuan berpikir kritis lebih tinggi, seperti mengevaluasi dan membuat keputusan, yang berkaitan dengan level

*evaluating* dan *creating* (Anderson et al., 2001b). Hal ini sejalan dengan pendapat (Ennis, 2011) yang menyatakan bahwa keterampilan berpikir kritis melibatkan kemampuan untuk menganalisis, mengevaluasi, dan membuat keputusan berdasarkan alasan yang logis. Dengan demikian, instrumen ini sudah mampu mengukur berbagai tingkat kemampuan berpikir kritis matematis siswa secara bertahap.

Berdasarkan hasil analisis parameter *slope* ( $a$ ), baik model GPCM maupun GRM menunjukkan bahwa sebagian besar item berada dalam kategori sedang hingga tinggi, yang berarti keduanya memiliki kemampuan diskriminatif yang cukup baik. Pada model GPCM, 7 dari 12 subitem berada dalam kategori sedang, dan 3 subitem (*clarity*, *overview*, *inference*) menunjukkan daya beda tinggi hingga sangat tinggi. Sementara itu, GRM menunjukkan pola serupa, dengan sebagian besar item berada dalam kategori sedang, dan beberapa indikator seperti *clarity*, *overview*, dan *inference* menunjukkan daya beda tinggi hingga sangat tinggi. Temuan ini sejalan dengan teori yang dikemukakan oleh Mutmainna et al., (2024), bahwa GPCM dan GRM merupakan model dua parameter (2PL) yang memperhitungkan parameter diskriminasi butir, sehingga lebih sensitif dalam membedakan peserta tes berdasarkan tingkat kemampuan. Dalam konteks keterampilan berpikir kritis matematis, kemampuan diskriminatif yang tinggi sangat penting karena indikator seperti *inference* dan *overview* menuntut pemrosesan kognitif yang kompleks dan reflektif. Oleh karena itu, baik GPCM maupun GRM dapat dianggap unggul dibanding PCM yang tidak mempertimbangkan daya beda, namun GPCM menunjukkan keunggulan tambahan

karena memberikan estimasi yang lebih stabil pada subitem dengan variasi kategori skor yang lebih kompleks.

Hasil distribusi pada Tabel 4.5, Tabel 4.6, dan Tabel 5.8 menunjukkan bahwa kemampuan berpikir kritis matematis siswa tersebar dalam berbagai kategori, dengan mayoritas berada pada tingkat cukup hingga sedang. Temuan ini mengindikasikan bahwa sebagian besar siswa mulai menunjukkan elemen berpikir kritis seperti analisis dan interpretasi, namun belum sepenuhnya stabil dalam melakukan Evaluasi dan inferensi matematis yang lebih kompleks. Hal tersebut sejalan dengan teori berpikir kritis yang dikemukakan oleh Ennis (2011), yang menekankan bahwa kemampuan berpikir kritis terdiri dari berbagai keterampilan kognitif yang berkembang bertahap. Menurut Ennis (1996), berpikir kritis adalah proses reflektif dan masuk akal yang digunakan untuk memutuskan apa yang harus diyakini atau dilakukan, yang dalam konteks matematis menuntut kemampuan bernalar secara sistematis dan logis. Dalam penelitian ini, indikator seperti *overview* memerlukan proses berpikir reflektif yang cukup kompleks. Namun, temuan menunjukkan bahwa sebagian besar siswa belum mampu menguasai keterampilan ini secara optimal. Kesenjangan ini mengindikasikan pentingnya strategi pengukuran yang lebih tepat sasaran untuk mengangkat dimensi *overview* dalam berpikir kritis siswa, khususnya pada tahap awal evaluasi.

Salah satu pendekatan yang dapat diterapkan adalah penyusunan instruksi soal yang lebih eksplisit dan menuntut, misalnya melalui perintah seperti “Buktikan bahwa jawabanmu benar.” Instruksi semacam ini tidak hanya memotivasi siswa untuk merefleksikan proses berpikirnya, tetapi juga mendorong mereka mengemukakan alasan dan bukti atas jawaban yang diberikan dua indikator penting

dalam berpikir kritis. Namun, dalam penelitian ini terdapat keterbatasan di mana instruksi soal hanya sebatas ajakan reflektif umum, seperti “Apakah kamu sudah memeriksa ulang jawaban dan yakin sudah benar?”, yang kurang menstimulasi elaborasi argumen atau justifikasi rasional. Meskipun pada awalnya penggunaan kalimat tersebut dimaksudkan untuk mengamati sejauh mana siswa secara spontan menunjukkan kemampuan berpikir kritis tanpa pancingan eksplisit, terdapat pandangan alternatif sebagaimana telah dijelaskan sebelumnya, yang menyarankan perlunya penyajian instruksi yang lebih tegas agar indikator berpikir kritis dapat muncul secara lebih sistematis dan terukur.

Penelitian terdahulu seperti yang dilakukan oleh Mardiyah et al. (2024) menunjukkan bahwa sebagian besar siswa SMP hanya menguasai satu hingga dua indikator berpikir kritis matematis, yang memperkuat temuan dalam penelitian ini. Selain itu, studi oleh Rahayu & Dewi (2022) menyoroti rendahnya peringkat Indonesia dalam asesmen berpikir tingkat tinggi, mempertegas perlunya penguatan pembelajaran berbasis pemecahan masalah. Oleh karena itu, hasil distribusi ini memberikan implikasi bahwa pendekatan pembelajaran yang lebih menekankan eksplorasi, refleksi, dan komunikasi matematis perlu diterapkan untuk mendorong siswa berpindah dari kategori sedang ke tinggi. Dengan demikian, hasil ini tidak hanya menggambarkan kondisi kemampuan siswa, tetapi juga menjadi dasar teoritis dan empiris untuk pengembangan strategi pembelajaran yang lebih efektif dalam meningkatkan keterampilan berpikir kritis matematis.

Temuan dalam penelitian ini memperkuat hasil yang dikemukakan oleh Amri (2022), bahwa indikator *Overview* merupakan aspek berpikir kritis yang paling jarang dicapai oleh siswa. Hal ini tercermin dari data yang menunjukkan

bahwa hanya sedikit peserta tes yang mampu melakukan refleksi menyeluruh terhadap proses berpikir dan strategi penyelesaian yang telah mereka gunakan. Rendahnya pencapaian pada indikator *Overview* mengindikasikan bahwa sebagian besar siswa belum terbiasa mengevaluasi kembali langkah-langkah berpikir mereka secara sistematis. Mereka cenderung berfokus pada hasil akhir tanpa mengembangkan kesadaran metakognitif terhadap proses yang dilalui. Dalam konteks keterampilan berpikir kritis matematis, kemampuan *Overview* sangat penting karena mencerminkan kedewasaan berpikir dan kemampuan untuk menilai efektivitas strategi yang digunakan. Oleh karena itu, hasil ini menegaskan perlunya pendekatan pembelajaran yang mendorong siswa untuk melakukan refleksi dan evaluasi diri secara lebih intensif.

#### **4.8.2 Perbandingan Akurasi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis antara Model PCM dan GPCM**

Pada rumusan masalah bagian pertama yaitu bagaimana variansi hasil estimasi parameter kemampuan berpikir kritis siswa dengan menggunakan model IRT PCM dan GPCM.

Model PCM digunakan sebagai pendekatan dasar untuk menganalisis data politomi dalam penelitian ini. PCM merupakan model dalam teori respons butir IRT yang tidak mengasumsikan adanya perbedaan daya beda antarbutir (parameter  $a$ ), sehingga model ini lebih sederhana dibandingkan model yang mempertimbangkan parameter tersebut. Analisis kecocokan model menunjukkan bahwa PCM memenuhi asumsi dasar penerapan model, dengan nilai  $p$ -value sebesar 0,264 dan  $p$ -value total sebesar 0,335. Hal ini menunjukkan bahwa model yang digunakan memiliki kecocokan yang tinggi dengan data empiris. Nilai  $p$ -value (PROB.) harus di atas batas signifikansi umum 0,05 yang mengindikasikan kecocokan model baik

secara keseluruhan (Niken et al., 2024). Hasil ini menandakan bahwa tidak terdapat penyimpangan signifikan antara data empiris dan model yang diestimasi. Dengan demikian, secara statistik, PCM dapat dianggap layak digunakan untuk mengukur kemampuan responden dalam konteks data yang dianalisis.

Namun demikian, ketika ditinjau dari aspek ketepatan estimasi, PCM menghasilkan nilai varians estimasi kemampuan sebesar 1,232 serta SEM sebesar 0,406. Nilai SEM ini mencerminkan tingkat ketelitian pengukuran pada estimasi kemampuan, dan dalam konteks ini tergolong baik. Hal ini menandakan bahwa meskipun model ini sesuai secara statistik, presisi estimasinya tidak setinggi model lain yang mengintegrasikan parameter daya beda. Ketidakterlibatan parameter  $\alpha$  dalam PCM menyebabkan model ini memberikan bobot yang sama terhadap informasi dari setiap butir, terlepas dari seberapa baik butir tersebut membedakan antara individu dengan kemampuan tinggi dan rendah.

Hasil analisis fungsi informasi tes dari PCM memperlihatkan bahwa model ini memberikan informasi paling tinggi di sekitar skor -1 dengan nilai puncak informasi sebesar 7. Ini berarti bahwa PCM memberikan pengukuran yang paling akurat untuk responden dengan kemampuan sedikit di bawah rata-rata. Distribusi informasi tes yang tidak merata ini dapat menjadi pertimbangan tambahan dalam pemilihan model, terutama jika tujuan pengukuran adalah untuk mendapatkan estimasi kemampuan yang presisi di seluruh rentang skala. Temuan ini sejalan dengan penelitian oleh (Aji & Retnawati, 2024), yang menyatakan bahwa GPCM memberikan distribusi informasi yang lebih merata dibandingkan PCM, sehingga lebih unggul dalam menghasilkan estimasi kemampuan yang presisi di berbagai rentang kemampuan. Model PCM hanya mempertimbangkan tingkat kesukaran

kategori dalam satu butir soal, sehingga informasi tes yang dihasilkan cenderung terpusat pada rentang kemampuan tertentu. Jika distribusi informasi tidak merata, maka model ini kurang optimal untuk mengestimasi kemampuan secara presisi di seluruh rentang skala (Sumintono & Widhiarso, 2015b). Oleh karena itu, meskipun PCM layak secara statistik, pertimbangan presisi pengukuran dapat mendorong eksplorasi model alternatif yang memasukkan parameter daya beda, seperti GPCM atau GRM.

Untuk menguji keunggulan presisi estimasi antar model, dilakukan pengujian hipotesis terhadap perbedaan varians estimasi kemampuan antara PCM dan GPCM. Hasilnya menunjukkan bahwa PCM menghasilkan varians sebesar 1,232, sedangkan GPCM sebesar 0,882. Uji F memberikan nilai F hitung sebesar 1,396 yang melampaui F tabel pada taraf signifikansi 5% (1,381), sehingga  $H_0$  ditolak. Ini menandakan bahwa GPCM secara signifikan memiliki varians estimasi yang lebih kecil, mencerminkan konsistensi yang lebih baik dalam mengukur kemampuan peserta. Hal ini diperkuat oleh nilai SEM GPCM yang lebih rendah (0,354) dibandingkan PCM (0,417), mengindikasikan bahwa GPCM memberikan estimasi yang lebih presisi dan andal dalam konteks data penelitian ini. Hal ini diperkuat oleh (Luo, 2018; Muraki, 1992) GPCM dapat memberikan gambaran lebih akurat mengenai kemampuan seseorang berdasarkan respons yang diberikan, sehingga hasil analisis lebih representatif dan informatif.

#### **4.8.3 Perbandingan Akurasi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis antara model GPCM dan GRM**

Pada rumusan masalah bagian kedua yaitu bagaimana variansi hasil estimasi parameter kemampuan berpikir kritis siswa dengan menggunakan model IRT GPCM.

Model GPCM, sebagai pengembangan dari PCM dengan memasukkan parameter daya beda ( $a$ ), menunjukkan performa paling optimal dalam penelitian ini. Nilai p-value sebesar 0,987 menunjukkan bahwa model ini sangat cocok dengan data empiris. Tingginya nilai p tersebut menandakan tidak terdapat penyimpangan signifikan antara data observasi dan prediksi model, sehingga GPCM layak secara statistik untuk digunakan dalam pengukuran kemampuan peserta.

