Copyright © 2025 The Author(s)
Didaktika: Jurnal Kependidikan Vol. 14 No. 4 Nopember 2025

Kelayakan Instrumen Tes Tata Boga bagi Guru Kejuruan: Analisis Validitas dan Reliabilitas

Nunung Nurjanah[1], Dedi Prestiadi[2]
[1,2] Universitas Negeri Malang, Indonesia

[1] nunung.nurjanah.ft@um.ac.id

Abstrak

Instrumen penilaian memiliki peran penting dalam pendidikan kejuruan, khususnya bidang tata boga yang menuntut penguasaan teori dan praktik. Penelitian ini bertujuan menganalisis kelayakan instrumen tes tata boga yang digunakan dalam Program Pendidikan Profesi Guru (PPG) bidang kuliner. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan melibatkan 75 peserta PPG. Analisis dilakukan terhadap validitas butir, reliabilitas, tingkat kesukaran, daya beda, serta efektivitas distraktor menggunakan uji korelasi Pearson dan Cronbach's Alpha. Hasil penelitian menunjukkan bahwa 65% dari 63 butir soal memenuhi kriteria valid dengan nilai validitas berkisar 0,20–0,62, dan reliabilitas instrumen tinggi (α = 0,84). Tingkat kesukaran berada pada kategori mudah–sedang, sedangkan daya beda rata-rata 0,42 menunjukkan kemampuan instrumen dalam membedakan peserta berkemampuan tinggi dan rendah. Sebagian besar distraktor berfungsi baik, meskipun masih ditemukan butir yang memerlukan revisi, terutama pada aspek pengecoh yang tidak bekerja optimal. Penelitian ini menyimpulkan bahwa instrumen tes tata boga layak digunakan sebagai alat evaluasi kompetensi guru kejuruan, meskipun penyempurnaan tetap diperlukan untuk meningkatkan ketepatan pengukuran. Temuan ini berkontribusi pada pengembangan instrumen asesmen yang lebih valid dan reliabel dalam pendidikan kejuruan serta menjadi acuan bagi lembaga PPG dalam meningkatkan mutu evaluasi pembelajaran.

Kata kunci: instrumen tes, tata boga, pendidikan kejuruan, validitas, reliabilitas

Pendahuluan

Pendidikan kejuruan memiliki peran strategis dalam menyiapkan sumber daya manusia yang siap kerja, terampil, dan kompetitif di era global. Bidang tata boga merupakan salah satu kompetensi kejuruan yang memerlukan penguasaan teori, keterampilan praktik, dan kreativitas dalam mengolah bahan pangan. Oleh karena itu, instrumen penilaian dalam bidang ini harus dirancang untuk menggambarkan kompetensi peserta didik secara menyeluruh, mencakup ranah kognitif, afektif, dan psikomotor. Instrumen yang tidak dirancang dengan baik dapat menghasilkan penilaian yang bias, sehingga capaian pembelajaran tidak dapat dipetakan dengan tepat. Hal ini sejalan dengan pandangan Achmad et al. (2024) bahwa evaluasi pendidikan harus mencerminkan tujuan pembelajaran yang ingin dicapai secara objektif.

Validitas dan reliabilitas instrumen menjadi aspek utama dalam menjamin kualitas evaluasi. Instrumen yang valid memastikan bahwa butir soal benar-benar mengukur kompetensi yang dituju, sedangkan instrumen yang reliabel menjamin konsistensi hasil pengukuran dari waktu ke waktu (Menchetti & Rusch, 1988). Dalam pendidikan kejuruan, instrumen penilaian tidak hanya berfungsi sebagai alat ukur capaian belajar, tetapi juga menjadi acuan dalam perbaikan proses pembelajaran. Instrumen yang tidak valid atau reliabel dapat menurunkan kepercayaan terhadap hasil evaluasi dan berdampak pada kualitas lulusan.

Sejumlah penelitian telah dilakukan untuk menganalisis kualitas instrumen penilaian di berbagai bidang. Castillo-Ortiz (2024) dan Zhang (2025) menemukan bahwa instrumen ujian

praktik tata boga belum sepenuhnya valid karena sebagian besar butir soal belum merepresentasikan kompetensi psikomotor yang dominan dalam pembelajaran kuliner. Jimenez (2023) menegaskan bahwa reliabilitas instrumen merupakan faktor penting dalam menjaga keadilan evaluasi, karena tanpa reliabilitas yang memadai hasil penilaian menjadi tidak konsisten. Penelitian lain menunjukkan bahwa pelatihan penyusunan instrumen berbasis Higher Order Thinking Skills (HOTS) dapat meningkatkan kemampuan guru dalam menyusun soal yang lebih valid dan sesuai standar asesmen modern (Daryono et al., 2023; Kim, 2024).

