



Analisis Kualitas Butir Soal Pada Instrumen Tryout Mahasiswa Program Studi PPG Perhotelan

Anggi Martiningtyas Januwati Saputri^{*1}, Lismi Animatul Chisbiyah²

^{1,2}Universitas Negeri Malang, Indonesia

E-mail: anggi.saputri.ft@um.ac.id, lismi.chisbiyah@um.ac.id

Article Info	Abstract
Article History Received: 2025-08-05 Revised: 2025-09-12 Published: 2025-10-18 Keywords: <i>Item Analysis;</i> <i>Competency Test;</i> <i>Hospitality Education;</i> <i>Validity;</i> <i>Reliability.</i>	This study aims to evaluate the quality of a tryout instrument used for the Hospitality Teacher Professional Education Program (PPG) by analyzing the psychometric characteristics of its test items. While tryouts are crucial for certification, their instrument quality is often assumed rather than empirically verified. This research provides the first systematic item-level analysis for the Hospitality PPG tryout test, addressing a critical gap in the program's quality assurance process. The analysis of 35 items from 109 participants revealed a mediocre reliability coefficient ($KR-20 = 0.655$). A significant weakness was found in item quality, with 49% of items being invalid and another 49% exhibiting poor discrimination power. Although the item difficulty distribution was good, the findings conclude that the instrument, in its current state, is not yet adequate for accurately measuring student competency. A major revision of the item bank is strongly recommended.
Artikel Info	Abstrak
Sejarah Artikel Diterima: 2025-08-05 Direvisi: 2025-09-12 Dipublikasi: 2025-10-18 Kata kunci: <i>Analisis Butir Soal;</i> <i>Uji Kompetensi;</i> <i>Pendidikan Perhotelan;</i> <i>Validitas;</i> <i>Reliabilitas.</i>	Penelitian ini bertujuan untuk mengevaluasi kualitas instrumen uji coba yang digunakan untuk Program Pendidikan Profesi Guru Perhotelan (PPG) dengan menganalisis karakteristik psikometrik butir-butir tesnya. Meskipun uji coba sangat penting untuk sertifikasi, kualitas instrumennya seringkali diasumsikan alih-alih diverifikasi secara empiris. Penelitian ini menyediakan analisis sistematis tingkat butir pertama untuk tes uji coba PPG Perhotelan, yang mengatasi kesenjangan kritis dalam proses penjaminan mutu program. Analisis terhadap 35 butir dari 109 peserta menunjukkan koefisien reliabilitas yang biasa-biasa saja ($KR-20 = 0,655$). Kelemahan signifikan ditemukan pada kualitas butir, dengan 49% butir tidak valid dan 49% lainnya menunjukkan daya pembeda yang buruk. Meskipun distribusi tingkat kesulitan butir baik, temuan menyimpulkan bahwa instrumen tersebut, dalam kondisi saat ini, belum memadai untuk mengukur kompetensi mahasiswa secara akurat. Revisi besar-besaran terhadap bank butir sangat disarankan.

I. PENDAHULUAN

Pendidikan Profesi Guru (PPG) merupakan program strategis pemerintah untuk memastikan pemenuhan standar kompetensi guru sebagaimana diamanatkan dalam Undang-Undang No. 14 Tahun 2005 tentang Guru dan Dosen (Pemerintah Republik Indonesia, 2005). Dalam konteks PPG Perhotelan, tantangan penjaminan mutu menjadi lebih kompleks karena calon guru tidak hanya dituntut menguasai kompetensi pedagogik, tetapi juga penguasaan konten kejuruan yang aplikatif dan relevan dengan kebutuhan industri (Fullan, 2016). Sebagai salah satu instrumen evaluasi formatif, *tryout* memegang peranan krusial untuk mengukur kesiapan mahasiswa sebelum menghadapi Uji Kompetensi Mahasiswa PPG (UKMPPG) dan memberikan umpan balik diagnostik terhadap kelemahan penguasaan materi.

