

EVALUASI BUTIR SOAL SECARA KLASIK DALAM EVALUASI PENDIDIKAN

Achmad Rasyid Ridha¹, Moh Luthfi², Mudijono³

^{1,2,3}Institut Islam Mambaul Ulum Surakarta

* Corresponding Email: afseedany@gmail.com

A B S T R A K

Analisis butir soal merupakan bagian penting dalam evaluasi pendidikan untuk menjamin kualitas instrumen penilaian yang valid, reliabel, dan adil. Penelitian ini membahas penerapan Teori Tes Klasik (Classical Test Theory/CTT) dalam menganalisis kualitas butir soal, khususnya pada evaluasi pembelajaran Pendidikan Agama Islam (PAI). Fokus analisis meliputi indeks kesukaran, daya pembeda, efektivitas pengecoh, serta reliabilitas tes. Metode yang digunakan adalah pendekatan kuantitatif dengan menganalisis data hasil tes peserta didik menggunakan prosedur analisis butir secara klasik yang dapat diterapkan melalui perangkat lunak sederhana seperti Microsoft Excel. Hasil pembahasan menunjukkan bahwa analisis butir soal secara klasik mampu mengidentifikasi butir soal yang berkualitas maupun yang perlu direvisi atau dibuang. Selain itu, pendekatan ini dinilai praktis, ekonomis, dan mudah diterapkan oleh guru, terutama dalam konteks Kurikulum Merdeka. Meskipun memiliki keterbatasan, seperti ketergantungan pada karakteristik sampel, analisis butir soal berdasarkan Teori Tes Klasik tetap relevan dan efektif dalam meningkatkan kualitas evaluasi dan pembelajaran PAI.

Kata Kunci : analisis butir soal, teori tes klasik, evaluasi pendidikan, PAI, kualitas tes

A B S T R A C T

Item analysis plays a crucial role in educational evaluation to ensure that assessment instruments are valid, reliable, and fair. This study discusses the application of Classical Test Theory (CTT) in analyzing the quality of test items, particularly in the evaluation of Islamic Education (PAI) learning. The analysis focuses on item difficulty index, discrimination index, distractor effectiveness, and test reliability. A quantitative approach was employed by analyzing students' test response data using classical item analysis procedures that can be implemented through simple tools such as Microsoft Excel. The discussion indicates that classical item analysis effectively identifies high-quality items as well as items that require revision or elimination. Moreover, this approach is considered practical, economical, and easy to apply for teachers, especially within the context of the Independent Curriculum (Kurikulum Merdeka). Despite its limitations, including sample dependency, Classical Test Theory remains a relevant and effective method for improving the quality of assessment and learning outcomes in Islamic Education.

Keywords : item analysis, classical test theory, educational evaluation, Islamic education, test quality

PENDAHULUAN

Dalam era pendidikan berbasis data saat ini, analisis butir soal secara klasik melalui Teori Tes Klasik (Classical Test Theory/CTT) menjadi fondasi utama dalam evaluasi pendidikan untuk memastikan instrumen asesmen memiliki validitas, reliabilitas, dan keadilan yang tinggi. Pendekatan ini memungkinkan pendidik mengukur kualitas soal

melalui indeks kesukaran (P), daya pembeda (D), dan efektivitas distraktor, yang diperoleh dari analisis respon siswa pasca-tes, sehingga mendukung revisi soal yang tepat sasaran dan peningkatan kualitas pembelajaran. Di Indonesia, di mana asesmen nasional seperti Asesmen Nasional (AN) dan Ujian Akhir Sekolah semakin menekankan pengukuran kompetensi, analisis klasik menawarkan solusi praktis dan ekonomis dibandingkan model modern seperti Item Response Theory (IRT), meskipun tetap memerlukan adaptasi kontekstual untuk mengatasi heterogenitas siswa.

Evaluasi pendidikan menghadapi tantangan kompleks, termasuk soal yang tidak sensitif membedakan kemampuan siswa atau distraktor yang gagal menjebak jawaban salah, yang sering kali mengakibatkan data hasil belajar yang bias. Teori Tes Klasik, yang dikembangkan sejak awal abad ke-20 oleh para ahli seperti Gulliksen (1950), menekankan bahwa skor observasi (X) merupakan fungsi skor sebenarnya (T) ditambah kesalahan pengukuran (E), sehingga analisis butir difokuskan pada estimasi parameter agregat dari kelompok tes. Penelitian di Indonesia menunjukkan bahwa hingga 30-40% soal tes sekolah memiliki indeks kesukaran di luar rentang ideal (0.30-0.70), yang berdampak pada reliabilitas tes secara keseluruhan (Arikunto, 2018). Oleh karena itu, kerangka analisis klasik menjadi krusial untuk bridging kesenjangan antara desain soal dan validasi empiris.