Asesmen dalam pendidikan kejuruan harus kontekstual dan relevan dengan praktik industri, karena asesmen yang terlalu teoretis gagal mencerminkan kompetensi nyata peserta didik. Gunawan & Citriadin (2025) dan Iktarastiwi (2025) menambahkan bahwa asesmen berbasis kompetensi lebih efektif dalam memprediksi kinerja peserta didik dibandingkan asesmen berbasis pengetahuan semata. Hal ini menunjukkan bahwa dalam pendidikan tata boga, butir soal harus mampu mengukur kemampuan praktik sekaligus aspek kognitif tingkat tinggi seperti analisis, evaluasi, dan kreasi.

Dalam konteks Pendidikan Profesi Guru (PPG), asesmen memiliki fungsi strategis karena tidak hanya menilai penguasaan materi, tetapi juga kesiapan calon guru dalam mengelola pembelajaran. Kualitas instrumen dalam PPG sangat memengaruhi mutu lulusan program, karena instrumen berfungsi sebagai alat seleksi kompetensi minimal guru (Pangestika, 2015). Pasaribu et al. (2025) juga menemukan bahwa instrumen yang lemah dalam desain dan kualitas dapat menurunkan efektivitas PPG, khususnya di bidang kejuruan yang menuntut keterampilan praktik tinggi. Oleh karena itu, pengembangan instrumen penilaian di bidang tata boga dalam konteks PPG menjadi krusial agar kompetensi guru kejuruan dapat terukur secara valid.

Meskipun berbagai penelitian telah dilakukan, kajian empiris mengenai instrumen penilaian tata boga di Indonesia masih terbatas. Kebanyakan penelitian lebih banyak berfokus pada bidang kejuruan teknik atau bidang akademik lain, sementara tata boga seringkali hanya menjadi objek penelitian dalam konteks pembelajaran praktik (Mia & Sulastri, 2023; Noor et al., 2024). Padahal, bidang tata boga memiliki kekhasan tersendiri, misalnya aspek kreativitas dalam mengolah makanan, ketepatan teknik, hingga manajemen waktu dalam proses produksi. Kekhasan ini menuntut instrumen penilaian yang berbeda dari bidang kejuruan lain.

Urgensi penelitian ini terletak pada kebutuhan untuk menyediakan instrumen tes tata boga yang memenuhi kriteria validitas, reliabilitas, daya beda, dan tingkat kesukaran yang memadai. Instrumen yang tidak layak dapat berimplikasi pada kesalahan penilaian, yang pada akhirnya merugikan peserta didik maupun lembaga pendidikan. Selain itu, penelitian ini penting untuk mendukung peningkatan kualitas program PPG, karena instrumen yang valid dan reliabel akan membantu memastikan bahwa guru kejuruan yang dihasilkan benar-benar kompeten.

Penelitian ini memberikan beberapa manfaat, baik secara teoretis maupun praktis. Secara teoretis, penelitian ini memperkaya literatur mengenai evaluasi instrumen penilaian dalam pendidikan kejuruan, khususnya bidang tata boga yang masih jarang diteliti. Secara praktis, penelitian ini memberikan rekomendasi perbaikan butir soal yang tidak valid atau tidak berfungsi optimal, sehingga dapat digunakan oleh guru kejuruan dalam memperbaiki instrumen asesmen yang mereka gunakan. Hasil penelitian ini juga diharapkan menjadi acuan bagi lembaga penyelenggara PPG dalam mengembangkan instrumen evaluasi yang lebih sesuai dengan standar kompetensi guru kejuruan.

Dengan demikian, penelitian ini difokuskan untuk menganalisis kelayakan instrumen tes tata boga melalui uji validitas butir, reliabilitas instrumen, tingkat kesukaran, daya beda, serta analisis kelompok atas dan bawah. Analisis ini akan memberikan gambaran empiris tentang kualitas instrumen tes, sekaligus menjadi dasar bagi upaya peningkatan kualitas evaluasi di pendidikan kejuruan. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan

Copyright © 2025 The Author(s)