Kualitas dari sebuah instrumen evaluasi adalah prasyarat mutlak untuk menghasilkan

data yang akurat dan keputusan yang adil. Sebuah tes yang berkualitas harus memenuhi serangkaian kriteria psikometrik yang ketat. Menurut Arikunto (2019), properti utama yang menentukan kualitas instrumen meliputi validitas, yang menjamin tes mengukur kompetensi yang seharusnya diukur, dan reliabilitas, yang memastikan konsistensi hasil pengukuran (Arikunto, 2019). Lebih lanjut, analisis butir soal secara spesifik mengkaji tingkat kesukaran untuk memastikan soal tidak terlalu mudah atau terlalu sulit, serta daya pembeda untuk melihat kemampuan soal dalam membedakan peserta berkemampuan tinggi dan rendah. Analisis ini menjadi semakin penting seiring tuntutan untuk mengukur keterampilan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS), seperti analisis dan evaluasi, bukan sekadar ingatan (Anderson & Krathwohl, 2001) dan (Brookhart, 2010).

Penyelenggaraan tryout sebagai bagian krusial dari persiapan UKMPPG acapkali belum disertai dengan analisis kualitas instrumen yang memadai. Akibatnya, kelayakan butir soal yang digunakan sebagai alat ukur lebih bersifat asumtif, bukan berbasis bukti empiris. Instrumen yang tidak teruji validitas dan reliabilitasnya berisiko memberikan gambaran yang keliru tentang kesiapan mahasiswa, sehingga intervensi pembelajaran yang dirancang menjadi tidak tepat sasaran. Adanya butir soal yang ambigu, terlalu sulit, atau memiliki daya pembeda yang lemah dapat merugikan peserta dan mendelegitimasi hasil evaluasi secara keseluruhan. Hingga saat ini, belum ditemukan adanya publikasi penelitian yang secara spesifik menganalisis kualitas butir soal pada instrumen *tryout* untuk mahasiswa PPG Perhotelan.

Penelitian ini bertujuan untuk menganalisis kualitas butir soal pada instrumen *tryout* yang digunakan oleh mahasiswa Program Studi PPG Perhotelan. Analisis ini difokuskan pada karakteristik psikometrik butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas fungsi distraktor. Hasil penelitian ini diharapkan dapat memberikan bukti empiris mengenai kelayakan instrumen serta menjadi dasar rekomendasi untuk perbaikan dan pengembangan soal di masa mendatang.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain deskriptif. Tujuan dari penelitian ini adalah untuk menganalisis dan mendeskripsikan kualitas butir soal pada instrumen *tryout* mahasiswa PPG Perhotelan. Kualitas instrumen dievaluasi berdasarkan karakteristik psikometriksnya, yang ditinjau dari aspek validitas butir, reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas fungsi distraktor (Creswell & Creswell, 2018). Partisipan dalam penelitian ini adalah 109 mahasiswa Program Profesi Guru (PPG) Bidang Perhotelan di Universitas Negeri Malang yang mengikuti kegiatan *tryout* pada Semester Gasal 2025 (Juli 2025). Seluruh populasi peserta *tryout* dijadikan sampel dalam penelitian, sehingga teknik pengambilan sampel yang digunakan adalah *total sampling*.

Proses penelitian diawali dengan pengumpulan data sekunder yang bersumber dari dokumentasi hasil pengerjaan instrumen *tryout* oleh 109 partisipan. Setiap respons jawaban kemudian dikonversi menjadi skor