Dari sisi ketepatan estimasi, GPCM menghasilkan varians estimasi keterampilan sebesar 0,882 dan nilai SEM sebesar 0,354 keduanya merupakan yang terendah dibandingkan dengan PCM maupun GRM. Nilai varians yang kecil mencerminkan konsistensi yang tinggi dalam estimasi kemampuan, sementara SEM yang rendah menunjukkan bahwa kesalahan pengukuran berada pada tingkat minimum (Huebner & Skar, 2021b). Kombinasi keduanya mengindikasikan bahwa GPCM memiliki keunggulan dalam memberikan estimasi kemampuan yang presisi dan stabil.

Selain itu, fungsi informasi tes dari GPCM Dalam IRT menggambarkan seberapa banyak informasi yang diberikan oleh suatu tes pada setiap tingkat kemampuan peserta. Semakin tinggi nilai fungsi informasi tes pada suatu titik kemampuan, semakin presisi estimasi kemampuan yang dihasilkan pada titik tersebut (Falani et al., 2020c). Berdasarkan hasil analisis, fungsi informasi tes dari model GPCM menunjukkan puncak informasi sebesar 20 pada rentang skor -1 hingga 0,5, yang berarti model ini sangat efektif dalam mengukur kemampuan peserta pada tingkat menengah ke bawah. Cakupan informasi yang luas dan puncak yang tinggi menunjukkan bahwa GPCM mampu menangkap variasi kemampuan

peserta dengan sangat baik, terutama pada kelompok yang berada di bawah rata-rata.

Untuk menguji konsistensi presisi estimasi antara GPCM dan GRM, dilakukan pengujian hipotesis kedua menggunakan uji F terhadap varians hasil estimasi. Berdasarkan Tabel 4.13, diketahui bahwa varians GPCM sebesar 0,882 lebih kecil dibandingkan GRM sebesar 1,219. Hasil uji F menunjukkan F hitung sebesar 1,382 yang sedikit melebihi F tabel sebesar 1,381 pada taraf signifikansi 5%. Karena  $F_{hitung} > F_{tabel}$ , maka hipotesis nol ditolak, yang berarti varians GPCM secara signifikan lebih kecil dibandingkan GRM. Ini memperkuat temuan bahwa GPCM lebih presisi dalam mengestimasi kemampuan responden. Temuan ini sejalan dengan penelitian Dai et al. (2021) yang menemukan bahwa GPCM memiliki performa yang lebih stabil dalam kondisi data hilang, sedangkan GRM cenderung menghasilkan estimasi yang kurang akurat saat banyak respon kosong. Selain itu, Wang et al., (2002) juga menunjukkan bahwa GPCM menghasilkan galat estimasi lebih kecil dibandingkan GRM, terutama pada instrumen polytomous. Penelitian lain oleh Ajia dan Retnawati (2024) juga membuktikan bahwa GPCM lebih unggul daripada GRM berdasarkan kriteria AIC, BIC, dan uji kesesuaian model, sehingga GPCM dinilai lebih akurat dan informatif dalam mengestimasi kemampuan responden.

Secara keseluruhan, hasil analisis menunjukkan bahwa GPCM unggul dari sisi kecocokan model, kestabilan estimasi, dan akurasi pengukuran. Dengan mempertimbangkan seluruh aspek tersebut, GPCM dapat direkomendasikan sebagai model paling informatif dan efektif dalam asesmen berbasis data politomi, khususnya dalam mengukur keterampilan berpikir kritis matematis.

#### **4.8.4 Perbandingan Akurasi Hasil Estimasi Parameter Keterampilan Berpikir Kritis Matematis antara Model PCM dan GRM**

Pada rumusan masalah bagian ketiga yaitu bagaimana variansi hasil estimasi parameter kemampuan berpikir kritis siswa dengan menggunakan model IRT GRM.

Model GRM menunjukkan performa yang cukup baik dalam hal kecocokan model, dengan nilai p-value sebesar 0,819 yang berada jauh di atas ambang signifikansi 0,05. Hal ini menunjukkan bahwa GRM cocok digunakan untuk data dalam penelitian ini, tanpa indikasi penyimpangan signifikan antara data empiris dan model teoritis. Sama seperti GPCM, GRM juga mempertimbangkan parameter daya beda ( $a$ ), sehingga mampu membedakan tingkat kemampuan peserta dengan sensitivitas yang lebih tinggi dibandingkan PCM yang tidak mempertimbangkan parameter ini.

Dalam konteks presisi estimasi, GRM menghasilkan variansi estimasi keterampilan sebesar 1,219, sedikit lebih rendah dibandingkan PCM (1,232), serta nilai SEM sebesar 0,406, yang juga sedikit lebih kecil dari PCM (0,417). Meskipun selisih antara keduanya relatif kecil, hasil ini menunjukkan bahwa GRM memberikan estimasi yang sedikit lebih konsisten dan akurat dibandingkan PCM. Fungsi informasi tes dari GRM memperlihatkan akurasi optimal di sekitar skor -1 dengan puncak informasi sekitar 7, identik dengan hasil yang ditunjukkan oleh PCM. Ini menunjukkan bahwa kedua model memiliki kontribusi informasi yang sebanding pada rentang kemampuan menengah ke bawah.

Untuk menguji perbedaan presisi estimasi secara statistik, dilakukan pengujian hipotesis ketiga dengan uji F terhadap varians hasil estimasi dari kedua model. Berdasarkan hasil pada Tabel 4.15, diperoleh nilai F hitung sebesar 1,0106, yang lebih kecil dari nilai F tabel sebesar 1,381 pada taraf signifikansi 5%. Karena  $F_{hitung} < F_{tabel}$ , maka hipotesis nol diterima, yang berarti tidak terdapat perbedaan yang signifikan antara varians hasil estimasi kemampuan peserta tes menggunakan PCM dan GRM. Dengan demikian, meskipun GRM memiliki beberapa keunggulan numerik, secara statistik tingkat presisi estimasinya tidak berbeda signifikan dari PCM. Oleh karena itu, pemilihan antara kedua model ini dapat dipertimbangkan berdasarkan kebutuhan praktis dan struktur data, bukan semata-mata berdasarkan presisi estimasi. Sejalan dengan hal tersebut Santoso et al., (2022) membandingkan estimasi kemampuan menggunakan beberapa model IRT politomus, termasuk PCM dan GRM. Hasil analisis mereka menunjukkan bahwa tidak terdapat perbedaan yang signifikan pada estimasi parameter kemampuan antara kedua model tersebut.

## **BAB V PENUTUP**

### **5.1 Kesimpulan**

Berdasarkan hasil analisis perbandingan model-model estimasi parameter keterampilan berpikir kritis matematis, diperoleh beberapa temuan penting.

1. Terdapat perbedaan varians hasil estimasi parameter antara model PCM dan GPCM. Varians estimasi keterampilan berpikir kritis matematis yang dihasilkan oleh model GPCM cenderung lebih kecil dibandingkan PCM, yang mengindikasikan efisiensi estimasi yang lebih tinggi dari model GPCM dalam keterampilan berpikir kritis matematis.
2. Perbandingan antara GPCM dan GRM juga menunjukkan bahwa model GPCM menghasilkan varians estimasi keterampilan berpikir kritis matematis yang lebih kecil. Hal ini semakin memperkuat posisi model GPCM sebagai pendekatan yang relatif lebih stabil dalam mengestimasi parameter keterampilan berpikir kritis matematis.
3. Analisis perbandingan antara model PCM dan GRM memperlihatkan bahwa GRM memiliki varians hasil estimasi parameter keterampilan berpikir kritis matematis yang lebih kecil dibandingkan PCM namun tidak signifikan. Dengan demikian, meskipun GRM banyak digunakan pada data politomus, dalam konteks penelitian ini, GPCM menunjukkan performa yang lebih baik dari sisi efisiensi estimasi keterampilan berpikir kritis matematis.

Secara keseluruhan, model GPCM dapat dipertimbangkan sebagai model yang paling optimal dalam menghasilkan estimasi parameter keterampilan berpikir kritis matematis dengan varians yang paling rendah di antara ketiga model yang dibandingkan.

## 5.2 Implikasi

1. Implikasi bagi Siswa: pemanfaatan model GPCM dalam penilaian memungkinkan siswa memperoleh estimasi kemampuan yang lebih akurat dan stabil. Hal ini penting karena hasil asesmen yang presisi dapat dijadikan dasar refleksi diri serta menjadi acuan bagi peningkatan kompetensi secara personal.
2. Implikasi bagi Guru: dengan model estimasi yang lebih efisien, guru dapat menginterpretasikan hasil tes dengan lebih adil dan akurat. Akurasi estimasi akan memudahkan guru dalam mengidentifikasi kebutuhan belajar siswa secara individual maupun kelompok, serta merancang strategi pembelajaran yang lebih responsif dan efektif.
3. Implikasi bagi sekolah: sekolah sebagai penyelenggara asesmen diharapkan mulai mempertimbangkan pemilihan model pengukuran yang sesuai dengan karakteristik data dan tujuan evaluasi. Implementasi model seperti GPCM berpotensi meningkatkan kualitas asesmen berbasis kompetensi serta mendukung pengambilan keputusan berbasis data (data-driven decision making) dalam pengembangan kurikulum dan kebijakan pendidikan.

## 5.3 Saran

Berikut merupakan saran untuk penelitian selanjutnya:

1. Penelitian lanjutan disarankan untuk mengeksplorasi lebih dalam aspek presisi estimasi kemampuan pada keterampilan HOTS melalui pendekatan kombinasi penyekoran, seperti penggabungan model skoring dikotomus dan

politomus, serta penerapan model IRT gabungan dengan jumlah responden yang lebih besar.

2. Penelitian selanjutnya disarankan untuk menganalisis tes format *essay* dengan penskoran politomi menggunakan perangkat lunak IRT lain yang memiliki fitur analisis lebih komprehensif untuk model IRT, seperti *software R* dan sebagainya.
3. Disarankan untuk mengembangkan sistem penilaian berbasis teknologi seperti aplikasi adaptif atau platform digital yang menggunakan model GPCM dalam memberikan umpan balik instan dan personalisasi pembelajaran.

## DAFTAR PUSTAKA

- Aini, N., Sari, A., Antia, V., Daimah, U. S., Muhakimah, I., & Dewanti, S. S. (2024). Konstruksi instrumen tes kemampuan pemecahan masalah menggunakan teori respon butir. *Teorema: Teori Dan Riset Matematika*, 09((02)), 193–206. <https://doi.org/DOI:http://dx.doi.org/10.25157/teorema.v9i2.14867>
- Aji, N. W., & Retnawati, H. (2024). Perbandingan kecocokan model analisis teori respons butir pada data politomi. *Jurnal Statistika Dan Sains Data*, 1(2), 31–47. <https://doi.org/10.21831/jssd.v1i2.18484>
- Akdon, & Riduwan. (2009). *Rumus dan Data dalam Analisis Statistik untuk Penelitian*. Bandung: Alfabeta.
- Alagoz, C. (2000). *Scoring tests with dichotomous and polytomous items*. Gazi University
- Amanda, N., & Nusantara, T. (2021). Analisis berpikir kritis siswa terhadap pemecahan masalah matematika di MTs Surya Buana Malang. *Jurnal Pendidikan Matematika Dan Sains*, 8(2), 89–92. <https://doi.org/10.21831/jpms.v8i2.19660>
- Ambarwati, R., & Alim, F. (2023). An analysis of students' critical thinking ability in solving jumping task questions based on learning style. *Journal of Innovative Mathematics Learning*, 6(4), 319–327. <https://doi.org/10.22460/jiml.v6i4.p21727>
- Amelia, R. N., & Kriswantoro, K. (2017). Implementation of item response theory for analysis of test items quality and students' ability in chemistry. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 2(1), 1. <https://doi.org/10.20961/jkpk.v2i1.8512>
- An Nabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis indeks aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *PAEDAGOGIA*, 25(2), 184. <https://doi.org/10.20961/paedagogia.v25i2.64566>
- Angraini, L., & Wahyuni, A. (2021). The effect of concept attainment model on mathematical critical thinking ability. *International Journal of Instruction*, 14, 727–742. <https://doi.org/10.29333/iji.2021.14144a>
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arini, R., Riansi, E. S., Fadillah, I., & Hasan, M. K. (2024). Peran tes objektif dan subjektif dalam mengukur keterampilan berfikir matematis siswa sekolah dasar. *Jurnal Inovasi Dan Riset Pendidikan Matematika*, 5(3), 236–244.