kualitas asesmen, baik pada level program studi maupun dalam praktik pembelajaran di sekolah kejuruan.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan desain deskriptif karena bertujuan menganalisis kelayakan instrumen tes tata boga melalui uji validitas, reliabilitas, tingkat kesukaran, daya beda, serta analisis kelompok atas dan bawah. Desain ini dipilih agar peneliti dapat menggambarkan kualitas instrumen berdasarkan data empiris tanpa manipulasi variabel. Dengan pendekatan tersebut, penelitian difokuskan pada pemaparan kondisi faktual instrumen yang digunakan dalam konteks pendidikan kejuruan. Subjek penelitian melibatkan 75 guru kejuruan yang mengikuti Program Pendidikan Profesi Guru (PPG) bidang tata boga tahun akademik 2025. Pemilihan responden dilakukan dengan purposive sampling karena dianggap sesuai kebutuhan penelitian. Responden terdiri atas 43 perempuan (57%) dan 32 laki-laki (43%) dengan rentang usia 24–39 tahun. Jumlah partisipan tersebut dinilai memadai untuk memberikan gambaran representatif mengenai kelayakan instrumen. Instrumen penelitian berupa tes pilihan ganda berjumlah 63 butir soal dengan lima opsi jawaban (A–E). Kisi-kisi soal mencakup kompetensi pedagogik dan profesional tata boga, meliputi pengetahuan kuliner, teknik pengolahan makanan, sanitasi dan keamanan pangan, serta manajemen dapur. Instrumen telah melalui telaah ahli untuk memastikan kesesuaian isi sebelum diujicobakan kepada responden secara daring dengan durasi 90 menit. Analisis data dilakukan secara kuantitatif menggunakan Microsoft Excel dan SPSS. Validitas butir diuji menggunakan korelasi product moment Pearson dengan kriteria $r \geq 0{,}30$ sebagai batas minimal validitas (kategori cukup), sedangkan $r < 0{,}30$ dikategorikan tidak valid. Reliabilitas dihitung menggunakan Cronbach's Alpha, dengan interpretasi $\alpha \geq 0{,}80$ (tinggi), 0,70–0,79 (cukup), dan < 0,70 (rendah).

Tingkat kesukaran ditentukan berdasarkan proporsi jawaban benar (p), dengan kategori: $p < 0{,}30$ (sulit), 0,30–0,70 (sedang), dan > 0,70 (mudah). Daya beda dianalisis menggunakan selisih proporsi jawaban benar antara kelompok atas dan bawah dengan kriteria: $D \geq 0{,}40$ (baik), 0,20–0,39 (cukup), dan < 0,20 (kurang). Efektivitas distraktor ditentukan berdasarkan persentase pemilihan opsi oleh responden. Distraktor dianggap berfungsi jika dipilih minimal 5% responden; kurang dari 5% dikategorikan tidak berfungsi. Penggunaan rubrik penilaian tersebut mengikuti prinsip analisis instrumen klasik (Classical Test Theory) untuk menentukan kualitas butir soal secara objektif sebelum merekomendasikan revisi atau retensi butir.

Hasil

Hasil penelitian ini disajikan untuk memberikan gambaran empiris mengenai kelayakan instrumen tes tata boga yang diujicobakan kepada 75 responden. Analisis dilakukan secara bertahap, dimulai dari karakteristik responden, dilanjutkan dengan statistik deskriptif dan screening butir soal, serta pengujian validitas dan reliabilitas instrumen. Penyajian hasil analisis dimaksudkan agar kualitas instrumen dapat diketahui secara menyeluruh, baik dari sisi kemampuan mengukur kompetensi maupun dari konsistensi hasil pengukuran.

Karakteristik Responden

Karakteristik responden merupakan dasar penting untuk memahami konteks data yang diperoleh dalam penelitian ini. Informasi mengenai jumlah responden, distribusi gender, capaian skor, serta waktu pengerjaan tes memberikan gambaran umum mengenai kondisi partisipan yang terlibat. Dengan mengetahui karakteristik responden, hasil analisis instrumen dapat

dipahami secara lebih komprehensif karena kualitas instrumen juga dipengaruhi oleh keragaman maupun homogenitas partisipan.

Table 1. Karaktersitik Responden

Karakteristik	Hasil
Jumlah Responden	75 orang
Laki-laki	32 (43%)
Perempuan	43 (57%)
Skor Tertinggi	98,46
Skor Terendah	72,00
Rata-rata Skor	87,50
Rata-rata Waktu Pengerjaan	92 Menit
Rentang Waktu Pengerjaan	53 - 150 Menit

Berdasarkan Tabel 1, jumlah responden yang dianalisis sebanyak 75 orang, terdiri atas 43 perempuan (57%) dan 32 laki-laki (43%). Skor tertinggi yang dicapai responden adalah 98,46 dan skor terendah adalah 72,00, dengan rata-rata sebesar 87,50 (SD = 5,2). Hasil ini menunjukkan bahwa responden memiliki kemampuan yang relatif homogen dengan pencapaian rata-rata tinggi. Selain itu, rata-rata waktu pengerjaan adalah 1 jam 32 menit dengan rentang antara 53 menit hingga 2 jam 30 menit. Hal ini mengindikasikan bahwa semua responden mampu menyelesaikan tes dalam durasi wajar, sehingga data yang diperoleh dapat dianggap representatif untuk menggambarkan kondisi nyata instrumen yang diujicobakan.

Statistik Deskriptif & Screening Soal

Analisis statistik deskriptif dan screening butir soal dilakukan untuk mengetahui kualitas instrumen secara menyeluruh dari aspek validitas, daya beda, tingkat kesukaran, serta distribusi kemampuan responden berdasarkan kelompok atas dan bawah. Tahap ini penting untuk menggambarkan sejauh mana butir soal mampu membedakan peserta didik dengan kemampuan berbeda, sekaligus memastikan bahwa soal tidak terlalu mudah ataupun terlalu sulit. Dengan demikian, instrumen dapat dinilai layak digunakan apabila memiliki proporsi butir soal yang bervariasi dan sesuai dengan kriteria standar pengukuran.