METODE PENELITIAN

Metode penelitian yang digunakan adalah metode kuantitatif dengan pendekatan deskriptif evaluatif, yang bertujuan untuk menganalisis kualitas butir soal berdasarkan Teori Tes Klasik (Classical Test Theory/CTT) melalui perhitungan indeks kesukaran, daya pembeda, efektivitas pengecoh, serta reliabilitas tes menggunakan data hasil jawaban peserta didik.

HASIL DAN PEMBAHASAN

Teori Tes Klasik (Classical Test Theory/CTT) menjadi dasar utama analisis butir soal dalam evaluasi pendidikan, dengan asumsi bahwa skor observasi (X) terdiri dari skor sebenar (T) plus kesalahan pengukuran (E), yaitu $X = T + E$. Pendekatan ini menekankan analisis agregat dari kelompok tes untuk menghitung indeks kesukaran (P), daya pembeda (D), dan reliabilitas tes melalui rumus seperti Cronbach's Alpha. CTT unggul dalam kemudahan komputasi menggunakan Excel untuk sampel kecil ($n > 30$), meskipun bergantung pada distribusi kemampuan siswa.

Analisis butir mencakup tiga parameter utama. Indeks kesukaran (P) dihitung sebagai $P = (\text{jumlah jawaban benar} / \text{total peserta})$, dengan kategori ideal 0.30-0.70; daya pembeda (D) sebagai $D = (\text{proporsi benar kelompok atas} - \text{proporsi benar kelompok bawah})/\text{ukuran kelompok}$ (27% atas/bawah), ideal >0.40 ; serta efektivitas distraktor melalui frekuensi pilihan salah. Reliabilitas tes diukur dengan $\alpha = [k/(k-1)] [1 - (\sum \sigma^2_i / \sigma^2_{\text{total}})]$, di mana nilai >0.70 menandakan tes reliabel.

Misalkan tes pilihan ganda 5 butir untuk 40 siswa Matematika SMA, dengan data respon berikut (kelompok atas 11 siswa skor tertinggi, bawah 11 terendah):

Butir	Benar (B)	A	B	C	D	E	Total Jawab
1	25	5	3	2	4	1	40
2	18	8	6	4	3	1	40
3	32	2	1	3	1	1	40
4	12	10	7	5	5	1	40
5	28	4	3	2	2	1	40

Untuk Butir 1: $P = 25/40 = 0.625$ (sedang). Kelompok atas: 10 benar ($P_{atas}=0.91$); bawah: 3 benar ($P_{bawah}=0.27$); $D = (0.91-0.27)/1 = 0.64$ (baik). Distraktor: E paling efektif ($1/40=0.025$, tapi rendah fungsi).

Pembahasan Hasil Contoh

Hasil menunjukkan Butir 1 berkualitas baik (P sedang, D tinggi, distraktor A dominan di bawah). Butir 4 bermasalah: $P=0.30$ (sukar), $D=0.18$ (rendah, hitung: atas $5/11=0.45$, bawah $1/11=0.09$), distraktor merata – rekomendasi revisi kurangi kesukaran atau perbaiki stem. Reliabilitas keseluruhan $\alpha=0.72$ (cukup), layak pakai tapi revisi Butir 4 tingkatkan $D>0.40$. Analisis ini mendeteksi bias tebakan di kelompok bawah, selaras studi UN Matematika di mana D rendah akibat menebak. Integrasi kualitatif (cek bahasa) perkuat validitas konten.

Implikasi dan Keterbatasan

CTT praktis untuk guru Indonesia dalam Kurikulum Merdeka, tapi terbatas ketergantungan sampel – disarankan hybrid dengan IRT untuk estimasi invariant. Penelitian ini implikasikan bank soal nasional butuh analisis rutin pasca-tes.