- Bahar, R., & Retnawati, H. (2022). Analisis karakteristik soal kemampuan koneksi matematika penskoran politimus. *Jurnal Tarbiyah*, 29, 195. <https://doi.org/10.30829/tar.v29i2.1650>
- Baker, F., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R*. Springer. <https://doi.org/10.1007/978-3-319-54205-8>
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD. [www.ascd.org/memberbooks](http://www.ascd.org/memberbooks)
- Ciptari, P. A., Purwanti, N. K. R., & Erawati, N. K. (2024). Comparison of the classical test theory and the rasch model in the analysis of mastery test of concept systems of linear (SPLDV). *Prima: Jurnal Pendidikan Matematika*, 8(2), 390. <https://doi.org/10.31000/prima.v8i2.8647>
- Dai, S., Vo, T., Kehinde, O., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of polytomous irt models with rating scale data: an investigation over sample size, instrument length, and missing data. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.721963>
- Demars. (2010). *Item response theory : understanding statistics measurement*. Oxford University Press. <https://doi.org/https://doi.org/10.15713/ins.mmj.3>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410605269>
- Ennis, R. H. (1996). *Critical Thinking Dispositions: Their Nature and Assessability*. Resnick.
- Ennis, R. H. (2011). *The nature of critical thinking: an outline of critical thinking dispositions and abilities*. [https://education.illinois.edu/docs/default-source/faculty-documents/robert-ennis/thenatureofcriticalthinking\\_51711\\_000.pdf](https://education.illinois.edu/docs/default-source/faculty-documents/robert-ennis/thenatureofcriticalthinking_51711_000.pdf)
- Ennis, R. H. (2018). Critical thinking across the curriculum: a vision. *Topoi*, 37(1), 165–184. <https://doi.org/10.1007/s11245-016-9401-4>
- Faiz, A., Putra, N. P., & Fajar, N. (2022). Memahami makna tes, pengukuran (measurement), penilaian (assessment), dan evaluasi (evaluation). *Jurnal Education and Development Institut Pendidikan Tapanuli Selatan*, 10(3). <https://doi.org/10.31004/obsesi.v5i2.972>
- 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>

- Falani, I., Akbar, M., & Naga, D. S. (2020a). Comparison of the accuracy of item response theory models in estimating student's ability. *Journal of Educational Science and Technology*, 6(2), 178–184. <https://doi.org/https://doi.org/10.26858/est.v6i2.13295>
- Falani, I., Akbar, M., & Naga, D. S. (2020b). The precision of students' ability estimation on combinations of item response theory models. *International Journal of Instruction*, 13(4), 545–558. <https://doi.org/https://doi.org/10.29333/iji.2020.13434a>
- Falani, I., Iriyadi, D., Ice, Y. W., Susanti, H., & Nasution, R. A. (2022). A Rasch analysis of perceived stigma of covid-19 among nurses in Indonesia questionnaire. *Psychological Thought*, 15(1), 12.
- Falani, I., & Irdiyansyah, I. (2018). The ability of estimation stability and item parameter characteristics reviewed by item response theory model. *Advances in Social Science, Education and Humanities Research*, 15. <https://doi.org/https://doi.org/10.2991/icems-17.2018.34>
- Falani, I., Nisraeni, N., & Irdiyansyah, I. (2017). The ability of estimation stability and item parameter characteristics reviewed by Item Response Theory model. *International Conference on Education in Muslim Society (ICEMS 2017)*, 175–178.
- Farman, F., Anjelina, S., Putri, Q. T., Mardiah, N. A., & Sari, K. (2021). Pengembangan instrumen penilaian pembelajaran matematika berbasis ispring suite. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(4), 2040. <https://doi.org/10.24127/ajpm.v10i4.4288>
- Fatkhudin, A., Surarso, B., & Subagio, A. (2016). Item response theory model empat parameter logistik pada computerized adaptive test. *Jurnal Sistem Informasi Bisnis*, 4. <https://doi.org/10.21456/vol4iss2pp121-129>
- Fernanda, J. W., & Hidayah, N. (2020). Analisis kualitas soal ujian statistika menggunakan classical test theory dan rasch model. *Square: Journal of Mathematics and Mathematics Education*, 2(1), 49. <https://doi.org/10.21580/square.2020.2.1.5363>
- Fitri, D., Mujahidawati, M., & Sabil, H. (2025). Deskripsi media video pembelajaran menggunakan problem based learning untuk meningkatkan kemampuan berpikir kritis matematis. *Kognitif: Jurnal Riset HOTS Pendidikan Matematika*, 5(2).
- Fitri, M., Mujahidawati, M., & Falani, I. (2025). Pengembangan e-modul etnomatematika 3d terintegrasi virtual reality berbasis tpack untuk meningkatkan kemampuan berpikir kritis siswa. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 14(2), 368. <https://doi.org/10.24127/ajpm.v14i2.10191>

- Frey, F. (2020). Test theory and classical test theory. *The International Encyclopedia of Media Psychology*, 1–6. <https://doi.org/10.1002/9781119011071.iemp0047>
- Friyatmi, F. (2018). Estimasi parameter tes dengan penskoran politomus menggunakan graded response model pada sampel kecil. *Jurnal Inovasi Pendidikan Ekonomi*, 8, 22. <https://doi.org/10.24036/01104490>
- Gustiawan, W., Niaga, A., & Padang, P. N. (2019). The theories of leadership: a review of papers. *Jurnal Ilmiah Poli Bisnis*, 11(1), 65–72.
- Habibah, S., & Marlina, R. (2023). Analisis kemampuan berpikir kritis matematis siswa kelas VIII SMP. *PHI: Jurnal Pendidikan Matematika*, 7(2), 141. <https://doi.org/10.33087/phi.v7i2.287>
- Halpern, D., & Dunn, D. (2022). *Thought and Knowledge: An Introduction to Critical Thinking*. New York, NY: Routledge. <https://doi.org/10.4324/9781003025412>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory* (Issue v. 2). Sage Publications.
- Handayani, R., Destania, Y., & Muhammadiyah Bengkulu, U. (2021). Soal essay materi aritmatika sosial untuk kemampuan berpikir kritis matematis siswa kelas VII. *Indiktika: Jurnal Inovasi Pendidikan Matematika*, 4(1). <https://doi.org/https://doi.org/10.31851/indiktika.v4i1.6791>
- Hartanti, A., & Purnomo, Y. W. (2023). Peningkatan keterampilan berpikir kritis matematis dan hasil belajar menggunakan model pembelajaran problem based learning. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(2), 1728. <https://doi.org/10.24127/ajpm.v12i2.6891>
- Hayat, B. (2021). Klasika: program analisis item dan tes dengan pendekatan klasik. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 10(1), 1–11. <https://doi.org/10.15408/jp3i.v10i1.20551>
- Huebner, A., & Skar, G. B. (2021). Conditional standard error of measurement: classical test theory, generalizability theory and many-facet rasch measurement with applications to writing assessment. *Practical Assessment, Research and Evaluation*, 26, 1–20. <https://doi.org/10.7275/vzmm-0z68>
- Indrawatiningsih, N., Purwanto, As'Ari, A. R., Dwiyanana, Sudirman, & Rahardi, R. (2019). The ability of high school students' critical thinking in solving trigonometric problems. *IOP Conference Series: Earth and Environmental Science*, 243(1). <https://doi.org/10.1088/1755-1315/243/1/012050>

- Inganah, S., Darmayanti, R., & Rizki, N. (2023). Problems, solutions, and expectations: 6c integration of 21 st century education into learning mathematics. *JEMS (Journal of Mathematics and Science Education)*, 11(1), 220–238.
- Keliat, B. A., Nasution, R. A., Handini, I. T., & Falani, I. (2023). Psychometric properties evaluation of the Indonesian version of the Beck Scale for Suicide Ideation (BSSI) questionnaire using a Rasch model. *European Review for Medical & Pharmacological Sciences*, 27(21).
- Kogar, E. Y., & Kogar, H. (2018). Examination of dimensionality and latent trait scores on mixed-format tests. *PUPIL: International Journal of Teaching, Education and Learning*, 2(1), 29–49. <https://doi.org/10.20319/pijtel.2018.21.2949>
- Komala, P., Kaniawati, I., & Efendi, R. (2020). Karakterisasi soal tes keterampilan berpikir kritis menggunakan analisis item response theory pada materi fluida statis. *WaPFI (Wahana Pendidikan Fisika)*, 5(2), 100–109. <https://doi.org/10.17509/wapfi.v5i2.7755>
- Lee, W.-C., Kim, S. Y., Choi, J., & Kang, Y. (2020). IRT approaches to modeling scores on mixed-format tests. *Journal of Educational Measurement*, 57(2), 230–254. <https://doi.org/https://doi.org/10.1111/jedm.12248>
- Leniati, B., & Indarini, E. (2021). Keefektifan model pembelajaran kooperatif tipe jigsaw dan tsts (two stay two stray) terhadap kemampuan berpikir kritis pada pembelajaran matematika siswa. *Mimbar Ilmu*. <https://ejournal.undiksha.ac.id/index.php/MI/article/view/33359>
- Luo, Y. (2018). Parameter recovery with marginal maximum likelihood and markov chain monte carlo estimation for the generalized partial credit model. *National Center for Assessment in Saudi Arabia*, 1–36. <https://doi.org/https://doi.org/10.48550/arXiv.1809.07359>
- LW, A., DR, K., PW, A., KA, C., Mayer, R., PR, P., Raths, J., & MC, W. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*.
- Marsigit, & Susilo, N. B. (2006). *Matematika 1 SMP Kelas VII*. Cimahi: Quadra.
- Maulana, A. (2022). Analisis validitas, reliabilitas, dan kelayakan instrumen penilaian rasa percaya diri siswa. *Jurnal Kualita Pendidikan*, 3(3), 2774–2156. <https://doi.org/https://doi.org/10.51651/jkp.v3i3.331>
- Mertasari, N. M. S., & Candiasa, I. M. (2023). Development and validation of metacognitive test in programming using graded response model. *IConVET 2022: Proceedings of the 5th International Conference on Vocational Education and Technology, IConVET 2022, 6 October 2022, Singaraja, Bali, Indonesia*, 70.

- Mujahidawati, Novferma, Simatupang, G. M., Frianto, A., & Romundza, F. (2023). Analysis of university students' higher order thinking skills (HOTS) in the differential equation courses in the time of Covid-19 pandemic. *AIP Conference Proceedings*, 2698(1), 060047.
- Muraki, E. (1992). A generalized partial credit model: application of an em algorithm index terms: item response model, national assessment of educational progress, nominal response model, partial credit model, polytomous response model, rating scale model. *Applied Psychological Measurement*, *Uj 1*, 159–176. <https://doi.org/https://hdl.handle.net/11299/115645>
- Mutmainna, M., Hasyim, M., Syamsuriwal, S., Setiaji, B., & Setiawati, F. (2024). Modeling of colorado learning attitude science survey in indonesian version: a study with applying item response theory. *JP3I (Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia)*, *13*, 80–104. <https://doi.org/10.15408/jp3i.v13i1.36745>
- Naga, D. S. (2012). *Teori Sekor pada Pengukuran Mental*. Jakarta: PT Nagarani Citrayasa.
- Nasihudin, N., & Hariyadin, H. (2021). Pengembangan keterampilan dalam pembelajaran. *Jurnal Pendidikan Indonesia*, *2(04)*, 733–743. <https://doi.org/10.59141/japendi.v2i04.150>
- Nasution, R. A., & Falani, I. (2024). Measuring cyberbullying levels among adolescents in integrated islamic junior high schools in jambi city: an implementation of the rasch model. *Journal of Educational, Health & Community Psychology (JEHCP)*, *13(1)*.
- Niken, Somayasa, W., Budiman, H., Ruslan, & Adi Wibawa, G. N. (2024). Uji goodness of fit dengan statistik pulkstenis - robinson dalam model regresi logistik ordinal. *Jurnal Matematika Komputasi Dan Statistika*, *4(1)*, 571–576. <https://doi.org/10.33772/jmks.v4i1.72>
- Nufus, H., & Kusaeri, A. (2020). Analisis tingkat kemampuan berpikir kritis siswa dalam memecahkan masalah geometri. *Jurnal Pendidikan Matematika Indonesia*, *5(2)*, 49–55.
- Nuryadi, N., Sukestiyarno, Y., Suyitno, H., & Kharisudin, I. (2024). Mathematical critical thinking profile of students on pisa framework space and shape content questions reviewed from self-efficacy. *Educational Administration: Theory and Practice*, *30(4)*, 1380–1400. <https://doi.org/10.53555/kuey.v30i4.1228>
- Osterlind, S. J. (2002). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats* (2nd ed). New York: Kluwer Academic.

