

# EKSPLORASI FAKTOR-FAKTOR KORELASI PENYEKORAN TEORI TES KLASIK DAN TEORI RESPONS BUTIR

*Abdurrahman Nadhif Al Fajri<sup>1</sup>, Wahyu Widhiarso<sup>2</sup>*

<sup>1,2</sup> Universitas Gadjah Mada

e-mail: <sup>1</sup>abdurrahman.nadhif.a@mail.ugm.ac.id, <sup>2</sup>wahyu\_psy@ugm.ac.id

**Abstract.** The main purpose of assessing abilities in the educational environment is to make various decisions based on the results of the exam. A scoring method must have the ability to produce high precision scores. To achieve this, test administrators usually use one of two popular scoring methods; Classical Test Theory and Item Response Theory. These Theories have their own advantages and disadvantages. IRT have high scoring precision, but the scoring process is difficult. On the other hand, CTT scoring comes with easy scoring processes but lacks precision. This study explores the factors that can lead to different scoring results through the CTT and IRT approaches while examining how far the scores produced by the two methods are related. The factors involved in this study were variations in item parameters, number of samples, number of items, and distribution of participants' abilities. Analysis of 216 factor combinations using one-way ANOVA found that there were significant differences in the correlation between combinations of factors. Meanwhile, the highest correlation between the two types of scores was found in conditions with irradiated item parameters, 1000 participants, 60 items, and normally distributed participants' abilities. Further research is needed to explore other factors beyond this research.

**Keywords:** (*Ability estimation, classical test theory, item response theory*)

**Abstrak.** Tujuan utama penilaian kemampuan di lingkungan pendidikan adalah untuk membuat berbagai keputusan berdasarkan hasil ujian. Sebuah metode penyekoran harus memiliki kemampuan menghasilkan skor dengan presisi tinggi. Untuk melakukan hal tersebut penyelenggara tes biasanya menggunakan salah satu dari dua metode penyekoran populer; *Classical Test Theory* dan *Item Response Theory*. Keunggulan dan kelemahan dua teori tersebut telah lama diteliti dan dibuktikan. Kelebihan IRT adalah presisi penyekoran yang tinggi, namun proses penyekoran sulit dilakukan. Keunggulan penyekoran CTT adalah proses penyekoran yang relatif mudah dilakukan namun memiliki akurasi yang kurang. Penelitian ini mengeksplorasi faktor-faktor yang dapat menyebabkan hasil penyekoran yang berbeda melalui pendekatan CTT dan IRT. Tujuan selanjutnya adalah menguji seberapa jauh skor yang dihasilkan kedua metode tersebut memiliki keterkaitan. Faktor-faktor yang dilibatkan dalam penelitian ini adalah variasi parameter butir, jumlah sampel, jumlah butir, dan distribusi kemampuan peserta. Analisis terhadap 216 kombinasi faktor dengan menggunakan ANOVA satu jalur menemukan bahwa terdapat perbedaan rerata korelasi yang signifikan antara kombinasi faktor. Di sisi lain, korelasi tertinggi antar kedua jenis skor ditemukan pada kondisi parameter soal yang tidak bervariasi, jumlah peserta sebesar 1.000, jumlah soal sebanyak 60, dan distribusi kemampuan peserta yang normal. Penelitian lebih lanjut diperlukan untuk mengeksplorasi faktor lain di luar penelitian ini.

**Kata kunci:** (*Penyekoran kemampuan, Teori Respons Butir, Teori Tes Klasik.*)