biner (1 untuk jawaban benar dan 0 untuk jawaban salah), dan selanjutnya seluruh data ditabulasi dalam format *spreadsheet* untuk persiapan analisis. Analisis data kuantitatif dilakukan secara sistematis untuk menguji setiap karakteristik butir soal dengan bantuan perangkat lunak Excel dan SPSS. Pertama, analisis validitas butir soal menggunakan teknik korelasi *Point Biserial*, di mana butir soal dinyatakan valid jika nilai koefisien korelasinya (r_{pbi}) melebihi nilai r tabel pada taraf signifikansi 5%. Selanjutnya, reliabilitas keseluruhan instrumen diestimasi menggunakan formula Kuder-Richardson 20 (KR-20), dengan koefisien di atas 0,70 dianggap reliabel. Adapun tingkat kesukaran setiap butir dihitung berdasarkan proporsi peserta yang menjawab benar dan dikategorikan sebagai sukar ($< 0,30$), sedang ($0,30 - 0,70$), atau mudah ($> 0,70$). Daya pembeda butir soal dianalisis dengan membandingkan proporsi jawaban benar antara kelompok atas (27% skor tertinggi) dan kelompok bawah (27% skor terendah), di mana indeks $\geq 0,30$ menunjukkan daya pembeda yang baik. Terakhir, efektivitas fungsi distraktor dievaluasi dengan memeriksa pola frekuensi pemilihannya oleh partisipan dari kelompok atas dan kelompok bawah.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Analisis kuantitatif terhadap data respons dari 109 partisipan menghasilkan gambaran kualitas instrumen *tryout* yang disajikan secara ringkas pada Tabel 1.

Tabel 1. Ringkasan Hasil Analisis Kualitas Instrumen Tryout

Analisis	Hasil	Indikator (Metrik & Kriteria)	Kategori	Rekomendasi
Reliabilitas Tes	Koefisien = 0,655	Koefisien KR-20	Cukup	Dapat digunakan dengan revisi mayor
Validitas Butir	Valid: 18 (51%) Tidak Valid: 17 (49%)	$r_{hitung} \geq r_{tabel}$ (0,195)	Kurang Baik	Buang atau revisi total butir tidak valid
Tingkat Kesukaran	Sukar: 1 (3%) Sedang: 22 (63%) Mudah: 12 (34%)	Indeks Kesukaran (P)	Baik	Pertahankan sebaran, revisi butir ekstrem
Daya Pembeda	Baik: 11 (31%) Cukup: 7 (20%) Jelek: 17	Indeks Pembeda (D)	Kurang Baik	Buang atau revisi total butir kategori jelek

(49%)				
Analisis Pengecoh	17 butir soal bermasalah	Pola Jawaban Kelompok Atas-Bawah	Kurang Baik	Ganti opsi pengecoh yang tidak fungsional

Hasil analisis reliabilitas instrumen menggunakan formula Kuder-Richardson 20 (KR-20) menunjukkan koefisien sebesar 0,655. Berdasarkan kriteria yang umum digunakan, nilai ini termasuk dalam kategori reliabilitas "Cukup". Sedangkan dari aspek validitas butir, ditemukan bahwa dari total 35 soal, hanya 18 soal (51,4%) yang dinyatakan valid, sementara 17 soal (48,6%) lainnya tidak valid. Menunjukkan bahwa hampir separuh dari butir soal tidak mengukur konstruk yang sama dengan keseluruhan tes. Analisis tingkat kesukaran menunjukkan distribusi soal yang tergolong "Baik". Sebagian besar soal, yaitu 22 butir (63%), berada pada kategori "Sedang". Sisanya terdistribusi pada 1 butir (3%) kategori "Sukar" dan 12 butir (34%) kategori "Mudah".

Namun, hasil yang kontras ditunjukkan pada analisis daya pembeda yang secara umum tergolong "Kurang Baik". Hanya 11 soal (31%) yang memiliki daya pembeda kategori "Baik", 7 soal (20%) berkategori "Cukup", dan mayoritas, yaitu 17 soal (49%), memiliki daya pembeda "Jelek". Kelemahan pada butir soal juga tercermin dari analisis pengecoh, di mana ditemukan 17 soal memiliki satu atau lebih opsi pengecoh (distraktor) yang tidak berfungsi secara efektif.

