

**IMPLEMENTASI *ITEM RESPONSE THEORY* (IRT) DALAM
PENINGKATAN AKURASI PENILAIAN KEMAMPUAN
PEMECAHAN MASALAH MATEMATIS SISWA SMP**

TESIS



**Disusun Oleh:
ADITYA PRAYOGI
NIM. P2A923011**

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

**IMPLEMENTASI *ITEM RESPONSE THEORY* (IRT) DALAM
PENINGKATAN AKURASI PENILAIAN KEMAMPUAN
PEMECAHAN MASALAH MATEMATIS SISWA SMP**

TESIS

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



Disusun Oleh:

ADITYA PRAYOGI

NIM. P2A923011

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

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

Tesis yang berjudul "Implementasi *Item Response Theory (IRT)* dalam Peningkatan Akurasi Penilaian Kemampuan Pemecahan Masalah Matematis Siswa SMP" merupakan Tesis Program Studi Magister Pendidikan Matematika Universitas Jambi yang disusun oleh Aditya Prayogi, Nomor Induk Mahasiswa P2A923011 telah di periksa oleh pembimbing.

Jambi, 2025
Pembimbing I



Dr. Dra. Mujahidawati, M.Si
NIP. 19641120 199001 2 001

Jambi, 2025
Pembimbing II



Dr. Ilham Falah, S.Pd., M.Si.
NIP. 19890518 202203 1 010

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Tesis yang berjudul "Implementasi *Item Response Theory* (IRT) Dalam Peningkatan Akurasi Penilaian Kemampuan Pemecahan Masalah Matematis Siswa SMP" Tesis Program Studi Magister Pendidikan Matematika yang disusun oleh Aditya Prayogi, Nomor Induk Mahasiswa P2A923011 telah dipertahankan di depan Dewan Penguji pada tanggal 04 Juli 2025.

Tim Penguji

Ketua	Dr. Dra. Mujahidawati, M.Si. NIP. 196411201990012001
Sekretaris	Dr. Ilham Falani, S.Pd., M.Si. NIP. 198905182022031010
Anggota	1. Prof. Dr. Drs. Syaiful, M.Pd NIP.195906011991021001 2. Dr. Eka Sastrawati, S.Pd., M.Pd NIP. 198112272006042012 3. Dr. Fiki Alghadari, S.Pd., M.P.d NIP. 198905242024061001

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 bertandatangan di bawah ini,

Nama : Aditya Prayogi
NIM : P2A923011
Program Studi : Magister (S2) Pendidikan Matematika

Dengan ini menyatakan dengan sesungguhnya bahwa tesis ini benar – benar karya sendiri dan bukan merupakan jiplakan dari hasil penelitian pihak lain. Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa tesis ini merupakan jiplakan atau plagiat, saya bersedia menerima sanksi dicabut gelar dan di tarik ijazah,

Demikian pernyataan ini dibuat dengan penuh kesadaran dan tanggung jawab.

Jambi, Juli 2025
Yang membuat pernyataan



Aditya Prayogi
NIM.P2A923011

ABSTRAK

Aditya Prayogi, 2025. Implementasi *Item Response Theory* (IRT) Dalam Akurasi Penilaian Kemampuan Pemecahan Masalah Matematika Siswa SMP, Program Pasca Sarjana Universitas Jambi, Pembimbing: (I) Dr. Dra. Mujahidawati, M.Si, (II) Dr. Ilham Falani, S.Pd., M.Si

Kata kunci: GPCM, GRM, IRT, kemampuan Pemecahan Masalah, PCM, Penilaian

Permasalahan dalam penilaian kemampuan pemecahan masalah matematis siswa di tingkat SMP masih diwarnai oleh penggunaan instrumen yang belum akurat dan belum sepenuhnya berdasarkan pendekatan pengukuran modern. Penilaian cenderung bersifat subjektif dan tidak mampu menangkap tahapan berpikir siswa secara menyeluruh, terutama dalam konteks data politomi yang muncul dari soal berbasis rubrik. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi penilaian kemampuan pemecahan masalah matematika siswa dengan mengimplementasikan pendekatan *Item Response Theory* (IRT) melalui tiga model, yaitu *Partial Credit Model* (PCM), *Generalized Partial Credit Model* (GPCM), dan *Graded Response Model* (GRM).

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Ex Post Facto*. Sampel penelitian adalah seluruh siswa kelas IX dari empat SMP di Kecamatan Bahar Utara, Kabupaten Muaro Jambi. Instrumen yang digunakan berupa tes esai berbasis rubrik yang telah divalidasi dengan indeks Aiken dan diuji reliabilitasnya menggunakan Alpha Cronbach. Data dianalisis dengan bantuan perangkat lunak PARSCALE dan Wingen untuk estimasi parameter kemampuan dan butir.

Hasil penelitian menunjukkan bahwa ketiga model IRT layak digunakan berdasarkan pengujian asumsi unidimensi dan kecocokan model. Di antara ketiganya, GPCM memberikan estimasi kemampuan yang paling akurat, dengan varians estimasi yang paling rendah dan fungsi informasi tertinggi dibandingkan PCM dan GRM. Hal ini menunjukkan bahwa GPCM lebih sensitif dalam membedakan kemampuan siswa berdasarkan respon parsial yang diberikan pada soal politomi.

Implikasi dari temuan ini menunjukkan bahwa penggunaan model IRT, khususnya GPCM, dalam penilaian kemampuan pemecahan masalah memberikan kontribusi signifikan dalam meningkatkan keakuratan pengukuran kemampuan siswa. Dengan demikian, hasil penilaian dapat lebih dipercaya sebagai dasar pengambilan keputusan pendidikan dan sebagai acuan dalam merancang intervensi pembelajaran yang tepat sasaran.

ABSTRACT

Aditya Prayogi, 2025. *Implementation of Item Response Theory (IRT) in the Accuracy of Assessment of Mathematical Problem-Solving Ability of Junior High School Students*, Postgraduate Program, University of Jambi, Supervisor: (I) Dr. Dra. Mujahidawati, M.Si, (II) Dr. Ilham Falani, S.Pd., M.Si

Keywords: *Assessment, GPCM, GRM, IRT, Problem-Solving Ability, PCM,*

Problems in assessing students' mathematical problem-solving ability at the junior high school level are still colored by the use of instruments that are not accurate and not fully based on modern measurement approaches. Assessment tends to be subjective and is not able to capture the stages of student thinking as a whole, especially in the context of polytomy data that arise from rubric-based questions. The purpose of this study is to improve the accuracy of students' assessment of mathematical problem-solving ability by implementing the Item Response Theory (IRT) approach through three models, namely Partial Credit Model (PCM), Generalized Partial Credit Model (GPCM), and Graded Response Model (GRM).

This study uses a quantitative approach with the Ex Post Facto method. The research sample was all grade IX students from four junior high schools in North Bahar District, Muaro Jambi Regency. The instrument used is a rubric-based essay test that has been validated with the Aiken index and tested for reliability using Alpha Cronbach. The data is analyzed with the help of PARSCALE and Wingen software for the estimation of capability parameters and items.

The results of the study show that the three IRT models are feasible to use based on testing one-dimensional assumptions and model suitability. Among the three, GPCM provides the most accurate estimation of capabilities, with the lowest estimation variance and the highest information function compared to PCM and GRM. This shows that GPCM is more sensitive in distinguishing students' abilities based on the partial responses given to the polytomy question.

The implications of these findings suggest that the use of IRT models, particularly GPCM, in the assessment of problem-solving abilities makes a significant contribution to improving the accuracy of student ability measurements. Thus, the results of the assessment can be more trusted as a basis for educational decision-making and as a reference in designing targeted learning interventions.

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 “**Implementasi *Item Response Theory (IRT)* Dalam Peningkatan Akurasi Penilaian Kemampuan Pemecahan Masalah Matematis Siswa SMP**”. 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.
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 Prof. Dr. Drs. Syaiful, M.Pd selaku dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.
7. Ibu. Dr. Eka Sastrawati, S.Pd., M.Pd selaku dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.

8. Bapak Dr. Fiki Alghadari, S.Pd., M.Pd selaku dosen penguji yang telah memberikan arahan dan masukan yang berharga dalam penyusunan Tesis ini.
9. 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.
10. 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.
11. Secara Khusus untuk kedua Orang Tuaku tercinta Ayahanda Baihaki dan Ibunda Eni zahrina, terima kasih telah memberikan dukungan penuh atas cita-citaku, cinta, dan rasa sabarnya untuk memberikan dukungan dari memulai kuliah hingga menyelesaikan tesis ini. .
12. 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.

Akhirnya, semoga Allah ﷻ berkenan membalas segala kebaikan dan amal semua pihak yang telah membantu penulia dalam menyelesaikan tugas akhir tesis di Magister Pendidikan Matematika Universitas Jambi. Semoga tesis ini bermanfaat bagi pengembangan ilmu pengetahuan dan teknologi.

Jambi, 05 Juli 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN	v
ABSTRAK.....	vi
ABSTRACT.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah.....	7
1.3 Rumusan Masalah.....	7
1.4 Batasan Masalah	8
1.5 Tujuan Penelitian	8
1.6 Manfaat Penelitian	9
1.7 Definisi Oprasional	9
BAB II KAJIAN TEORI.....	11
2.1 Deskripsi Teoritik.....	11
2.2 Penelitian yang Relevan.....	29
2.3 Kerangka Berfikir	31
2.4 Hipotesis Penelitian	33
BAB III METODOLOGI PENELITIAN.....	34
3.1 Waktu dan Tempat Penelitian.....	34
3.2 Metode Penelitian	34
3.3 Data Penelitian	35
3.4 Populasi dan Sampel.....	35
3.5 Prosedur Penelitian	35
3.6 Teknik Analisis Data	39
BAB IV HASIL PENELITIAN DAN PEMBAHAN	44
4.1 Hasil Validasi Tes oleh Ahli Pakar	44
4.2 Hasil uji coba tes.....	45
4.3 Deskriptif Statistik Kemampuan Pemecahan Masalah	46
4.4 Hasil Uji Asumsi Prasyarat IRT	50
4.5 Estimasi Parameter Butir	53
4.6 Estimasi Parameter Kemampuan Pemecahan Masalah Siswa.....	57
4.7 Fungsi Informasi Tes.....	60
4.7 Pengujian hipotesis penelitian	64
4.8 Pembahasan Hasil Penelitian	69
BAB V PENUTUP.....	80

5.1	Simpulan	80
5.2	Implikasi	81
5.3	Saran	81
DAFTAR PUSTAKA		83

DAFTAR TABEL

Tabel	halaman
1. 1 Rekapitulasi Hasil Survei.....	3
3. 1 Rancangan Penelitian.....	34
3. 2 Biodata Ahli untuk Validasi	36
4. 1 Validasi Para Ahli.....	44
4. 2 Uji Reliabilitas.....	45
4. 3 Deskriptif Statistik Identifikasi Masalah.....	46
4. 4 Deskriptif Statistik Merumuskan strategi	47
4. 5 Deskriptif Statistik Menjalankan Strategi	48
4. 6 Deskriptif Statistik Memeriksa Kembali.....	49
4. 7 Total Variance Explained	51
4. 8 Estimasi Parameter Butir GPCM	53
4. 9 Estimasi Parameter Butir GRM	54
4. 10 Estimasi Parameter Butir PCM	55
4. 11 Distribusi Frekuensi Estimasi Parameter Kemampuan Model GPCM	57
4. 12 Distribusi Frekuensi Hasil Estimasi Kemampuan Model GRM.....	58
4. 13 Distribusi Frekuensi Hasil Estimasi Kemampuan Model PCM.....	58
4. 14 Ringkasan Deskriptif Estimasi Kemampuan Siswa	59
4. 15 Ringkasan Fungsi Informasi Tes.....	63
4. 16 Uji Normalitas.....	64
4. 17 Pengujian Hipotesis Pertama	65
4. 18 Pengujian Hipotesis Kedua	66
4. 19 Pengujian Hipotesis Ketiga.....	67
4. 20 Rekapitulasi Uji Hipotesis	68

DAFTAR GAMBAR

Gambar	halaman
2. 1 Kerangka Pikir	32
4. 1 Scree Plot	52
4. 2 Fungsi Informasi Tes PCM	60
4. 3 Fungsi Informasi Tes GPCM	61
4. 4 Fungsi Informasi Tes GRM.....	62

DAFTAR LAMPIRAN

Lampiran	halaman
1 Sintak GPCM.....	87
2 Sintak PCM.....	88
3 Sintak GRM.....	89
4 form Validasi tes Matematika	90
5 Hasil Estimasi Parameter GPCM.....	115
6 Hasil Estimasi Parameter GRM.....	116
7 Hasil Estimasi Parameter PCM.....	117

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Penilaian memiliki peran krusial dalam pendidikan sebagai alat untuk mengukur, memantau, dan meningkatkan proses serta hasil belajar peserta didik. Melalui penilaian yang tepat dan akurat, pendidik dapat memperoleh gambaran objektif mengenai pencapaian kompetensi siswa, mengidentifikasi kelemahan dan kekuatan belajar, serta merancang intervensi pembelajaran yang lebih efektif (Septikasari et al. 2023). Selain itu, penilaian berfungsi sebagai umpan balik bagi siswa untuk memahami sejauh mana penguasaan materi dan keterampilan yang telah dicapai, sekaligus memotivasi mereka dalam mencapai tujuan pembelajaran. Dalam konteks yang lebih luas, penilaian juga menjadi dasar pengambilan keputusan pendidikan, mulai dari perencanaan kurikulum hingga pengembangan kebijakan, sehingga mutunya sangat menentukan arah dan keberhasilan sistem pendidikan secara keseluruhan (Brookhart 2014).

Penilaian merupakan proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar siswa. Penilaian hasil belajar adalah bagian integral yang tidak dapat dipisahkan dari kegiatan mengajar. Penilaian hasil belajar juga berfungsi sebagai alat untuk memberikan umpan balik yang konstruktif, baik bagi siswa maupun guru, sehingga proses pembelajaran dapat terus ditingkatkan (Septiani, Pratiwi, and Rossy 2023). Proses ini bertujuan untuk mengukur efektivitas metode pengajaran, tingkat pemahaman siswa, serta pencapaian kompetensi yang telah ditetapkan dalam kurikulum.

Kemampuan pemecahan masalah adalah suatu proses pembelajaran yang mendorong siswa untuk berpartisipasi secara aktif dalam proses pembelajaran sehingga mereka dapat menerima dan merespon pertanyaan dengan baik serta mengatasi masalah dan tantangan yang muncul (Nisrina, Simatupang, and Mujahidawati 2021). Oleh karena itu perlu suatu pengukuran terkait kemampuan pemecahan siswa. Hal ini memastikan bahwa dalam

prakteknya siswa mampu memaksimalkan kemampuan pemecahan masalah untuk menghadapi tantangan yang akan mereka hadapi di kehidupan sehari-hari (Fitri, Mujahidawati, and Sabil 2025). Untuk mendapatkan gambaran yang akurat maka tingkat akurasi penilaian kemampuan pemecahan masalah siswa sangat penting.

Kemampuan pemecahan masalah adalah suatu proses pembelajaran yang mendorong siswa untuk berpartisipasi secara aktif dalam proses pembelajaran sehingga mereka dapat menerima dan merespon pertanyaan dengan baik serta mengatasi masalah dan tantangan yang muncul (Herliana, Mujahidawati, and Pasaribu 2021). Namun demikian, dalam praktiknya masih banyak persoalan yang dihadapi terkait kemampuan pemecahan masalah siswa. Beberapa siswa menunjukkan kesulitan dalam mengidentifikasi informasi penting dari soal, merumuskan strategi penyelesaian yang tepat, serta melakukan refleksi terhadap solusi yang telah diberikan. Masalah ini tidak hanya berdampak pada rendahnya hasil belajar matematika, tetapi juga menghambat pengembangan keterampilan berpikir kritis dan logis yang sangat diperlukan dalam kehidupan sehari-hari (Nuriyati and Supriadi 2022). Selain itu, guru juga menghadapi tantangan dalam menilai kemampuan ini secara objektif karena keterbatasan dalam instrumen penilaian yang tersedia, yang sering kali belum mampu mengakomodasi variasi tingkat kemampuan siswa dalam menyelesaikan masalah secara bertahap sesuai tahapan Polya. Ketiadaan instrumen yang mampu mengungkap secara komprehensif aspek-aspek dari proses berpikir siswa membuat hasil penilaian menjadi kurang akurat dan kurang bermanfaat sebagai dasar pengambilan keputusan pembelajaran (Agustina 2016).

Kemampuan pemecahan masalah merupakan salah satu indikator penting dalam pembelajaran matematika sebagaimana tercantum dalam kurikulum. Namun, berdasarkan hasil survei terhadap guru-guru SMP di Kabupaten Muaro Jambi, ditemukan beberapa temuan yang menunjukkan masih lemahnya implementasi pengukuran kemampuan pemecahan masalah secara sistematis. Seperti yang terlihat pada table 1.1 berikut

Tabel 1. 1 Rekapitulasi Hasil Survei

No	Aspek yang Disurvei	Persentase Responden (%)	Keterangan
1	Guru pernah mengukur kemampuan pemecahan masalah	80	Sebagian besar guru pernah melakukan pengukuran
2	Guru menggunakan rubrik/indikator penilaian	40	Sebagian kecil guru menggunakan indikator yang jelas
3	Guru mengetahui pendekatan pengukuran modern (IRT)	20	Sebagian besar belum mengetahui pendekatan IRT
4	Guru tertarik mendapatkan pelatihan tentang IRT	100	Seluruh responden menunjukkan ketertarikan
5	Soal 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 kemampuan pemecahan masalah 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 hanya beberapa guru yang mengaku 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.

Selain itu, sekitar 70% guru menyatakan bahwa soal tes tertulis yang mereka gunakan belum dirancang berdasarkan indikator kemampuan pemecahan masalah, yang tentunya berdampak pada keakuratan hasil penilaian. Hanya sebagian kecil guru yang mencoba memastikan validitas soal melalui diskusi informal dengan rekan sejawat, tanpa ada proses validasi yang baku.

Pentingnya seorang guru memiliki kemampuan penilaian yang baik tidak dapat diremehkan, karena ini berperan sentral dalam mengukur dan mengevaluasi kemajuan belajar siswa secara akurat dan adil. Selain itu, kemampuan penilaian yang baik membantu guru dalam mengidentifikasi kebutuhan belajar siswa secara individual, memungkinkan intervensi pendidikan yang tepat waktu dan sesuai. Dengan penilaian yang efektif, guru dapat memastikan bahwa setiap siswa mendapatkan kesempatan untuk berkembang secara optimal, serta berkontribusi terhadap peningkatan kualitas pendidikan secara keseluruhan.

Ketepatan akurasi dalam penilaian atau pengukuran kemampuan pemecahan masalah siswa memegang peranan yang sangat krusial dalam dunia pendidikan. Penilaian yang tepat dapat memberikan gambaran yang akurat tentang kemampuan siswa dalam menyelesaikan masalah, yang pada akhirnya dapat membantu pendidik dalam merumuskan strategi pembelajaran yang lebih efektif. Selain itu, penilaian yang tepat juga sangat penting untuk memastikan bahwa siswa mendapatkan bantuan yang mereka perlukan untuk mengembangkan kemampuan ini. Oleh karenanya, ketepatan akan penilaian tidak hanya membantu dalam meningkatkan standar pendidikan, tetapi juga dalam mempersiapkan siswa untuk menghadapi tantangan di masa mendatang dengan lebih baik.

Distribusi dan sebaran hasil penilaian memberikan wawasan tentang sejauh mana siswa memahami materi yang diajarkan, sehingga dapat menjadi dasar untuk menentukan apakah perlu ada perbaikan dalam proses pembelajaran. Keakuratan informasi yang diperoleh dari penilaian sangat bergantung pada instrumen yang digunakan. Oleh karena itu, soal-soal dalam

tes harus dikalibrasi untuk memastikan validitas dan reliabilitasnya. Kalibrasi soal adalah metode yang digunakan untuk memprediksi apakah suatu soal layak untuk mengukur kemampuan atau prestasi (Ul Hassan and Miller 2019). Dalam pengukuran, terdapat dua pendekatan utama yang sering digunakan untuk menganalisis butir soal, yaitu teori tes klasik (classical test theory, CTT) dan teori respons butir (item response theory, IRT) (Sarea and Ruslan 2019).

Teori tes klasik (CTT) adalah teori dasar dalam pengukuran kemampuan mental yang menggambarkan hubungan antara skor yang diamati pada tes dan skor sebenarnya yang tidak terlihat. CTT bersifat tergantung pada kelompok dan item, yang berarti bahwa indeks daya pembeda, tingkat kesulitan, dan koefisien reliabilitas tes bergantung pada siapa yang mengerjakan tes tersebut serta soal atau item yang digunakan (Retnawati, Hadi, and Nugraha 2016). Penelitian yang dilakukan oleh (Hayat 2021; Fernanda and Hidayah 2020; Ciptari, Purwanti, and Erawati 2024) 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* (teori tes modern, yang dikenal juga sebagai teori respon butir/aitem (TRA) atau *item response theory* (IRT) dan dikenal pula dengan nama *latent traits theory* (LTT). 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).

Sementara itu, teori respons butir (IRT) adalah kerangka umum dari fungsi matematika yang menjelaskan interaksi antara individu dan butir soal (Sumintono and Widhiarso 2014). IRT tidak bergantung pada sampel soal tertentu atau individu yang mengikuti ujian. Salah satu model IRT yang paling populer, diperkenalkan oleh Georg Rasch pada tahun 1960-an adalah Model Rasch. Model ini terus berkembang, tidak hanya untuk analisis dikotomi tetapi juga untuk data politomi, salah satunya dikembangkan oleh David Andrich dari Australia (Sumintono and Widhiarso 2014). Untuk analisis data politomi, model Rasch yang dikembangkan disebut Partial Credit Model (PCM). Selain PCM,

ada juga Generalized Partial Credit Model (GPCM), (Muraki 1992), model IRT yang dirancang untuk menangani butir soal dengan penilaian parsial. Artinya, daripada memberikan penilaian benar/salah secara mutlak, evaluasi dapat memberikan poin parsial untuk jawaban yang menyelesaikan beberapa bagian soal dengan benar. Misalnya, soal pilihan ganda biasanya dinilai 0 poin jika salah dan 1 poin jika benar. Namun, dengan model GPCM, soal bisa memiliki 3 aspek, di mana peserta yang menjawab salah mendapatkan 0 poin, benar sepenuhnya mendapatkan 3 poin, dan 1 atau 2 poin jika hanya menjawab 1 atau 2 aspek dengan benar, tapi tidak semua.

Model IRT lainnya untuk data politomi adalah Graded Response Model (GRM). GRM juga dikenal sebagai Model Respons Kategoris Terurut karena menangani kategori politomus yang terurut, yang dapat terkait dengan item respons terkonstruksi atau respons terpilih, di mana peserta ujian diharapkan mendapatkan berbagai tingkat skor seperti 0-4 poin. Dalam hal ini, kategorinya adalah 0, 1, 2, 3, dan 4, yang diurutkan. 'Diurutkan' berarti ada urutan atau peringkat tertentu dari respons tersebut.

Penelitian ini menitikberatkan pada ketepatan penilaian kemampuan pemecahan masalah matematika dengan menggunakan pendekatan IRT untuk analisis data politomi. Penerapan IRT dapat meningkatkan ketepatan penilaian variabel laten, yang dalam penelitian ini adalah kemampuan pemecahan masalah matematika siswa.

Pendekatan IRT memiliki beberapa keunggulan yang menjadikannya pilihan yang baik untuk analisis data tes dan kuesioner. Salah satu keunggulan utamanya adalah kemampuannya untuk memprediksi data yang hilang berdasarkan pola respon yang sistematis (Kim and Wilson 2020). Model ini juga cocok untuk analisis data yang dikumpulkan menggunakan skala penilaian, skala respon tipe Likert, atau data respon lainnya dengan kategori yang berurutan. Selain itu, model ini mempertahankan karakteristik utama analisis Rasch untuk respon biner (Sözer and Kahraman 2021). Model Rasch Polytomous memberikan fleksibilitas dan ketepatan yang tinggi dalam menganalisis data dengan berbagai format respon. Oleh karena itu, penerapan

pendekatan IRT dapat menjadi langkah penting dalam meningkatkan ketepatan penilaian kemampuan pemecahan masalah matematika siswa.