Table 2. Statistik Deskriptif & Screening Soal

Indikator	Rata-rata	Minimum	Maksimum	Interpretasi Umum
Validitas butir (r)	0,36	0,20	0,62	Sebagian besar valid
Daya beda	0,42	0,18	0,68	Mayoritas sedang–baik
Tingkat kesukaran	0,71	0,45	0,92	soal mudah–sedang
Skor total responden	87,50	72,00	98,46	kategori tinggi
Kelompok Rendah (0–29)	–	–	20 orang	24% responden
Kelompok Sedang (30–32)	–	–	41 orang	55% responden
Kelompok Tinggi (33–35)	–	–	14 orang	19% responden

Berdasarkan Tabel 2, rata-rata validitas butir soal adalah 0,36 dengan rentang antara 0,20 hingga 0,62. Hal ini menunjukkan bahwa sebagian besar butir soal berada pada kategori cukup valid, meskipun terdapat sejumlah butir yang perlu direvisi agar kualitas instrumen

Copyright © 2025 The Author(s)

meningkat. Rata-rata daya beda sebesar 0,42 dengan rentang 0,18–0,68, yang berarti sebagian besar soal mampu membedakan dengan baik antara peserta berkemampuan tinggi dan rendah. Tingkat kesukaran soal berada pada rata-rata 0,71, yang menandakan dominasi soal dalam kategori mudah–sedang, sehingga sesuai untuk mengukur variasi kemampuan responden. Distribusi skor responden memperlihatkan bahwa mayoritas berada pada kelompok sedang (55%), diikuti kelompok rendah (24%) dan kelompok tinggi (19%). Temuan ini memperlihatkan bahwa instrumen memiliki sebaran kesukaran dan daya pembeda yang cukup baik, meskipun masih ada butir yang perlu perbaikan agar lebih seimbang.

Validitas dan Reliabilitas Instrumen

Uji validitas dan reliabilitas dilakukan untuk menentukan sejauh mana instrumen dapat digunakan secara sahih dan konsisten dalam mengukur kompetensi responden. Validitas butir soal menunjukkan kemampuan setiap item untuk merepresentasikan konstruk yang hendak diukur, sedangkan reliabilitas mencerminkan konsistensi hasil pengukuran jika instrumen digunakan secara berulang. Instrumen yang valid dan reliabel menjadi syarat penting agar data yang diperoleh benar-benar mencerminkan kemampuan peserta didik, bukan hasil kebetulan atau bias pengukuran.

Table 3. Validitas dan Reliabilitas Instrumen

Aspek	Hasil Ringkasan
Jumlah item keseluruhan	63 butir
Item valid	41 butir (65%)
Item tidak valid	22 butir (35%)
Rentang validitas (r)	0.20 – 0.62
Rata-rata validitas	0.36 (kategori cukup)
Reliabilitas instrumen	α = 0.84 (kategori tinggi)

Berdasarkan Tabel 3, dari total 63 butir soal, sebanyak 41 butir (65%) memenuhi kriteria valid, sedangkan 22 butir (35%) dinyatakan tidak valid. Nilai validitas berada pada rentang 0,20–0,62 dengan rata-rata 0,36. Meskipun nilai rata-rata tersebut masih berada tepat pada batas minimal kelayakan (cut-off $r \geq 0{,}30$), kondisi ini menunjukkan bahwa kualitas representasi konstruk pada sebagian butir soal masih lemah. Pada taraf ini, instrumen memang dapat digunakan, tetapi tingkat akurasi pengukuran cenderung moderat dan belum sepenuhnya ideal.

Reliabilitas instrumen menunjukkan nilai $\alpha = 0{,}84$ yang termasuk kategori tinggi, sehingga konsistensi internal antar-butir tercapai dengan baik. Namun, proporsi butir tidak valid yang cukup besar (35%) berimplikasi langsung terhadap efektivitas instrumen: (1) kemampuan tes dalam mengukur kompetensi secara stabil dapat menurun apabila butir tidak valid tetap dipertahankan; (2) interpretasi skor peserta dapat menjadi bias karena adanya butir yang tidak merefleksikan kompetensi yang seharusnya diukur; dan (3) instrumen membutuhkan revisi struktural agar pemetaan kompetensi lebih akurat dan proporsional. Dengan demikian, instrumen tes tata boga layak digunakan, tetapi perbaikan signifikan tetap diperlukan agar memenuhi standar asesmen kejuruan secara optimal.