B. Pembahasan

Pembahasan hasil penelitian ini menyoroti sebuah paradoks antara reliabilitas instrumen yang "Cukup" (0,655) dengan kualitas butir soal individual yang secara umum "Kurang Baik". Koefisien reliabilitas yang berada di ambang batas kriteria menunjukkan bahwa tes ini memiliki konsistensi internal yang minimal, namun belum bisa dikatakan handal untuk membuat keputusan penting mengenai kesiapan mahasiswa. Konsistensi yang sedang ini kemungkinan besar terganggu oleh banyaknya butir soal yang bermasalah.

Akar masalah dari kualitas instrumen ini terletak pada hasil analisis validitas dan daya pembeda. Dengan hampir separuh butir soal (49%) tidak valid dan memiliki daya pembeda yang jelek, instrumen ini secara signifikan gagal dalam fungsi utamanya: membedakan

secara akurat antara mahasiswa yang kompeten dan yang kurang kompeten. Butir-butir soal yang jelek ini tidak memberikan kontribusi informasi yang berguna, bahkan berpotensi menjadi "noise" yang mengganggu akurasi skor total.

Kelemahan pada butir soal juga tercermin dari analisis pengecoh, di mana ditemukan 17 soal memiliki satu atau lebih opsi pengecoh yang tidak berfungsi secara efektif. Sebagai contoh, pada butir soal nomor 25, ditemukan bahwa distraktor pilihan (B) tidak dipilih oleh satu pun partisipan. Sebaliknya, pada soal nomor 10, distraktor (C) justru lebih banyak dipilih oleh kelompok atas daripada kunci jawaban, mengindikasikan adanya potensi ambiguitas. Akibatnya, mahasiswa dapat menebak jawaban benar dengan lebih mudah atau justru terkecoh oleh soal yang ambigu, bukan karena kurangnya penguasaan materi.

Meskipun sebaran tingkat kesukaran soal tergolong "Baik", dengan mayoritas soal berada di kategori "Sedang", keunggulan ini menjadi kurang bermakna. Soal dengan tingkat kesulitan ideal menjadi sia-sia jika tidak mampu membedakan peserta secara efektif dan tidak valid dalam mengukur kompetensi yang dituju.

Implikasinya, penggunaan instrumen ini dalam kondisi sekarang berisiko memberikan hasil evaluasi yang tidak akurat dan umpan balik yang keliru kepada mahasiswa maupun penyelenggara PPG. Keputusan yang didasarkan pada skor tes ini mungkin tidak sepenuhnya mencerminkan kemampuan mahasiswa yang sebenarnya. Oleh karena itu, rekomendasi utama dari penelitian ini adalah melakukan revisi besar-besaran terhadap bank soal. Butir-butir soal yang tidak valid dan berdaya pembeda jelek harus dibuang atau ditulis ulang sepenuhnya, dengan perhatian khusus pada perancangan distraktor yang fungsional berdasarkan analisis respons mahasiswa.

IV. SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan analisis kualitas butir soal terhadap instrumen *tryout* mahasiswa PPG Perhotelan, dapat disimpulkan bahwa instrumen yang diuji memiliki kualitas yang belum sepenuhnya memadai. Meskipun instrumen ini menunjukkan tingkat konsistensi internal yang cukup (reliabilitas cukup) dan memiliki sebaran tingkat

kesukaran yang baik, ditemukan kelemahan fundamental pada kualitas butir soal individual. Hampir separuh dari butir soal tidak valid dan memiliki daya pembeda yang jelek, serta menunjukkan masalah pada efektivitas fungsi pengecoh. Hal ini mengindikasikan bahwa instrumen dalam kondisinya saat ini belum dapat diandalkan secara maksimal untuk mengukur kompetensi mahasiswa secara akurat dan adil.

B. Saran

Saran dalam penelitian ini adalah dilakukan perbaikan pada butir Soal yang tidak Valid.

DAFTAR RUJUKAN

- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.
- Arikunto, S. (2019). *Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi)*. Bumi Aksara.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE Publications.
- Fullan, M. (2016). *The New Meaning of Educational Change*. Teachers College Press.
- Pemerintah Republik Indonesia. (2005). *Undang-Undang Nomor 14 Tahun 2005 tentang Guru dan Dosen*.