Berdasarkan latar belakang yang dipaparkan di atas penelitian ini bertujuan untuk meningkatkan akurasi penilaian kemampuan pemecahan masalah pada siswa dengan mengimplemetasikan pendekatan IRT

1.2 Identifikasi Masalah

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

1. Penilaian kemampuan pemecahan masalah masih belum optimal. Hal ini baru didasarkan pada asumsi dari guru ketika siswa mampu menjawab soal cerita tetapi belum dinilai dengan menggunakan indikator yang ada.
2. Masih rendahnya pemanfaatan pendekatan pengukuran modern seperti IRT dalam penilaian kemampuan pemecahan masalah matematika.
3. Kemampuan pemecahan masalah matematika siswa masih tergolong rendah, ditandai dengan kesulitan siswa dalam memahami informasi soal, merumuskan strategi penyelesaian, serta melakukan refleksi terhadap hasil penyelesaiannya.
4. Diperlukan pendekatan yang mampu meningkatkan akurasi dalam pengukuran kemampuan pemecahan masalah, khususnya pendekatan IRT yang mampu menangani data politomi dan memberikan estimasi parameter kemampuan siswa yang lebih presisi.

1.3 Rumusan Masalah

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

1. Apa saja bentuk kesulitan yang dialami siswa dalam menyelesaikan soal pemecahan masalah matematika berdasarkan tahapan Polya?
2. Bagaimana keakuratan penilaian kemampuan pemecahan masalah matematis siswa menggunakan model PCM dibandingkan dengan model GRM?

3. Bagaimana keakuratan penilaian kemampuan pemecahan masalah matematis siswa menggunakan model GPCM dibandingkan dengan model GRM?
4. Bagaimana keakuratan penilaian kemampuan pemecahan masalah matematis siswa menggunakan model PCM dibandingkan dengan model GPCM?

1.4 Batasan Masalah

Adapun yang menjadi batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan merupakan data hasil tes kemampuan pemecahan masalah matematika untuk siswa Sekolah Menengah Pertama
2. Tes yang dilakukan berupa tes Sumatif pada materi kesebangunan dan kekongruenan.
3. Data hasil tes dianalisis dengan 3 model IRT yaitu PCM, GPCM dan GRM.
4. Penelitian ini hanya berfokus pada hasil estimasi parameter kemampuan pemecahan masalah matematika siswa yang dihasilkan oleh ketiga model IRT tersebut

1.5 Tujuan Penelitian

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

1. Untuk mengidentifikasi bentuk-bentuk kesulitan yang dialami siswa dalam menyelesaikan soal pemecahan masalah matematika berdasarkan tahapan Polya.
2. Mengetahui keakuratan penilaian kemampuan pemecahan masalah matematis siswa menggunakan model PCM dibandingkan dengan model GPCM
3. Mengetahi keakuratan penilaian kemampuan pemecahan masalah matematis siswa menggunakan model PCM dibandingkan dengan model GPCM.

4. Mengetahui keakuratan penilaian kemampuan pemecahan masalah matematis siswa menggunakan model PCM dibandingkan dengan model GPCM.

1.6 Manfaat Penelitian

Adapun manfaat penelitian ini:

1. Hasil penelitian ini diharapkan dapat dijadikan manfaat bagi siswa. Dimana siswa memperoleh informasi tentang kemampuan pemecahan masalah matematis.
2. Hasil penelitian ini diharapkan dapat menjadi bahan bagi guru agar mempermudah dalam melakukan penilaian terhadap kemampuan siswa
3. Hasil penelitian ini diharapkan dapat memberikan sumbangan berupa kajian teori tentang kemampuan pemecahan masalah matematika serta pengaplikasian rasch model untuk penelitian selanjutnya dalam bidang matematika.

1.7 Definisi Operasional

1. Akurasi

Akurasi dalam konteks penelitian ini merujuk pada sejauh mana model IRT (dalam hal ini GPCM) mampu memberikan estimasi kemampuan siswa yang mendekati nilai kemampuan sesungguhnya.

2. Penilaian

Penilaian dalam penelitian ini dilakukan secara kuantitatif dengan menggunakan rubrik penskoran berdasarkan indikator dari setiap tahap Polya.

3. Kemampuan Pemecahan Masalah

Yang dimaksud dengan kemampuan pemecahan masalah matematika dalam penelitian ini adalah kemampuan siswa dalam menyelesaikan soal-soal matematika berdasarkan tahapan pemecahan masalah menurut Polya, yaitu: (1) memahami masalah, (2) merencanakan penyelesaian, (3) melaksanakan rencana, dan (4) melakukan pemeriksaan kembali (*review/refleksi*). Kemampuan ini diukur melalui soal berbasis masalah

(*problem solving*) pada topik tertentu (misalnya SPLDV) dan dianalisis secara kuantitatif berdasarkan respons siswa terhadap butir-butir soal tersebut.

4. Item Response Teory

Item Response Theory dalam penelitian ini merujuk pada pendekatan statistik yang digunakan untuk memodelkan hubungan antara kemampuan laten siswa (θ) dan probabilitas menjawab item dengan benar.

5. Item

Item mengacu pada satuan pertanyaan yang dirancang untuk menilai kemampuan Pemecahan Masalah. Setiap item dapat terdiri dari satu atau beberapa subitem yang merepresentasikan indikator tertentu dari keterampilan yang diukur. Dengan demikian, subitem berperan sebagai bagian analitis yang mencerminkan aspek kognitif atau prosedural spesifik 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 peserta didik 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 peserta didik (PP No.19 Th 2005). Melalui penilaian, pendidik dapat menilai sejauh mana siswa memahami materi, menguasai kemampuan, dan mampu menerapkan pengetahuan mereka dalam situasi nyata.

Penilaian merupakan bagian integral dari proses pendidikan yang berfungsi untuk mengetahui sejauh mana tujuan pembelajaran telah tercapai. Penilaian tidak hanya berfungsi sebagai alat untuk mengukur hasil belajar siswa, tetapi juga sebagai sarana untuk memperbaiki proses pembelajaran secara keseluruhan. Menurut Brookhart (2014), penilaian dalam pendidikan adalah proses sistematis untuk mengumpulkan, menganalisis, dan menginterpretasi informasi guna mengetahui sejauh mana siswa telah mencapai tujuan belajar.

Salah satu aspek penting dari penilaian adalah validitas dan reliabilitas instrumen. Validitas menunjukkan sejauh mana suatu instrumen benar-benar mengukur apa yang seharusnya diukur, sedangkan reliabilitas menunjukkan sejauh mana hasil pengukuran konsisten dari waktu ke waktu. Kedua aspek ini sangat menentukan kualitas keputusan yang dihasilkan dari suatu penilaian (Nitko 1996).

Selain itu, penilaian juga memiliki fungsi diagnostik, yaitu untuk mengidentifikasi kesulitan belajar siswa dan memberikan intervensi yang sesuai. Penilaian yang baik akan memberikan gambaran objektif tentang

kekuatan dan kelemahan siswa, serta membantu guru dalam merancang pembelajaran yang lebih efektif dan personal (Black and Wiliam 1998).

Akurasi penilaian merujuk pada sejauh mana suatu instrumen dapat mengukur konstruk yang dimaksud secara valid dan reliabel. Validitas mengacu pada ketepatan instrumen dalam mengukur aspek yang seharusnya diukur, sementara reliabilitas menyangkut konsistensi hasil ukur dalam berbagai kesempatan pengukuran (Brookhart 2014). Dalam konteks penilaian kemampuan pemecahan masalah, akurasi menjadi hal penting karena kemampuan ini tidak hanya mencakup hasil akhir, tetapi juga mencerminkan proses berpikir, strategi yang digunakan, serta refleksi terhadap solusi yang diambil.

Dengan berkembangnya ilmu pengukuran dan evaluasi, pendekatan penilaian juga mengalami evolusi. Dari yang awalnya menggunakan teori tes klasik (Classical Test Theory/CTT), kini banyak beralih ke pendekatan teori respons butir (Item Response Theory/IRT). IRT memungkinkan analisis butir soal yang lebih akurat dan independen terhadap sampel peserta tes, serta memberikan estimasi kemampuan siswa (θ) yang lebih presisi (Hambleton, Swaminathan, and Rogers 1991).

2.1.2 Teori Belajar Dan Implikasinya Terhadap Akurasi Penilaian

Akurasi penilaian merupakan aspek krusial dalam proses pembelajaran karena menentukan sejauh mana informasi yang diperoleh dari hasil tes mencerminkan kemampuan siswa secara nyata. Dalam pendekatan konstruktivisme yang dikembangkan oleh Piaget dan Bruner, proses belajar dipandang sebagai aktivitas membangun pengetahuan secara aktif berdasarkan pengalaman dan interaksi (Slavin 2018). Oleh karena itu, penilaian yang akurat tidak hanya mengukur jawaban benar atau salah, tetapi juga harus menangkap proses berpikir siswa, termasuk strategi dan refleksi yang digunakan dalam menyelesaikan masalah (Anderson and Krathwohl 2001).

Teori Taksonomi Bloom yang direvisi oleh Anderson dan Krathwohl (2001) menekankan pentingnya pengukuran terhadap kemampuan berpikir

tingkat tinggi seperti menganalisis, mengevaluasi, dan mencipta. Oleh karena itu, penilaian yang akurat harus memiliki instrumen yang mampu mengidentifikasi tingkat kognitif siswa secara valid dan reliabel. Ini sejalan dengan prinsip validitas konstruk dalam penilaian, yang menyatakan bahwa suatu instrumen harus benar-benar mengukur apa yang dimaksudkan untuk diukur (Nitko 1996).

Dalam teori pemrosesan informasi yang dikemukakan oleh Newell and Simon (1972), pemecahan masalah dipandang sebagai proses pencarian solusi dari kondisi awal ke kondisi tujuan melalui serangkaian langkah sistematis. Akurasi penilaian dalam konteks ini ditentukan oleh kemampuan instrumen untuk merekam jejak berpikir siswa secara bertahap, bukan hanya jawaban akhir. Model pengukuran seperti Generalized Partial Credit Model (GPCM) dalam kerangka Item Response Theory (IRT) dinilai tepat karena memungkinkan pemberian skor parsial yang mencerminkan proses berpikir tersebut (Muraki 1992; Embretson and Reise 2000).

Di sisi lain, pendekatan behavioristik seperti yang dikembangkan oleh Thorndike dan Skinner lebih menitikberatkan pada penguatan terhadap jawaban benar melalui stimulus, sehingga fokus utamanya adalah pada hasil akhir (Slavin 2018). Meskipun dapat meningkatkan reliabilitas pengukuran, namun validitas konstruksinya cenderung terbatas karena tidak menangkap proses kognitif yang mendalam.

Berdasarkan berbagai teori tersebut, dapat disimpulkan bahwa akurasi penilaian sangat dipengaruhi oleh pemilihan model pengukuran yang sesuai dengan teori belajar yang mendasari proses pembelajaran. Ketika pembelajaran menekankan proses berpikir tingkat tinggi, maka penilaian pun harus mampu menangkap dinamika proses tersebut dengan model pengukuran yang mendalam dan presisi seperti IRT.

Model teori belajar memiliki pengaruh signifikan terhadap pendekatan dan alat penilaian yang digunakan. Misalnya, dalam pandangan behavioristik, penilaian lebih berfokus pada penguatan respon yang benar, sehingga bentuk tes yang digunakan cenderung bersifat objektif dan berbasis hasil akhir (Slavin 2018). Model ini dapat meningkatkan reliabilitas, tetapi berisiko rendah dalam validitas konstruk, terutama jika yang dinilai adalah proses kognitif kompleks seperti pemecahan masalah (Anderson and Krathwohl 2001).

Model IRT memberikan kontribusi penting terhadap akurasi penilaian karena memperhitungkan karakteristik item dan kemampuan siswa secara simultan (Embretson and Reise 2000). Dalam model ini, kemampuan pemecahan masalah dapat dipetakan lebih akurat dalam skala kontinu (logit), serta dilengkapi dengan informasi seperti standard error of measurement (SEM) pada setiap titik kemampuan (θ). Hal ini memungkinkan pendidik mengetahui tingkat ketidakpastian dalam estimasi skor siswa secara lebih presisi (F. B. Baker and Kim 2017).

Misalnya, dalam Partial Credit Model (PCM) atau Generalized Partial Credit Model (GPCM), skor parsial dari setiap indikator pemecahan masalah (seperti memahami masalah, merancang strategi, melaksanakan rencana, dan mengevaluasi hasil) dapat dianalisis secara lebih mendalam. Setiap indikator dinilai berdasarkan kontribusinya terhadap tingkat kemampuan keseluruhan, sehingga akurasi penilaian menjadi lebih tajam dan sesuai dengan pendekatan teori belajar yang menekankan proses (Masters 1982; Muraki 1992).

Dengan demikian, akurasi penilaian kemampuan pemecahan masalah tidak bisa dilepaskan dari landasan teori belajar yang digunakan dalam proses pembelajaran. Ketika pendekatan konstruktivistik diterapkan, maka penggunaan model penilaian yang mampu menangkap proses berpikir menjadi penting untuk menjaga validitas dan keakuratan hasil asesmen. Hal ini menunjukkan bahwa keselarasan antara teori belajar dan model pengukuran merupakan kunci utama dalam menghasilkan informasi asesmen yang bermakna dan dapat dipertanggungjawabkan.

2.1.3 Kemampuan Pemecahan Masalah

Masalah merupakan suatu kondisi ketika terdapat kesenjangan antara keadaan yang diharapkan dengan keadaan nyata, dan cara untuk menjembatani kesenjangan tersebut belum diketahui secara pasti. Dalam konteks kehidupan sehari-hari maupun dunia pendidikan, masalah menjadi stimulus penting yang mendorong berpikir kritis dan kreatif. Oleh karena itu, kemampuan pemecahan masalah (*problem solving*) menjadi kompetensi esensial dalam pengembangan individu abad ke-21.

Kemampuan pemecahan masalah adalah kapasitas individu untuk mengidentifikasi masalah, merumuskan dan mengembangkan strategi, serta mengeksekusi solusi yang efektif. Menurut Mayer (1992), *problem solving* adalah proses kognitif yang melibatkan penemuan kombinasi tindakan yang dapat mengarahkan dari kondisi awal menuju kondisi tujuan. Pemecahan masalah tidak hanya membutuhkan pengetahuan konseptual tetapi juga keterampilan berpikir tingkat tinggi.

Maesuri (2002) menekankan pentingnya kemampuan pemecahan masalah dalam pembelajaran matematika, dengan menyatakan bahwa siswa harus dihadapkan pada situasi dunia nyata yang membutuhkan kemampuan ini. Gagne dalam (Mulyasa 2008) juga menekankan bahwa pemecahan masalah tidak hanya tentang menemukan solusi, tetapi juga tentang proses pembelajaran yang menghasilkan pemahaman baru. Oleh karena itu, pemecahan masalah menjadi elemen kunci dalam pembelajaran matematika di semua tingkatan, karena memungkinkan pembelajaran yang lebih fleksibel dan menantang (Mulyasa 2008)

Gagne (dalam Ruseffendi, 1980) menyatakan bahwa pemecahan masalah adalah bentuk pembelajaran yang paling tinggi dan kompleks. Wilson (1993) menekankan pentingnya pemecahan masalah dalam pembelajaran matematika, sementara Anderson (1985) menunjukkan bahwa metode pemecahan masalah memiliki sifat heuristik yang dapat diterapkan untuk

menyelesaikan berbagai masalah. Oleh karena itu, pemecahan masalah tidak hanya penting dalam matematika, tetapi juga menjadi tujuan utama dalam pendidikan secara umum (Orton, 1992).

Proses pemecahan masalah dimulai dengan mengidentifikasi masalah, di mana Rosyada (2007) menekankan pentingnya mengidentifikasi masalah yang relevan dengan fokus yang akan dicari melalui penemuan, kajian, dan penelitian mendalam. Karena tidak semua masalah dapat diselesaikan, siswa diarahkan untuk memilih satu masalah yang dapat dijadikan fokus pembahasan. Setelah masalah ditetapkan, berbagai strategi yang dapat digunakan untuk menyelesaikannya dikaji. Setelah masalah diidentifikasi, langkah berikutnya adalah memilih strategi yang tepat untuk menyelesaikannya. Namun, pemecahan masalah juga bergantung pada pengetahuan tentang substansi masalah, seperti pemahaman terhadap inti masalah, langkah-langkah yang tepat, dan aturan yang relevan (Polya, 1981; Anderson, 1985). Oleh karena itu, pemecahan masalah tidak hanya melibatkan kemampuan kognitif, tetapi juga memerlukan pemahaman yang mendalam tentang masalah yang dihadapi.

Dalam konteks pendidikan, terutama pembelajaran matematika, berbagai teori telah dikembangkan untuk memahami bagaimana siswa memecahkan masalah dan bagaimana kemampuan ini dapat dikembangkan serta dinilai secara efektif. Berikut ini adalah beberapa teori utama yang menjadi landasan dalam kajian kemampuan pemecahan masalah

1. Teori Tahapan Polya (1945)

Polya membagi proses pemecahan masalah ke dalam empat tahapan utama, yaitu:

- 1) Memahami masalah: Mengidentifikasi apa yang diketahui, apa yang ditanyakan, serta informasi apa yang tersedia.
- 2) Membuat rencana: Menentukan strategi penyelesaian, seperti menggunakan pola, membuat gambar, atau menyusun persamaan.

- 3) Melaksanakan rencana: Menerapkan strategi yang dipilih untuk memperoleh solusi.
- 4) Melakukan pengecekan kembali: Mengevaluasi hasil dan proses untuk memastikan kebenaran solusi.

2. Teori Metakognitif Flavell (1976)

Flavell menyatakan bahwa metakognisi adalah kesadaran dan pengaturan terhadap proses berpikir sendiri. Dalam pemecahan masalah, metakognisi meliputi, Perencanaan strategi sebelum menyelesaikan masalah, Pemantauan proses selama pengerjaan, evaluasi terhadap hasil dan proses yang dilakukan. Siswa yang memiliki metakognisi tinggi cenderung lebih efisien dan reflektif dalam memecahkan masalah.

3. Teori Pemrosesan Informasi (Newell & Simon, 1972)

Newell dan Simon memandang pemecahan masalah sebagai suatu proses pencarian dari keadaan awal (*initial state*) ke tujuan (*goal state*) dengan menggunakan strategi atau aturan tertentu (*operators*). Model ini menekankan bahwa strategi berpikir efektif sangat penting dalam mencapai solusi optimal.

4. Teori Skema Pengetahuan (Anderson, 1983)

Anderson menyatakan bahwa seseorang menyelesaikan masalah dengan cara mengaktifkan skema atau struktur pengetahuan yang telah tersimpan dalam ingatan jangka panjang. Kesuksesan dalam pemecahan masalah bergantung pada kecocokan antara skema yang dimiliki dengan karakteristik masalah yang dihadapi.

2.1.4 Teori Tes Klasik (*classical test theory*)

Teori tes klasik atau *classical thruescape theory* adalah pengukuran behavior tertua di dunia. Di Indonesia, teori ini disebut sebagai teori tes klasik. Teori ini terkenal dengan kemudahannya dalam penerapan serta keefektifitasnya dalam menjelaskan bagaimana kesalahan pengukuran dapat mempengaruhi hasil yang diamati (Sarea and Ruslan 2019).

Teori klasik berusaha menjelaskan kesalahan dalam pengukuran menggunakan model berdasarkan koefisien korelasi. Charle Spearman yang menemukan koefisien korelasi ini, menggunakan dua komponen utama untuk menjelaskan kesalahan : korelasi yang sebenarnya dan korelasi yang diamati.

Teori tes klasik mendasarkan pada konsep bahwa skor yang diamati merupakan hasil dari penjumlahan skor sebenarnya dan kesalahan pengukuran(Hardianti, Liliawati, and Tayubi 2021). Kualitas Buir soal dipengaruhi oleh tingkat kesulitan dan daya pembeda. Dalam teori klasik koefisien korelasi didasarkan pada gagasan bahwa rata rata pengukuran dari setiap kemungkinan hasil akan sama dengan pengukuran sebenarnya dalam populasi. Oleh karna itu teori tes klasik menyatakan bahwa: 1) kesalahan bersifat acak dan 2) pengukuran terdiri dari tiga komponen yaitu : indikator yang diamati, indikator hipotesis yang menunjukkan nilai populasi murni, dan konsep hipotesis yang menunjukkan tingkat ketidaksesuaian antara indikator yang diamati dan yang sebenarnya. Rumus yang digunakan adalah $X = T + E$

Menurut (Hableton dan Jons 1993)” Model tes klasik sering disebut dengan model lemah karna asumsi asumsi dari model ini cukup mudah di penuh oleh data tes”. Teori tes klasis memiliki beberapa asumsi: pertama instrumen pengukurann hanya satu dimensi; kedua kesalahan pengukuran tidak berinteraksi dengan skor sebenarnya; ketiga kesalahan pengukuran tidak berkorelasi dengan skor sebenarnya maupun dengan kesalahan pada tes lainnya untuk peserta yang sama; keempat rata rata kesalahan pengukuran adalah nol. Asumsi asumsi ini menjadi dasar dalam pengembangan rumus untuk menghitung reliabilitas tes.

2.1.5 Item Response Theory

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) (Dali S. Naga, 1992). 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 (Samsul Hadi, 2013).

IRT adalah sebuah model probabilitas yang berusaha menjelaskan hubungan antara respon seseorang terhadap sebuah butir soal dengan variabel laten (kemampuan/*ability* atau sifat/*trait*) yang diukur oleh tes tersebut (Fajrianti, Hendriani, and Septarini 2016). Dalam hal ini, respon 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 (Embretson & Reise, 2000). Variabel laten (misalnya kemampuan) dalam IRT disebut sebagai theta (θ) dimana semakin tinggi θ yang dimiliki seseorang maka semakin tinggi pula probabilitasnya untuk menjawab soal dengan benar (Baker, 2001). Probabilitas seorang peserta tes untuk menjawab benar sebuah butir merupakan fungsi dari tingkat kemampuan pemecahan masalah yang dimiliki dengan tingkat kesulitan dari butir tersebut.

Menurut Hambleton, Swaminathan, & Rogers (1991), secara umum ciri-ciri IRT adalah sebagai berikut:

1. Karakteristik butir tidak tergantung pada peserta ujian.
2. Skor yang digambarkan peserta ujian tidak tergantung pada tes.
3. Merupakan model yang lebih menekankan pada tingkat butir dari pada tingkat tes,
4. Merupakan model yang tidak mensyaratkan secara ketat tes paralel untuk menaksir reliabilitas, dan
5. Merupakan model yang menguraikan sebuah ukuran keputusan untuk tiap skor kemampuan yakni ada hubungan fungsional antara peserta tes dengan tingkat kemampuan yang dimiliki.

Salah satu aspek utama dari teori respons butir adalah probabilitas peserta tes memberikan jawaban yang benar. Parameter butir dan parameter peserta tes dihubungkan melalui suatu model matematis. Dalam model ini, kemungkinan peserta tes menjawab soal dihitung sebagai fungsi logistik dari perbedaan parameter yang dimasukkan ke dalam model tersebut.

Dalam teori respons butir, model matematis menunjukkan bahwa kemungkinan seorang subjek menjawab butir dengan benar bergantung pada tingkat kemampuan subjek dan sifat-sifat butir tersebut. Ini mengimplikasikan bahwa peserta dengan kemampuan tinggi cenderung memiliki peluang lebih besar untuk menjawab dengan benar dibandingkan dengan peserta dengan kemampuan lebih rendah. Hambleton et al. (1991) dalam (Sarea and Ruslan 2019) menyatakan bahwa ada tiga asumsi yang mendasari teori respon butir, yaitu unidimensi, independensi lokal dan invariansi parameter.