Analisis Distraktor

Analisis distraktor dilakukan untuk menilai efektivitas setiap opsi pengecoh pada butir soal. Distraktor yang baik seharusnya dipilih oleh sebagian responden, terutama pada kelompok berkemampuan rendah, karena hal ini menunjukkan bahwa pengecoh bekerja sebagaimana mestinya dan mampu mengurangi peluang peserta menjawab benar secara kebetulan.

Sebaliknya, distraktor yang tidak dipilih sama sekali mengindikasikan adanya kelemahan konstruksi butir, baik dari sisi redaksi, tingkat kesukaran, maupun ketidaktepatan konsep. Oleh karena itu, analisis distraktor menjadi penting untuk menentukan sejauh mana kualitas butir perlu ditingkatkan, serta memastikan bahwa setiap item memenuhi fungsi pengukuran secara optimal.

Table 4. Analisis Distraktor

Kategori Distraktor	Jumlah Butir	Persentase	Karakteristik
Tidak berfungsi sama sekali	15 butir	24%	Hampir semua respon memilih benar
Sebagian berfungsi	32 butir	51%	1–2 opsi dipilih ≥10% responden
Berfungsi baik	16 butir	25%	Distraktor efektif, >30% salah pilih
Total Butir	63 butir	100%	–

Berdasarkan Tabel 4, sebanyak 45 butir soal (71%) memiliki distraktor yang berfungsi baik karena dipilih oleh lebih dari 5% responden. Sebanyak 12 butir soal (19%) memiliki distraktor yang kurang berfungsi optimal karena dipilih oleh kurang dari 5% responden, sedangkan 6 butir soal (10%) memiliki distraktor yang sama sekali tidak berfungsi. Hasil ini menunjukkan bahwa mayoritas distraktor sudah efektif, namun terdapat 18 butir soal (29%) yang masih memerlukan perbaikan, baik melalui revisi redaksi maupun penggantian opsi jawaban. Perbaikan ini penting dilakukan agar kualitas instrumen semakin meningkat, khususnya dalam mengurangi kemungkinan jawaban benar yang diperoleh secara kebetulan.

Secara keseluruhan, hasil penelitian menunjukkan bahwa instrumen tes tata boga yang dianalisis memiliki kualitas yang cukup baik. Sebagian besar butir soal memenuhi kriteria validitas, daya beda, tingkat kesukaran, serta reliabilitas yang tinggi. Instrumen juga mampu membedakan responden berkemampuan tinggi dan rendah, serta mayoritas distraktor berfungsi dengan efektif. Namun demikian, masih terdapat sejumlah butir soal yang tidak valid dan beberapa distraktor yang tidak berfungsi optimal, sehingga memerlukan revisi untuk meningkatkan kualitas instrumen secara menyeluruh. Temuan ini memberikan dasar yang kuat untuk dilakukan pembahasan lebih lanjut mengenai implikasi teoretis maupun praktis dari kelayakan instrumen penilaian dalam pendidikan kejuruan bidang tata boga.

Discussion

Hasil penelitian ini menunjukkan bahwa instrumen tes tata boga yang dianalisis memiliki kualitas yang secara umum dapat dikategorikan baik, meskipun masih memerlukan perbaikan pada sejumlah butir soal. Dari aspek validitas, sebanyak 41 butir soal (65%) dinyatakan valid, sementara 22 butir (35%) dinyatakan tidak valid. Temuan ini mengindikasikan bahwa sebagian besar butir soal telah mampu mengukur kompetensi yang dituju, namun masih terdapat kelemahan pada butir tertentu. Kondisi ini sejalan dengan pernyataan (Puspitasari, 2025) bahwa instrumen yang baik jarang seluruhnya valid; selalu ada butir yang tidak sesuai dengan konstruk yang hendak diukur. Leixiao (2024) dan Pan (2023) juga menegaskan bahwa validitas merupakan syarat mutlak dalam penyusunan instrumen, karena instrumen yang tidak valid tidak akan mampu memberikan gambaran nyata tentang kemampuan peserta didik. Dengan demikian, meskipun proporsi butir valid cukup dominan, perbaikan terhadap butir yang tidak valid tetap penting dilakukan agar instrumen semakin representatif.

Reliabilitas instrumen dalam penelitian ini tergolong tinggi dengan koefisien Cronbach's Alpha sebesar 0,84. Nilai ini menunjukkan konsistensi internal antar butir soal yang kuat,

Copyright © 2025 The Author(s)

sehingga instrumen dapat memberikan hasil pengukuran yang stabil jika digunakan kembali dalam konteks berbeda. Temuan ini konsisten dengan standar reliabilitas yang dikemukakan oleh Abdellatif (2019) dan Winther & Klotz (2013) yang menyatakan bahwa nilai α di atas 0,80 dapat dianggap sangat baik. Penelitian sejenis oleh (Sutadji et al., 2021) juga menemukan bahwa instrumen dengan reliabilitas tinggi mampu meningkatkan keadilan dalam proses penilaian, karena setiap peserta didik akan mendapatkan hasil yang konsisten. Hal ini penting dalam konteks pendidikan kejuruan, karena hasil evaluasi berfungsi sebagai dasar penentuan kelulusan dan juga refleksi kualitas pembelajaran.

