ABSTRAK

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

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

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

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

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

ABSTRACT

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

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

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

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

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