1. *Unidimensi*, artinya setiap butir tes hanya mengukur satu kemampuan. Contohnya, pada tes prestasi belajar bidang studi matematika, butir-butir yang termuat di dalamnya hanya mengukur kemampuan peserta didik dalam bidang studi matematika saja, bukan bidang yang lainnya. Pada praktiknya, asumsi unidimensi sulit dilakukan karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Oleh karena itu, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung satu komponen dominan yang mengukur prestasi subjek.
2. *Independensi lokal* menyatakan bahwa sikap kemampuan yang mempengaruhi suatu tes adalah konstan, maka respon peserta tes pada setiap butir soal adalah independen secara statistik. Sebagaimana dijelaskan oleh van der Linden & Hambleton (2013):

The term “local” in the local independence assumption is used to indicate that responses are assumed independent at the level of individual persons with the same value of, but the assumption does not generalize to the case of variation in.

Asumsi independensi lokal menyatakan bahwa tidak ada korelasi antara respon peserta tes pada butir soal yang berbeda. Hal ini menunjukkan bahwa kemampuan yang dinyatakan dalam model adalah satu-satunya faktor yang mempengaruhi respon peserta tes pada butir-butir soal. Bukti yang lain adalah peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal.

3. *Invariansi* parameter artinya bahwa karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal.

2.1.6 Penyeoran

Pemberian skor (*scoring*) merupakan langkah pertama dalam proses pengolahan 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 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*) test (Sudijono, 1996). 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 (*weithing*) 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.

Penyekoran politomi adalah pengklasifikasian yang menggambarkan respons peserta tes terhadap suatu butir, di mana respons tersebut terdiri dari lebih dari dua kategori yang mungkin. Menurut Mentasari dan Candisa (2021), Penskoran politomi adalah pemberian skor pada hasil tes yang terdiri dari dua nilai atau lebih, dimana penskorannya dilakukan langkah perlangkah dalam suatu butir soal, sehingga dalam proses analisisnya memperhitungkan tingkat kesulitan pada tiap langkah dalam menyelesaikan butir soal tersebut. Skor tertinggi tentu saja didapatkan ketika peserta tes mampu menyelesaikan dengan benar soal hingga langkah akhir.

Tingkat kategori dalam penyekoran ini ditentukan sesuai dengan kriteria yang telah dirancang oleh pembuat tes. Menurut ilham falani (2020) Semakin banyak gradasi dalam kategori, memungkinkan penyekoran ini untuk dapat mengukur kemampuan peserta tes dalam mendemonstrasikan pengetahuannya secara parsial pada tiap butir. Penyekoran politomi dapat dilakukan dengan menggunakan rubrik penilaian yang disusun oleh pembuat tes.

2.1.7 Model Respon Butir Politomi

2.1.7.1 *Partial Credit Model (PCM)*

Pemodelan *Rasch* juga terus berkembang tidak hanya untuk analisis dikotomi, namun juga untuk jenis data politomi yang salah satunya dilakukan oleh David Andrich dari Australia (Sumintono and Widhiarso 2014). Untuk analisis data politomi pemodelan Rasch yang dikembangkan dinamakan *Partial Credit Model*.

Partial Credit Model (PCM) merupakan pengembangan dari model rasch butir dikotomi yang diterapkan pada butir politomi, jika i adalah butir politomi dalam kategori skor, $0, 1, 2, \dots, m$, probabilitas dari individu n skor x pada butir i yang di wujudkan dalam persamaan berikut.

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

Keterangan:

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

θ_n = level trait individu (lokasi trait individu pada kontinum trait laten)

δ_{ij} = persimpangan garis antar kategori (j) pada Butir (i).

(Sumintono, B. 2015)

2.1.7.2 *Generalized Partial Credit Model (GPCM).*

Selain model PCM model lainnya dalam pendekatan IRT untuk data politomi adalah *Generalized Partial Credit Model (GPCM)*. Model ini disebut *Generalized Partial Credit Model* karena Muraki (1992) mengembangkan model ini sebagai bentuk umum dari *Partial Credit Model (PCM)* yang dikembangkan oleh Masters (1982) yang merupakan pengembangan dari Model Rasch (Dai et al. 2021) GPCM merupakan Model IRT yang dikembangkan untuk menganalisis data politomus, di mana respon disekor 0, 1, ..., k, dengan k adalah kategori skor tertinggi untuk butir.

Generalized Partial Credit Model (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 *Partial Credit Model (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.

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

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

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

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

$$d_{j0} = 0$$

Dengan,

$P_{jk}(\theta)$ = probabilitas peserta tes dengan kemampuan θ yang memperoleh nilai skor kategori k pada butir j.

aj = merupakan indeks daya beda pada butir ke j.

θ = 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, 2024)

2.1.7.3 Graded Response Model (GRM)

Graded Response Model. Graded Response Model (GRM) merupakan sebuah model politomi IRT yang dikembangkan untuk butir dengan karakteristik kategorinya terurut seperti skala Likert (C. Baker, Wuest, and Stern 1992; Herwin et al. 2023). 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 respon k_i . Tiap butir dideskripsikan oleh parameter a dan parameter β ambang batas (*threshold*) m_i , di mana $m_i = k - 1$ adalah jumlah kategori dalam butir dikurangi satu. Misalkan kemampuan θ , X adalah variable bilangan random yang menunjukkan respon bertingkat butir ke I dan $x = 1, \dots, k_i$ dinotasikan sebagai aktual respon. Peluang $P_{ix}(\theta) = P_i(X = x|\theta)$ dari peserta tes dengan kemampuan θ mendapatkan score x untuk butir I, dengan $x = 1, \dots, k_i$ dapat dirumuskan sebagai berikut,

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

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,

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

(Samejima, 2011)

Untuk $x = 2, \dots, k_i$. Untuk melengkapi definisi model, ditambahkan bahwa peluang untuk respon skor sama dengan atau lebih dari nilai terendah $P_{i1}^*=1$. Sedangkan peluang untuk respon sama dengan atau lebih dari nilai kategori tertinggi $P_{i(k_i+1)}^* = 0$.

2.1.8 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 respon yang telah diberikan pada setiap butir tes.

Demars (2010), Toit (2003), 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 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.

Seorang peserta tes dengan kemampuan θ 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 respon untuk semua butir. Jika asumsi independensi lokal terpenuhi, Likelihood dari dapat dihitung sebagai

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

Di mana P_{iu_i} peluang peningkatan skor u_i pada butir politomus ke i . nilai dari θ yang memaksimalkan fungsi likelihood dari θ 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,

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

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

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

(Falani and Kumala 2017)

2.1.9 Sistem Persamaan Linear Dua Variabel

2.1.9.1 Pengertian SPLDV

Sistem Persamaan Linear Dua Variabel (SPLDV) adalah dua persamaan linear yang memiliki dua variabel, biasanya x dan y , dan diselesaikan secara bersamaan.

Contoh SPLDV:

$$\begin{cases} 2x + 3y = 12 \\ x - y = 4 \end{cases}$$

2.1.9.2 Bentuk Umum SPLDV

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases}$$

dengan $a_1, a_2, b_1, b_2, c_1, c_2$ adalah bilangan real, dan x dan y adalah variabel.

2.1.9.3 Metode Penyelesaian

2.1.9.3.1 Metode Substitusi

Langkah:

- Ubah salah satu persamaan menjadi bentuk $x =$ atau $y =$
- Substitusikan ke persamaan lain

Contoh:

$$\begin{cases} x + y = 4 \\ 2x + 4y = 6 \end{cases}$$

Langkah:

- Dari persamaan (1): $x = 4 - y$
- Substitusi ke persamaan (2):

$$2(4 - y) + 4y = 6$$

$$8 - 2y + 4y = 6$$

$$2y = 6 - 8$$

$$2y = -2$$

$$y = -1$$

- Kemudian substitusi nilai y ke salah satu persamaan:

$$x + (-1) = 4$$

$$x - 1 = 4$$

$$x = 4 + 1$$

$$x = 5$$

2.1.9.3.2 Metode Eliminasi

Langkah:

- Samakan koefisien salah satu variabel
- Kurangkan atau jumlahkan kedua persamaan untuk mengeliminasi salah satu variabel

contoh

$$\begin{cases} x + y = 4 \\ 2x + 4y = 6 \end{cases}$$

- Dari persamaan 1 kita kalikan dengan 2 agar koefisien variable x sama

$$(x + y = 4) \times 2 = 2x + 2y = 8$$

- Kemudian kurangi kedua persamaan

$$(2x + 2y) - (2x + 4y) = 8 - 6$$

$$-2y = 2$$

$$y = \frac{2}{-2}$$

$$y = -1$$

- kita kalikan lagi persamaan 1 dengan 4 agar koefisien y sama

$$(x + y = 4) \times 4 = 4x + 4y = 16$$

- kemudian kurangi kedua persamaan

$$(4x + 4y) - (2x + 4y) = 16 - 6$$

$$2x = 10$$

$$x = \frac{10}{2}$$

$$x = 5$$

2.2 Penelitian yang Relevan

Penelitian terdahulu telah mencoba meneliti lebih lanjut terkait penggunaan model IRT untuk data politomus. Penelitian yang relevan dengan penelitian ini diantaranya yaitu:

Mariati 2009 melakukan penelitian yang bertujuan untuk melakukan perbandingan model IRT dengan indikator pada teori uji klasik dan memilih model IRT yang sesuai untuk memvisualisasikan karakteristik butir soal dan kemampuan peserta tes. Data yang dipakai adalah data dari Olimpiade Sains Provinsi bidang Informatika Tahun 2009, sebanyak 60 item pilihan ganda dengan sampel 500 peserta tes. Langkah yang dilakukan yaitu melakukan pendugaan karakteristik butir soal dan kemampuan peserta, serta pemeriksaan kesesuaian model. Proses analisis dilakukan dengan berbantuan *software* TAP dan Bilog-MG. Dari hasil analisis, diperoleh perbedaan antara daya pembeda dan tingkat kesukaran pada metode teori klasik dan model IRT. Daya pembeda dan tingkat kesukaran pada teori klasik dipengaruhi oleh kemampuan kelompok sedangkan pada model IRT

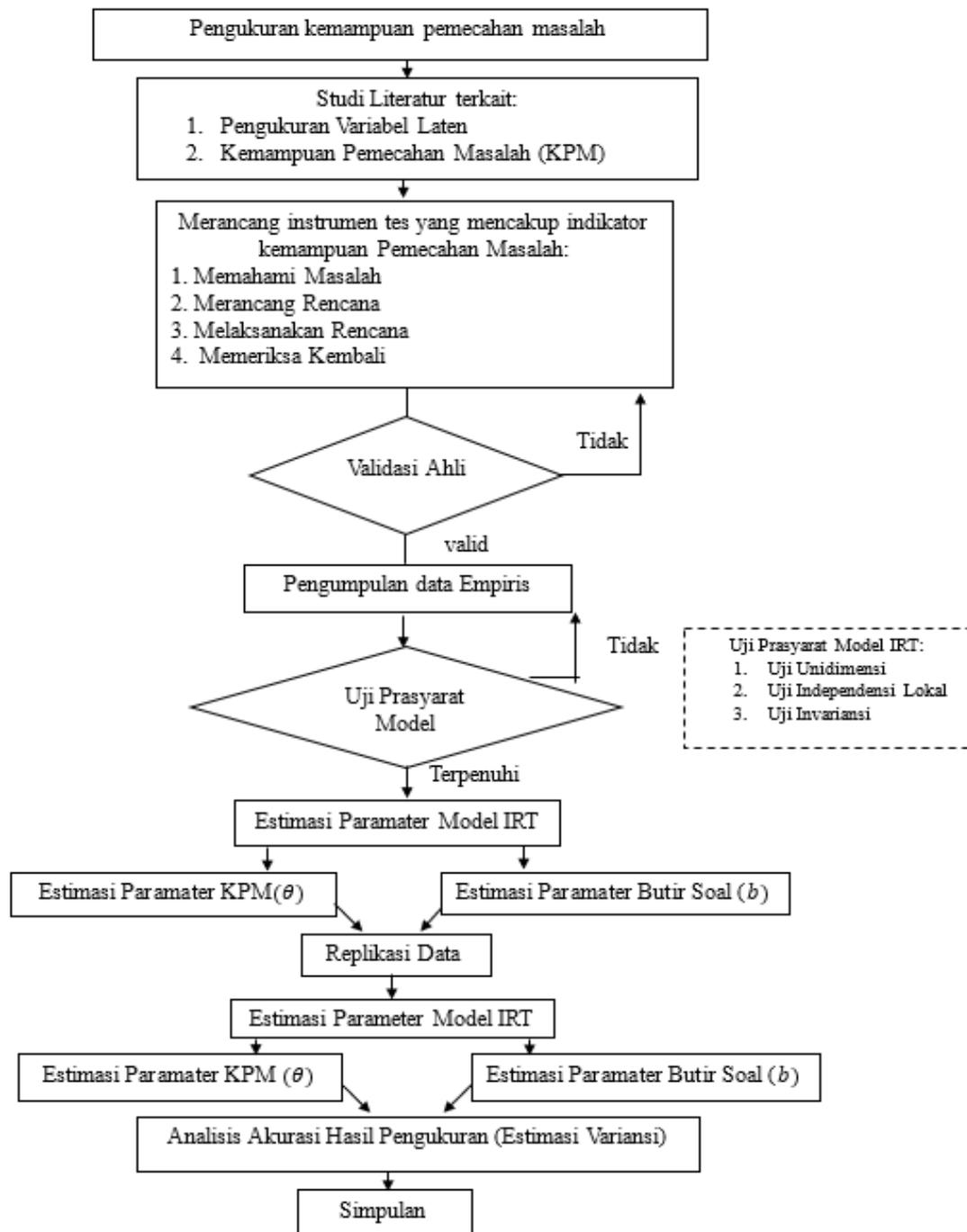
dipengaruhi oleh kemampuan individu. Dengan teori klasik, didapatkan tingkat kesukaran dari soal OSP yang diujikan tergolong sedang, dengan reliabilitas skor tes sebesar 0,910 yang menunjukkan bahwa tingkat ketepatan dan kekonsistenan peserta tes dalam menjawab soal sudah sangat baik. Adapun model yang sangat sesuai dalam menggambarkan butir-butir soal pada soal OSP yaitu model IRT tiga parameter logistik (IRT 3PL), yakni sebesar 85% butir soal sudah sesuai atau dapat digambarkan oleh model.

Falani 2020 The Precision of Students' Ability Estimation on Combinations of Item Response Theory Models. Membandingkan ketepatan estimasi kemampuan pada berbagai jenis model teori respons item untuk data format campuran. Peserta dalam penelitian ini adalah 1625 siswa SMP di Depok, Indonesia. Tes format campuran digunakan untuk mengukur kemampuan siswa dalam matematika. Tes yang digunakan terdiri dari pilihan ganda dan respons yang dikonstruksi. Item pilihan ganda dinilai secara dikotomi, sedangkan item respons yang dikonstruksi dinilai secara politomi. Selanjutnya, data respons campuran dianalisis menggunakan kombinasi model teori respons item. Penelitian ini menggunakan kombinasi Multiple-Choice Model untuk data dikotomi dan model respons bergradasi untuk data politomi (MCM+GRM). Analisis kombinasi model ini belum pernah dilakukan secara bersamaan. Data respons pengujian dianalisis menggunakan PARSCALE. Selanjutnya, hasil estimasi dibandingkan dengan hasil estimasi dari kombinasi Model Logistik 3 Parameter dan Generalized Partial Credit Model (3PLM+GPCM). Ada dua kriteria evaluasi untuk tingkat presisi estimasi: Root Mean Squared Error (RMSE) dan metode korelasi. Berdasarkan hasil yang diperoleh, estimasi nilai RMSE untuk MCM+GRM lebih kecil dari estimasi nilai RMSE dengan 3PLM+GPCM. Selain itu, hasil perkiraan kemampuan dengan MCM+GRM menghasilkan nilai korelasi yang lebih tinggi daripada 3PLM+GPCM. Jadi, dapat disimpulkan bahwa presisi tingkat model MCM+GRM lebih tinggi dari 3PLM+GPCM. Oleh karena itu, MCM+GRM lebih direkomendasikan untuk memperkirakan kemampuan matematika siswa dalam tes format campuran.

Ogunsakin and Shogbesan 2018 Item Response Theory (IRT): A Modern Statistical Theory for Solving Measurement Problem in 21st Century" Penelitian ini merupakan kajian teoritis yang mengulas berbagai model, asumsi, dan aplikasi IRT dalam mengatasi masalah pengukuran. Fokus penelitian terletak pada perbandingan IRT dengan Classical Test Theory (CTT) serta bagaimana IRT dapat digunakan untuk data kognitif dan non-kognitif. Studi literatur yang menjelaskan keunggulan IRT sebagai kerangka pengukuran dalam berbagai disiplin ilmu, termasuk pendidikan, psikologi, dan sosiolog. IRT diidentifikasi sebagai kerangka yang lebih unggul dibandingkan CTT karena menghasilkan parameter item dan kemampuan individu yang independen dari sampel pengujian. Model IRT, seperti 1PL, 2PL, dan 3PL, dapat memprediksi hubungan antara kemampuan laten siswa dan probabilitas menjawab benar suatu item. Aplikasi IRT mencakup analisis data kognitif dalam pendidikan serta pengukuran non-kognitif seperti kepribadian, sikap, dan psikopatologi.

2.3 Kerangka Berfikir

Kerangka berpikir dapat digambarkan sebagai berikut:



Gambar 2. 1 Kerangka Pikir

2.4 Hipotesis Penelitian

Hipotesis penelitian pada penelitian ini disusun sebagai berikut:

Hipotesis penelitian disusun sebagai berikut:

1. Varians hasil estimasi parameter keterampilan berpikir kritis matematis dengan model GPCM secara signifikan lebih kecil daripada varians hasil estimasi menggunakan model PCM.
2. Varians hasil estimasi parameter keterampilan berpikir kritis matematis dengan model GPCM secara signifikan lebih kecil daripada varians hasil estimasi menggunakan model GRM.
3. Varians hasil estimasi parameter keterampilan berpikir kritis matematis dengan model PCM secara signifikan lebih kecil daripada varians hasil estimasi menggunakan model GRM.

BAB III

METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Sekolah Menengah Pertama (SMP) sekecamatan Bahar Utara Kabupaten Muaro Jambi kelas IX tahun ajaran 2024/2025. Penelitian ini dilaksanakan pada bulan Januari 2024 sampai selesai

3.2 Metode Penelitian

3.2.1 Jenis Penelitian

Penelitian ini menggunakan metode *Ex Post Fakto* dimana peneliti mengamati hubungan sebab-akibat yang tidak dimanipulasi atau tidak diberi perlakuan terhadap sampel. Terdapat variabel

3.2.2 Desain penelitian

Penelitian ini didesain dengan membandingkan tingkat presisi parameter kemampuan pemecahan masalah matematis. Adapun rancangan yang digunakan dalam penelitian ini yaitu terdapatada table berikut:

Tabel 3. 1 Rancangan Penelitian

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

Keterangan:

$\sigma^2 GRM$ = nilai varians hasil estimasi parameter kemampuan pemecahan masalah matematika siswa dengan pendekatan GRM

$\sigma^2 PCM$ = nilai varians hasil estimasi parameter kemampuan pemecahan masalah matematika siswa dengan pendekatan PCM

$\sigma^2 GPCM$ = nilai varians hasil estimasi parameter kemampuan pemecahan masalah matematika siswa dengan pendekatan GPCM

3.3 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data hasil respon peserta didik 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 seluruh peserta didik kelas IX semester genap tahun 2024/2025 di Sekolah Menengah Pertama (SMP) di wilayah Kecamatan Bahar Utara Kabupaten Muaro Jambi yang terdiri dari 4 SMP Negeri. Untuk sampel menggunakan teknik total sampling yang artinya seluruh populasi dijadikan sampel.

3.5 Prosedur Penelitian

3.5.1 Penyusunan Instrumen Penelitian

Instrumen penelitian ini berupa tes esai untuk mata pelajaran matematika yang disusun berdasarkan capaian pembelajaran (CP) kurikulum Merdeka kelas 9 semester genap. Matematika adalah bidang studi yang selalu ada 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 saling terkait dan mengikuti struktur hierarkis, sehingga pemahaman yang kuat pada konsep dasar sangat penting sebelum melanjutkan ke materi yang lebih lanjut. Untuk memahami materi secara keseluruhan, siswa perlu memahami konsep-konsep prasyarat secara bertahap. Sifat ini memungkinkan penyusunan soal yang langkah penyelesaiannya 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 esai ini dibuat berdasarkan capaian pembelajaran (CP) pada fase D dan dijabarkan menjadi alur tujuan pembelajaran (ATP) yang terdiri dari 5 soal untuk mengukur kemampuan pemecahan masalah matematis siswa. Penyusunan butir tes ini merupakan hasil diskusi dan masukan dari para ahli serta guru matematika di sekolah tempat penelitian dilakukan.

Soal-soal tes yang telah disusun tidak langsung digunakan untuk pengumpulan data penelitian. Sebelum itu, beberapa langkah diambil terlebih dahulu untuk memastikan bahwa instrumen tes tersebut valid dan reliabel dalam mengukur kemampuan pemecahan masalah 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. 2 Biodata Ahli untuk Validasi

No	Nama	Bidang Keahlian	Institusi	Jabatan
1.	Dr. Syamsir Sainuddin, S. Pd, M.Pd	Matematika	Universitas Jambi	Dosen
2.	Aliffiya Teja Prasasasty, M.Pd	Matematika	Universitas Indraprasta PGRI	Dosen
3.	Yuni Kuspihana, S.Pd	Matematika	SMPN 14 Muaro Jambi	Guru
4.	Nanang Sriadi, S.Pd	Matematika	SMPN 48 Muaro Jambi	Guru
5.	Sahnar, S.Pd	Matematika	SMPN 12 Muaro Jambi	Guru

Penilaian yang dilakukan oleh para ahli tersebut meliputi: kesesuaian Capaian 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 (Lewis R, 1980, 1985; Retnawati, 2015). Perhitungan indeks Aiken dengan menggunakan rumus,

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

V adalah indeks kesepakatan ahli atau pakar mengenai validitas butir, s adalah sekor yang ditetapkan setiap penilai r dikurangi sekor terendah dalam kategori penilaian l_0 yang digunakan. n adalah banyaknya penilai, dan c adalah banyaknya kategori penilaian.

Tahap kedua adalah uji coba tes. Tes yang telah disesuaikan berdasarkan penilaian indeks Aiken diujikan kepada 28 siswa kelas 9 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. Diharapkan, tes dengan format essay yang digunakan sebagai instrumen pengumpulan data dapat berfungsi dengan optimal.

3.5.1 Penyekoran Data Hasil Tes

Setelah tes diujikan kepada sampel penelitian, hasilnya kemudian dinilai secara politomi menggunakan rubrik penilaian kemampuan pemecahan masalah yang telah disusun oleh peneliti. Butir soal diskor secara politomi

dengan empat kategori sekor yaitu 0, 1, 2, dan 3. Pemberian sekor dilakukan berdasarkan tahapan respon jawaban benar pada penyelesaian butir soal.

3.5.2 Pengolahan Data

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. Langkah berikutnya adalah memeriksa tingkat kecocokan model IRT dengan data lapangan. Kecocokan model terhadap data dievaluasi melalui teknik pemeriksaan prediksi model. Teknik ini diterapkan pada hasil uji empiris serta hasil uji harapan matematika dari model. Proses pencocokan model dimulai dengan menghitung residu mentah, yaitu selisih antara hasil uji empiris dan hasil uji harapan matematika dari model tersebut. Residu mentah kemudian diubah menjadi residu baku (z). Dengan asumsi bahwa jumlah kuadrat dari residu baku berdistribusi chi-kuadrat, pengujian kecocokan model dilakukan dengan menggunakan taraf signifikansi, di mana nilai chi-kuadrat dari residu dibandingkan dengan nilai kritis distribusi probabilitas chi-kuadrat. Jika nilai chi-kuadrat melebihi nilai kritis atau Prob nya diatas 0,05, ini menunjukkan bahwa butir tidak sesuai dengan model yang digunakan, dan sebaliknya.