Dari segi tingkat kesukaran, instrumen ini didominasi oleh soal kategori mudah–sedang, dengan rata-rata 0,71. Temuan ini sesuai dengan pandangan (Warju et al., 2020) yang menyatakan bahwa tes yang baik sebaiknya terdiri dari kombinasi butir mudah, sedang, dan sulit agar dapat mengukur kemampuan peserta secara komprehensif. Namun, proporsi soal mudah yang cukup tinggi perlu diantisipasi agar tidak menurunkan daya tantang instrumen. Dominasi soal mudah dapat menyebabkan peserta dengan kemampuan tinggi tidak terbedakan secara optimal (Jatmoko et al., 2025). Oleh karena itu, penyusunan instrumen selanjutnya perlu menambah jumlah soal kategori sulit, khususnya yang menuntut keterampilan berpikir tingkat tinggi *(higher order thinking skills)*.

Daya beda soal dalam penelitian ini berada pada rata-rata 0,42, yang berarti mayoritas soal memiliki kemampuan membedakan peserta berkemampuan tinggi dan rendah dengan cukup baik. Hal ini sejalan dengan temuan Chao (2016 dan Farmer (1933) yang menyatakan bahwa daya beda merupakan indikator penting dalam asesmen berbasis kompetensi, karena instrumen yang baik harus mampu menyeleksi dan mengklasifikasikan peserta berdasarkan level pencapaiannya. Temuan ini juga konsisten dengan penelitian lain yang menemukan bahwa butir soal dengan daya beda sedang–tinggi lebih efektif dalam memprediksi kemampuan praktik peserta didik tata boga (Jannah et al., 2025). Dengan demikian, meskipun masih terdapat beberapa butir soal dengan daya beda rendah, secara keseluruhan instrumen telah memenuhi kriteria untuk membedakan kemampuan responden.

Analisis kelompok atas dan bawah semakin memperkuat temuan daya beda. Rata-rata skor kelompok atas sebesar 95,4 dibandingkan dengan 76,8 pada kelompok bawah menunjukkan adanya perbedaan capaian yang cukup jelas. Hal ini menunjukkan bahwa instrumen berhasil mengidentifikasi variasi kemampuan responden, sesuai dengan teori pengukuran klasik yang menyebutkan bahwa tes harus mampu memberikan distribusi skor yang memadai (Holst, 2024). Temuan ini juga mendukung penelitian (Subroto, 2019)yang menunjukkan bahwa instrumen dengan perbedaan jelas antara kelompok atas dan bawah dapat meningkatkan objektivitas dalam asesmen PPG. Dengan kata lain, instrumen tata boga dalam penelitian ini memiliki potensi besar untuk digunakan sebagai alat seleksi maupun evaluasi pencapaian kompetensi guru kejuruan.

Analisis distraktor juga memberikan temuan menarik. Sebanyak 71% butir soal memiliki distraktor yang berfungsi baik, sementara sisanya (29%) masih kurang berfungsi atau tidak berfungsi sama sekali. Hasil ini menunjukkan bahwa sebagian besar opsi pengecoh telah efektif dalam memengaruhi responden, khususnya kelompok berkemampuan rendah. Hal ini penting karena distraktor yang tidak berfungsi membuat soal menjadi lebih mudah ditebak dan mengurangi validitas instrument (Hadi et al., 2024). Penelitian oleh Márquez-Herrera (2022) menemukan bahwa revisi distraktor yang tidak berfungsi dapat meningkatkan kualitas soal secara signifikan, karena memperbesar peluang soal untuk menyeleksi kemampuan peserta secara lebih akurat. Oleh karena itu, revisi terhadap distraktor dalam instrumen ini menjadi langkah penting dalam pengembangan lebih lanjut.

Jika dilihat secara keseluruhan, hasil penelitian ini mengindikasikan bahwa instrumen tes tata boga sudah memenuhi sebagian besar kriteria instrumen yang baik, meskipun masih ada aspek yang perlu ditingkatkan. Dari segi validitas, instrumen sudah cukup baik tetapi memerlukan revisi pada 22 butir soal agar lebih sesuai dengan konstruk kompetensi yang diukur. Dari segi reliabilitas, instrumen sudah sangat baik sehingga dapat diandalkan sebagai alat evaluasi. Dari segi daya beda dan tingkat kesukaran, instrumen sudah mampu mengklasifikasikan responden dengan memadai, tetapi tetap perlu memperhatikan keseimbangan proporsi soal mudah, sedang, dan sulit. Dari segi distraktor, meskipun mayoritas berfungsi, perbaikan terhadap distraktor yang lemah akan semakin memperkuat kualitas instrumen.

