

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. 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. 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. 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.

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/item (*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 penelitian 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. Mengetahui 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 Theory

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.