Setelah dilakukan uji prasayat IRT, slenjutnya dilakukan uji kalibrasi antara data yang diolah dengandengan pendekatan IRT. Hasil kalibrasi bertujuan untuk melakukan perbandingan tingkat akurasi yang lebih detail terkait data kemampuan pemecahan masalah siswa.

Pemilihan model yang sesuai diharapkan mampu menghasilkan estimasi parameter kemampuan dengan tingkat ketelitian yang tinggi. Ketepatan ini sangat berpengaruh terhadap hasil evaluasi dan penilaian yang bertujuan untuk meningkatkan mutu pembelajaran. Dalam penelitian ini, indikator yang digunakan untuk mengukur tingkat ketelitian adalah varians dari hasil estimasi.

Tahap selanjutnya dalam pengolahan data adalah menentukan nilai fungsi informasi tes untuk setiap model. Nilai fungsi informasi tes merupakan total dari nilai fungsi informasi tiap butir soal pada tes. Oleh karena itu, nilai

fungsi informasi untuk setiap butir tes pada setiap model dihitung terlebih dahulu. Setelah itu, nilai fungsi informasi tes dihitung dengan menjumlahkan seluruh nilai fungsi informasi dari butir-butir tes. Hasil analisis nilai fungsi informasi tes dari ketiga model kemudian dibandingkan untuk mengetahui model mana yang menghasilkan nilai fungsi informasi tes 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.

3.6 Teknik Analisis Data

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

3.6.1 Uji Prasyarat Model

3.6.1.1 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. Kalau kurva karakteristik butir yang dipilih sesuai dengan data dari lapangan, maka tujuan unidimensi sudah tercapai.

3.6.1.2 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.

3.6.1.3 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, dapat dilakukan dengan bantuan software yang bernama Parscale. Perangkat lunak Parscale digunakan untuk melakukan pengujian prasyarat dalam model IRT. Parscale merupakan salah satu program analisis yang berbasis pada teori respons butir (*Item Response Theory*), yaitu pendekatan dalam psikometri yang digunakan untuk mengevaluasi validitas dan reliabilitas suatu alat ukur. Penggunaan Parscale memungkinkan estimasi parameter yang lebih akurat serta hasil pengukuran yang stabil dan dapat direplikasi. Selain itu, perangkat ini mampu mengatasi permasalahan pada data yang tidak lengkap. Salah satu keunggulan utama dari Parscale adalah kemampuannya menghasilkan skala pengukuran dengan interval yang setara (linier), yang sangat membantu dalam proses interpretasi hasil. Dalam praktik penelitian, Parscale banyak dimanfaatkan untuk menganalisis kemampuan kognitif atau sikap responden, serta dalam proses validasi alat ukur seperti kuesioner maupun tes. Kemudahan penggunaan dan keakuratan output menjadikan Parscale sebagai salah satu perangkat analisis yang banyak digunakan dalam studi-studi berbasis IRT.

Setelah pengujian prasyarat model IRT terpenuhi, selanjutnya akan dilakukan replikasi. Replikasi dilakukan dengan bantuan *software* yang

bernama Wingen. Wingen merupakan aplikasi untuk menghasilkan replikasi atau tiruan data hasil analisis IRT (Falani and Kumala 2017). Wingen membantu pengguna untuk menjalankan simulasi pengujian sehingga menghasilkan data yang diperlukan untuk analisis IRT dengan lebih efisien. Hal ini tentu mempermudah proses perolehan data yang baru tanpa harus turun lapangan untuk melakukan pengujian instrumen tes secara langsung. 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 kemampuan peserta yang mengikuti tes. Dengan demikian, Wingen mampu memberikan kontribusi yang signifikan terhadap pengembangan metode pengukuran yang lebih baik untuk berbagai aplikasi penelitian.

3.6.2 Estimasi Parameter Butir

Estimasi parameter butir dilakukan untuk mendapatkan informasi mengenai karakteristik masing-masing soal. Parameter yang diestimasi tergantung pada model IRT yang digunakan. Pada model PCM dan GRM, parameter threshold antar kategori diestimasi. Pada model GPCM, selain threshold, juga diestimasi parameter diskriminasi (slope). Estimasi ini dilakukan dengan perangkat lunak berupa Parscale.

3.6.3 Estimasi Parameter Kemampuan

Estimasi parameter kemampuan (θ) siswa dilakukan untuk mengetahui letak kemampuan siswa dalam skala logit berdasarkan hasil jawaban terhadap butir soal. Estimasi ini bersifat individual dan tidak tergantung pada bentuk distribusi skor total. Nilai θ digunakan untuk membandingkan ketepatan estimasi antar model

3.6.4 *Standar Error of Measurement*

Standard Error of Measurement (SEM) menggambarkan tingkat ketidakpastian atau kesalahan pengukuran dalam estimasi kemampuan siswa.

SEM dihitung pada setiap titik estimasi theta dan digunakan untuk membandingkan ketepatan antar model. Model yang menghasilkan SEM lebih kecil dianggap lebih akurat

3.6.5 Uji Persyaratan Hipotesis

Sebelum melakukan pengujian hipotesis penelitian, diperlukan uji prasyarat pengujian hipotesis, 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 (Kadir, 2017). Uji normalitas pada penelitian ini menggunakan SPSS 27.

3.7 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 $F_{\text{hitung}} > F_{\text{tabel}}$ maka H_0 ditolak (Riduwan, 2003)

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.

$$\text{Jika } \sigma_x^2 = \sigma_y^2 = \sigma$$

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

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

$$\text{Jika } \sigma_x^2 \neq \sigma_y^2$$

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

Selanjutnya nilai t_{hitung} tersebut kemudian dibandingkan dengan t_{tabel} (n_1+n_2-2). Kriteria uji yang digunakan adalah tolak H_0 jika $t_{hitung} > t_{tabel}$ atau $t_{hitung} < -t_{tabel}$ (Riduwan, 2004).

3.8 Hipotesis Statistik Data Empirik

Hipotesis statistik penelitian ini adalah:

1. $H_0 = \sigma^2_{PCM} = \sigma^2_{GRM}$ $H_1 = \sigma^2_{PCM} < \sigma^2_{GRM}$
2. $H_0 = \sigma^2_{GPCM} = \sigma^2_{GRM}$ $H_1 = \sigma^2_{GPCM} < \sigma^2_{GRM}$
3. $H_0 = \sigma^2_{GPCM} = \sigma^2_{PCM}$ $H_1 = \sigma^2_{GPCM} < \sigma^2_{PCM}$

Keterangan:

σ^2_{PCM} = nilai varians hasil estimasi parameter kemampuan dengan PCM

σ^2_{GPCM} = nilai varians hasil estimasi parameter kemampuan dengan GPCM

σ^2_{GRM} = nilai varians hasil estimasi parameter kemampuan dengan GRM

BAB IV

HASIL PENELITIAN DAN PEMBAHAN

4.1 Hasil Validasi Tes oleh Ahli Pakar

Validasi tes dalam penelitian, peneliti memberikan 3 buah soal matematika dalam bentuk esai yang telah disusun kepada para ahli di bidang matematika.

Para ahli memberikan berbagai masukan dan saran guna penyempurnaan tes yang mencakup beberapa aspek, yaitu: 1) keterkaitan antara kompetensi dasar dengan indikator, 2) keselarasan antara indikator dan butir soal, 3) kejelasan dalam redaksi soal, 4) penggunaan bahasa yang sesuai dengan kaidah baku, dan 5) kecocokan antara soal dan sistem penskoran. Berdasarkan masukan dan saran tersebut, peneliti kemudian melakukan revisi terhadap instrumen tes.

Pada tahap berikutnya, peneliti memberikan tes yang telah direvisi berdasarkan masukan dan saran sebelumnya kepada para ahli untuk dilakukan penilaian. Para ahli diminta memberikan penilaian secara objektif terhadap setiap butir soal dengan mempertimbangkan lima aspek yang telah dijelaskan sebelumnya. Penilaian dilakukan dengan sistem skor sebagai berikut: skor 1 diberikan jika hanya satu kriteria terpenuhi, skor 2 untuk dua kriteria, skor 3 untuk tiga kriteria, skor 4 untuk empat kriteria, dan skor 5 jika seluruh kriteria terpenuhi. Hasil rekapitulasi penilaian dari para ahli disajikan pada Tabel 4.1 berikut.

Tabel 4. 1 Validasi Para Ahli

Butir	Pakar					Nilai V	Nilai V Tabel	Keterangan
	1	2	3	4	5			
1	5	5	5	4	5	0,95	0,8	Valid
2	5	4	5	5	4	0,9	0,8	Valid
3	5	4	4	5	5	0,9	0,8	Valid

Berdasarkan hasil penilaian butir oleh para ahli yang disajikan dalam tabel di atas, proses validasi dilakukan dengan menggunakan metode Aiken. Nilai V untuk masing-masing butir ditampilkan pada kolom ketujuh. Selanjutnya, nilai V

tersebut dibandingkan dengan nilai minimum yang ditetapkan berdasarkan V_{tabel} , yaitu 0,8, sebagaimana terlihat pada kolom kedelapan. Hasil perbandingan pada Tabel 4.1 menunjukkan bahwa seluruh butir memiliki nilai V yang melampaui batas minimum, sehingga dapat disimpulkan bahwa semua soal memiliki tingkat validitas yang baik.

4.2 Hasil Uji coba Tes

Instrumen tes yang telah melalui proses validasi oleh para pakar kemudian diuji cobakan kepada 31 siswa kelas 9 di salah satu SMP yang berlokasi di Kecamatan Bahar Utara. Pengujian reliabilitas terhadap data respons siswa dilakukan dengan menggunakan metode Cronbach's Alpha. Hasil uji reliabilitas ditampilkan berdasarkan output dari SPSS versi 26.

Tabel 4. 2 Uji Reliabilitas

Case Processing Summary			
		N	%
Cases	Valid	108	100.0
	Excluded ^a	0	.0
	Total	108	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics	
Cronbach's Alpha	N of Items
.816	2

Berdasarkan tabel di atas, diperoleh nilai Cronbach's Alpha sebesar 0,816, yang melebihi batas minimum yaitu 0,6. Hal ini menunjukkan bahwa instrumen tes yang digunakan dalam penelitian ini tergolong andal atau reliabel. Berdasarkan hasil uji statistik terhadap data respons siswa, dapat disimpulkan bahwa instrumen

tes dalam penelitian ini layak digunakan untuk pengumpulan data karena telah memenuhi syarat validitas dan reliabilitas.

4.3 Deskriptif Statistik Kemampuan Pemecahan Masalah

4.3.1 Identifikasi masalah

Tabel 4. 3 Deskriptif Statistik Identifikasi Masalah

Identifikasi Masalah					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	37	34.3	34.3	34.3
	1.00	22	20.4	20.4	54.6
	2.00	37	34.3	34.3	88.9
	3.00	12	11.1	11.1	100.0
	Total	108	100.0	100.0	

Berdasarkan hasil analisis data terhadap kemampuan siswa dalam tahap identifikasi masalah, diperoleh distribusi frekuensi yang menunjukkan variasi tingkat penguasaan siswa dalam mengidentifikasi inti persoalan dari soal yang diberikan. Dari total 108 responden, sebanyak 37 siswa (34,3%) memperoleh skor 0.00, yang menunjukkan bahwa mereka belum mampu mengidentifikasi permasalahan secara tepat. Proporsi yang sama juga terlihat pada skor 2.00 (34,3%), yang mengindikasikan bahwa sejumlah siswa mulai menunjukkan pemahaman yang cukup terhadap permasalahan yang diberikan, meskipun belum sepenuhnya lengkap.

Sementara itu, 22 siswa (20,4%) mendapatkan skor 1.00, mencerminkan kemampuan parsial dalam mengidentifikasi permasalahan; mereka hanya mampu mengenali sebagian dari unsur penting soal. Hanya 12 siswa (11,1%) yang mencapai skor tertinggi (3.00), yaitu kategori mampu mengidentifikasi masalah secara utuh dan tepat sesuai dengan maksud soal.

Secara kumulatif, sebanyak 65,7% siswa berada di bawah tingkat penguasaan penuh (skor < 3.00), yang menunjukkan bahwa mayoritas peserta didik masih mengalami kesulitan pada tahap awal pemecahan masalah, yaitu memahami dan mengidentifikasi inti dari persoalan matematika. Padahal,

tahapan ini merupakan langkah awal yang sangat menentukan keberhasilan dalam menyelesaikan persoalan secara menyeluruh (Polya, 1973; Lester, 1994). Polya (1973) menekankan bahwa memahami masalah adalah tahap pertama dari proses berpikir matematis yang sistematis, dan apabila tahap ini tidak dilakukan dengan baik, maka tahapan selanjutnya seperti merencanakan dan melaksanakan penyelesaian akan terhambat

4.3.2 Merumuskan Strategi

Tabel 4. 4 Deskriptif Statistik Merumuskan strategi

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	19	17.6	17.6	17.6
	1.00	33	30.6	30.6	48.1
	2.00	50	46.3	46.3	94.4
	3.00	6	5.6	5.6	100.0
	Total	108	100.0	100.0	

Tahap kedua dalam pemecahan masalah menurut Polya (1973) adalah merumuskan strategi penyelesaian, yaitu proses memilih pendekatan atau langkah-langkah yang akan digunakan untuk menyelesaikan masalah berdasarkan informasi yang tersedia. Berdasarkan hasil analisis data, diperoleh bahwa mayoritas siswa berada pada tingkat kemampuan cukup (skor 2.00), yaitu sebanyak 50 siswa (46,3%), yang menunjukkan bahwa sebagian besar siswa dapat menyusun strategi dengan cukup baik, meskipun belum sepenuhnya lengkap.

Namun demikian, sekitar 48,2% siswa lainnya masih menunjukkan kelemahan dalam aspek ini, dengan rincian 30,6% siswa hanya mampu menyusun sebagian strategi, dan 17,6% siswa tidak mampu menyusun strategi sama sekali. Hanya 5,6% siswa yang berada pada kategori tertinggi, yaitu mampu merumuskan strategi secara lengkap, logis, dan sistematis.

Temuan ini menunjukkan bahwa kemampuan siswa dalam merancang langkah-langkah penyelesaian masih perlu ditingkatkan. Padahal, kemampuan ini sangat penting karena menjadi jembatan antara pemahaman masalah dengan pelaksanaan penyelesaian. Sejalan dengan pendapat Lester (1994), kegagalan dalam merumuskan strategi sering kali mengakibatkan siswa tidak dapat melanjutkan ke tahap eksekusi atau mengalami kebuntuan dalam penyelesaian.

Oleh karena itu, guru perlu memberikan penekanan lebih pada latihan-latihan yang menuntut siswa untuk menyusun dan memilih strategi secara sadar, seperti soal terbuka atau proyek berbasis masalah. Selain itu, penggunaan scaffolding atau pembimbingan strategi secara bertahap juga direkomendasikan (Nisrina et al., 2021), agar siswa terbiasa berpikir sistematis dan memilih pendekatan yang tepat dalam menyelesaikan masalah matematika.

4.3.3 Menjalankan Strategi

Tabel 4. 5 Deskriptif Statistik Menjalankan Strategi

Menjalankan Strategi					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	40	37.0	37.0	37.0
	1.00	35	32.4	32.4	69.4
	2.00	26	24.1	24.1	93.5
	3.00	7	6.5	6.5	100.0
Total		108	100.0	100.0	

Tahap menjalankan strategi merupakan implementasi dari rencana penyelesaian masalah matematika yang telah disusun. Berdasarkan hasil analisis, diketahui bahwa mayoritas siswa (69,4%) belum mampu mengeksekusi strategi penyelesaian dengan baik. Tercatat 37% siswa tidak dapat menjalankan strategi sama sekali, dan 32,4% siswa hanya melaksanakan sebagian langkah penyelesaian yang mungkin tidak logis atau tidak selesai. Hanya 6,5% siswa yang mampu melaksanakan strategi secara lengkap, menunjukkan kemampuan eksekusi yang matang.

Data ini menunjukkan bahwa eksekusi strategi merupakan titik lemah utama siswa, meskipun beberapa dari mereka mampu menyusun strategi secara memadai pada tahap sebelumnya. Hal ini mengindikasikan adanya kemungkinan gap antara pemahaman dan keterampilan prosedural dalam menyelesaikan masalah. Menurut Polya (1973), menjalankan strategi membutuhkan kombinasi antara ketepatan rencana dan keterampilan teknis untuk mengeksekusinya secara logis dan sistematis. Lester (1994) juga menekankan bahwa kegagalan pada tahap ini dapat berasal dari kesalahan perhitungan, miskonsepsi, atau ketidaktepatan menerapkan rumus dan prosedur.

Kelemahan siswa pada tahap ini menunjukkan perlunya pembelajaran yang tidak hanya berfokus pada pemahaman konsep, tetapi juga pada latihan intensif menjalankan strategi melalui latihan bertingkat dan refleksi proses. Guru perlu memberikan umpan balik spesifik terhadap cara siswa menyelesaikan masalah, bukan hanya pada jawaban akhir.

4.3.4 Memeriksa Kembali

Tabel 4. 6 Deskriptif Statistik Memeriksa Kembali

Memeriksa Kembali					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	17	15.7	15.7	15.7
	1.00	33	30.6	30.6	46.3
	2.00	29	26.9	26.9	73.1
	3.00	29	26.9	26.9	100.0
Total		108	100.0	100.0	

Refleksi atau memeriksa kembali hasil penyelesaian masalah merupakan tahap krusial dalam pemecahan masalah matematika karena pada tahap ini siswa meninjau kembali langkah-langkah penyelesaian dan mengevaluasi kebenaran jawaban mereka. Berdasarkan tabel, terlihat bahwa sebagian besar siswa sudah mulai menunjukkan kesadaran dalam melakukan

refleksi. Sebanyak 53,8% siswa (gabungan skor 2.00 dan 3.00) melakukan pemeriksaan terhadap hasil kerja mereka secara cukup atau menyeluruh.

Namun demikian, masih terdapat 46,3% siswa yang tidak melakukan refleksi secara optimal, bahkan 15,7% di antaranya sama sekali tidak melakukan pemeriksaan. Kondisi ini menunjukkan bahwa meskipun siswa mampu memahami dan menyelesaikan masalah, kesadaran metakognitif mereka dalam mengevaluasi proses berpikirnya masih perlu ditingkatkan.

Tahap refleksi sering kali diabaikan oleh siswa karena tidak langsung menghasilkan skor atau nilai, padahal menurut Polya (1973), memeriksa kembali solusi adalah esensi dari pembelajaran bermakna, karena siswa belajar dari kesalahan dan menyadari strategi mana yang berhasil. Nisrina et al. (2021) juga menunjukkan bahwa siswa dengan kemampuan reflektif tinggi cenderung memiliki solusi yang lebih benar dan konsisten.

Oleh karena itu, dalam proses pembelajaran, guru perlu menekankan pentingnya tahapan refleksi, misalnya dengan memberikan ruang dalam lembar kerja siswa untuk menjelaskan kembali langkah-langkah atau menjawab pertanyaan seperti "Apakah jawaban ini masuk akal?" atau "Apakah ada cara lain yang lebih efisien?"

4.4 Hasil Uji Asumsi Prasyarat IRT

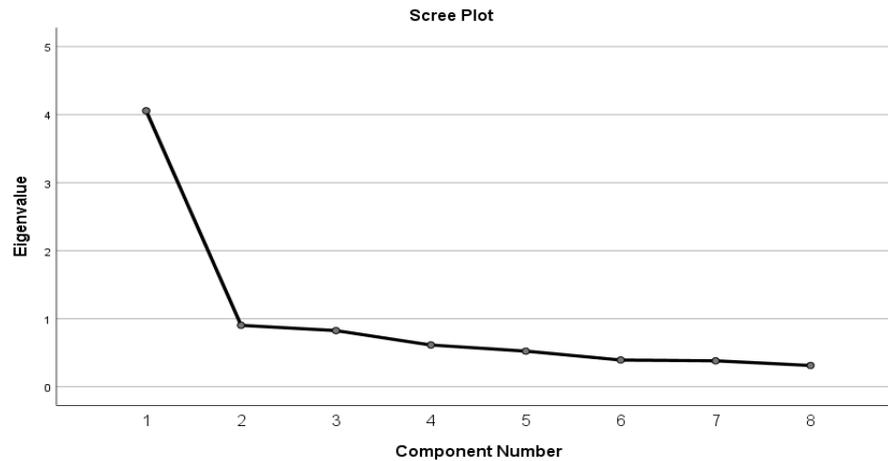
Sebelum dilakukan analisis data lebih lanjut menggunakan PARSCALE 4.1. Perlu adanya pengujian asumsi prasyarat IRT. Menurut Embretson & Reise, (2000), asumsi unidimensi dapat diuji menggunakan analisis faktor, dengan bantuan *software Statistical Package for Social Science* (SPSS) 26. Uji unidimensi dilakukan terhadap data-data sebelum digunakan untuk estimasi parameter kemampuan peserta tes. Data-data yang diuji tersebut adalah data politomi untuk model GPCM.

Tabel 4. 7 Total Variance Explained

Total Variance Explained						
<i>Factor</i>	<i>Initial Eigenvalues</i>			<i>Extraction Sums of Squared Loadings</i>		
	<i>Total</i>	<i>% of Variance</i>	<i>Cumulative %</i>	<i>Total</i>	<i>% of Variance</i>	<i>Cumulative %</i>
1	4.056	50.699	50.699	4.056	50.699	50.699
2	0.902	11.278	61.977			
3	0.824	10.300	72.277			
4	0.613	7.657	79.934			
5	0.522	6.524	86.458			
6	0.392	4.898	91.356			
7	0.380	4.749	96.106			
8	0.312	3.894	100.000			

Tabel 4.3 di atas menunjukkan hasil uji unidimensi melalui analisis komponen utama (*Principal Component Analysis*). Terlihat bahwa komponen pertama memiliki nilai eigen (Total) sebesar 4.056, yang setara dengan 50,699% dari total varians yang dijelaskan oleh data. Nilai ini secara signifikan lebih besar dibandingkan dengan komponen-komponen lainnya, yang masing-masing memiliki nilai eigen kurang dari 1.

Komponen kedua hanya menjelaskan 11,278% dari varians, dan komponen ketiga sebesar 10,300%. Secara kumulatif, tiga komponen pertama menjelaskan sekitar 72,277% dari total varians, namun hanya komponen pertama yang memenuhi kriteria Kaiser, yaitu nilai eigen lebih besar dari 1. Berdasarkan kriteria ini, hanya satu komponen yang dianggap layak untuk dipertahankan karena mampu menjelaskan proporsi terbesar dari varians total. Berikut Gambar 1 merupakan grafik scree plot data.



Gambar 4. 1 Scree Plot

Gambar 4.1 *Scree Plot* di atas digunakan untuk menguji dimensionalitas suatu instrumen melalui pendekatan analisis komponen utama (PCA). Uji ini bertujuan untuk mengidentifikasi jumlah dimensi atau faktor utama yang mendasari sekumpulan item atau variabel.

Dari grafik tersebut terlihat bahwa komponen pertama memiliki eigenvalue yang paling tinggi, yaitu lebih dari 4, sementara komponen kedua dan seterusnya menunjukkan penurunan tajam dan nilai eigenvalue yang berada di bawah angka 1. Pola ini menunjukkan adanya titik siku (*elbow*) yang jelas setelah komponen pertama, yang mengindikasikan bahwa hanya satu dimensi utama yang signifikan.

Berdasarkan kriteria Kaiser (retensi komponen dengan eigenvalue > 1), hanya komponen pertama yang layak dipertahankan. Hal ini memperkuat dugaan bahwa instrumen yang diuji bersifat unidimensional, atau hanya mengukur satu konstruk utama. Dengan kata lain, data cenderung konsisten dalam mengukur satu aspek psikologis atau karakteristik yang sama, sehingga instrumen dapat dikatakan memiliki struktur satu dimensi yang kuat.