PROSEDUR ANALISIS BUTIR SOAL SECARA KLASIK

Prosedur analisis butir soal secara klasik meliputi tahapan sebagai berikut:

1. Pengumpulan Data Jawaban: Data dikumpulkan dari hasil jawaban peserta didik terhadap soal yang telah diberikan.
2. Pengolahan Data: Menghitung nilai tingkat kesukaran, daya beda, dan efektivitas pengecoh dengan rumus-rumus tertentu dalam teori tes klasik.
3. Interpretasi Hasil: Hasil pengolahan dianalisis untuk menentukan soal yang layak dan tidak layak dipakai.
4. Rekomendasi Perbaikan: Soal yang memiliki tingkat kesukaran terlalu rendah atau tinggi, daya beda yang rendah, serta pengecoh yang tidak efektif maka perlu dilakukan perbaikan atau pembuangan soal tersebut (Elviana, 2019).

MANFAAT ANALISIS BUTIR SOAL SECARA KLASIK DALAM PENDIDIKAN PAI

Dalam pembelajaran Pendidikan Agama Islam (PAI), analisis butir soal klasik memiliki beberapa manfaat yang penting, antara lain:

1. Meningkatkan Validitas dan Reliabilitas Soal Analisis ini membantu memastikan bahwa soal benar-benar mengukur kompetensi yang diharapkan dengan tingkat keandalan yang tinggi.

2. Menjamin Keadilan Evaluasi Soal yang sudah dianalisis dan disesuaikan tingkat kesukaran dan daya bedanya dapat memberikan hasil evaluasi yang adil bagi seluruh peserta didik (Nurul Muchlizani, 2021).
3. Memperoleh Informasi Diagnostik Guru dapat mengetahui bagian materi mana yang sudah dikuasai siswa dan mana yang belum, sehingga evaluasi tidak hanya sebagai pengukuran akhir namun juga sebagai umpan balik untuk pembelajaran berikutnya.
4. Meningkatkan Kualitas Pembelajaran PAI. Dengan soal yang berkualitas, proses pembelajaran dapat lebih terarah dan efektif dalam mencapai tujuan pendidikan agama.

SIMPULAN DAN SARAN

Analisis butir soal secara klasik dalam evaluasi pendidikan, khususnya PAI, memiliki peran penting untuk memastikan kualitas butir soal yang valid, reliabel, dan mampu membedakan tingkat pemahaman peserta didik. Proses analisis ini meliputi pengujian tingkat kesukaran, daya beda, dan efektivitas pengecoh sehingga evaluasi yang digunakan dapat memberikan hasil yang objektif dan akurat. Meski memiliki keterbatasan, analisis butir soal secara klasik masih merupakan metode yang banyak digunakan karena kesederhanaannya dan kemudahan penerapannya di berbagai jenjang pendidikan.

DAFTAR PUSTAKA

- Arikunto, S. (2018). Dasar-dasar evaluasi pendidikan (Edisi 2). Bumi Aksara.
- Dwipayani, S. (2013). Analisis validitas dan reliabilitas butir soal ulangan akhir semester bidang studi Bahasa Indonesia. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 5(1), 1–6.
- Elviana, S. A. (2019). Analisis butir soal evaluasi pembelajaran PAI menggunakan program Anates (Skripsi). Universitas Islam Negeri Ar-Raniry Banda Aceh.
- Nurul Muchlizani, A. (2021). Analisis kualitas butir soal ujian akhir semester mata pelajaran Akidah Akhlak kelas V MI Radhiyatul Adawiyah Makassar (Tesis). Universitas Islam Negeri Alauddin Makassar.
- Retnawati, H. (2015). Analisis statistik dalam pendidikan. Nuha Literasi.
- Retnawati, H. (2017). Analisis butir soal CTT dan IRT. *Pythagoras*, 12(2), 1–10.
- Sumaryanta. (2021). Analisis daya diskriminan berdasarkan teori tes klasik. *Bio-Pedagogi*, 14(1), 1–9.
- Triana. (2024). Keterbatasan analisis butir soal berdasarkan teori tes klasik. *Jurnal Pendidikan*, 10(2), 45–52.
- Tyas, F., Rahmawati, D., & Suryani, L. (2020). Evaluasi tingkat kesukaran butir soal dalam pembelajaran. *Jurnal Pendidikan dan Pembelajaran*, 7(3), 112–120.
- Widyaningsih, S., & Yusuf, M. (2018). Model Rasch pada pengukuran pendidikan. *Jurnal Pendidikan*, 6(2), 50–58.