- Partono, P., Wardhani, H. N., Setyowati, N. I., Tsalitsa, A., & Putri, S. N. (2021). Strategi meningkatkan kompetensi 4c (critical thinking, creativity, communication, & collaborative). *Jurnal Penelitian Ilmu Pendidikan*, 14(1), 41–52. <https://doi.org/10.21831/jpipfip.v14i1.35810>
- Ramalisa, Y., Falani, I., & Pasaribu, F. T. (2023). Rasch analysis in developing Jambi culture-based ethnomathematics test for prospective mathematics teachers. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, 243–257.
- Retnawati, H. (2015). Perbandingan estimasi kemampuan laten antara metode maksimum likelihood dan metode bayes [latent capability between maximum likelihood method and bayes method]. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(2), 145–155.
- Rosidah, N. A., Ramalis, T. R., & Suyana, I. (2018). Karakteristik tes keterampilan berpikir kritis ( KBK ). *Jurnal Inovasi Dan Pembelajaran Fisika*, January 2018, 54–63.
- Santoso, A., Pardede, T., Djidu, H., Apino, E., Rafi, I., Rosyada, M. N., & Abd Hamid, H. S. (2022). The effect of scoring correction and model fit on the estimation of ability parameter and person fit on polytomous item response theory. *REID (Research and Evaluation in Education)*, 8(2), 140–151. <https://doi.org/10.21831/reid.v8i2.54429>
- Sappaile, B. I. (2010). Konsep penelitian ex-post facto. *Jurnal Pendidikan Matematika*, 1(2), 105–113. <https://doi.org/10.36709/jpm.v1i2.1946>
- Saputri, H. A., Zuhijrah, Larasati, N. J., & Shaleh. (2023). Analisis instrumen assesmen: validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09(05), 2986–2995.
- Satria, M. R. (2024). Transformasi standar penilaian pendidikan dan revitalisasi penilaian pembelajaran di indonesia. *Jurnal Penelitian Kebijakan Pendidikan*, 17(1), 57–66. <https://doi.org/10.24832/jpkp.v17i1.930>
- Septiani, A. N., Pratiwi, D., & Rossy, R. (2023). Evaluasi pembelajaran dalam peningkatan mutu pendidikan di sekolah dasar. *Masaliq*, 3(5), 824–832. <https://doi.org/10.58578/masaliq.v3i5.1380>
- Setiana, D. S., & Purwoko, R. Y. (2020). Analisis kemampuan berpikir kritis ditinjau dari gaya belajar matematika siswa. *Jurnal Riset Pendidikan Matematika*, 7(2), 163–177. <https://doi.org/10.21831/jrpm.v7i2.34290>

- Siburian, J., Sinaga, E., & Murni, P. (2023). Kemampuan berpikir kritis melalui implementasi flipped classroom pada siswa sma. *INKUIRI: Jurnal Pendidikan IPA*, 12(1), 71. <https://doi.org/10.20961/inkuiri.v12i1.68213>
- Sugianto, R., Cholily, Y. M., Darmayanti, R., Rahmah, K., & Hasanah, N. (2022). Development of rainbow mathematics card in tgt learning for increasing mathematics communication ability. *Kreano, Jurnal Matematika Kreatif-Inovatif*, 13(2), 221–233. <https://doi.org/10.15294/kreano.v13i2.38068>
- Sugiyono. (2015). *Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, dan R&D)*. Bandung: Alfabeta.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Jakarta: Trim Komunikata Publishing House.
- Supriyati, Y., Falani, I., & Maulana, S. (2021). The information function of mixed-format test of physics learning outcomes measurement. *AIP Conference Proceedings*, 2320(1).
- Susanti, D., & Hernawati, C. (2022). Keterampilan berpikir kritis mahasiswa dalam menyelesaikan soal higher order thinking skills. *EDU-MAT: Jurnal Pendidikan Matematika*, 10(1), 115. <https://doi.org/10.20527/edumat.v10i1.12190>
- Tao, Jian, Shi, Ning-Zhong, & Chang, Hua-Hua. (2012). Item-weighted likelihood method for ability estimation in tests composed of both dichotomous and polytomous items. *Journal of Educational and Behavioral Statistics*, 37(2), 298–315. <https://doi.org/10.3102/1076998610393969>
- Teluma, M., & Rivaie, W. (2019). *Penilaian* (1st ed.). PGRI Prov Kalbar.
- Wang, S., Asi, C., & Wang, T. (2002). *Relative precision of ability estimation in polytomous CAT: A comparison under the generalized partial credit model and graded response model*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. <https://eric.ed.gov/?id=ED477926>
- Wisnu Deni Arif, Tri Rijanto, Rina Harimurti, & Yulia Fransisca. (2023). Penerapan Partial Credit Model (PCM) untuk meningkatkan kemampuan berpikir kritis siswa pada mata pelajaran perbaikan peralatan listrik pada kelas XI di SMK Negeri 1 Kediri. *Jurnal Elektronika Dan Teknik Informatika Terapan (JENTIK)*, 1(3), 249–262. <https://doi.org/10.59061/jentik.v1i3.405>

- Yilmaz, H. B. (2019). A comparison of irt model combinations for assessing fit in a mixed format elementary school science test. *International Electronic Journal of Elementary Education*, *11*(5 SE-Articles), 539–545.
- Yokia Marsita, & Ishaq Nuriadin. (2024). Analisis kemampuan berpikir kritis matematis siswa dalam menyelesaikan soal cerita spldv ditinjau dari disposisi matematis. *Jurnal Math-UMB.EDU*, *11*(2), 129–136. <https://doi.org/10.36085/mathumbedu.v11i2.6143>
- Zahri, T. N., Yusuf, A. M., & S, N. (2017). Hubungan gaya belajar dan keterampilan belajar dengan hasil belajar mahasiswa serta implikasinya dalam pelayanan Bimbingan dan Konseling di Fakultas Ilmu Pendidikan Universitas Negeri Padang. *Konselor*, *6*(1), 18. <https://doi.org/10.24036/02017615734-0-00>
- Zain, A. Z., Ramalis, T. R., & Muslim, M. (2022). Karakterisasi instrumen tes keterampilan berpikir kreatif berdasarkan analisis partial credit model. *Jurnal Ilmiah Pendidikan Fisika*, *6*(1), 176. <https://doi.org/10.20527/jipf.v6i1.4806>
- Zainal, N. F. (2020). Pengukuran, assessment dan evaluasi dalam pembelajaran matematika. *Laplace: Jurnal Pendidikan Matematika*, *3*(1), 8–26. <https://doi.org/10.31537/laplace.v3i1.310>
- Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: are they mutually exclusive?. *Journal of the Learning Sciences*, *12*(2), 145–181. [https://doi.org/10.1207/S15327809JLS1202\\_1](https://doi.org/10.1207/S15327809JLS1202_1)

**LAMPIRAN****Lampiran 1. Syntax PARSCALE PCM**

```
>COMMENTS  
  
PCM  
  
>FILE  DFNAME='PCM.DAT', SAVE;  
  
>SAVE  SCORE='PCM.SCO', PARM='PCM.PAR';  
  
>INPUT  NIDW=4, NTOTAL=12, NTEST=1, LENGTH=12;  
  
(4A1,1X,12A1)  
  
>TEST  TNAME='SCALE1', ITEM=(1(1)12), NBLOCK=12;  
  
>BLOCK1  REPEAT=12, NITEMS=1, NCAT=4, ORIGINAL=(0,1,2,3),  
MODIFIED=(1,2,3,4);  
  
>CALIB  LOGISTIC, PARTIAL, NQPT=15, CYCLES=(50,1,1,1,1), NEWTON=2,  
        CRIT=0.01, SCALE=1.7, ITEMFIT=10, CSlope, ridge=(0.5,5);  
  
>SCORE  MLE, SMEAN=0.0, SSD=1.0, NAME=SCALE1, PFQ=5;
```

**Lampiran 2. Syntax PARSCALE GPCM**

```
>COMMENTS  
  
GPCM  
  
>FILES  DFNAME='GPCM.DAT', SAVE;  
  
>SAVE  SCORE='GPCM.SCO', PARM='GPCM.PAR';  
  
>INPUT  NIDW=4, NTOTAL=12, NTEST=1, LENGTH=12;  
  
(4A1,1X,12A1)  
  
>TEST  TNAME='SCALE1', ITEM=(1(1)12), NBLOCK=12;  
  
>BLOCK1  REPEAT=12, NITEMS=1, NCAT=4, ORIGINAL=(0,1,2,3),  
MODIFIED=(1,2,3,4);  
  
>CALIB  PARTIAL, LOGISTIC, SCALE=1.7, NQPTS=15, CYCLE=(100,1,1,1,1),  
NEWTON=2,CRIT=0.01, ITEMFIT=10, ridge=(0.5,5);  
  
>SCORE  MLE, SMEAN=0.0, SSD=1.0, NAME='PCR_MLE', PFQ=5;
```

**Lampiran 3. Syntax PARSCALE GRM**

```
>COMMENTS  
  
GPCM  
  
>FILE  DFNAME='GRM.DAT', SAVE;  
  
>SAVE  PARM='GRM.PAR', SCORE='GRM.SCO';  
  
>INPUT  NIDW=4, NTOTAL=12, NTEST=1, LENGTH=(12), NFMT=1;  
  
(4A1,1X,12A1)  
  
>TEST  TNAME='SCALE1', ITEM=(1(1)12), NBLOCK=12;  
  
>BLOCK1  REPEAT=12, NITEMS=1, NCAT=4, ORIGINAL=(0,1,2,3),  
MODIFIED=(1,2,3,4);  
  
>CALIB  GRADED, LOGISTIC, SCALE=1.7, NQPTS=15,  
CYCLES=(100,1,1,1,1), NEWTON=2, CRIT=0.01, ITEMFIT=10, SPRIOR,  
GPRIOR;  
  
>SCORE  MLE, NQPTS=30, SMEAN=0.0, SSD=1.0, NAME=EAP, PFQ=5;
```

**Lampiran 4. Form Validasi Instrumen Tes Keterampilan Berpikir Kritis Matematis**

**FORM VALIDASI INSTRUMEN TES KETERAMPILAN BERPIKIR KRITIS MATEMATIS**



**PROGRAM MAGISTER PENDIDIKAN MATEMATIKA  
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN  
UNIVERSITAS JAMBI  
2025**

## A. Tentang Penelitian

Judul	: Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP Dengan Pendekatan <i>Item Response Theory</i> (IRT)
Format Tes	: Essay (3)
Level	: Kelas VII
Kurikulum	: Merdeka
Materi	: Rasio
Sampel Penelitian	: 200 Responden
Model Penyekoran	: 1. <i>Partial Credit Model</i> (PCM) 2. <i>Generalized Partial Credit Model</i> (GPCM) 3. <i>Graded Response Model</i> (GRM)
Fokus Penelitian	: 1. Kecocokan Model 2. Hasil Estimasi Karakteristik Butir dan Keterampilan Berpikir Kritis Matematis 3. <i>Standard Error Measurement</i> (SEM) Pengukuran Keterampilan Berpikir Kritis

## B. Petunjuk Penilaian Instrumen Penelitian

Berikut ini adalah petunjuk dalam melakukan penilaian:

1. Bapak/Ibu yang terhormat, mohon memberikan penilaian secara objektif dan memberikan saran-saran untuk merevisi instrumen tes yang telah disusun.
2. Memberikan penilaian terhadap setiap butir dengan meninjau beberapa aspek sebagai berikut:
  - a. Kesesuaian capaian pembelajaran dan indikator.
  - b. Kesesuaian indikator dan soal.
  - c. Soal dirumuskan secara jelas.
  - d. Soal menggunakan bahasa yang baku, mudah dipahami, dan tidak menimbulkan penafsiran ganda.
  - e. Kesesuaian soal dan penyekoran.
3. Memberikan penilaian dengan memberikan tanda cek ( $\checkmark$ ) terhadap setiap butir dengan menggunakan skor pada rentangan 1 sampai 5 dengan ketentuan:
  - a. Skor 1 jika hanya ada satu kriteria yang muncul
  - b. Skor 2 jika hanya ada dua kriteria yang muncul
  - c. Skor 3 jika hanya ada tiga kriteria yang muncul
  - d. Skor 4 jika ada empat kriteria yang muncul
  - e. Skor 5 jika semua kriteria muncul

**FORM BIODATA VALIDATOR**

**Nama Validator** :.....

**Jenis Kelamin** :    L-laki        Peempuan

**Pendidikan Terakhir** :    S1    S2    S3

**Bidang Keahlian** :.....

**Institusi** :.....

....., ...../...../2025

(.....)