Dengan terpenuhinya uji unidimensi, maka secara tidak langsung uji independensi lokal dan uji invariansi parameter juga dianggap telah terpenuhi, sehingga cukup memusatkan perhatian pada uji unidimensi saja (Hambleton, Swaminathan, and Rogers 1991)

Setelah dilakukan uji prasyarat unidimensi, selanjutnya dilakukan uji kecocokan model. Berdasarkan hasil uji kecocokan pada butir tes dengan bantuan PARSCALE 4.1 diperoleh *item fit statistics* nilai χ^2 GPCM sebesar 65,69 (p-value = 0,686), nilai χ^2 GRM sebesar 77,39 (p-value = 0,311), nilai χ^2 PCM sebesar 54,99 (p-value = 0,872).

Berdasarkan tabel *Item Fit Statistics* di atas, seluruh butir soal memiliki nilai p-value (Prob.) yang lebih besar dari 0,05. Hal ini menunjukkan bahwa tidak ada butir yang menyimpang secara signifikan dari model, sehingga dapat disimpulkan bahwa seluruh item memiliki kecocokan yang baik terhadap semua model.

Dengan demikian, butir-butir soal dalam instrumen ini valid digunakan untuk mengukur kemampuan siswa secara akurat.

4.5 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 Estimasi Parameter Butir GPCM

Item	Daya Beda (a)	S.E	Tingkat Kesuran	S.E
1	1.267	0.099	0.512	0.031
2	1.553	0.111	0.272	0.032
3	0.938	0.075	0.757	0.041
4	0.995	0.065	-0.290	0.036
5	1.036	0.083	-0.468	0.038
6	1.641	0.128	-0.445	0.029
7	1.067	0.076	-0.222	0.037
8	0.736	0.056	-0.093	0.048

Berdasarkan hasil analisis parameter daya beda (*discrimination parameter*, a) dan tingkat kesukaran (*difficulty parameter*) dari delapan item menggunakan model IRT khususnya model GPCM) yang dianalisis melalui perangkat lunak PARSCALE, diperoleh informasi sebagai berikut. Nilai daya beda item berkisar antara 0.736 hingga 1.641. Mengacu pada klasifikasi yang dikemukakan oleh Amelia dan Kriswanto (2017), daya beda item dikategorikan sebagai sangat

rendah apabila $a < 0.25$, rendah (0.25–0.64), sedang (0.65–1.34), tinggi (1.35–1.69), dan sangat tinggi (≥ 1.70). Berdasarkan klasifikasi tersebut, empat item yakni item 2 ($a = 1.553$), item 6 ($a = 1.641$), item 1 ($a = 1.267$), dan item 5 ($a = 1.036$) menunjukkan kualitas yang baik dalam membedakan kemampuan peserta didik, dengan kategori sedang hingga tinggi.

Selanjutnya, item 3 ($a = 0.938$), item 4 ($a = 0.995$), item 7 ($a = 1.067$), dan item 8 ($a = 0.736$) tergolong dalam kategori sedang. Ini mengindikasikan bahwa sebagian besar item dalam instrumen ini memiliki kemampuan diskriminatif yang cukup untuk membedakan peserta berdasarkan tingkat kemampuan Pemecahan Masalah matematis yang berbeda.

Dari sisi tingkat kesukaran, nilai logit berkisar antara -0.468 hingga 0.757. Menurut Linacre (2006), rentang logit untuk klasifikasi tingkat kesukaran adalah sebagai berikut: sangat mudah (logit < -2.00), mudah (-2.00 s.d. -0.50), sedang (-0.49 s.d. +0.49), sulit (+0.50 s.d. +2.00), dan sangat sulit ($\geq +2.01$). Berdasarkan rentang ini, item 5 dan item 6 tergolong mudah (logit -0.468 dan -0.445), begitu pula dengan item 4 dan item 7 yang berada pada logit -0.290 dan -0.222. Sementara itu, item 8 (logit -0.093) berada dalam kategori sedang, sedangkan item 1 (logit 0.512) dan item 3 (logit 0.757) termasuk kategori sulit. Item 2 (logit 0.272) berada dalam batas atas kategori sedang, mendekati kategori sulit.

Tabel 4. 9 Estimasi Parameter Butir GRM

Item	Daya Beda (a)	S.E	Tingkat Kesukaran	S.E
1	1.347	0.208	0.658	0.138
2	1.399	0.208	0.326	0.135
3	0.979	0.142	0.994	0.158
4	1.131	0.140	-0.368	0.143
5	1.079	0.142	-0.629	0.149
6	1.593	0.221	-0.554	0.125
7	1.040	0.142	-0.342	0.148
8	0.765	0.087	-0.159	0.178

Berdasarkan hasil analisis menggunakan model IRT dengan pendekatan GRM, diperoleh estimasi parameter daya beda (a) dan tingkat kesukaran (b) dari

delapan item yang dianalisis. Nilai parameter daya beda berkisar antara 0.765 hingga 1.593. Berdasarkan klasifikasi dari Amelia dan Kriswantoro (2017), nilai a antara 0.65 hingga 1.34 termasuk dalam kategori sedang, sedangkan nilai antara 1.35 hingga 1.69 termasuk kategori tinggi. Hasil ini menunjukkan bahwa sebagian besar item memiliki kemampuan yang cukup baik dalam membedakan peserta berdasarkan tingkat kemampuan pemecahan masalah matematis.

Tiga item, yaitu item 1 ($a = 1.347$), item 2 ($a = 1.399$), dan item 6 ($a = 1.593$), masuk dalam kategori daya beda tinggi. Ini menunjukkan bahwa ketiga item tersebut sangat efektif dalam membedakan peserta dengan tingkat kemampuan yang berbeda secara tajam. Sementara itu, lima item lainnya, yaitu item 3 ($a = 0.979$), item 4 ($a = 1.131$), item 5 ($a = 1.079$), item 7 ($a = 1.040$), dan item 8 ($a = 0.765$), berada pada kategori daya beda sedang, yang berarti item-item ini tetap berfungsi secara optimal dalam proses identifikasi perbedaan kemampuan antar peserta.

Dari sisi tingkat kesukaran, nilai logit berkisar antara -0.629 hingga 0.994. Berdasarkan rentang interpretasi logit yang dikemukakan oleh Linacre (2006), nilai logit antara -2.00 hingga -0.50 termasuk kategori mudah, -0.49 hingga +0.49 termasuk sedang, dan +0.50 hingga +2.00 tergolong sulit. Dalam konteks ini, dua item yaitu item 5 ($b = -0.629$) dan item 6 ($b = -0.554$) termasuk dalam kategori mudah, tiga item yaitu item 4 ($b = -0.368$), item 7 ($b = -0.342$), dan item 8 ($b = -0.159$) tergolong sedang, serta tiga item lainnya yaitu item 1 ($b = 0.658$), item 2 ($b = 0.326$), dan item 3 ($b = 0.994$) tergolong dalam kategori sulit.

Tabel 4. 10 Estimasi Parameter Butir PCM

Item	Daya Beda (a)	S.E	Tingkat Kesuran	S.E
1	1.027	0.026	0.554	0.037
2	1.027	0.026	0.332	0.045
3	1.027	0.026	0.750	0.039
4	1.027	0.026	-0.308	0.036
5	1.027	0.026	-0.493	0.038
6	1.027	0.026	-0.530	0.042
7	1.027	0.026	-0.240	0.039
8	1.027	0.026	-0.098	0.036

Berdasarkan hasil analisis menggunakan model *Partial Credit Model* (PCM), diperoleh bahwa seluruh item memiliki nilai daya beda (a) yang sama, yaitu sebesar 1.027. Hal ini sesuai dengan asumsi dasar PCM, yang menganggap bahwa semua item memiliki daya beda yang setara. Variasi kemampuan peserta dalam menjawab item lebih banyak ditentukan oleh perbedaan tingkat kesukaran item, bukan oleh kemampuan item dalam membedakan peserta. Tingkat kesukaran (*difficulty parameter*) pada masing-masing item menunjukkan nilai logit yang bervariasi, dari -0.530 hingga 0.750. Nilai ini mengindikasikan bahwa item dalam instrumen memiliki sebaran tingkat kesukaran dari kategori mudah hingga sulit. Dengan demikian, PCM memberikan estimasi yang cukup baik dalam menggambarkan penyebaran kesukaran soal, meskipun tidak menangkap perbedaan daya diskriminasi antar item. Model ini cocok digunakan dalam konteks pengukuran yang menuntut konsistensi struktur soal, seperti rubrik kompetensi atau asesmen bertingkat yang seragam (Masters, 1982; Linacre, 2006).

Tabel 4.4 hingga 4.6 menampilkan informasi mengenai parameter daya beda dan tingkat kesukaran untuk masing-masing model. Pada model PCM, seluruh nilai daya beda bersifat seragam karena model ini mengasumsikan bahwa setiap butir memiliki tingkat kesulitan yang berbeda, namun daya bedanya dianggap sama. Dengan kata lain, model PCM tidak mempertimbangkan variasi dalam parameter daya beda. Sebaliknya, model GPCM dan GRM memperhitungkan kedua parameter tersebut, yaitu tingkat kesulitan dan daya beda, yang menggambarkan sejauh mana suatu butir mampu membedakan individu berdasarkan perbedaan tingkat kemampuan mereka.

Daya beda (a) dianggap berada dalam kategori baik jika berada dalam rentang 0,25 hingga 1,5 sebagai kisaran yang umum (*typical range*), dan nilai yang melebihi 1,5 juga masih termasuk dalam kategori baik. Sebaliknya, nilai yang berada di bawah 0,25 diklasifikasikan sebagai rendah (Hambleton et al., 1991). Berdasarkan data dalam tabel, tidak ditemukan butir soal dengan nilai daya beda (a) di bawah 0,25. Dengan demikian, dapat disimpulkan bahwa parameter butir dari setiap model telah memenuhi kriteria kualitas yang baik.

Tingkat kesulitan (b) dikategorikan baik apabila nilainya berada dalam rentang logit antara -3 hingga 3 (Hambleton et al., 1991). Berdasarkan tabel, seluruh nilai tingkat kesulitan pada instrumen ini berada dalam kisaran tersebut, sehingga dapat disimpulkan bahwa butir-butir soal dalam semua model memiliki tingkat kesulitan yang sesuai dengan kriteria yang baik.

4.6 Estimasi Parameter Kemampuan Pemecahan Masalah Siswa

Estimasi parameter kemampuan dilakukan secara simultan dan terpisah dengan menggunakan bantuan PARSCALE 4.1. Untuk menganalisis keterampilan siswa dapat dilihat di Phase 3 pada output PARSCALE yang dipakai untuk mengestimasi keterampilan siswa yang diukur melalui respon terhadap butir instrumen yang dinyatakan dalam satuan logit.

Nilai ini merepresentasikan posisi relatif subjek pada skala interval yang sama dengan soal tes yang memungkinkan perbandingan objektif antar individu. Dari 108 responden yang dianalisis menggunakan pendekatan IRT, diperoleh klasifikasi kemampuan sebagai berikut:

Tabel 4. 11 Distribusi Frekuensi Estimasi Parameter Kemampuan Model GPCM

Kategori	Rentang Theta	Jumlah Siswa	Persentase (%)
Tinggi	$\theta > 1$	21 siswa	19,44%
Sedang	$-1, \leq \theta \leq 1$	70 siswa	64,81%
Rendah	$\theta < -1$	17 siswa	15,74%

Dari tabel diatas estimasi kemampuan pemecahan masalah model GPCM dengan nilai rata rata -0,0037 dan variase sebesar 0,904. Distribuki semampuan model GPCM terbagi menjadi tiga kategori. Kategori rendah dengan jumlah responden 17 siswa (15,74%), kategori kemampuan sedang dengan jumlah responden terbanyak yaitu 70 siswa (64,81%), kategori kemampuan tinggi dengan jumlah responden 21 siswa(19,44%). Lebih dari 50 % responden memiliki sekor kemampuan padsa level

sedang. Dengan demikian mayoritas responden berada pada kategori kemampuan sedang. Dan hanya sebagian kecil berada pada kategori rendah atau tinggi.

Tabel 4. 12 Distribusi Frekuensi Hasil Estimasi Kemampuan Model GRM

Kategori	Rentang Theta	Jumlah Siswa	Persentase (%)
Tinggi	$\theta > 1$	19 siswa	17,59%
Sedang	$-1, \leq \theta \leq 1$	72 siswa	66,67%
Rendah	$\theta < -1$	17 siswa	15,74%

Dari tabel diatas estimasi kemampuan pemecahan masalah model GPCM dengan nilai rata rata -0,0009 dan varianse sebesar 1,271. Distribuki semampuan model GPCM terbagi menjadi tiga kategori. Kategori rendah dengan jumlah responden 17 siswa (15,74%), kategori kemampuan sedang dengan jumlah responden terbanyak yaitu 72 siswa (66,67 %), kategori kemampuan tin ggi dengan jumlah responden 19 siswa (19,44%). Lebih dari 50 % responden memiliki sekor kemampuan pada level sedang. Dengan demikian mayoritas responden berada pada kategori kemampuan sedang. Dan hanya sebagian kecil berada pada kategori rendah atau tinggi.

Tabel 4. 13Distribusi Frekuensi Hasil Estimasi Kemampuan Model PCM

Kategori	Rentang Theta	Jumlah Siswa	Persentase (%)
Tinggi	$\theta > 1$	17 siswa	17,59%
Sedang	$-1, \leq \theta \leq 1$	73 siswa	67,59%
Rendah	$\theta < -1$	18 siswa	16,67%

Dari tabel diatas estimasi kemampuan pemecahan masalah model GPCM dengan nilai rata rata -0,0074 dan varianse sebesar 1,257. Distribuki semampuan model GPCM terbagi menjadi tiga kategori. Kategori rendah dengan jumlah responden 18 siswa (16,67%), kategori kemampuan sedang dengan jumlah responden terbanyak yaitu 73 siswa (67,59 %), kategori kemampuan tinggi dengan jumlah responden 17

siswa (17,59%). Lebih dari 50 % responden memiliki skor kemampuan pada level sedang. Dengan demikian mayoritas responden berada pada kategori kemampuan sedang. Dan hanya sebagian kecil berada pada kategori rendah atau tinggi.

Berikut merupakan deskriptif statistik kemampuan pemecahan masalah pada siswa berdasarkan tiga model yang digunakan.

Tabel 4. 14 Ringkasan Deskriptif Estimasi Kemampuan Siswa

Statistik	PCM	GRM	GPCM
<i>Mean</i>	0,0074	-0,0009	-0,0037
<i>Median</i>	0,1093	0,0561	0,0802
<i>Standar Deviasi</i>	1,0044	1,0046	1,0045
<i>varians</i>	1,257	1,271	0,904
<i>Minimum</i>	-2,41	-2,35	-2,47
<i>Maximum</i>	2,01	2,26	2,14

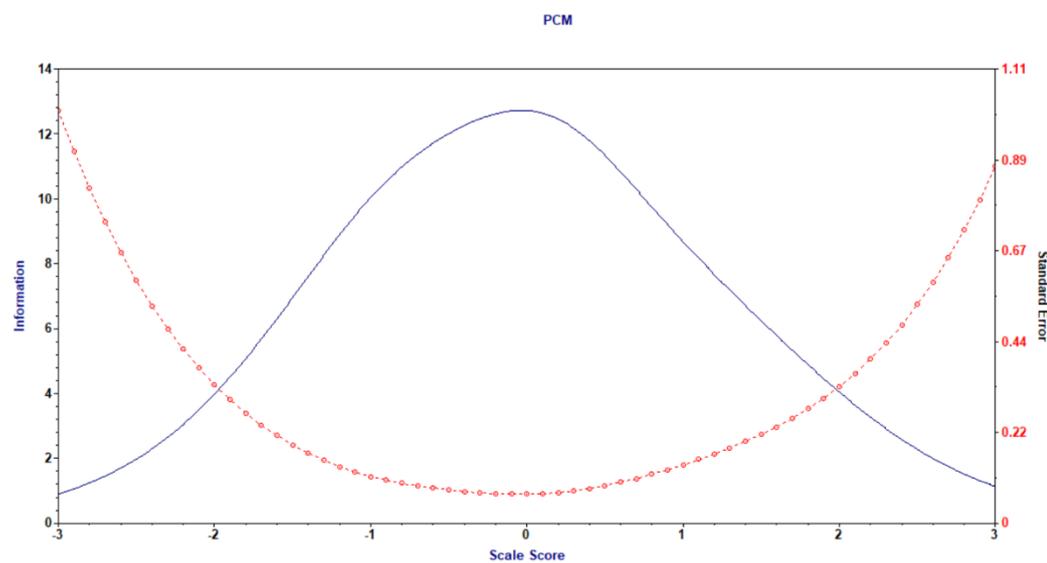
Tabel 4.7 tersebut menyajikan ringkasan deskriptif dari hasil estimasi parameter kemampuan peserta tes berdasarkan berbagai model yang digunakan dalam penelitian ini. Secara umum, nilai rata-rata (mean) dari ketiga model mendekati nol, menunjukkan bahwa kemampuan peserta tersebar merata di sekitar nilai tengah. Nilai mean tertinggi ditunjukkan oleh model PCM (0,0074), sedangkan GPCM memiliki mean sedikit negatif (0,0037). Nilai median pada ketiga model relatif sama, mengindikasikan bahwa lebih dari separuh peserta memiliki kemampuan sedikit di bawah rata-rata. Dari sisi standar deviasi dan varians, ketiga model juga menunjukkan nilai yang hampir sama, yaitu standar deviasi sekitar 1 dan varians mendekati 1. Ini menandakan bahwa sebaran kemampuan peserta cukup seragam antar model. Rentang nilai kemampuan peserta juga tidak berbeda jauh, dengan nilai minimum berkisar antara -2,47 (GPCM) hingga -2,35 (GRM), dan nilai maksimum antara 2,01 (PCM) hingga 2,26 (GRM). Ini menunjukkan bahwa ketiga model mampu mengukur rentang kemampuan yang hampir setara.

Secara keseluruhan, ketiga model estimasi kemampuan menghasilkan distribusi yang seimbang, dengan karakteristik penyebaran data yang serupa, sehingga dapat

disimpulkan bahwa ketiganya memiliki konsistensi dalam mengukur kemampuan peserta.

4.7 Fungsi Informasi Tes

Fungsi informasi tes menyajikan data yang berkaitan dengan karakteristik kemampuan siswa. Terdapat hubungan terbalik antara fungsi informasi butir dan kesalahan baku pengukuran (*Standard Error Measurement*); artinya, semakin kecil tingkat kesalahan pengukuran, maka semakin tinggi informasi yang dapat diberikan oleh butir tes tersebut(Akbar, and Naga 2020).

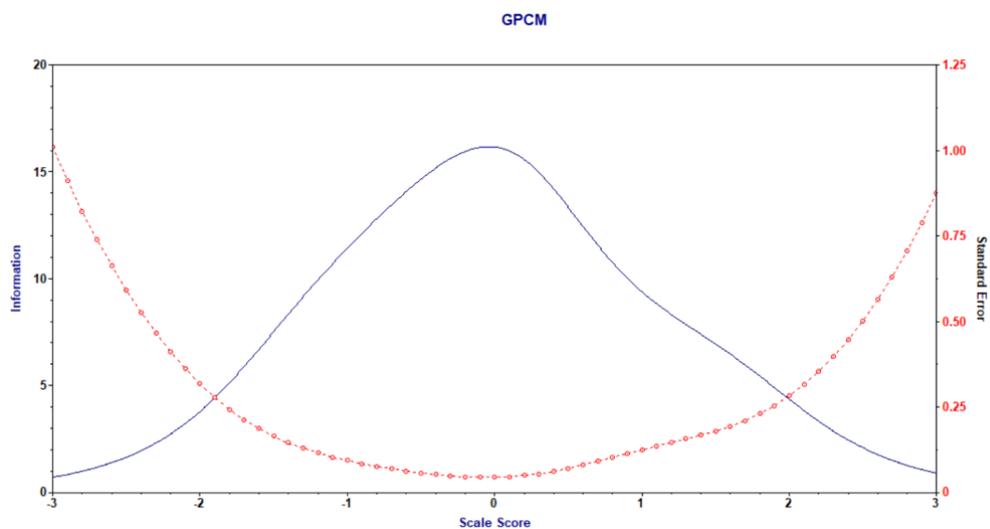


Gambar 4. 2 Fungsi Informasi Tes PCM

Gambar di atas merupakan kurva fungsi informasi dan standar kesalahan pengukuran dari model Partial Credit Model (PCM). Kurva berwarna biru menunjukkan fungsi informasi, sedangkan kurva berwarna merah dengan titik-titik menggambarkan standar kesalahan pengukuran (Standard Error) pada berbagai nilai skor kemampuan (Scale Score).

Berdasarkan Gambar di atas, terlihat bahwa nilai informasi tertinggi diberikan oleh tes pada rentang skor kemampuan sekitar 0. Hal ini menunjukkan bahwa instrumen paling akurat dalam mengukur peserta dengan kemampuan rata-rata. Sebaliknya, standar kesalahan pengukuran mencapai nilai terendah pada titik yang sama, yang mengindikasikan bahwa estimasi kemampuan peserta paling tepat terjadi di sekitar

nilai tersebut. Semakin jauh dari nilai 0, informasi yang diberikan tes menurun, dan standar kesalahan meningkat, yang berarti bahwa akurasi pengukuran lebih rendah untuk peserta dengan kemampuan sangat rendah atau sangat tinggi. Informasi mencapai puncaknya pada skor kemampuan sekitar 0, dengan nilai informasi mendekati 13. Hal ini menunjukkan bahwa instrumen tes paling akurat dalam mengukur peserta dengan kemampuan rata-rata. Lebarnya kurva yang tetap berada pada tingkat informasi yang cukup baik antara skor -2 hingga +2 mengindikasikan bahwa tes ini masih efektif dalam mengukur kemampuan dalam rentang tersebut. Di luar rentang itu, kurva informasi menurun dan standar kesalahan meningkat, menandakan penurunan akurasi pengukuran.

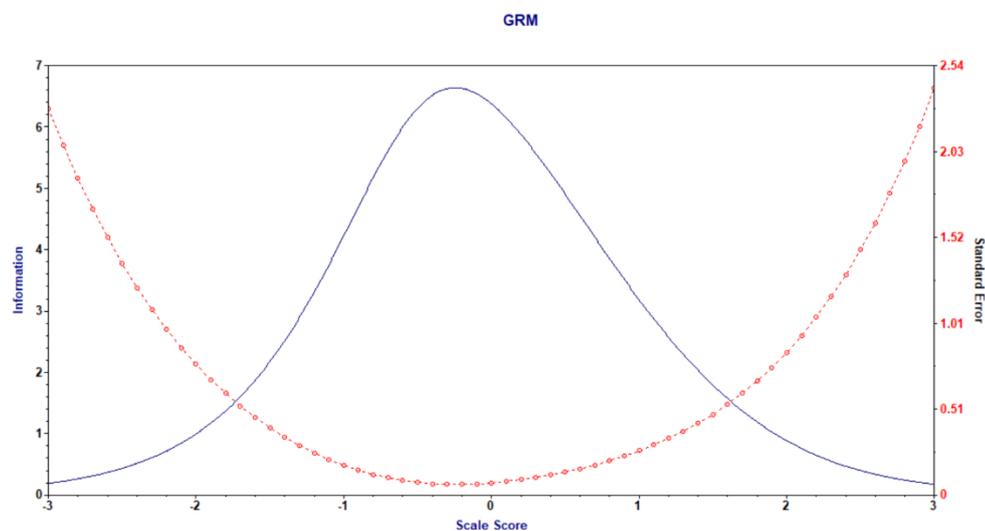


Gambar 4. 3 Fungsi Informasi Tes GPCM

Gambar di atas merupakan kurva fungsi informasi dan standar kesalahan pengukuran dari GPCM. Kurva berwarna biru menunjukkan fungsi informasi, sedangkan kurva berwarna merah dengan titik-titik menggambarkan standar kesalahan pengukuran (Standard Error) pada berbagai nilai skor kemampuan (Scale Score).

