

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

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

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

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

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

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

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

Tabel 1.1 Rekap Survei Guru

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.2 Identifikasi Masalah

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

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

1.3 Rumusan Masalah

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

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

1.4 Tujuan Penelitian

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

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

1.5 Manfaat Hasil Penelitian

Diharapkan setelah melakukan penelitian ini dapat memberikan manfaat sebagai berikut:

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

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

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

1.6 Batasan Penelitian

Adapun yang menjadi batasan masalah dalam penelitian ini adalah:

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

1.7 Definisi Operasional

1. Akurasi

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

2. Penilaian

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

3. Keterampilan Berpikir Kritis Matematis

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

4. *Item Response Theory*

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

5. *Item*

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