### KISI-KISI INSTRUMEN TES KETERAMPILAN BERPIKIR KRITIS (KBK) MATEMATIS

Satuan Pendidikan : SMP/MTs

Alokasi Waktu : 80 menit

Mata Pelajaran : Matematika

Jumlah Soal : 3 butir (3 Essay)

Materi : Rasio

Bentuk Tes : Tertulis

Kelas/Semester : VII (Tujuh)/Genap

Penyusun : Muziyati Hanim

Kurikulum : Merdeka

No .	Capaian Pembelajaran (CP)	Tujuan Pembelajaran (TP)	Indikator KBK	Indikator Soal	Level Kognitif	No. Soal
1	Pada akhir fase D, peserta didik dapat menyelesaikan masalah kontekstual peserta didik dengan menggunakan konsep-konsep dan keterampilan matematika yang dipelajari pada fase ini. Peserta didik dapat menggunakan faktorisasi prima dan pengertian rasio (skala, proporsi, dan laju perubahan) dalam penyelesaian masalah	Peserta didik dapat menggunakan rasio (dan laju perubahan yang terkait) untuk menyelesaikan masalah sehari-hari	<i>Focus</i> : Siswa dapat memahami masalah dan dapat menuliskan apa yang diketahui dari soal. <i>Reason</i> : Siswa dapat memahami masalah dan dapat menuliskan apa yang ditanya dari soal. <i>Inference</i> : Siswa membuat kesimpulan terkait langkah yang tepat untuk penyelesaian soal.	Disajikan suatu permasalahan terkait jejak kaki seorang anak laki-laki, peserta didik diminta untuk dapat menganalisis dan menyelesaikan permasalahan menggunakan konsep rasio.	C4	1
2		Peserta didik dapat menghubungkan rasio ekuivalen dengan proporsi dalam menyelesaikan masalah sehari-hari	<i>Situation</i> : Siswa menyelesaikan permasalahan dengan mengumpulkan semua informasi yang sesuai dengan masalah <i>Clarity</i> : Siswa dapat memberikan penjelasan (membenarkan atau mengklarifikasi) lebih lanjut tentang kesimpulan akhir	Disajikan suatu permasalahan terkait dibutuhkan instruktur kegiatan membuat disekolah, peserta didik diminta untuk menyelesaikan permasalahan terkait total instruktur yang harus ditambah dengan menggunakan konsep rasio ekuivalen	C4	2
3		Peserta didik dapat menggunakan konsep rasio untuk menyelesaikan permasalahan sehari-hari	<i>Overview</i> : Siswa memeriksa kembali langkah-langkah yang dikerjakan atau menguji jawaban	Peserta didik diminta untuk membuat kelompok berenang dan mencari apakah ada siswa yang tidak mendapat kelompok dari formasi yang ditentukan.	C4	3

## FORM VALIDASI INSTRUMEN TES KETERAMPILAN BERPIKIR KRITIS MATEMATIS

Satuan Pendidikan : SMP/MTs  
 Mata Pelajaran : Matematika  
 Materi : Rasio  
 Kelas/Semester : VII (Tujuh)/Genap  
 Kurikulum : Merdeka

Alokasi Waktu : 80 menit  
 Jumlah Soal : 3 butir (3 Essay)  
 Bentuk Tes : Tertulis  
 Penyusun : Muziyati Hanim

SOAL	PENYELESAIAN	INDIKATOR KBK	DESKRIPSI	PENYEKORAN	1	2	3	4	5	Saran
<b>1. Berjalan</b> Gambar tersebut menunjukkan jejak kaki seorang anak laki-laki yang sedang berjalan. Panjang langkah dilambangkan dengan $p$ . $p$ adalah jarak antara bagian dua jarak kaki yang berurutan. Jika rumus yang terdapat pada gambar diatas berlaku	Tahap 1 Diketahui: $\frac{n}{p} = 70$ $n$ = jumlah langkah per menit, dan $p$ = panjang langkah dalam meter	<i>Focus (Fokus)</i>	Siswa dapat menuliskan: $\frac{n}{p} = 70$ $n$ = jumlah langkah per menit, dan $p$ = panjang langkah dalam meter	3						
			Siswa hanya menuliskan 2 unsur yang diketahui dari soal misalnya: $n$ = jumlah langkah per menit, dan $p$ = panjang langkah dalam meter	2						
			Siswa hanya menuliskan satu unsur yang diketahui dari soal, misalkan $p$ = panjang langkah dalam meter	1						
			Tidak ada jawaban	0						
	Tahap 2 Ditanya: Kita perlu mencari nilai $p$ : Berapakah panjang langkah Budi? Untuk mencari $p$ , kita perlu mengganti $n$ dengan 56 dan menyelesaikan persamaan	<i>Reason (Alasan)</i>	Siswa menuliskan apa yang ditanya dengan lengkap: Berapakah panjang langkah Budi?	3						
			Siswa menuliskan apa yang ditanya namun kurang lengkap. Berapakah panjang langkah?	2						
			Siswa menuliskan apa yang ditanya tidak lengkap atau kurang tepat. • Berapakah langkah Budi?	1						

<p>untuk langkah Budi dan Budi berjalan 56 langkah per menit, berapakah panjang langkah Budi? (Nyatakan hasil akhir dalam bentuk desimal )</p>	<p>Tahap 3 Dari rumus yang ada, kita dapat menyusun ulang untuk mencari p:</p> $p = \frac{n}{70}$ <p>Dengan substitusi <math>n = 56</math></p> $p = \frac{56}{70}$	<p><i>Inference</i> (Kesimpulan)</p>	<p>• Berapakah panjang?</p>					
			Tidak ada jawaban	0				
			Siswa dapat mensubstitusi dengan benar kedalam rumus yang sudah diketahui	3				
			Siswa salah menempatkan simbol namun perbandingannya benar:	2				
			$n = \frac{56}{70}$					
	<p>Tahap 4 Mari kita hitung nilai p:</p> $p = \frac{56}{70} = 0.8 \text{ meter}$	<p><i>Situation</i> (Situasi)</p>	Siswa benar dalam menempatkan simbol namun salah dalam membuat perbandingan:	1				
			$p = \frac{70}{56}$					
			Tidak ada jawaban	0				
			Siswa dapat menjawab dengan tepat yaitu 0.8 meter	3				
			Siswa dapat menjawab dengan tepat namun masih dalam bentuk pecahan sederhana	2				
	<p>Tahap 5 Panjang langkah Budi dapat disimpulkan jelas sebagai berikut: Budi berjalan dengan panjang langkah sekitar 0.8 meter untuk setiap langkahnya.</p>	<p><i>Clarity</i> (Kejelasan)</p>	Siswa dapat menjawab namun hasil operasi tidak tepat	1				
			Tidak ada jawaban	0				
			Siswa dapat menuliskan kesimpulan dengan tepat	3				
			Siswa dapat menuliskan kesimpulan dengan tepat namun dalam bentuk pecahan	2				
			Siswa dapat menuliskan kesimpulan namun kurang lengkap dan tepat.	1				
Tidak ada jawaban	0							

	<p>Tahap 6</p> <p>Panjang langkah Budi sudah dihitung dan hasilnya adalah 0.8 meter. Kita bisa memeriksa kembali dengan substitusi kedalam rumus asal:</p> $\frac{56}{0.8} = 70$ <p>Hasilnya sesuai dengan rumus yang diberikan, jadi penyelesaian benar dan lengkap</p>	<p><i>Overview</i> (memeriksa kembali)</p>	<p>Siswa dapat memeriksa kembali dengan mensubstitusi hasil kedalam rumus asal dengan benar</p>	3					
			<p>Siswa dapat memeriksa kembali dengan mensubstitusi hasil kedalam rumus asal namun salah</p>	2					
			<p>Siswa hanya menjawab ya</p>	1					
			<p>Tidak ada jawaban</p>	0					
<p>2. Proyek Batik di Sekolah</p> <p>Kegiatan membatik telah menjadi bagian dari budaya di Provinsi Jambi. Di SMPN 53 Muaro Jambi, motif batik Angso Duo menjadi ciri khas yang diajarkan dan dikembangkan dalam kegiatan ekstrakurikuler. Murid-murid belajar teknik membatik dari pengrajin batik yang</p>	<p>Tahap 1</p> <p>Diketahui:</p> <ul style="list-style-type: none"> <li>• 12 mesin batik menghasilkan 95 kain batik Angso Duo dalam satu bulan</li> <li>• Sekolah mengupayakan menghasilkan 285 batik dalam satu bulan</li> </ul>	<p><i>Focus</i> (Fokus)</p>	<p>Siswa dapat menuliskan yang diketahui dari soal secara lengkap</p>	3					
			<p>Siswa hanya menuliskan dua unsur yang diketahui.</p>	2					
			<p>Siswa hanya menuliskan satu unsur yang diketahui</p>	1					
			<p>Tidak ada jawaban</p>	0					
		<p>Tahap 2</p> <p>Ditanya:</p> <p>Berapakah mesin batik yang dibutuhkan agar upaya sekolah terealisasi?</p>	<p><i>Reason</i> (Alasan)</p>	<p>Siswa menuliskan apa yang ditanya dengan lengkap:</p> <p>Berapakah mesin batik yang dibutuhkan agar upaya sekolah terealisasi?</p>	3				
				<p>Siswa menuliskan apa yang ditanya namun kurang lengkap:</p> <p>Berapakah mesin batik yang dibutuhkan?</p>	2				
				<p>Siswa menuliskan apa yang ditanya namun tidak lengkap:</p> <p>Berapakah mesin batik?</p>	1				
				<p>Tidak ada jawaban</p>	0				
		<p>Tahap 3</p>	<p><i>Inference</i> (Kesimpulan)</p>	<p>Siswa dapat membuat perbandingan untuk mencari jumlah mesin batik</p>	3				

berpengalaman . Pada tahun ajaran 2025, sekolah memiliki 8 pengrajin batik yang mengajarkan seni membatik dengan menggunakan 12 mesin batik. Dalam sebulan, pengrajin batik bersama murid menghasilkan 95 kain Angso Duo. Melihat antusiasme dan minat tinggi murid terhadap kegiatan membatik, sekolah berusaha menambah mesin batik agar produksi kain batik Angso Duo bisa mencapai 285 buah setiap bulan. Berapa mesin batik yang dibutuhkan	Maka kita dapat membuat perbandingan dari hal yang telah diketahui untuk mencari jumlah mesin batik (mb) yang dibutuhkan: $\frac{12 \text{ mb}}{95 \text{ kain}} = \frac{x \text{ mb}}{285 \text{ kain}}$		yang dibutuhkan secara tepat, misalnya: $\frac{12 \text{ mb}}{95 \text{ kain}} = \frac{x \text{ mb}}{285 \text{ kain}}$					
			Siswa salah membuat perbandingan, tetapi bentuknya masih logis atau mendekati benar, misalnya: $\frac{12 \text{ mb}}{x \text{ mb}} = \frac{95 \text{ kain}}{285 \text{ kain}}$	2				
			Siswa salah menempatkan simbol $x$ dan salah membuat perbandingan, seperti: $\frac{12}{x} = \frac{285}{95}$	1				
			Tidak ada jawaban	0				
	Tahap 4 Menghitung nilai $x$ dengan cara menyelesaikan perbandingan senilai: $\frac{12}{95} = \frac{x}{285}$ Sekarang kita dapat menyelesaikan perbandingan tersebut dengan cara silang ( <i>cross multiplication</i> ) $12 \times 285 = 95 \times x$ $3420 = 95x$ $x = \frac{3420}{95} = 36$	<i>Situation</i> (Situasi)	Siswa dapat menjawab dengan tepat dengan langkah-langkah yang tepat dengan hasil 36	3				
			Siswa dapat menjawab dengan tepat dengan langkah-langkah yang tepat namun masih dalam bentuk pecahan sederhana	2				
			Siswa dapat menjawab namun hasil operasi tidak tepat	1				
			Tidak ada jawaban	0				
	Tahap 5 Karena saat ini ada 12 mesin batik, maka jumlah mb tambahan yang dibutuhkan adalah:	<i>Clarity</i> (Kejelasan)	Siswa dapat menyebutkan kesimpulan dengan langkah-langkah yang tepat yaitu 24	3				
			Siswa dapat menyebutkan kesimpulan dengan langkah-langkah yang tepat namun hasil operasi kurang tepat	2				