Berdasarkan Gambar kurva fungsi informasi dan standar kesalahan dari model GPCM, terlihat bahwa kurva informasi mencapai puncak pada skor kemampuan sekitar 0 dengan nilai informasi tertinggi mendekati 16. Hal ini menunjukkan

bahwa instrumen tes sangat akurat dalam mengukur peserta didik dengan kemampuan rata-rata. Kurva informasi tetap berada pada level yang cukup tinggi pada rentang skor -2 hingga +2, yang menunjukkan bahwa tes ini efektif dalam mengukur kemampuan peserta pada rentang tersebut. Sebaliknya, standar kesalahan pengukuran berada pada titik terendah di sekitar skor 0 dan meningkat tajam di luar rentang tersebut, yang mengindikasikan bahwa estimasi kemampuan menjadi kurang akurat untuk peserta dengan kemampuan sangat rendah atau sangat tinggi.



Gambar 4. 4 Fungsi Informasi Tes GRM

Berdasarkan kurva fungsi informasi dan standar kesalahan dari model GRM, terlihat bahwa puncak informasi terjadi pada skor kemampuan sekitar 0 dengan nilai informasi maksimum sekitar 6,5. Hal ini menunjukkan bahwa instrumen tes cukup akurat dalam mengukur peserta dengan kemampuan rata-rata, namun tingkat informasi yang diberikan relatif lebih rendah dibandingkan dengan model PCM dan GPCM. Kurva informasi menyebar dalam rentang skor -2 hingga +2, namun dengan tingkat informasi yang menurun secara signifikan di luar rentang tersebut. Sebaliknya, standar kesalahan pengukuran menunjukkan nilai minimum di sekitar skor 0, dan meningkat tajam pada area kemampuan yang sangat rendah maupun sangat tinggi, mencapai nilai lebih dari 2,5. Hal ini mengindikasikan bahwa estimasi kemampuan kurang akurat bagi peserta dengan kemampuan ekstrem.

Dengan demikian, model GRM cocok digunakan untuk mengukur kemampuan peserta dalam rentang sedang, namun memiliki keterbatasan dalam memberikan informasi yang cukup pada rentang kemampuan yang lebih rendah atau lebih tinggi dari rata-rata.

Gambar 4.5 hingga 4.7 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. 15 Ringkasan Fungsi Informasi Tes

Aspek	PCM	GPCM	GRM
Nilai Informasi	13	16	6.5
Posisi Puncak	Skor 0	Skor 0	Skor 0
Lebar Kurva	-2 s.d. +2	-2 s.d. +2	-2 s.d. +2
Akurasi Pengukuran	Tinggi	Paling tinggi	Paling rendah

Berdasarkan kurva fungsi informasi dan standar kesalahan dari ketiga model, terlihat bahwa model GPCM memiliki puncak informasi tertinggi, yaitu mendekati 16, diikuti oleh PCM dengan nilai sekitar 13, dan GRM yang hanya mencapai sekitar 6,5. Ketiga model menunjukkan bahwa informasi tertinggi dan standar kesalahan terendah terjadi pada skor kemampuan sekitar 0, yang berarti ketepatan pengukuran tertinggi terdapat pada peserta dengan kemampuan rata-rata. 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 GRM

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 PCM dan GRM.

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 kemampuan dari tiga model terlebih dahulu di replikasi menggunakan *software* WinGen sebanyak 10 replika.

Pemeriksaan terhadap normalitas distribusi dilakukan terlebih dahulu pada hasil estimasi parameter, baik secara keseluruhan maupun secara terpisah. Proses ini menggunakan program SPSS versi 26 dengan uji normalitas Kolmogorov-Smirnov

Tabel 4. 16 Uji Normalitas

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
GRM	.057	108	.200 [*]	.991	108	.679
PCM	.060	108	.200 [*]	.987	108	.402
GPCM	.050	108	.200 [*]	.992	108	.751

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Berdasarkan tabel di atas, seluruh data yang dianalisis menunjukkan nilai signifikansi (sig) lebih besar dari 0,05. Sesuai dengan kriteria uji Kolmogorov-Smirnov, nilai sig > 0,05 mengindikasikan bahwa data memiliki distribusi normal (Kadir, 2017). Oleh karena itu, data ini digunakan dalam proses analisis dan pengujian hipotesis dalam penelitian ini.

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. Hasil dari pengujian hipotesis tersebut disajikan sebagai berikut.

4.7.1 Pengujian Hipotesis pertama

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

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

Keterangan:

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

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

Tabel 4. 17 Pengujian Hipotesis Pertama

statistik	Hasil estimasi parameter	
	PCM	GRM
n	$n_1 = 108$	$n_2 = 108$
df	$df_1 = n_1 - 1$ $= 107$	$df_2 = n_2 - 1$ $= 107$
σ^2	1,257	1,271
F_{hitung}	$\frac{1,271}{1,257} = 1,011$	
F_{tabel}	$F_{(0,05;107;107)} = 1,376$	

Berdasarkan hasil pada tabel di atas, 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 kemampuan Pemecahan masalah matematis, temuan ini menunjukkan bahwa kedua model memberikan tingkat presisi yang relatif setara dalam mengestimasi kemampuan siswa.

4.7.2 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 kemampuan dengan GPCM

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

Tabel 4. 18 Pengujian Hipotesis Kedua

Statistic	Hasil estimasi parameter	
	GRM	GPCM
n	$n_1 = 108$	$n_2 = 108$
df	$df_1 = n_1 - 1$ $= 107$	$df_2 = n_2 - 1$ $= 107$
σ^2	1,271	0,902
F_{hitung}	$\frac{1,257}{0,902} = 1,409$	
F_{tabel}	$F_{(0,05;107;107)} = 1,376$	

Berdasarkan hasil pada tabel di atas, diperoleh bahwa nilai $F_{hitung} > F_{tabel}$, sehingga H_0 ditolak. Ini menunjukkan bahwa varians estimasi parameter kemampuan peserta tes menggunakan GPCM lebih kecil dibandingkan dengan varians estimasi menggunakan GRM. 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 kemampuan pemecahan masalah 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 pemecahan masalah matematis.

4.7.2 Pengujian Hipotesis ketiga

Hipotesis ketiga 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 kemampuan dengan PCM

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

Tabel 4. 19 Pengujian Hipotesis Ketiga

statistik	Hasil estimasi parameter	
	GPCM	PCM
n	$n_1 = 108$	$n_2 = 108$
df	$df_1 = n_1 - 1$ $= 107$	$df_2 = n_2 - 1$ $= 107$
σ^2	0,902	1,257
F_{hitung}	$\frac{1,271}{0,902} = 1,393$	
F_{tabel}	$F_{(0,05;107;107)} = 1,376$	

Berdasarkan hasil pada tabel di atas, diperoleh bahwa nilai $F_{hitung} > F_{tabel}$, sehingga H_0 ditolak. Ini menunjukkan bahwa varians estimasi parameter

kemampuan peserta tes menggunakan GPCM lebih kecil dibandingkan dengan varians estimasi menggunakan PCM. 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 kemampuan pemecahan masalah 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 pemecahan masalah matematis siswa.

Tabel 4. 20 Rekapitulasi Uji Hipotesis

hipotesis	Uji Beda varian		Hasil uji	kesimpulan
	UJI F			
1	PCM	GRM	Terima H_0	Perbedaan tidak signifikan
2	GPCM	GRM	Terima H_1	GPCM lebih Presisi
3	GPCM	PCM	Terima H_1	GPCM lebih presisi

Selanjutnya, dilakukan analisis *Standard Error of Measurement* (SEM) dalam kerangka IRT untuk mengevaluasi tingkat ketepatan estimasi kemampuan peserta berdasarkan respons mereka terhadap item item yang dianalisis menggunakan tiga model, yaitu PCM, GPCM, dan GRM. SEM di sini bukan hanya pelengkap, melainkan menjadi indikator penting yang mendukung pengujian hipotesis. Nilai SEM yang lebih rendah menandakan bahwa model memberikan estimasi kemampuan pemecahan masalah matematis yang lebih stabil dan akurat. Oleh karena itu, perbandingan SEM antar model memberikan dasar empiris yang signifikan dalam menilai efektivitas masing-masing model dalam mengukur kemampuan tersebut. Tingginya presisi estimasi sangatlah penting, mengingat

kemampuan pemecahan masalah matematis merupakan kemampuan kompleks yang membutuhkan pengukuran yang sensitif terhadap variasi individu.

1. Model PCM

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} = 0,396$$

2. Model GPCM

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} = 0,334$$

3. Model GRM

$$SEM = \sqrt{\frac{1}{N} \sum_{i=1}^N SE_i^2} = 0,407$$

Berdasarkan hasil perhitungan SEM dari ketiga model, Model GPCM menunjukkan nilai terkecil sebesar 0,334, diikuti oleh PCM dengan 0,396, dan GRM dengan 0,407. Semakin rendah nilai SEM, maka semakin kecil tingkat kesalahan pengukuran yang menyertai estimasi kemampuan, yang berarti model tersebut menghasilkan estimasi yang lebih presisi (Huebner & Skar, 2021b). Presisi semacam ini sangat krusial dalam konteks penilaian kemampuan pemecahan masalah matematis yang melibatkan berbagai proses kognitif tingkat tinggi.

4.8 Pembahasan Hasil Penelitian

4.8.1 Parameter Butir

Dalam penelitian ini, setiap soal yang digunakan untuk mengukur kemampuan pemecahan masalah matematika siswa tidak hanya dinilai dari benar atau salahnya jawaban. Tetapi juga dianalisis lebih dalam, yaitu seberapa sulit soalnya dan seberapa baik soal itu bisa membedakan siswa yang pintar dan yang belum menguasai materi. Inilah yang disebut sebagai parameter butir. Ada dua hal

penting dari parameter butir: Tingkat kesulitan soal (b) Ini menunjukkan apakah soal itu termasuk mudah, sedang, atau sulit. Kalau nilainya tinggi, berarti soal itu sulit dan hanya bisa dijawab oleh siswa dengan kemampuan tinggi. Kalau nilainya rendah, berarti soal itu mudah dan bisa dijawab oleh banyak siswa. Kemampuan membedakan siswa atau daya pembeda (a) Ini menunjukkan apakah soal itu bisa benar-benar membedakan siswa yang sudah paham dengan yang belum. Soal dengan daya pembeda tinggi sangat bagus karena bisa menyeleksi mana siswa yang benar-benar mengerti materi.

Kedua hal ini sangat berkaitan dengan kemampuan siswa. Sistem IRT yang digunakan dalam penelitian ini bekerja dengan cara mencocokkan tingkat kesulitan soal dengan kemampuan siswa (θ). Jadi, semakin cocok soal dengan kemampuan siswa, semakin akurat hasil penilaiannya. Misalnya begini ada seorang siswa yang berada pada kemampuan sedang (θ mendekati 0), maka dia akan lebih mungkin menjawab benar soal yang tingkat kesulitannya juga sedang. Tapi jika dia diberi soal yang sangat sulit, kemungkinan besar dia akan salah. Nah, dari pola jawabannya itu, sistem akan menghitung dan memperkirakan sejauh mana sebenarnya kemampuan siswa tersebut.

Hasil penelitian ini menunjukkan bahwa soal-soal yang digunakan dalam tes sudah sesuai dengan kemampuan siswa SMP. Hal ini terlihat dari rata-rata dan median kemampuan siswa (θ) mendekati angka nol, yang artinya sebagian besar siswa berada pada kemampuan sedang. Di sisi lain, tingkat kesulitan soal juga berada di kisaran yang sama Dimana parameter kesulitan soal (b) yang juga tersebar di kisaran $-0,5$ hingga $+1,0$. Dalam pendekatan IRT, kesesuaian antara nilai theta dan b menunjukkan bahwa soal berada pada tingkat kesulitan yang tepat untuk mengukur kemampuan siswa secara optimal(Hambleton et al., 1991).

Jadi, secara sederhana, parameter butir membantu kita memastikan bahwa soal-soal yang diberikan benar-benar bisa menggambarkan kemampuan siswa secara adil dan tepat. Kalau soal terlalu mudah atau terlalu sulit, hasil penilaian bisa tidak akurat. Tapi dengan bantuan parameter butir, kita bisa menyesuaikan soal

dengan kemampuan siswa sehingga hasil yang diperoleh benar-benar mencerminkan apa yang mereka kuasai.

4.8.2 Fungsi Informasi Tes

Fungsi ini menggambarkan seberapa besar informasi yang diberikan oleh suatu tes dalam mengukur kemampuan siswa pada tingkat tertentu. Semakin tinggi nilai informasi tes di suatu titik kemampuan (θ), maka semakin akurat pula hasil pengukuran di titik tersebut (Hambleton et al 1991).

Dalam penelitian ini, fungsi informasi tes tertinggi ditemukan pada model GPCM yaitu mendekati nilai 16, dan berada stabil pada kisaran kemampuan siswa logit -2 hingga $+2$. Ini menunjukkan bahwa tes paling akurat dalam mengukur siswa dengan kemampuan sedang, karena pada rentang inilah fungsi informasi bekerja secara optimal. Hal ini sangat sesuai dengan kondisi siswa dalam penelitian ini, yang mayoritas berada pada kemampuan rata-rata (θ mendekati 0). Hal ini terjadi Karena fungsi informasi tes sangat dipengaruhi oleh karakteristik butir soal, terutama daya pembeda (a) dan tingkat kesulitan soal (b). Soal-soal yang memiliki daya pembeda tinggi dan tingkat kesulitan yang bervariasi akan menghasilkan informasi yang lebih besar dan akurat. Tes seperti ini mampu membedakan siswa berkemampuan tinggi dan rendah secara lebih jelas (Embretson & Reise, 2000).

Selain itu, fungsi informasi tes juga berkaitan langsung dengan standard error of measurement. Ketika fungsi informasi tinggi, maka standard error menjadi rendah, artinya hasil pengukuran kemampuan siswa menjadi lebih presisi dan dapat dipercaya. Ini sangat penting bagi guru dan pengambil keputusan pendidikan, karena data hasil tes digunakan sebagai dasar untuk menentukan program pembelajaran atau intervensi siswa (Widhiarso 2011).

Dengan kata lain, fungsi informasi tes bukan sekadar indikator statistik, melainkan cerminan dari sejauh mana suatu instrumen benar-benar mampu menilai kemampuan siswa dengan tepat. Tes dengan informasi tinggi memberikan gambaran kemampuan siswa yang lebih tajam dan tidak bias, sehingga hasilnya

lebih adil dan bermakna dalam konteks pembelajaran matematika, khususnya dalam kemampuan pemecahan masalah yang menuntut proses berpikir bertingkat.

4.8.3 Perbandingan Hasil Estimasi Kemampuan Pemecahan Masalah Antara Model PCM dan GRM

Model PCM menunjukkan performa yang layak dalam hal kesesuaian model, dengan p-value sebesar 0,872, yang secara signifikan berada di atas batas signifikansi 0,05. Ini mengindikasikan bahwa PCM merupakan model yang tepat untuk dianalisis dalam penelitian ini, karena tidak ditemukan perbedaan signifikan antara data observasi dan model yang diharapkan secara teoritis.

Dari sisi akurasi estimasi, PCM menghasilkan varians estimasi sebesar 1,257, sedikit lebih rendah dibandingkan GRM yang memiliki varians 1,271. Selain itu, nilai SEM yang diperoleh PCM adalah 0,396, sedikit lebih rendah dari GRM yang sebesar 0,407. Meski perbedaannya kecil, hasil ini mengindikasikan bahwa PCM mampu memberikan estimasi kemampuan yang sedikit lebih stabil dan presisi dibandingkan GRM. Analisis fungsi informasi tes pada PCM juga menunjukkan puncak informasi dengan nilai maksimum informasi sekitar 13 dibandingkan GRM sekitar 6,5. Hal ini menunjukkan bahwa model PCM memiliki akurasi lebih baik dari model GRM.

Untuk mengetahui apakah terdapat perbedaan presisi estimasi secara statistik antara kedua model, dilakukan uji hipotesis ketiga menggunakan uji F terhadap varians estimasi. Berdasarkan hasil pada Tabel 4.15, diperoleh nilai F hitung sebesar 1,011 yang lebih kecil dari F tabel sebesar 1,376 pada taraf signifikansi 5%. Karena $F_{hitung} < F_{tabel}$, maka hipotesis nol diterima, yang berarti tidak ada perbedaan signifikan dalam varians estimasi kemampuan antara PCM dan GRM. Oleh sebab itu, meskipun PCM secara numerik menunjukkan performa yang sedikit lebih baik, secara statistik keduanya setara dalam hal presisi estimasi. Dengan demikian, pemilihan model dapat disesuaikan dengan kebutuhan praktis dan karakteristik data, bukan hanya berdasarkan nilai presisi. Temuan ini sejalan dengan penelitian oleh Santoso et al. (2022), yang juga menemukan bahwa

perbedaan estimasi kemampuan antara PCM dan GRM tidak signifikan secara statistik.

4.8.4 Perbandingan Hasil Estimasi Kemampuan Pemecahan Masalah antara Model GPCM dan GRM

Dalam hal ketepatan estimasi, GPCM mencatatkan varians estimasi kemampuan sebesar 0,902 dan SEM sebesar 0,335, keduanya merupakan yang terendah dibandingkan model PCM dan GRM. Varians yang rendah mencerminkan tingkat konsistensi yang tinggi dalam estimasi, sedangkan SEM yang kecil mengindikasikan tingkat kesalahan pengukuran yang minimal (Huebner & Skar, 2021). Kombinasi keduanya menunjukkan bahwa GPCM memberikan estimasi kemampuan yang lebih presisi dan stabil.

Selain itu, fungsi informasi tes dari GPCM menggambarkan jumlah informasi yang disediakan tes pada berbagai tingkat kemampuan peserta. Fungsi ini sangat penting karena semakin tinggi nilai informasi pada suatu titik kemampuan, semakin presisi pula estimasi yang dihasilkan (Falani et al., 2020). Hasil analisis menunjukkan bahwa fungsi informasi GPCM mencapai puncaknya pada skor antara -2 hingga +2 dengan nilai informasi maksimum sebesar 16. Ini menandakan bahwa GPCM sangat efektif dalam mengukur kemampuan peserta pada rentang kemampuan menengah ke bawah, serta mampu menangkap variasi kemampuan secara lebih komprehensif, khususnya pada kelompok di bawah rata-rata.

Untuk menilai konsistensi presisi estimasi antara GPCM dan GRM, dilakukan pengujian hipotesis kedua menggunakan uji F terhadap varians estimasi. Berdasarkan Tabel 4.13, diketahui bahwa varians estimasi pada GPCM (0,902) lebih kecil daripada GRM (1,271). Uji F menghasilkan nilai F hitung sebesar 1,409, lebih besar dari F tabel sebesar 1,376 pada taraf signifikansi 5%. Karena F hitung $>$ F tabel, maka hipotesis nol ditolak, yang berarti terdapat perbedaan signifikan pada varians estimasi, dengan GPCM secara statistik lebih presisi dibandingkan GRM. Temuan ini didukung oleh hasil studi Dai et al. (2021), yang menunjukkan bahwa GPCM lebih stabil dalam kondisi data tidak lengkap, sementara GRM cenderung kurang akurat pada respons yang hilang. Studi oleh Wang et al. (2002)

juga memperlihatkan bahwa GPCM menghasilkan kesalahan estimasi yang lebih kecil pada instrumen politomus.

Secara keseluruhan, temuan ini menegaskan bahwa GPCM lebih unggul dalam aspek kesesuaian model, stabilitas estimasi, dan ketepatan pengukuran. Dengan mempertimbangkan keunggulan-keunggulan tersebut, GPCM dapat direkomendasikan sebagai model paling efektif dan informatif dalam asesmen berbasis data politomus, terutama dalam mengukur kemampuan Pemecahan Masalah.

4.8.5 Perbandingan Hasil Estimasi Kemampuan Pemecahan Masalah Antara Model GPCM dan PCM

Model PCM dipilih sebagai pendekatan awal untuk menganalisis data politomus. PCM merupakan model dasar dalam teori respons butir yang tidak memasukkan parameter daya beda (a), sehingga seluruh butir dianggap memiliki kemampuan diskriminasi yang sama. Karakteristik ini membuat PCM lebih sederhana dibandingkan model yang mengakomodasi perbedaan daya beda antarbutir. Hasil uji kecocokan model menunjukkan bahwa PCM sesuai untuk digunakan, dengan nilai p -value sebesar 0,876. Nilai tersebut melampaui batas signifikansi umum sebesar 0,05, yang berarti tidak terdapat perbedaan signifikan antara model dan data empiris (Niken et al., 2024), sehingga secara statistik PCM layak digunakan dalam konteks ini.

Namun, dari aspek akurasi estimasi, PCM menghasilkan varians estimasi sebesar 1,257 dan nilai SEM sebesar 0,396. Meskipun nilai SEM ini menunjukkan tingkat kesalahan pengukuran yang masih dalam kategori baik, akurasi model ini tidak setinggi model yang mempertimbangkan parameter daya beda. Karena tidak memperhitungkan parameter a , PCM memberikan bobot informasi yang sama untuk semua butir, terlepas dari seberapa efektif butir tersebut membedakan peserta dengan kemampuan tinggi dan rendah.

Fungsi informasi tes PCM menunjukkan bahwa informasi tertinggi diberikan pada skor sekitar -2 dan +2 dengan puncak nilai sebesar 13. Artinya,

akurasi pengukuran terbaik berada pada peserta dengan kemampuan sedikit di bawah rata-rata. Pola distribusi informasi yang tidak merata ini menjadi pertimbangan penting, khususnya ketika tujuan pengukuran adalah untuk memperoleh estimasi yang akurat di seluruh rentang kemampuan. Hal ini sejalan dengan temuan Aji & Retnawati (2024), yang menyatakan bahwa GPCM menyajikan distribusi informasi yang lebih merata dibanding PCM, sehingga lebih efektif dalam menghasilkan estimasi yang presisi di berbagai tingkat kemampuan. Karena PCM hanya mempertimbangkan tingkat kesulitan antar kategori dalam satu butir, maka informasi yang dihasilkan cenderung terkonsentrasi pada rentang tertentu. Dalam konteks ini, PCM menjadi kurang optimal untuk mengukur kemampuan secara merata di seluruh skala (Sumintono & Widhiarso, 2015). Oleh karena itu, meskipun secara statistik PCM dapat diterapkan, kebutuhan terhadap presisi pengukuran dapat mendorong penggunaan model alternatif seperti GPCM atau GRM yang memasukkan parameter diskriminasi.

Untuk menilai keunggulan presisi estimasi antara PCM dan GPCM, dilakukan uji hipotesis terhadap perbedaan varians hasil estimasi kemampuan. Diketahui bahwa PCM menghasilkan varians sebesar 1,257, sedangkan GPCM hanya sebesar 0,902. Uji F menghasilkan nilai F hitung sebesar 1,393 yang lebih besar dari F tabel pada taraf signifikansi 5% (1,376), sehingga hipotesis nol ditolak. Ini menunjukkan bahwa GPCM memiliki varians yang secara signifikan lebih kecil, menandakan konsistensi yang lebih baik dalam mengestimasi kemampuan peserta. Hasil ini diperkuat dengan nilai SEM GPCM yang lebih rendah (0,335), menegaskan bahwa GPCM lebih presisi dan andal dalam konteks data penelitian ini. Hal ini juga didukung oleh temuan Luo (2018) dan Muraki (1992), yang menyatakan bahwa GPCM mampu memberikan gambaran yang lebih akurat tentang kemampuan individu berdasarkan pola respons mereka, sehingga hasil estimasi menjadi lebih representatif dan informatif.

4.8.6 Deskriptif Statistik Kemampuan Pemecahan Masalah Matematis

Siswa

Kemampuan pemecahan masalah matematika siswa dianalisis berdasarkan empat indikator utama menurut tahapan Polya (1973), yaitu identifikasi masalah, merumuskan strategi, menjalankan strategi, dan memeriksa kembali. Analisis dilakukan terhadap 108 siswa dan ditampilkan secara kuantitatif untuk menunjukkan proporsi penguasaan kemampuan secara nyata.

4.8.6.1 Identifikasi Masalah

Berdasarkan hasil analisis, terlihat bahwa sebanyak 37 siswa (34,3%) tidak mampu mengidentifikasi permasalahan dengan benar (skor 0). Sementara itu, 22 siswa (20,4%) hanya mampu mengenali sebagian unsur penting (skor 1). Di sisi lain, 37 siswa (34,3%) mulai menunjukkan pemahaman cukup (skor 2), dan hanya 12 siswa (11,1%) mampu mengidentifikasi masalah secara utuh (skor 3).