Dibandingkan dengan penelitian terdahulu, temuan ini memperlihatkan beberapa kesamaan sekaligus perbedaan. Sejalan dengan Sensus et al. (2022), penelitian ini juga menemukan bahwa tidak semua butir soal valid, dan revisi diperlukan agar instrumen lebih representatif. Konsistensi dengan penelitian Sutadji et al. (2021) terlihat pada tingginya reliabilitas instrumen yang menjamin keadilan penilaian. Namun, berbeda dengan penelitian (Anggraini et al., 2024) yang menekankan pengembangan soal berbasis HOTS, penelitian ini masih menunjukkan dominasi soal kategori mudah, sehingga pengembangan lebih lanjut perlu memperhatikan integrasi soal analitis dan evaluatif. Perbandingan ini memperkaya pemahaman bahwa instrumen tes tata boga memiliki karakteristik khusus yang harus terus disempurnakan agar sesuai dengan kebutuhan pendidikan kejuruan di Indonesia.

Implikasi dari penelitian ini cukup luas, baik secara teoretis maupun praktis. Secara teoretis, penelitian ini memperkuat literatur mengenai pentingnya validitas, reliabilitas, daya beda, dan distraktor dalam instrumen penilaian pendidikan kejuruan. Penelitian ini menegaskan bahwa keempat aspek tersebut saling melengkapi untuk menjamin kualitas instrumen. Secara praktis, penelitian ini memberikan rekomendasi bagi guru kejuruan dan lembaga penyelenggara PPG untuk memperbaiki butir soal yang tidak valid serta distraktor yang lemah. Instrumen yang lebih baik akan menghasilkan evaluasi yang lebih objektif dan akurat, sehingga mutu lulusan program PPG juga meningkat.

Selain itu, penelitian ini juga menegaskan urgensi penyusunan instrumen penilaian tata boga yang sesuai dengan standar asesmen modern. Nitko dan Brookhart (2014) menekankan bahwa asesmen kejuruan harus kontekstual, relevan dengan dunia industri, dan mengukur keterampilan nyata yang dibutuhkan peserta didik. Dengan instrumen yang valid, reliabel, dan memiliki daya beda baik, evaluasi pembelajaran tata boga akan lebih mampu mencerminkan kesiapan peserta didik menghadapi tantangan industri kuliner. Oleh karena itu, penelitian ini memberikan kontribusi dalam mendorong pengembangan asesmen yang tidak hanya berorientasi pada pengetahuan teoretis, tetapi juga pada keterampilan praktik yang menjadi ciri khas pendidikan tata boga.

Conclusion

Penelitian ini menunjukkan bahwa instrumen tes tata boga yang digunakan pada program Pendidikan Profesi Guru (PPG) memiliki kualitas yang secara umum dapat dikategorikan baik. Hasil analisis memperlihatkan bahwa mayoritas butir soal memenuhi kriteria validitas, memiliki daya beda yang memadai, serta tingkat kesukaran yang bervariasi dengan dominasi kategori mudah–sedang. Reliabilitas instrumen juga tinggi dengan nilai Cronbach's Alpha 0,84, sehingga instrumen dapat diandalkan untuk memberikan hasil pengukuran yang konsisten. Selain itu, instrumen terbukti mampu membedakan responden berkemampuan tinggi dan rendah, serta sebagian besar distraktor berfungsi dengan baik dalam memengaruhi pilihan jawaban peserta.

Copyright © 2025 The Author(s)

Meskipun demikian, penelitian ini juga menemukan bahwa masih terdapat sejumlah butir soal yang tidak valid dan beberapa distraktor yang tidak berfungsi optimal, sehingga perlu dilakukan revisi dan penyempurnaan. Perbaikan instrumen diharapkan dapat meningkatkan representasi kompetensi yang diukur, sekaligus memastikan bahwa hasil asesmen benar-benar mencerminkan kemampuan peserta didik. Temuan ini menegaskan pentingnya penyusunan instrumen yang tidak hanya valid dan reliabel, tetapi juga seimbang dari segi tingkat kesulitan serta didukung oleh opsi jawaban yang efektif.

Secara praktis, penelitian ini memberikan kontribusi bagi guru kejuruan dan lembaga penyelenggara PPG dalam upaya meningkatkan kualitas instrumen evaluasi di bidang tata boga. Secara teoretis, hasil penelitian memperkaya literatur mengenai asesmen pendidikan kejuruan dan menegaskan bahwa instrumen yang baik harus memenuhi standar pengukuran modern yang kontekstual dan relevan dengan kebutuhan industri. Dengan demikian, instrumen tes tata boga yang layak tidak hanya berfungsi sebagai alat evaluasi pembelajaran, tetapi juga menjadi sarana penting untuk mendukung peningkatan kualitas pendidikan kejuruan secara berkelanjutan.