agar upaya sekolah dapat terealisasi?	$36 - 12 = \text{mesin batik}$ Dengan demikian, 24 mesin batik tambahan yang diperlukan sekolah.		Siswa dapat menyebutkan kesimpulan namun kurang tepat	1					
	Tahap 6 Untuk membuktikan nilai $x$ tepat. Kita dapat memeriksa dengan mensubstitusi nilai $x$ keperbandingan awal yang telah dibuat.  $\frac{12}{95} = \frac{x}{285}$  $\frac{12}{95} = \frac{36}{285}$	<i>Overview</i> (memeriksa kembali)	Siswa dapat memeriksa kembali dengan mensubstitusi nilai $x$ kedalam perbandingan awal	3					
			Siswa hanya menuliskan ulang langkah-langkah penyelesaian yang tepat	2					
			Siswa memeriksa kembali namun jawabannya tidak tepat	1					
			Tidak ada jawaban	0					
	<b>3. Pembentukan Kelompok</b>  Untuk membentuk kelompok senam diperlukan 3 siswa laki-laki dan 2 siswa perempuan. Jumlah siswa yang berminat untuk ikut senam adalah 9 siswa	Tahap 1 Diketahui : Membuat pembagian kelompok senam 1 Kelompok = 3 siswa laki-laki dan 2 siswa perempuan Yang berminat = 9 siswa perempuan dan 12 siswa laki-laki.	<i>Focus (Fokus)</i>	Siswa membuat poin ketiga yang diketahui	3				
Siswa membuat poin kedua yang diketahui				2					
Siswa membuat poin pertama yang diketahui				1					
Siswa tidak sama sekali menuliskan apa yang diketahui				0					
Tahap 2 Ditanya : A. Berapa banyak kelompok senam yang dapat terbentuk ? B. Apakah ada siswa yang tidak mendapat kelompok? Jika iya berapa orang?		<i>Reason (Alasan)</i>	Siswa membuat poin ketiga yang ditanya	3					
			Siswa membuat poin kedua yang ditanya	2					
			Siswa membuat poin pertama yang ditanya	1					
			Tidak ada jawaban	0					
Tahap 3		<i>Inference</i> (Kesimpulan)	Siswa dapat membuat formula dengan benar	3					

<p>perempuan dan 12 siswa laki-laki.</p> <p>a. Berapa banyak kelompok senam yang dapat terbentuk ?</p> <p>b. Apakah ada siswa yang tidak mendapat kelompok? Jika iya berapa orang?</p>	$\frac{3}{2} = \frac{12}{X}$		Siswa dapat membuat formula tetapi kurang tepat	2					
			Siswa dapat membuat formula tetapi salah	1					
			Tidak ada jawaban	0					
	<p>Tahap 4</p> $X = \frac{2 \times 12}{3}$ $X = 8$ <p>Jumlah siswa perempuan yang masuk kelompok senam adalah 8 orang</p>	<p><i>Situation</i> (Situasi)</p>	Siswa dapat membuat langkah-langkah dengan jawaban yang benar yaitu 8	3					
			Siswa hanya menuliskan jawaban akhir yaitu 8	2					
			Siswa salah menstubtitusi dalam rumus	1					
			Tidak ada jawaban	0					
	<p>Tahap 5</p> <p>Perbandingan Senilai</p> <p>1 Kelompok senam : 3 siswa laki-laki : 2 siswa Perempuan</p> <p>X Kelompok senam : 12 siswa laki-laki : 8 siswa 2perempuan</p> <p>A. Jadi, jumlah ke0lompok senam yang terbentuk adalah 4</p> <p>Siswa yang tidak mendapat kelompok</p> <ul style="list-style-type: none"> <li>- Laki-laki = <math>12 - 12 ( 3 \times 4 ) = 0</math></li> <li>- Perempuan = <math>9 - 8 ( 2 \times 4 ) = 1</math></li> </ul>	<p><i>Clarity</i> (Kejelasan)</p>	Siswa dapat membuat jawaban poin a dan b dengan tepat	3					
			Siswa hanya bisa menulis satu dari dua point	2					
			Siswa dapat menjawab tetapi kurang tepat	1					
Tidak ada jawaban			0						

	B. Jadi, siswa yang tidak mendapat kelompok adalah satu siswa perempuan									
	Tahap 6 Siswa dapat menyajikan kembali dalam bentuk yang berbeda dengan hasil yang sama Siswa dapat mengalikan perbandingan dengan 4 dan 2 didapatkan hasil bahwa terdapat satu orang yang tidak mendapat kelompok yaitu siswa perempuan	<i>Overview</i> (memeriksa kembali)	Siswa dapat memeriksa kembali dengan mensubstitusi nilai ke dalam rumus asal	3						
			Siswa hanya menuliskan ulang langkah-langkah penyelesaian yang tepat	2						
			Siswa memeriksa kembali namun jawabannya tidak tepat	1						
			Tidak ada jawaban	0						

Berdasarkan hasil penilaian yang dilakukan, maka dapat ditetapkan:

- a) Perangkat tes dapat digunakan tanpa revisi
- b) Perangkat tes dapat digunakan dengan sedikit revisi
- c) Perangkat tes dapat digunakan dengan banyak revisi
- d) Perangkat tes tidak dapat digunakan dan masih memerlukan konsultasi.

\*lingkari pada poin yang dipilih

## LEMBAR SOAL ESSAY

Mata Pelajaran : Matematika

Kelas : VII

Materi : Rasio

Alokasi Waktu : 80 menit

### PETUNJUK UMUM

1. Tulislah nama anda pada lembar jawaban yang telah disediakan
2. Periksa dan bacalah dengan teliti sebelum anda mengerjakan
3. Jangan lupa berdoa sebelum mengerjakan soal
4. Gunakan waktu yang efektif dan efisien dalam mengerjakan soal
5. Periksa kembali jawaban anda sebelum dikumpulkan

### Soal 1: Berjalan



*Sumber: PISA Assessments 2009*

Gambar tersebut menunjukkan jejak kaki seorang anak laki-laki yang sedang berjalan. Panjang langkah dilambangkan dengan  $p$ .  $p$  adalah jarak antara bagian dua jejak kaki yang berurutan.

Jika rumus yang terdapat pada gambar diatas berlaku untuk langkah Budi dan Budi berjalan 56 langkah per menit, berapakah panjang langkah Budi? Apakah kamu sudah memeriksa ulang jawaban dan yakin sudah benar?

(Nyatakan hasil akhir dalam bentuk desimal )

Soal 2: Proyek Batik di Sekolah



*Sumber: Instagram.@muziyatihanim16*

Kegiatan membatik telah menjadi bagian dari budaya di Provinsi Jambi. Di SMPN 53 Muaro Jambi, motif batik Angso Duo menjadi ciri khas yang diajarkan dan dikembangkan dalam kegiatan ekstrakurikuler. Murid-murid belajar teknik membatik dari pengrajin batik yang berpengalaman.

Pada tahun ajaran 2025, sekolah memiliki 8 pengrajin batik yang mengajarkan seni membatik dengan menggunakan 12 mesin batik. Dalam sebulan, pengrajin batik bersama murid menghasilkan 95 kain Angso Duo. Melihat antusiasme dan minat tinggi murid terhadap kegiatan membatik, sekolah berusaha menambah mesin batik agar produksi kain batik Angso Duo bisa mencapai 285 buah setiap bulan.

Berapa mesin batik yang harus ditambahkan agar upaya sekolah dapat terealisasi? Apakah kamu sudah memeriksa ulang jawaban dan yakin sudah benar?

## Lampiran 5. Jawaban Siswa

Nama: Adana angraeni

Kelas: 7A

Sekolah: SMPN 14 Muara Jambi

① Diketahui:  $\frac{n}{p} = 70$

$n$  = Jumlah langkah per menit

Ditanya: Berapakah mesin yang harus ditambah? 3

Jawab:

$$\frac{n}{p} = 70 \quad 3$$

$$p = \frac{n}{70}$$

$$p = \frac{56}{70} \quad 3$$

$$p = 0,8$$

Jadi Panjang langkah yaitu 0,8 meter 3

→ di Periksa:  $\frac{56}{0,8} = 70$

$$70 = 70 \text{ benar} \quad 3$$

② Diketahui: 8 pengrajin batik  
12 mesin batik  
95 kain 3

Produksi kain oleh sekolah 285 kain

Ditanya: Berapakah mesin yang di butuhkan

Jawab:  $\frac{a}{b} = \frac{c}{d} = \frac{12}{95} = \frac{x}{285}$  3

$$x = \frac{12 \cdot 285}{95}$$

$$= \frac{3420}{95}$$

$$= 36 - 12$$

$$= 24$$

Jadi mesin batik yang di butuhkan  $36 - 12 = 24$  3

di Periksa:  $\frac{12}{95} = \frac{x}{285}$  3

$$\frac{12}{95} = \frac{36}{285}$$

$$0,1 = 0,1 \text{ (benar)}$$

Siswa Berkemampuan Tinggi

nama: Tsalia Ngeliva Tsemana Siano | 18/2025  
 kelas: VII C | 03  
 sekolah: SMP 14 Muaro Jambi | Selesai

1) Di ketahui = Rumus =  $\frac{n}{p} = 70$  3

n = jumlah langkah per menit

p = panjang langkah dalam meter

Ditanya: Berapa panjang langkah Budi? 3

Dijawab:  $\frac{n}{p} = 70$  3

$p = \frac{n}{70}$

$p = \frac{56}{70}$  3

$p = 0,8$

Jadi panjang langkah Budi adalah 0,8 3

2) Diketahui: 12 mesin dapat menghasilkan 95 kain batik angso 3  
 200

Ditanya: jika setiapbulan menghasilkan 285 kain, maka berapa 3  
 mesin pembatik yang dibutuhkan?