Realitasnya, hanya 11,1% siswa yang menunjukkan kemampuan optimal pada tahap awal pemecahan masalah, sedangkan selebihnya masih berada di bawah tingkat penguasaan penuh. Hal ini mengindikasikan bahwa sebagian besar siswa belum memiliki keterampilan dasar yang kuat untuk memahami inti permasalahan matematika. Padahal, tanpa pemahaman yang tepat di tahap ini, siswa akan kesulitan merumuskan strategi penyelesaian yang efektif.

4.8.6.2 Merumuskan Strategi

Pada tahap ini, 50 siswa (46,3%) menunjukkan kemampuan cukup dalam menyusun strategi penyelesaian (skor 2), namun 33 siswa (30,6%) hanya mampu menyusun sebagian strategi, dan 19 siswa (17,6%) tidak mampu merumuskan strategi sama sekali. Hanya 6 siswa (5,6%) yang menunjukkan kemampuan merancang strategi secara lengkap dan logis (skor 3).

Data ini mencerminkan bahwa sebanyak 5,6 % siswa yang mampu merancang strategi dengan benar selebihnya siswa berada pada kategori rendah dalam merumuskan strategi, mengindikasikan adanya hambatan dalam menghubungkan

pemahaman masalah ke tahap rencana solusi. Realitas ini menunjukkan bahwa kebanyakan siswa belum mampu menyusun langkah-langkah penyelesaian secara sistematis dan efektif, sehingga sangat mungkin mereka akan mengalami kesulitan pada tahap pelaksanaan.

4.8.6.3 Menjalankan Strategi

Hasil menunjukkan bahwa 37% siswa tidak dapat menjalankan strategi sama sekali, dan 32,4% hanya mampu melaksanakan sebagian langkah penyelesaian. Hanya 7 siswa (6,5%) yang mampu mengeksekusi strategi secara logis dan lengkap (skor 3).

Realitasnya, kebanyakan siswa gagal menjalankan strategi secara memadai, menjadikan tahap ini sebagai titik lemah utama dalam pemecahan masalah matematika siswa. Meskipun beberapa siswa bisa menyusun strategi, masih terjadi kesenjangan antara perencanaan dan pelaksanaan. Hal ini mengindikasikan perlunya penguatan pada aspek keterampilan teknis dan prosedural dalam menyelesaikan masalah.

4.8.6.4 Memeriksa Kembali

Tahap refleksi menunjukkan sedikit perbaikan. Sebanyak 58 siswa (53,8%) menunjukkan kesadaran dalam memeriksa kembali hasil kerja mereka (gabungan skor 2 dan 3). Namun demikian, 50 siswa (46,3%) masih belum melakukan refleksi secara memadai, bahkan 17 siswa (15,7%) sama sekali tidak memeriksa kembali hasil pekerjaannya (skor 0).

Realitas di lapangan menunjukkan bahwa hampir separuh siswa masih lemah dalam kesadaran metakognitif, meskipun refleksi merupakan elemen penting untuk memastikan akurasi dan efektivitas strategi yang digunakan. Hal ini menunjukkan bahwa keterampilan berpikir tingkat tinggi seperti evaluasi dan refleksi belum menjadi kebiasaan dalam menyelesaikan masalah matematika.

Secara umum, mayoritas siswa mengalami kesulitan di hampir semua tahapan pemecahan masalah, terutama pada tahap pelaksanaan strategi. Hanya sebagian kecil siswa yang mampu melewati seluruh tahapan secara utuh dan benar. Fakta ini

menjadi dasar penting bahwa pendekatan pembelajaran yang digunakan saat ini belum sepenuhnya mengembangkan keterampilan pemecahan masalah secara menyeluruh, dan perlu dilakukan intervensi melalui pembelajaran berbasis masalah, scaffolding, serta pelatihan berpikir reflektif.

4.8.7 Peningkatan akurasi kemampuan Pemecahan Masalah Berdasarkan Realitas Data

Estimasi kemampuan siswa (θ) yang diperoleh melalui pendekatan *Item Response Theory* (IRT) menunjukkan distribusi yang konsisten di antara ketiga model yang digunakan (PCM, GPCM, dan 2PL). Mayoritas siswa berada pada kategori kemampuan sedang, yakni sekitar ± 1 dari nilai tengah, dengan proporsi lebih dari 65%, sebagaimana ditunjukkan dalam Tabel 4.10.

Nilai rata-rata estimasi kemampuan yang sangat dekat dengan nol ($\mu \approx 0$) dan standar deviasi mendekati 1 ($\sigma \approx 1$) mencerminkan bahwa skala kemampuan siswa telah terstandarisasi secara proporsional. Kondisi ini menunjukkan bahwa instrumen pengukuran bersifat adil (*fair*), tidak memberikan keuntungan maupun kerugian terhadap kelompok siswa dengan kemampuan ekstrem, baik sangat rendah maupun sangat tinggi. Dengan demikian, instrumen memiliki distribusi normal yang simetris, yang menjadi salah satu indikator kualitas pengukuran dalam IRT (Hambleton, Swaminathan, & Rogers, 1991).

Berbeda dengan pendekatan klasik yang mengandalkan total skor mentah, pendekatan IRT menghasilkan estimasi kemampuan individual (θ) yang tidak tergantung pada skor total, tetapi berdasarkan pola respons siswa terhadap butir soal. Oleh karena itu, siswa yang mampu menjawab soal dengan tingkat kesulitan tinggi akan mendapatkan nilai θ yang lebih tinggi, meskipun secara jumlah benar mungkin setara dengan siswa lain yang hanya menjawab soal mudah. Hal ini memberikan keunggulan IRT dalam mengungkap kemampuan sejati siswa (*true ability*) secara lebih akurat dan adaptif (Embretson & Reise, 2000).

Sebagai contoh, pada responden No 1 dan 3 mereka menjawab dengan sekor untuk tiap item (2,2,2,2,2,3,2,1) untuk responden 1 dan (1,1,2,2,3,3,2,2) untuk responden

3, jika kita menggunakan pendekatan klasik ditemukan bahwa siswa memperoleh skor total yang serupa yaitu 66,66 didapatkan dengan membagi jumlah skor dengan skor maksimal kemudian kita kalikan dengan 100, namun pada pendekatan IRT nilai estimasi θ berbeda secara signifikan bisa dilihat pada lambiran 5 hasil estimasi kemampuan dengan model GPCM dimana responden 1 nilai estimasi θ sekitar 0,65 setelah di konversikan sekornya menjadi 58,13 dan responden 3 nilai estimasi θ sekitar 0,50 setelah di konvresikan menjadi 56,25. Hal ini ini terjadi karanal IRT mempertimbangkan kualitas butir soal yang dijawab (berdasarkan parameter kesulitan dan daya beda), bukan sekadar kuantitas jawaban benar. Dengan demikian, model IRT bisa memberikan pengukuran yang lebih akurat, dan juga lebih adil secara pengukuran, terutama dalam konteks pembelajaran yang menekankan keunikan individu.

Temuan ini mendukung hasil penelitian sebelumnya oleh DeMars (2010) yang menyatakan bahwa model IRT memberikan estimasi kemampuan yang stabil dan tidak bias, terutama dalam populasi heterogen. Selain itu, hasil estimasi ini juga menunjukkan bahwa alat ukur memiliki kecocokan yang baik terhadap populasi sampel, karena distribusi kemampuan siswa mencerminkan normalitas, sebagaimana yang diharapkan dalam pengukuran berbasis model IRT.

Secara keseluruhan, karakteristik butir yang baik dengan daya beda optimal dan distribusi tingkat kesulitan yang proporsional akan menghasilkan pengukuran yang valid, reliabel, dan adil. Butir-butir yang mampu membedakan siswa secara tepat menghasilkan skor yang lebih bermakna dan lebih relevan dengan kemampuan kognitif siswa, serta meningkatkan kualitas instrumen penilaian secara keseluruhan. Temuan ini konsisten dengan studi oleh Reise & Waller (2009) yang menekankan pentingnya pemilihan butir dengan daya beda tinggi dalam merancang instrumen psikometrik yang sensitif dan informatif.

BAB V

PENUTUP

5.1 Simpulan

Berdasarkan hasil analisis dan pembahasan dalam penelitian ini mengenai implementasi pendekatan Item Response Theory (IRT) terhadap akurasi penilaian kemampuan pemecahan masalah matematis siswa SMP, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Siswa mengalami berbagai kesulitan dalam setiap tahap pemecahan masalah menurut Polya. Pada tahap memahami masalah, beberapa siswa tidak mampu mengidentifikasi informasi penting dalam soal. Pada tahap merencanakan penyelesaian, siswa mengalami kendala dalam merumuskan strategi yang tepat. Pada tahap melaksanakan rencana, kesalahan prosedural masih sering ditemukan. Sementara pada tahap melakukan pengecekan kembali, mayoritas siswa tidak mampu melakukan refleksi atau verifikasi terhadap solusi yang telah dibuat. Hal ini menunjukkan bahwa kemampuan pemecahan masalah siswa masih perlu ditingkatkan secara menyeluruh.
2. Model Generalized Partial Credit Model (GPCM) menghasilkan estimasi kemampuan yang paling akurat dibandingkan dengan model Partial Credit Model (PCM) dan Graded Response Model (GRM). GPCM menunjukkan nilai varians dan standard error terendah serta fungsi informasi tertinggi. Hal ini menunjukkan bahwa GPCM lebih tepat digunakan dalam menilai kemampuan pemecahan masalah karena dapat menangkap keragaman tingkat kemampuan siswa secara lebih sensitif dan rinci.
3. Distribusi kemampuan siswa yang berada di sekitar logit 0 menunjukkan bahwa mayoritas siswa memiliki kemampuan dalam rentang rata-rata, dan model GPCM sangat efektif dalam mengukur kemampuan pada rentang tersebut. Selain itu, perbedaan hasil klasifikasi kemampuan antar model menunjukkan bahwa pemilihan model IRT yang digunakan dalam penilaian sangat memengaruhi hasil interpretasi kemampuan siswa. GPCM dinilai

paling sesuai karena dapat menangkap proses berpikir siswa secara bertahap melalui penskoran politomi.

4. Akurasi penilaian yang tinggi memberikan dasar yang kuat untuk diagnosis kemampuan siswa secara individual. Guru dapat mengetahui dengan tepat bagian mana dari proses pemecahan masalah yang belum dikuasai siswa, sehingga strategi pembelajaran atau remediasi dapat disesuaikan secara lebih efektif.

Dengan mempertimbangkan keseluruhan hasil tersebut, penelitian ini menyimpulkan bahwa implementasi IRT, khususnya GPCM, mampu meningkatkan akurasi penilaian kemampuan pemecahan masalah siswa dan sejalan dengan tuntutan penilaian modern berbasis proses serta pengembangan kompetensi abad ke-21.

5.2 Implikasi

Penelitian ini menunjukkan bahwa guru dapat memperoleh informasi kemampuan siswa yang lebih akurat jika menggunakan model IRT, khususnya GPCM. Dengan informasi ini, guru bisa memberikan intervensi pembelajaran yang lebih tepat, misalnya membantu siswa yang lemah di tahap merancang strategi atau memeriksa kembali jawabannya. Dengan adanya model penilaian yang akurat, siswa mendapatkan hasil asesmen yang lebih adil dan sesuai dengan kemampuan mereka. Ini akan meningkatkan kepercayaan diri dan motivasi belajar siswa karena mereka dinilai berdasarkan proses berpikir, bukan hanya hasil akhir.

5.3 Saran

Berdasarkan Kesimpulan dan implikasi diatas, maka dapat disusun beberapa saran sebagai berikut:

1. Untuk penelitian selanjutnya disarankan untuk memperluas objrk penelitian dengan menggunakan variasi materi Matematika lainnya, atau menggunakan Model IRT lainnya pada Tingkat Pendidikan yang berbeda agar terlihat kosistensi hasilnt

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. Untuk guru diperlukan pelatihan penggunaan model IRT secara teknis agar guru dapat melakukan analisis butir dan kemampuan siswa secara mandiri, hal ini penting untuk memperbaiki kualitas penilaian.

DAFTAR PUSTAKA

- Agustina, Lisna. 2016. "Upaya Meningkatkan Kemampuan Pemahaman Konsep Dan Pemecahan Masalah Matematika Siswa SMP Negeri 4 Sipirok Kelas VII Melalui Pendekatan Matematika Realistik (PMR)." *EKSAKTA: Jurnal Penelitian Dan Pembelajaran MIPA* 1 (1).
- Anderson, Lorin W, and David R Krathwohl. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. Addison Wesley Longman, Inc.
- Baker, Cynthia, Judith Wuest, and Phyllis Noerager Stern. 1992. "Method Slurring: The Grounded Theory/Phenomenology Example." *Journal of Advanced Nursing* 17 (11): 1355–60.
- Baker, Frank B, and Seock-Ho Kim. 2017. *The Basics of Item Response Theory Using R*. Vol. 10. Springer.
- Black, Paul, and Dylan Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice* 5 (1): 7–74.
- Brookhart, Susan M. 2014. *How to Design Questions and Tasks to Assess Student Thinking*. ASCD.
- Ciptari, Putu Aulia, Ni Kadek Rini Purwanti, and Ni Ketut Erawati. 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, Shenghai, Thao Thu Vo, Olasunkanmi James Kehinde, Haixia He, Yu Xue, Cihan Demir, and Xiaolin Wang. 2021. "Performance of Polytomous IRT Models with Rating Scale Data: An Investigation over Sample Size, Instrument Length, and Missing Data." In *Frontiers in Education*, 6:721963. Frontiers Media SA.
- Embretson, Susan E, and Steven Paul Reise. 2000. "Item Response Theory for Psychologists Multivariate." London, UK: Erlbaum Publishers.
- Fajrianti, Fajrianti, Wiwin Hendriani, and Berlian Gressy Septarini. 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, Ilham. 2020. "Kombinasi Model Penyekoran Item Response Theory Pada Tes Format Campuran." Universitas Negeri Jakarta.
- Falani, Ilham, Maruf Akbar, and Dali S Naga. 2020. "The Precision of Students' Ability Estimation on Combinations of Item Response Theory Models."

International Journal of Instruction 13 (4): 545–58.

- Falani, Ilham, and Siti Ayu Kumala. 2017. “Kestabilan Estimasi Parameter Kemampuan Pada Model Logistik Item Response Theory Ditinjau Dari Panjang Tes.” *SAP (Susunan Artikel Pendidikan)* 2 (2).
- Fernanda, Jerhi Wahyu, and Noer Hidayah. 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, Daniati, Mujahidawati Mujahidawati, and Husni Sabil. 2025. “Deskripsi Media Video Pembelajaran Menggunakan Pobleem Based Learning Untuk Meningkatkan Kemampuan Berpikir Kritis Matematis.” *Kognitif: Jurnal Riset HOTS Pendidikan Matematika* 5 (2).
- Frey, Felix. 2020. “Test Theory and Classical Test Theory.” *The International Encyclopedia of Media Psychology*, 1–6. <https://doi.org/10.1002/9781119011071.iemp0047>.
- Hambleton, Ronald K, Hariharan Swaminathan, and H Jane Rogers. 1991. *Fundamentals of Item Response Theory*. Vol. 2. Sage.
- Hardianti, Hardianti, Winny Liliawati, and Yuyu Rachmat Tayubi. 2021. “Karakteristik Tes Kemampuan Berpikir Kritis Siswa SMA Pada Materi Momentum Dan Impuls: Perbandingan Classical Theory Test (CTT) Dan Model Rasch.” *WaPFI (Wahana Pendidikan Fisika)* 8 (1): 21–28.
- Hayat, Bahrul. 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>.
- Herliana, Herliana, Dra Mujahidawati, and Feri Tiona Pasaribu. 2021. “Pengembangan LKPD Berbasis Problem Based Learning Berbantuan Cabri 3D Untuk Mendukung Kemampuan Pemecahan Masalah Pada Materi Bangun Ruang Sisi Datar Kelas VIII SMP.” Universitas Jambi.
- Herwin, Herwin, Riana Nurhayati, Aprilia Tina Lidyasari, and Augusto da Costa. 2023. “Graded Response Models on the Curiosity Measurement of Elementary School Students.” *International Journal of Educational Methodology* 9 (1): 53–62.
- Kim, Jinho, and Mark Wilson. 2020. “Polytomous Item Explanatory IRT Models with Random Item Effects: Concepts and an Application.” *Measurement* 151:107062.
- Mariati, Ida. 2009. “Analisis Butir Soal Dengan Teori Tes Klasik (Classical Test Theory) Dan Teori Respons Butir (Item Response Theory)(Studi Kasus: Soal Ujian Olimpiade Sains Provinsi Bidang Informatika 2009).” *PYTHAGORAS Jurnal Pendidikan Matematika* 5 (2).

- Masters, Geoff N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47 (2): 149–74.
- Mulyasa, Enco. 2008. *Implementasi Kurikulum Tingkat Satuan Pendidikan: Kemandirian Guru Dan Kepala Sekolah*. Bumi Aksara.
- Muraki, Eiji. 1992. "A Generalized Partial Credit Model: Application of an EM Algorithm." *Applied Psychological Measurement* 16 (2): 159–76.
- Newell, Allen, and Herbert Alexander Simon. 1972. *Human Problem Solving*. Vol. 104. Prentice-hall Englewood Cliffs, NJ.
- Nisrina, Desy, Gugun M Simatupang, and Dra Mujahidawati. 2021. "Pengaruh Model Problem Solving Dalam Pembelajaran Jarak Jauh Terhadap Kemampuan Pemecahan Masalah Matematis Siswa Pada Materi Pola Bilangan Di Kelas VIII MTs Negeri 5 Kota Jambi." Universitas Jambi.
- Nitko, Anthony J. 1996. *Educational Assessment of Students*. ERIC.
- Nuriyati, Nuriyati, and Supriadi Supriadi. 2022. "Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa Sekolah Dasar Melalui Pembelajaran Etnomatematika Sunda Dengan Media Congklak." *Jurnal Pendidikan Tambusai* 6 (2): 15473–81.
- Ogunsakin, I B, and Y O Shogbesan. 2018. "Item Response Theory (IRT): A Modern Statistical Theory for Solving Measurement Problem in 21st Century." *International Journal of Scientific Research in Education (IJSRE)* 11:627–35.
- Retnawati, Heri, Samsul Hadi, and Ariadie Chandra Nugraha. 2016. "Vocational High School Teachers' Difficulties in Implementing the Assessment in Curriculum 2013 in Yogyakarta Province of Indonesia." *International Journal of Instruction* 9 (1): 33–48.
- Sarea, Muh Syahrul, and Rosnia Ruslan. 2019. "KARAKTERISTIK BUTIR SOAL: CLASSICAL TEST THEORY VS ITEM RESPONSE THEORY?" *Didaktika: Jurnal Kependidikan* 13 (1): 1–16.
- Septiani, Astrid Nur, Desti Pratiwi, and Rossy Rossy. 2023. "Evaluasi Pembelajaran Dalam Meningkatkan Mutu Pendidikan Di Sekolah Dasar." *Masaliq* 3 (5): 824–32. <https://doi.org/10.58578/masaliq.v3i5.1380>.
- Septikasari, Resti, Fajeri Inayah, Nova Amirotul Husniyah, and Riska Marfia Rini. 2023. "Teknik Penilaian Tes Dan Non Tes." *Madani: Jurnal Ilmiah Multidisiplin* 1 (11).
- Slavin, Robert E. 2018. *Educational Psychology: Theory and Practice*. Pearson.
- Sözer, Esra, and Nilüfer Kahraman. 2021. "Investigation of Psychometric Properties of Likert Items with Same Categories Using Polytomous Item Response Theory Models." *Journal of Measurement and Evaluation in Education and Psychology* 12 (2): 129–46.

- Sumintono, Bambang, and Wahyu Widhiarso. 2014. *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial (Edisi Revisi)*. Trim Komunikata Publishing House.
- Ul Hassan, Mahmood, and Frank Miller. 2019. "Optimal Item Calibration for Computerized Achievement Tests." *Psychometrika* 84 (4): 1101–28.
- Widhiarso, Wahyu. 2011. "Melibatkan Rater Dalam Pengembangan Alat Ukur." *Available at SSRN 2593452*.

Lampiran 1 Sintak GPCM

SINTAK GPCM

```
>FILES  DFNAME='GPCM.DAT', SAVE;  
  
>SAVE  SCORE='GPCM.SCO', PARM='GPCM.PAR';  
  
>INPUT  NIDW=4, NTOTAL=8, NTEST=1, LENGTH=8;  
  
(4A1,5X,8A1)  
  
>TEST  TNAME='SCALE1', ITEM=(1(1)8), NBLOCK=8;  
  
>BLOCK1  REPEAT=8, 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 2 Sintak PCM

SINTAK PCM

>COMMENTS

PCM

>FILE DFNAME=PCM1.DAT', SAVE;

>SAVE SCORE='PCM1.SCO', PARM='PCM1.PAR';

>INPUT NIDW=4, NTOTAL=8, NTEST=1, LENGTH=8;

(4A1,5X,8A1)

>TEST TNAME='SCALE1', ITEM=(1(1)8), NBLOCK=8;

>BLOCK1 REPEAT=8, 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 3 Sintak GRM

SINTAK GRM

```
>FILE  DFNAME='GRM1.DAT', SAVE;  
  
>SAVE  PARM='GRM1.PAR', SCORE='GRM1.SCO';  
  
>INPUT  NIDW=4, NTOTAL=8, NTEST=1, LENGTH=8, NFMT=1;  
  
(4A1,5X,8A1)  
  
>TEST  TNAME='SCALE1', ITEM=(1(1)8), NBLOCK=8;  
  
>BLOCK1  REPEAT=8, 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 tes Matematika

FORM VALIDASI INSTRUMEN TES MATEMATIKA



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

A. Tentang Penelitian

- Judul penelitian : Implementasi *Item Response Theory* (IRT) dalam Peningkatan Akurasi Penilaian Kemampuan Pemecahan Masalah Matematis Siswa SMP
- Format Tes : Essay (3)
- Level : Kelas IX
- Kurikulum : Merdeka
- Materi : Sistem Persamaan Linear Dua Variabel
- Sampel Penelitian : 108 responden.
- Model Penyekoran : 1. *Partial Credit Model* (PCM)
2. *Generalized Partial Credit Model* (GPCM)
3. *Graded Response Model* (GRM)
- Fokus Penelitian : 1. Kecocokan Model
2. Hasil Estimasi Karakteristik Butir dan Kemampuan pemecahan masalah matematis
3. *Standard Error Measurement* (SEM) Pengukuran Kemampuan pemecahan masalah

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.
 - e. Kesesuaian soal dan penyekoran.
3. Memberikan penilaian dengan memberikan tanda cek (√) 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 AHLI/PAKAR

Nama Ahli/Pakar :

Jenis Kelamin : **Laki-laki** **Perempuan**

Pendidikan Terakhir : **S1** **S2** **S3**

Bidang Keahlian :

Institusi :

.....,/...../202

(.....)