Dijawab: 12 mesin, 95 kain/bulan

$\frac{12}{95} = \frac{n}{285}$  3

$95 \cdot x = 12 \cdot 285$

$95 \cdot x = 3420$

$x = \frac{3420}{95}$  3

$x = 36 - 12 = 24$  mesin batik 3

Nama = Rahmad dwi Julyanto  
 Kls = 7c

1) di ketahui:  $\frac{n}{p} = 70$  3

n = jumlah langkah per menit

p = panjang langkah dalam meter

n = 56

ditanya: berapa panjang langkah Budi? 3

di jawab:  $\frac{n}{p} = 70$  3

$p = \frac{n}{70}$

$p = \frac{56}{70}$

$p = 0,8$

2) diketahui: 12. mesin batik bisa menghasilkan 95 kain 3  
 - Sekolah menerima 285 kain batik

ditanya: berapa mesin yang dibutuhkan? 3

di jawab:  $\frac{12}{95} = \frac{x}{285}$  3

$95x = 12 \cdot 285$

$95x = 3420$

$x = \frac{3420}{95}$

$x = 36$

$x = 36 - 12 = 24$

$x = 24$

$$\begin{array}{r} 11 \\ 285 \\ \times 12 \\ \hline 570 \\ 2850 \\ \hline 3420 \end{array}$$

Siswa Berkemampuan Sedang

Siswa Berkemampuan Rendah

## Lampiran 6. Uji Kecocokan Model Luaran PARSCALE ph.2

ITEM FIT STATISTICS

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	11.30503	10.	0.334
BLOCK	0002	13.95080	6.	0.030
BLOCK	0003	12.91464	9.	0.166
BLOCK	0004	7.76546	8.	0.457
BLOCK	0005	7.67881	7.	0.362
BLOCK	0006	15.39692	9.	0.080
BLOCK	0007	3.14789	9.	0.958
BLOCK	0008	3.04032	8.	0.932
BLOCK	0009	9.03702	6.	0.170
BLOCK	0010	7.10838	9.	0.627
BLOCK	0011	10.45706	9.	0.314
BLOCK	0012	4.60770	8.	0.800
TOTAL		106.41002	98.	0.264

68668 BYTES OF NUMERICAL WORKSPACE USED OF 819;  
 592 BYTES OF CHARACTER WORKSPACE USED OF 204;  
 NORMAL END

ITEM FIT STATISTICS

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	0.99839	2.	0.613
BLOCK	0002	0.00048	1.	0.931
BLOCK	0003	1.19224	2.	0.556
BLOCK	0004	4.68230	4.	0.321
BLOCK	0005	1.66293	6.	0.947
BLOCK	0006	4.31279	8.	0.829
BLOCK	0007	1.32560	1.	0.248
BLOCK	0008	1.18694	1.	0.276
BLOCK	0009	2.33982	2.	0.311
BLOCK	0010	3.73494	5.	0.590
BLOCK	0011	1.75091	4.	0.784
BLOCK	0012	3.36497	9.	0.948
TOTAL		26.55230	45.	0.987

72364 BYTES OF NUMERICAL WORKSPACE USED OF 819;  
 592 BYTES OF CHARACTER WORKSPACE USED OF 204;  
 NORMAL END

### PCM

### GPCM

ITEM FIT STATISTICS

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	0.80088	2.	0.676
BLOCK	0002	0.56604	1.	0.458
BLOCK	0003	2.77018	2.	0.249
BLOCK	0004	8.43738	5.	0.133
BLOCK	0005	4.65031	7.	0.704
BLOCK	0006	4.86990	10.	0.900
BLOCK	0007	0.37818	1.	0.546
BLOCK	0008	1.40496	1.	0.234
BLOCK	0009	4.04880	3.	0.255
BLOCK	0010	6.22443	7.	0.515
BLOCK	0011	2.94760	6.	0.817
BLOCK	0012	7.37149	9.	0.600
TOTAL		44.47016	54.	0.819

72364 BYTES OF NUMERICAL WORKSPACE USED OF 8192000 A  
 592 BYTES OF CHARACTER WORKSPACE USED OF 2048000 A  
 NORMAL END

### GRM

Lampiran 7. Estimasi Keterampilan Berpikir Kritis Matematis Siswa Luaran  
PARSCALE ph. 3

PCM

SUBJECT IDENTIFICATION				WEIGHT/FREQUENCY			ABILITY	S.E.
SCORE	NAME	GROUP	WEIGHT	MEAN	CATEGORY	ATTEMPTS		
M001			1		GROUP 01	1.00		
1	SCALE1	1	1.00	3.75		1.00	1.4825	0.3437
M002			2		GROUP 01	1.00		
1	SCALE1	1	1.00	3.75		1.00	1.4825	0.3437
M003			3		GROUP 01	1.00		
1	SCALE1	1	1.00	3.08		1.00	0.1621	0.2215
M004			4		GROUP 01	1.00		
1	SCALE1	1	1.00	3.42		1.00	0.6753	0.2556
M005			5		GROUP 01	1.00		
1	SCALE1	1	1.00	3.33		1.00	0.5313	0.2421
M006			6		GROUP 01	1.00		
1	SCALE1	1	1.00	2.67		1.00	-0.3797	0.2101
M007			7		GROUP 01	1.00		
1	SCALE1	1	1.00	3.83		1.00	1.8062	0.4091
M008			8		GROUP 01	1.00		
1	SCALE1	1	1.00	3.42		1.00	0.6753	0.2556
M009			9		GROUP 01	1.00		
1	SCALE1	1	1.00	3.58		1.00	1.0223	0.2910
M010			10		GROUP 01	1.00		
1	SCALE1	1	1.00	3.58		1.00	1.0223	0.2910

	1	SCALE1	1	1.00	2.00	1.00	-1.1374	0.1996
M097				97	GROUP 01	1.00		
1	SCALE1	1	1.00	2.58		1.00	-0.4813	0.2075
M098				98	GROUP 01	1.00		
1	SCALE1	1	1.00	3.75		1.00	1.4825	0.3437
M099				99	GROUP 01	1.00		
1	SCALE1	1	1.00	2.58		1.00	-0.4813	0.2075
M100				100	GROUP 01	1.00		
1	SCALE1	1	1.00	1.75		1.00	-1.4351	0.2166
M101				101	GROUP 01	1.00		
1	SCALE1	1	1.00	1.83		1.00	-1.3300	0.2084
M102				102	GROUP 01	1.00		
1	SCALE1	1	1.00	2.08		1.00	-1.0454	0.1979
M103				103	GROUP 01	1.00		
1	SCALE1	1	1.00	1.25		1.00	-2.5953	0.4054
M104				104	GROUP 01	1.00		
1	SCALE1	1	1.00	3.08		1.00	0.1621	0.2215
M105				105	GROUP 01	1.00		
1	SCALE1	1	1.00	1.75		1.00	-1.4351	0.2166
M106				106	GROUP 01	1.00		
1	SCALE1	1	1.00	1.92		1.00	-1.2316	0.2030

## GPCM

SUBJECT IDENTIFICATION				WEIGHT/FREQUENCY				
SCORE	NAME	GROUP	WEIGHT	MEAN	CATEGORY	ATTEMPTS	ABILITY	S.E.
M001	1	PCR_MLE	1	1.00	GROUP 01	1.00	1.2737	0.2476
M002	1	PCR_MLE	1	1.00	GROUP 01	1.00	1.3856	0.2659
M003	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.3056	0.1858
M004	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.6486	0.1989
M005	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.5041	0.1873
M006	1	PCR_MLE	1	1.00	GROUP 01	1.00	-0.3509	0.2033
M007	1	PCR_MLE	1	1.00	GROUP 01	1.00	1.6517	0.3342
M008	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.6348	0.1976
M009	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.8705	0.2184
M010	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.9443	0.2227

M096	1	PCR_MLE	1	1.00	GROUP 01	1.00	-1.3077	0.1874
M097	1	PCR_MLE	1	1.00	GROUP 01	1.00	-0.5745	0.1911
M098	1	PCR_MLE	1	1.00	GROUP 01	1.00	1.4469	0.2788
M099	1	PCR_MLE	1	1.00	GROUP 01	1.00	-0.5626	0.1919
M100	1	PCR_MLE	1	1.00	GROUP 01	1.00	-1.4060	0.2014
M101	1	PCR_MLE	1	1.00	GROUP 01	1.00	-1.1949	0.1760
M102	1	PCR_MLE	1	1.00	GROUP 01	1.00	-1.0800	0.1697
M103	1	PCR_MLE	1	1.00	GROUP 01	1.00	-2.9185	0.5495
M104	1	PCR_MLE	1	1.00	GROUP 01	1.00	0.5356	0.1893
M105	1	PCR_MLE	1	1.00	GROUP 01	1.00	-1.0950	0.1702
M106	1	PCR_MLE	1	1.00	GROUP 01	1.00	-1.3641	0.1950

## GRM

SUBJECT IDENTIFICATION			WEIGHT/FREQUENCY			ABILITY	S.E.
SCORE	NAME	GROUP	WEIGHT	MEAN CATEGORY	ATTEMPTS		
M001			1	GROUP 01	1.00		
1	EAP	1	1.00	3.75	1.00	1.3570	0.3267
M002			2	GROUP 01	1.00		
1	EAP	1	1.00	3.75	1.00	1.2201	0.2847
M003			3	GROUP 01	1.00		
1	EAP	1	1.00	3.08	1.00	0.2330	0.1972
M004			4	GROUP 01	1.00		
1	EAP	1	1.00	3.42	1.00	0.5936	0.1931
M005			5	GROUP 01	1.00		
1	EAP	1	1.00	3.33	1.00	0.4590	0.1949
M006			6	GROUP 01	1.00		
1	EAP	1	1.00	2.67	1.00	-0.3098	0.1894
M007			7	GROUP 01	1.00		
1	EAP	1	1.00	3.83	1.00	1.5288	0.3190
M008			8	GROUP 01	1.00		
1	EAP	1	1.00	3.42	1.00	0.6067	0.2324
M009			9	GROUP 01	1.00		
1	EAP	1	1.00	3.58	1.00	1.0181	0.2616
M010			10	GROUP 01	1.00		
1	EAP	1	1.00	3.58	1.00	0.9584	0.2348
M096			96	GROUP 01	1.00		
1	EAP	1	1.00	2.00	1.00	-1.4055	0.2326
M097			97	GROUP 01	1.00		
1	EAP	1	1.00	2.58	1.00	-0.5920	0.1784
M098			98	GROUP 01	1.00		
1	EAP	1	1.00	3.75	1.00	1.1740	0.2762
M099			99	GROUP 01	1.00		
1	EAP	1	1.00	2.58	1.00	-0.5177	0.1898
M100			100	GROUP 01	1.00		
1	EAP	1	1.00	1.75	1.00	-1.4765	0.1976
M101			101	GROUP 01	1.00		
1	EAP	1	1.00	1.83	1.00	-1.2127	0.1911
M102			102	GROUP 01	1.00		
1	EAP	1	1.00	2.08	1.00	-1.0958	0.1814
M103			103	GROUP 01	1.00		
1	EAP	1	1.00	1.25	1.00	-2.7715	0.3913
M104			104	GROUP 01	1.00		
1	EAP	1	1.00	3.08	1.00	0.3161	0.1984
M105			105	GROUP 01	1.00		
1	EAP	1	1.00	1.75	1.00	-1.1114	0.2231
M106			106	GROUP 01	1.00		
1	EAP	1	1.00	1.92	1.00	-1.4901	0.2253

## Lampiran 8. Penyekoran Tes Keterampilan Berpikir Kritis Matematis

M001 333331333233	M048 033300333330	M093 331100333300
M002 333333333231	M047 333111333331	M094 333211322121
M003 333221333200	M049 333300333330	M095 333211222211
M004 333221333231	M050 333331333221	M096 332200200000
M005 333220333231	M051 333331333331	M097 333200222110
M006 233110332200	M052 333321333332	M098 333333333222
M007 333333333232	M053 333300333200	M099 332000332210
M008 333330333230	M054 332000333000	M100 311000211000
M009 333330333331	M055 332000332000	M101 112200111100
M010 333331333231	M056 333300333300	M102 321100321000
M011 233330333230	M057 333332333332	M103 110000100000
M012 333331333331	M058 300000330000	M104 11222333222
M013 333330333230	M059 333000333000	M105 000000332100
M014 233333333333	M060 333000333000	M106 333000110000
M015 333330333330	M061 333330333332	
M016 233322333331	M062 333000000000	
M017 333231332230	M063 333000230000	
M018 333200333330	M064 333331333321	
M019 203331333332	M065 333311333332	
M020 333300332220	M066 333000333000	
M021 333000332220	M067 333331333310	
M022 333000332220	M068 233330333331	
M023 233333333332	M069 333000333100	
M024 333200321000	M070 333331233300	
M025 333300333200	M071 133331333300	
M026 333300333211	M072 133331233311	
M027 333321333331	M073 112000300000	
M028 330000333330	M074 131100333331	
M029 333300333333	M075 111100331000	
M030 333333333330	M076 221111311000	
M031 333330333330	M077 111100333311	
M032 333332333220	M078 111100123300	
M033 333200333000	M079 131100333300	
M034 333333333221	M080 131100333300	
M035 333300333200	M081 231100333100	
M036 111000100000	M082 231100333300	
M037 333200333200	M083 112100333110	
M038 333300333200	M084 131100333300	
M039 333300302200	M085 333322333322	
M040 333300333200	M086 132100333331	
M041 333300333210	M087 112000200000	
M042 333000333332	M088 211000112000	
M043 333000333320	M089 222000211000	
M044 333300333330	M090 111000110000	
M045 333322333321	M091 333333333200	
M046 333211333332	M092 111000100000	