KISI-KISI INSTRUMEN TES KEMAMPUAN PEMECAHAN MASALAH

No	Capaian Pembelajaran (CP)	Tujuan Membelajaran (TP)	Indikator kemampuan pemecahan masalah	Indikator soal	Level kognitif	No soal
	Pada akhir fase D, peserta didik dapat menyelesaikan masalah kontekstual peserta didik dengan menggunakan konsep-konsep dan keterampilan matematika yang dipelajari pada fase ini. Mereka dapat menyajikan dan	Siswa mampu menyelesaikan permasalahan sehari hari dngn menggunakan pendekatan SPLDV dengan menggunakan metode yang telah di tentukan	<ul style="list-style-type: none"> • Memahami masalah (understand the problem). • Membuat rencana pemecahan masalah (make a plan) • Melaksanakan rencana (carry out our plan) 	Diberikan sebuah soal permasalahan mengenai SPLDV kemudian siswa minta untuk menyelesaikan masalah dengan menggunakan metode eliminasi, substitusi, ataupun gabungan	C3	1,2

	menyelesaikan persamaan dan pertidaksamaan linier satu variabel dan sistem persamaan linier dengan dua variabel dengan beberapa cara, memahami dan menyajikan relasi dan fungsi		<ul style="list-style-type: none"> • Memeriksa kembali jawaban 			
	menyelesaikan persamaan dan pertidaksamaan linier satu variabel dan sistem persamaan linier dengan dua variabel dengan beberapa cara, memahami dan menyajikan relasi dan fungsi	Siswa mampu menerapkan SPLDV pada bidang datar.	<ul style="list-style-type: none"> • Memahami masalah (understand the problem). • Membuat rencana pemecahan masalah (make a plan) • Melaksanakan rencana (carry out our plan) • Memeriksa kembali jawaban 	diberikan dua buah persamaan titik kemudian siswa di minta untuk mencari titik potong dengan penerapan SPLDV	C4	3

FORM PENILAIAN INSTRUMEN TES

Satuan Pendidikan : SMP/MTs

Alokasi Waktu : 120 menit

Mata Pelajaran : Matemátika.

Jumlah Soal : 3 butir (Essay)

Kelas/Semester : IX(sembilan)/genap

Bentuk Tes : Tertulis

Kurikulum : 2025

Penyusun : Aditya Prayogi

Tujuan pembelajaran	Indikator soal	Soal	1	2	3	4	5	Saran
Siswa mampu menyelesaikan permasalahan sehari hari dngn menggunakan pendekatan SPLDV dengan	Diberikan sebuah soal permasalahan seorang petani, kemudian siswa diminta untuk menghitung	Seorang petani memiliki dua jenis tanaman, yaitu tanaman jagung dan tanaman padi. Dalam satu bulan, petani tersebut menghasilkan total 800 kilogram hasil panen. Setiap tanaman jagung menghasilkan 5 kilogram per pohon dan tanaman padi menghasilkan 3 kilogram per pohon. Jika						

menggunkan metode eliminasi	banyak pohon yang ditanami oleh pentane	jumlah total pohon jagung dan padi yang ditanam oleh petani tersebut adalah 180 pohon, tentukan banyak pohon jagung dan pohon padi yang ditanam oleh petani tersebut!						
-----------------------------	---	---	--	--	--	--	--	--

Tujuan pembelajaran	Indakor soal	Soal	1	2	3	4	5	Saran
Siswa mampu menyelesaikan masalah SPLDV dengan menggunakan metode substitusi	Diberikan sebuah permasalahan di took buku kemudizn siswa diminta untuk mencari banyak	Sebuah toko buku menjual dua jenis buku, Buku Fiksi dan Buku Non-Fiksi. Buku Fiksi dijual dengan harga Rp. 80.000 per buah dan Buku Non-Fiksi dijual dengan harga Rp. 120.000 per buah. Jika total penjualan kedua jenis buku tersebut sebanyak 50 buku						

	masing masing buku yang terjual	menghasilkan pendapatan Rp. 4.400.000, berapa banyak masing-masing buku yang terjual?					
Siswa mampu menerapkan SPLDV pada bidang dan ruang	diberikan duan buah persamaan titik kemudian siswa di minta untuk mencari titik potong dengan penerapan SPLDV	<p>Dua garis memiliki persamaan berikut:</p> $y = 2x + 3$ $y = -x + 6$ <p>Tentukan titik perpotongan kedua garis tersebut!</p> <p>Petunjuk Penyelesaian:</p> <ul style="list-style-type: none"> Gunakan metode eliminasi atau substitusi untuk menemukan nilai x dan y. 					

		<ul style="list-style-type: none"> Titik perpotongan adalah solusi dari SPLDV tersebut 						
--	--	---	--	--	--	--	--	--

Rubrik penyekoran kemampuan Pemecahan Masalah Matematis

NO	soal	Penyelesaian	Indicator pemecahan masalah	deskripsi	pensekoran
1	Seorang petani memiliki dua jenis tanaman, yaitu tanaman jagung dan tanaman padi. Dalam	<p>Memahami masalah (Understand the problem)</p> <p>Diketahui bahwa:</p>	<p>Memahami masalah (Understand the problem)</p>	Menuliskan apa yang diketahui dan ditanyakan dengan benar	3

<p>satu bulan, petani tersebut menghasilkan total 800 kilogram hasil panen. Setiap tanaman jagung menghasilkan 5 kilogram per pohon dan tanaman padi menghasilkan 3 kilogram per pohon. Jika jumlah total pohon jagung dan padi yang ditanam oleh petani tersebut adalah 180 pohon, tentukan banyak pohon jagung dan pohon padi yang</p>	<ul style="list-style-type: none"> • Setiap pohon jagung menghasilkan 5 kg dan setiap pohon padi menghasilkan 3 kg. • Total hasil panen adalah 800 kg. • Jumlah pohon jagung dan padi yang ditanam totalnya adalah 180 pohon. <p>Ditannya: jumlah pohon jagung dan pohon padi</p>		Menuliskan apa yang diketahui dan ditanyakan, tetapi salah satunya salah	2
			Menuliskan apa yang diketahui dan ditanya tetapi keduanya salah	1
			Tidak menuliskan apa yang diketahui dan ditanyakan	0
	<p>Menyusun rencana (Devise a plan)</p> <p>Misalkan</p> <ul style="list-style-type: none"> • jumlah pohon jagung adalah x • jumlah pohon padi adalah y. 	<p>Menyusun rencana (Devise a plan)</p>	Menuliskan rencana penyelesaian dengan benar	3
			Menuliskan rencana	2

	ditanam oleh petani tersebut!	<p>Maka, kita dapat membuat dua persamaan:</p> <ul style="list-style-type: none"> Persamaan pertama: total jumlah pohon jagung dan padi adalah 180 pohon. $x + y = 180$ Persamaan kedua: total hasil panen adalah 800 kg. $5x + 3y = 800$ 		penyelesaian, tetapi hanya sebagian yang benar	
				Menuliskan rencana penyelesaian, tetapi semuanya tidak tepat	1
				Tidak ada rencana sama sekali	0
		<p>Melaksanakan rencana (Carry out the plan)</p> <p>Kita akan mengeliminasi variable y terlebih dahulu:</p> <p>Samakan koefisien variable Y</p>	<p>Melaksanakan rencana (Carry out the plan)</p>	Membuat penyelesaian dengan benar, tetapi tidak lengkap	3

		<p>Dari persamaan pertama:</p> $x + y = 180 \text{ (dikalika 3)}$ $3x + 3y = 540$ <p>Dari persamaan kedua</p> $5x + 3y = 800$ <p>Eliminasi kedua persamaan</p> $(5x + 3y) - (3x + 3y) = 800 - 540$ $2x = 260$ $x = 130$ <p>Kemudian kita eliminasi lagi untuk variable x</p> <p>Samakan koefisien variable x</p> <p>Dari persamaan pertama:</p>			
				Membuat penyelesaian, tetapi sebagian jawaban salah	2
				Membuat penyelesaian, tetapi seluruh jawaban salah	1

		$x + y = 180$ (<i>dikalika 5</i>) $5x + 5y = 900$ Dari persamaan kedua $5x + 3y = 800$ Eliminasi kedua persamaan $(5x + 3y) - (5x + 5y) = 800 - 900$ $-2y = -100$ $y = 50$ Jadi, jumlah pohon jagung yang ditanam adalah 130 pohon, dan jumlah pohon padi yang ditanam adalah 50 pohon.			
				Tidak ada proses penyelesaian masalah sama sekali	0
				Ada pemeriksaan hasil penyelesaian	3

		<p>Melakukan pemeriksaan (Look back)</p> <p>Periksa hasilnya dengan substitusi ke dalam kedua persamaan:</p> <ul style="list-style-type: none"> • Jumlah pohon jagung dan padi: $130 + 50 = 180$ (benar). • Total hasil panen: $5(130) + (50) = 650 + 150 = 800$. 	<p>Melakukan pemeriksaan (Look back)</p>	<p>masalah yang relevan dan benar</p>	
				<p>Ada pemeriksaan hasil penyelesaian masalah, tetapi hanya sebagian yang relevan</p>	2
				<p>Ada pemeriksaan hasil penyelesaian masalah, tetapi tidak semuanya relevan</p>	1
				<p>Tidak ada pemeriksaan hasil penyelesaian</p>	0

2	Sebuah toko buku menjual dua jenis buku, Buku Fiksi dan Buku Non-Fiksi. Buku Fiksi dijual dengan harga Rp. 80.000 per buah dan Buku Non-Fiksi dijual dengan harga Rp. 120.000 per buah. Jika total penjualan kedua jenis buku tersebut sebanyak 50 buku menghasilkan pendapatan Rp. 4.400.000, berapa	Memahami Masalah	Memahami masalah (Understand the problem)	Menuliskan apa yang diketahui dan ditanyakan dengan benar	3
		Diketahui:		Menuliskan apa yang diketahui dan ditanyakan, tetapi salah satunya salah	2
		• Buku Fiksi dijual dengan harga Rp. 80.000 per buku		Menuliskan apa yang diketahui dan ditanya tetapi keduanya salah	1
		• Buku Non- Fiksi dijual dengan harga Rp. 120.000 per buku		Tidak menuliskan apa yang diketahui dan ditanyakan	0
	• Total penjualan kedua jenis buku: 50 buku				
		• Total pendapatan: Rp. 4.400.000			
		Ditanyakan:			
		• Berapa banyak masing-masing buku yang terjual?			
		Menyusun Rencana	Menyusun rencana	Menuliskan rencana	3

banyak masing-masing buku yang terjual?	Buat persamaan berdasarkan informasi yang diberikan: Misalkan: <ul style="list-style-type: none"> • x adalah jumlah Buku Fiksi yang terjual • y adalah jumlah Buku Non-Fiksi yang terjual Buat dua persamaan: Persamaan jumlah buku: $x + y = 5$ Persamaan total pendapatan: $80000x + 120000y = 4400000$	(Devise a plan)	penyelesaian dengan benar	
			Menuliskan rencana penyelesaian, tetapi hanya sebagian yang benar	2
			Menuliskan rencana penyelesaian, tetapi semuanya tidak tepat	1
			Tidak ada rencana sama sekali	0
	Melaksanakan Rencana		Membuat penyelesaian	3

		<p>Gunakan metode substitusi untuk menyelesaikan sistem persamaan tersebut:</p> <p>Dari persamaan $x + y = 50$</p> $y = 50 - x$ <p>Substitusikan y ke dalam persamaan total pendapatan:</p> $80000x + 120000(50 - x) = 4400000$ $80000x + 6000000 - 120000x = 4400000$ $80.000x - 120.000x = 4,400.000 - 6.000.00$ $-40,000x = -1.600.000$ $x = 40$	<p>Melaksanakan rencana (Carry out the plan)</p>	<p>dengan benar, tetapi tidak lengkap</p>	
				<p>Membuat penyelesaian, tetapi sebagian jawaban salah</p>	2
				<p>Membuat penyelesaian, tetapi seluruh jawaban salah</p>	1

	<p>Subtitusikan nilai x ke persamaan $y = 50 - x$</p> $y = 50 - x$ $y = 50 - 40$ $y = 10$ <p>Jumlah Buku Fiksi yang terjual adalah 40 buku, dan jumlah Buku Non-Fiksi yang terjual adalah 10 buku</p>		<p>Tidak ada proses penyelesaian masalah sama sekali</p>	0
	<p>Memeriksa Kembali</p> <p>Jumlah buku:</p> $40 + 10 = 50$ <p>Total pendapatan:</p> $80,000(40) + 120,000(10)$ $= 3.200.000 + 1.200.000$ $= 4.400.000$	<p>Melakukan pemeriksaan (Look back)</p>	<p>Ada pemeriksaan hasil penyelesaian masalah yang relevan dan benar</p>	3
			<p>Ada pemeriksaan hasil penyelesaian masalah, tetapi hanya sebagian yang relevan</p>	2

				Ada pemeriksaan hasil penyelesaian masalah, tetapi tidak semuanya relevan	1
				Tidak ada pemeriksaan hasil penyelesaian	0
3	Dua garis memiliki persamaan berikut: $y = 2x + 3$ $y = -x + 6$	Memahami Masalah Diketahui dua persamaan garis: $y = 3x + 3$ $y = -x + 6$ ditanya	Memahami masalah (Understand the problem)	Menuliskan apa yang diketahui dan ditanyakan dengan benar	3
				Menuliskan apa yang diketahui dan ditanyakan, tetapi salah satunya salah	2

	Tentukan titik perpotongan kedua garis tersebut!	titik perpotongan kedua garis , yaitu nilai x dan y yang memenuhi kedua persamaan secara bersamaan.		Menuliskan apa yang diketahui dan ditanya tetapi keduanya salah	1
				Tidak menuliskan apa yang diketahui dan ditanyakan	0
	Petunjuk Penyelesaian: <ul style="list-style-type: none"> Gunakan metode eliminasi atau substitusi untuk menemukan nilai x dan y. Titik perpotongan adalah solusi dari SPLDV tersebut. 	Menyusun rencana Untuk menemukan titik perpotongan dua garis, kita dapat menggunakan metode substitusi karena kedua persamaan sudah dalam bentuk $y = f(X)$ <ul style="list-style-type: none"> Kita akan menyamakan kedua persamaan, karena kedua ruas bernilai y. 	Menyusun rencana (Devise a plan)	Menuliskan rencana penyelesaian dengan benar	3
				Menuliskan rencana penyelesaian, tetapi hanya sebagian yang benar	2

		<ul style="list-style-type: none"> • Setelah itu, kita menyelesaikan persamaan untuk mendapatkan nilai x. • Setelah mendapatkan x, kita substitusikan kembali ke salah satu persamaan untuk mendapatkan y. 		Menuliskan rencana penyelesaian, tetapi semuanya tidak tepat	1
				Tidak ada rencana sama sekali	0
		<p>Melaksanakan Rencana</p> <p>Langkah 1: menyamakan kedua persamaan Karna $y = 2x + 3$ dan $y = -x + 6$ maka $2x + 3 = -x + 6$</p>	Melaksanakan rencana (Carry out the plan)	Membuat penyelesaian dengan benar, tetapi tidak lengkap	3
		<p>Langkah ke 2 : menyelesaikan persamaan untuk x</p> $2x + 3 = -x + 6$ $2x + x = 6 - 3$		Membuat penyelesaian, tetapi sebagian jawaban salah	2

		$3x = 3$ $x = 1$ <p>Langkah ke 3 substitusikan x untuk mencari y</p> $y = 2(1) + 3 = 2 + 3 = 5$ <p>Maka titik perpotongan kedua garis adalah (1,5)</p>		Membuat penyelesaian, tetapi seluruh jawaban salah	1
				Tidak ada proses penyelesaian masalah sama sekali	0
		<p>Memeriksa Kembali</p> <p>Cek apakah titik (1,5) memenuhi persamaan</p> <ul style="list-style-type: none"> • Substitusi ke persamaan pertama $y = 2(1) + 3$ $y = 5 \text{ (benar)}$ • Substitusi ke persamaan kedua $y = -1 + 6$ $y = 5 \text{ (benar)}$ 	<p>Melakukan pemeriksaan (Look back)</p>	Ada pemeriksaan hasil penyelesaian masalah yang relevan dan benar	3
				Ada pemeriksaan hasil penyelesaian masalah, tetapi hanya sebagian yang relevan	2

				Ada pemeriksaan hasil penyelesaian masalah, tetapi tidak semuanya relevan	1
				Tidak ada pemeriksaan hasil penyelesaian	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

Lampiran 5 Hasil Estimasi Parameter GPCM

A001	1	PCR_MLE	1	1	GROUP 01	1.00		
	1			1.00	3.00	1.00	0.6535	0.2413
A002	1	PCR_MLE	1	2	GROUP 01	1.00		
	1			1.00	2.62	1.00	0.1422	0.2144
A003	1	PCR_MLE	1	3	GROUP 01	1.00		
	1			1.00	3.00	1.00	0.5024	0.2304
A004	1	PCR_MLE	1	4	GROUP 01	1.00		
	1			1.00	1.50	1.00	-1.4936	0.2871
A005	1	PCR_MLE	1	5	GROUP 01	1.00		
	1			1.00	2.75	1.00	0.3291	0.2207
A006	1	PCR_MLE	1	6	GROUP 01	1.00		
	1			1.00	1.62	1.00	-1.1133	0.2524
A007	1	PCR_MLE	1	7	GROUP 01	1.00		
	1			1.00	2.88	1.00	0.4115	0.2249
A008	1	PCR_MLE	1	8	GROUP 01	1.00		
	1			1.00	3.00	1.00	0.7235	0.2468
A009	1	PCR_MLE	1	9	GROUP 01	1.00		
	1			1.00	3.00	1.00	0.4945	0.2299
A010	1	PCR_MLE	1	10	GROUP 01	1.00		
	1			1.00	1.75	1.00	-0.8765	0.2377
A011	1	PCR_MLE	1	11	GROUP 01	1.00		
	1			1.00	1.75	1.00	-0.9339	0.2409
A012	1	PCR_MLE	1	12	GROUP 01	1.00		
	1			1.00	3.50	1.00	1.3740	0.2994
A013	1	PCR_MLE	1	13	GROUP 01	1.00		
	1			1.00	1.88	1.00	-0.8843	0.2381
A014	1	PCR_MLE	1	14	GROUP 01	1.00		
	1			1.00	3.38	1.00	1.1674	0.2833
A015	1	PCR_MLE	1	15	GROUP 01	1.00		
	1			1.00	3.50	1.00	1.3740	0.2994
A016	1	PCR_MLE	1	16	GROUP 01	1.00		
	1			1.00	3.75	1.00	2.1754	0.3988
A017	1	PCR_MLE	1	17	GROUP 01	1.00		
	1			1.00	2.25	1.00	-0.3950	0.2176
A018	1	PCR_MLE	1	18	GROUP 01	1.00		
	1			1.00	2.00	1.00	-0.6879	0.2283
A019	1	PCR_MLE	1	19	GROUP 01	1.00		
	1			1.00	1.88	1.00	-0.5631	0.2232
A020	1	PCR_MLE	1	20	GROUP 01	1.00		
	1			1.00	1.62	1.00	-1.0879	0.2507
A021	1	PCR_MLE	1	21	GROUP 01	1.00		
	1			1.00	2.00	1.00	-0.6544	0.2269
A022				22	GROUP 01	1.00		

Lampiran 6 Hasil Estimasi Parameter GRM

A001	1	PCR_MLE	1	1	GROUP 01	1.00		
	1	PCR_MLE	1	1.00	3.00	1.00	0.6535	0.2413
A002	1	PCR_MLE	1	1.00	2.62	1.00	0.1422	0.2144
A003	1	PCR_MLE	1	1.00	3.00	1.00	0.5024	0.2304
A004	1	PCR_MLE	1	1.00	1.50	1.00	-1.4936	0.2871
A005	1	PCR_MLE	1	1.00	2.75	1.00	0.3291	0.2207
A006	1	PCR_MLE	1	1.00	1.62	1.00	-1.1133	0.2524
A007	1	PCR_MLE	1	1.00	2.88	1.00	0.4115	0.2249
A008	1	PCR_MLE	1	1.00	3.00	1.00	0.7235	0.2468
A009	1	PCR_MLE	1	1.00	3.00	1.00	0.4945	0.2299
A010	1	PCR_MLE	1	1.00	1.75	1.00	-0.8765	0.2377
A011	1	PCR_MLE	1	1.00	1.75	1.00	-0.9339	0.2409
A012	1	PCR_MLE	1	1.00	3.50	1.00	1.3740	0.2994
A013	1	PCR_MLE	1	1.00	1.88	1.00	-0.8843	0.2381
A014	1	PCR_MLE	1	1.00	3.38	1.00	1.1674	0.2833
A015	1	PCR_MLE	1	1.00	3.50	1.00	1.3740	0.2994
A016	1	PCR_MLE	1	1.00	3.75	1.00	2.1754	0.3988
A017	1	PCR_MLE	1	1.00	2.25	1.00	-0.3950	0.2176
A018	1	PCR_MLE	1	1.00	2.00	1.00	-0.6879	0.2283
A019	1	PCR_MLE	1	1.00	1.88	1.00	-0.5631	0.2232
A020	1	PCR_MLE	1	1.00	1.62	1.00	-1.0879	0.2507
A021	1	PCR_MLE	1	1.00	2.00	1.00	-0.6544	0.2269

Lampiran 7 Hasil Estimasi Parameter PCM

A001	1	SCALE1	1	1.00	1	GROUP 01	3.00	1.00	0.5693	0.2427
A002	1	SCALE1	1	1.00	2	GROUP 01	2.62	1.00	0.1093	0.2281
A003	1	SCALE1	1	1.00	3	GROUP 01	3.00	1.00	0.5693	0.2427
A004	1	SCALE1	1	1.00	4	GROUP 01	1.50	1.00	-1.3569	0.2810
A005	1	SCALE1	1	1.00	5	GROUP 01	2.75	1.00	0.2568	0.2307
A006	1	SCALE1	1	1.00	6	GROUP 01	1.62	1.00	-1.1504	0.2628
A007	1	SCALE1	1	1.00	7	GROUP 01	2.88	1.00	0.4091	0.2355
A008	1	SCALE1	1	1.00	8	GROUP 01	3.00	1.00	0.5693	0.2427
A009	1	SCALE1	1	1.00	9	GROUP 01	3.00	1.00	0.5693	0.2427
A010	1	SCALE1	1	1.00	10	GROUP 01	1.75	1.00	-0.9658	0.2509
A011	1	SCALE1	1	1.00	11	GROUP 01	1.75	1.00	-0.9658	0.2509
A012	1	SCALE1	1	1.00	12	GROUP 01	3.50	1.00	1.3769	0.3025
A013	1	SCALE1	1	1.00	13	GROUP 01	1.88	1.00	-0.7951	0.2428
A014	1	SCALE1	1	1.00	14	GROUP 01	3.38	1.00	1.1384	0.2815
A015	1	SCALE1	1	1.00	15	GROUP 01	3.50	1.00	1.3769	0.3025
A016	1	SCALE1	1	1.00	16	GROUP 01	3.75	1.00	2.0143	0.3836
A017	1	SCALE1	1	1.00	17	GROUP 01	2.25	1.00	-0.3286	0.2300
A018	1	SCALE1	1	1.00	18	GROUP 01	2.00	1.00	-0.6337	0.2371
A019	1	SCALE1	1	1.00	19	GROUP 01	1.88	1.00	-0.7951	0.2428
A020	1	SCALE1	1	1.00	20	GROUP 01	1.62	1.00	-1.1504	0.2628
A021	1	SCALE1	1	1.00	21	GROUP 01	2.00	1.00	-0.6337	0.2371

Lampiran 8 Item Fit Statistik

GPCM

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	9.11286	10.	0.522
BLOCK	0002	6.84087	11.	0.813
BLOCK	0003	16.31309	11.	0.129
BLOCK	0004	5.93017	7.	0.549
BLOCK	0005	4.25907	8.	0.834
BLOCK	0006	5.26625	7.	0.629
BLOCK	0007	11.76055	8.	0.161
BLOCK	0008	6.21622	10.	0.798
TOTAL		65.69908	72.	0.686

GRM

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	9.25008	11.	0.600
BLOCK	0002	19.81534	10.	0.131
BLOCK	0003	12.66864	11.	0.315
BLOCK	0004	11.58060	8.	0.170
BLOCK	0005	8.61492	7.	0.281
BLOCK	0006	4.80841	6.	0.570
BLOCK	0007	4.16005	8.	0.843
BLOCK	0008	6.49347	11.	0.839
TOTAL		77.39150	72.	0.311

PCM

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	10.80353	9.	0.289
BLOCK	0002	6.45945	11.	0.842
BLOCK	0003	6.52867	9.	0.687
BLOCK	0004	6.80547	7.	0.450
BLOCK	0005	4.38853	8.	0.822
BLOCK	0006	5.08875	7.	0.651
BLOCK	0007	4.88565	8.	0.771
BLOCK	0008	10.03817	9.	0.347
TOTAL		54.99823	68.	0.872