Lampiran 9. Surat Selesai Bimbingan Validasi Instrumen Keterampilan Berpikir Kritis Matematis



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN,  
RISET DAN TEKNOLOGI  
UNIVERSITAS JAMBI**

FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN  
JURUSAN PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PROGRAM STUDI MAGISTER PENDIDIKAN MATEMATIKA  
Alamat: Kampus UNJA Pasar, Jl. Raden Mattaheer No. 16- Jambi 0741- 34058  
-Laman: mpmat.unja.ac.id

---

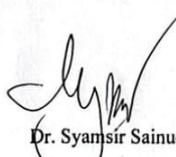
**PERNYATAAN PEMBIMBING VALIDATOR**

Nama : Dr. Syamsir Sainuddin, M.Pd  
Kedudukan Sebagai : Validator I

Menyatakan Bahwa  
Nama : Muziyati Hanim  
NIM : P2A923004  
Judul Tesis : Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)

Telah melakukan bimbingan Instrumen Penelitian berupa Tes Keterampilan Berpikir Kritis Matematis dan divalidasi pada tanggal 18 Februari 2025

Jambi, 18 Februari 2025  
Mengetahui,  
Pembimbing Validator I

  
 Dr. Syamsir Sainuddin, M.Pd  
 NIP-198810292024061001

---

*"Magister Pendidikan Matematika UNJA Menjadi Pusat Unggulan Dalam Pembelajaran dan Penelitian Pendidikan Matematika di Indonesia Pada Tahun 2025"*

CS Dipindai dengan CamScanner



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN,  
RISET DAN TEKNOLOGI  
UNIVERSITAS JAMBI**  
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN  
JURUSAN PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PROGRAM STUDI MAGISTER PENDIDIKAN MATEMATIKA  
Alamat: Kampus UNJA Pasar, Jl. Raden Mattaher No. 16- Jambi 0741- 34058  
-Laman: mpmat.unja.ac.id

### PERNYATAAN PEMBIMBING VALIDATOR

Nama : Aliffia Teja Prasasty, M.Pd.  
Instansi : Universitas Indraprasta PGRI  
Kedudukan Sebagai : Validator II

#### Menyatakan Bahwa

Nama : Muziyati Hanim  
NIM : P2A923004  
Judul Tesis : Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa  
SMP Dengan Pendekatan *Item Response Theory (IRT)*

Telah selesai melakukan bimbingan Instrumen Penelitian berupa Tes Keterampilan Berpikir Kritis Matematis dan divalidasi pada tanggal 06 Maret 2025

Jambi, 06 Maret 2025  
Mengetahui,  
Pembimbing Validator II

Aliffia Teja Prasasty, M.Pd.  
NIDN 0314098902

*"Magister Pendidikan Matematika UNJA Menjadi Pusat Unggulan Dalam Pembelajaran dan Penelitian Pendidikan Matematika di Indonesia Pada Tahun 2025"*

## FORM BIODATA VALIDATOR

Nama Validator : YUNI ASTRIANA, S.Pd  
 Jenis Kelamin :  Laki-laki  Perempuan  
 Pendidikan Terakhir :  S1  S2  S3  
 Bidang Keahlian : GURU MATEMATIKA  
 Institusi : SMPN 14 MUARO JAMBI

MUARO JAMBI, 10 03, 2025

  
 (.....YUNI ASTRIANA, S.Pd.....)

## FORM BIODATA VALIDATOR

Nama Validator : SAHNIAR, S.Pd  
 Jenis Kelamin :  Laki-laki  Perempuan  
 Pendidikan Terakhir :  S1  S2  S3  
 Bidang Keahlian : Guru Matematika  
 Institusi : SMP NEGERI 12 MUARO JAMBI

Muaro Jambi, 10 Maret, 2025

  
 (.....SAHNIAR, S.Pd.....)

## FORM BIODATA VALIDATOR

Nama Validator : Nanang Sriyadi, S.Pd. Gr  
 Jenis Kelamin :  Laki-laki  Perempuan  
 Pendidikan Terakhir :  S1  S2  S3  
 Bidang Keahlian : Guru Matematika  
 Institusi : SMP NEGERI 54 MUARO JAMBI

Muaro Jambi, 10 Maret, 2025

  
 (.....Nanang Sriyadi, S.Pd.....)

## Lampiran 10. Surat Izin Selesai Penelitian



PEMERINTAH KABUPATEN MUARO JAMBI  
DINAS PENDIDIKAN DAN KEBUDAYAAN  
**SMP NEGERI 49 MUARO JAMBI**  
KECAMATAN BAHAR UTARA

Desa Sumber Jaya, Sungai Bahar XIX, Kecamatan Bahar Utara Kode Pos 36365  
Laman Pos -el : [smpn1ataps@gmail.com](mailto:smpn1ataps@gmail.com)

**SURAT KETERANGAN**

Nomor: 421.3/0 /SMPN49MJ/2025

Yang bertandatangan dibawah ini, Kepala SMP Negeri 49 Muaro Jambi dengan ini menerangkan:

Nama : Muziyati Hanim  
NIM : P2A923004  
Tempat/Tanggal Lahir : Balai Jaya, 16 Mei 2000  
Prodi : Megister Pendidikan Matematika  
Jurusan : Pendidikan MIPA  
Alamat : Desa Balai Jaya RT 02, RW 00 Kec. Pelepat, Kabupaten Bungo

Telah melaksanakan penelitian dan pengambilan data di SMPN 49 Muaro Jambi yang berhubungan dengan tesis yang bersangkutan yakni "Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)" dari tanggal 23 April s/d 20 Mei 2025.

Sumber Jaya, 20 Mei 2025

Kepala SMPN 49 Muaro Jambi

  
**EDGA SIANIPAR, S.Pd**  
NIP. 196811272007011005



**DINAS PENDIDIKAN DAN KEBUDAYAAN  
SMP NEGERI 14 KABUPATEN MUARO JAMBI  
NSS.20.1.10.09.07.003  
KECAMATAN BAHAR UTARA**

Alamat: Desa Talang Datar, Sungai Bahar VI

Kode pos : 36365

SURAT KETERANGAN

Nomor: 800/56/SMPN.14/MJ/PDD.2025

Yang bertandatangan dibawah ini, Kepala SMP Negeri 14 Muaro Jambi dengan ini menerangkan:

Nama	: Muziyati Hanim
NIM	: P2A923004
Tempat/Tanggal Lahir	: Balai Jaya, 16 Mei 2000
Prodi	: Megister Pendidikan Matematika
Jurusan	: Pendidikan MIPA
Alamat	: Desa Balai Jaya RT 02, RW 00 Kec. Pelepat, Kabupaten Bungo

Telah melaksanakan penelitian dan pengambilan data di SMPN 14 Muaro Jambi yang berhubungan dengan tesis yang bersangkutan yakni "Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)" dari tanggal 23 April s/d 20 Mei 2025.

Talang Datar, 20 Mei 2025

Kepala SMPN 14 Muaro Jambi



**HELMIRZAD HIDAYAT, S.Ag**  
NIP. 197112212007011003



**PEMERINTAH KABUPATEN MUARO JAMBI  
DINAS PENDIDIKAN DAN KEBUDAYAAN  
SMP NEGERI 45 MUARO JAMBI  
KECAMATAN BAHAR UTARA**

Alamat: Jl Flamboyan No. 11 Desa Bukit Mulya Kec. Bahar Utara

Kode pos : 36365

SURAT KETERANGAN

Nomor: 800/ 71/SMPN45/MJ/2025

Yang bertandatangan dibawah ini, Kepala SMP Negeri 45 Muaro Jambi dengan ini menerangkan:

Nama	: Muziyati Hanim
NIM	: P2A923004
Tempat/Tanggal Lahir	: Balai Jaya, 16 Mei 2000
Prodi	: Megister Pendidikan Matematika
Jurusan	: Pendidikan MIPA
Alamat	: Desa Balai Jaya RT 02, RW 00 Kec. Pelepat, Kabupaten Bungo

Telah melaksanakan penelitian dan pengambilan data di SMPN 45 Muaro Jambi yang berhubungan dengan tesis yang bersangkutan yakni "Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)" dari tanggal 23 April s/d 20 Mei 2025.

Bukit Mulya, 20 Mei 2025

Kepala SMPN 45 Muaro Jambi



*[Signature]*  
**ROSNITA, S.Ag**

NIP. 197102082008012001



PEMERINTAH KABUPATEN MUARO JAMBI  
DINAS PENDIDIKAN DAN KEBUDAYAAN  
**SMP NEGERI 53 MUARO JAMBI**  
KECAMATAN BAHAR UTARA



Alamat : Jl. Lintas sungai Bahar- Jambi, KM.56, RT.04, RW.01 Desa Markanding (36365)  
Email : markandingsmpnsatap@gmail.com

SURAT KETERANGAN

Nomor: 421.2/63/SMPN53/2025

Yang bertandatangan dibawah ini, Kepala SMP Negeri 53 Muaro Jambi dengan ini menerangkan:

Nama : Muziyati Hanim  
NIM : P2A923004  
Tempat/Tanggal Lahir : Balai Jaya, 16 Mei 2000  
Prodi : Megister Pendidikan Matematika  
Jurusan : Pendidikan MIPA  
Alamat : Desa Balai Jaya RT 02, RW 00 Kec. Pelepat, Kabupaten Bungo

Telah melaksanakan penelitian dan pengambilan data di SMPN 53 Muaro Jambi yang berhubungan dengan tesis yang bersangkutan yakni "Perbandingan Akurasi Penilaian Keterampilan Berpikir Kritis Matematis Siswa SMP dengan Pendekatan *Item Response Theory* (IRT)" dari tanggal 23 April s/d 20 Mei 2025.

Markanding, 20 Mei 2025

PIK Kepala SMPN 53 Muaro Jambi



**NOVI KURNIAWAN, S.Pd**  
NIP. 198902092019031002

### Lampiran 11. Dokumentasi Penelitian



SMPN 14 Muaro Jambi



SMPN 53 Muaro Jambi



SMPN 49 Muaro Jambi



SMPN 45 Muaro Jambi

**Lampiran 12. Link Drive Langkah-langkah GPCM**

<https://drive.google.com/drive/folders/1JbQCzMkHEa1hQ3NqeLKmIoV0IsfghNRR>

## RIWAYAT HIDUP

Muziyati Hanim, perempuan yang dilahirkan di Balai Jaya, 16 Mei 2000. Anak pertama dari tiga bersaudara, pasangan dari Bapak Hamdan dan Ibu Halimah Tusaqdhah. Penulis menyelesaikan pendidikan di SD Negeri 170 II Balai Jaya pada tahun 2012. Pada tahun 2015, penulis menyelesaikan pendidikan di Madrasah Tsanawiyah Negeri 1 Muara Bungo dan kemudian melanjutkan pendidikan di Madrasah Aliyah Negeri 1 Bungo dan tamat pada tahun 2018.



Pada tahun 2018 penulis melanjutkan pendidikan di perguruan tinggi negeri, tepatnya di Universitas Jambi (UNJA) Fakultas Keguruan dan Ilmu Pendidikan pada program studi Pendidikan Matematika. Penulis merupakan ketua penerima pendanaan PBNP UNJA pada kegiatan PKM-Riset 2021. Semasa perkuliahan penulis banyak mendapatkan ilmu baru dan bertambah pengalaman dengan mengikuti pengabdian program Kampus Mengajar Angkatan 1 yang dilaksanakan oleh Kementerian Pendidikan dan Kebudayaan (KEMENDIKBUD). Pada Tahun 2023 Penulis melanjutkan Studi Magister di Program Magister Pendidikan Matematika UNJA. Pada tahun yang sama, melalui beasiswa penulis juga mengambil Profesi Pendidikan Guru Pra-Jabatan di UNJA. Penulis juga merupakan awardee Beasiswa Unggulan Tahun 2024. Semua pengalaman dan ilmu yang penulis dapatkan tidak terlepas dari adanya kemauan, kegigihan, dan rasa percaya diri serta iringan do'a kepada Al-Khaliq.