Acknowledgment

Penulis menyampaikan ucapan terima kasih kepada seluruh guru kejuruan bidang tata boga yang telah berpartisipasi dalam penelitian ini sebagai responden. Apresiasi juga diberikan kepada tim penyelenggara Program Pendidikan Profesi Guru (PPG) yang telah memfasilitasi pelaksanaan tes sehingga penelitian dapat berjalan dengan lancar. Secara khusus, penulis juga menyampaikan penghargaan kepada Universitas Negeri Malang, khususnya Sekolah Pascasarjana, yang telah memberikan dukungan akademik dan kesempatan untuk melaksanakan penelitian ini sebagai bagian dari pengembangan ilmu di bidang pendidikan kejuruan.

References

Abdellatif, H. (2019). Effect of blueprinting methods on test difficulty, discrimination, and reliability indices: Cross-sectional study in an integrated learning program. *Advances in Medical Education and Practice*, *10*, 23–30. https://doi.org/10.2147/AMEP.S190827

Achmad, I. A., Asmas, M. A., & Sukawati, Y. (2024). Evaluasi Program Pendidikan Kecapakan Kerja Pelatihan Otomotif Dalam Meningkatkan Keterampilan Peserta Didik. *Continuing Learning Society Journal*, *2*(1), 40–58.

Amelung, D. (2022). Considering vocational training as selection criterion for medical students: evidence for predictive validity. *Advances in Health Sciences Education*, *27*(4), 933–948. https://doi.org/10.1007/s10459-022-10120-y

Anggraini, S., Dalimunthe, N. S., Sari, D. N., & Husna, H. (2024). Analisis Pemahaman Siswa Terhadap Konsep Vektor di Bidang dalam Pembelajaran Matematika. *Kognitif: Jurnal Riset HOTS Pendidikan Matematika*, *4*(4), 1687–1699. https://doi.org/10.51574/kognitif.v4i4.2313

Castillo-Ortiz, I. (2024). Evaluating culinary skill transfer: A deep learning approach to comparing student and chef dishes using image analysis. *International Journal of Gastronomy and Food Science*, *38*. https://doi.org/10.1016/j.ijgfs.2024.101070

Chao, T. (2016). Construction of the teacher situational judgment test. *Journal of Research in Education Sciences*, *61*(2), 85–117. https://doi.org/10.6209/JORIES.2016.61(2).04

Daryono, R. W., Ramadhan, M. A., Kholifah, N., Isnantyo, F. D., & Nurtanto, M. (2023). An empirical study to evaluate the student competency of vocational education. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(2), 1079–1086.

Davis, M. H., Hall, J. A., & Mayer, P. S. (2016). Developing a new measure of entrepreneurial mindset: Reliability, validity, and implications for practitioners. *Consulting Psychology Journal: Practice and Research*, *68*(1), 21.

Farmer, E. (1933). The reliability of the criteria used for assessing the value of vocational tests. *British Journal of Psychology*, *24*(1), 109.

Gunawan, L. E., & Citriadin, Y. (2025). STRATEGI MANAJEMEN MUTU PEMBELAJARAN DALAM MENGEMBANGKAN KOMPETENSI TATA BOGA DI SMK IT NW PUJUT. *… Manajemen Pendidikan Dan …*. https://www.ejournal.unma.ac.id/index.php/madinasika/article/view/15529

Hadi, S., Purnomo, Imron, A., Pristian, R., Ranu, M. E., & Jamil, A. S. (2024). Adapted Integrated Transformational-Instructional Leadership Behaviors: Construct Validity and Perceptions among Teachers and Principals in Indonesian Vocational Schools. *Eurasian Journal of Educational Research*, *2024*(109), 238–259. https://doi.org/10.14689/ejer.2024.109.013

Holst, J. (2024). Whole Institution Approach: measurable and highly effective in empowering learners and educators for sustainability. *Sustainability Science*, *19*(4), 1359–1376. https://doi.org/10.1007/s11625-024-01506-5

Iktarastiwi, N. (2025). EVALUASI PENERAPAN TUGAS MINI RISET PADA MATA KULIAH PASTRY DI PENDIDIKAN TATA BOGA. *Journal Central Publisher*. https://centralpublisher.co.id/jurnalcentralpublisher/index.php/Publish/article/view/373

Jannah, Y. R., Siregar, J., Gusnita, W., & ... (2025). A Descriptive Analysis of Implementing Basic Culinary Principles in Teaching the Main Course Topic of Continental Cuisine at SMK Negeri 6 Padang. *… Pendidikan Tata Boga …*. http://boga.ppj.unp.ac.id/index.php/jptb/article/view/26896

Jatmoko, D., Basuki, B., Pambudi, A., Setiyono, K. A., & Nugroho, B. (2025). Analisis Butir Soal Teknik Dasar Otomotif Berbasis Diagnostik Kognitif dengan Anates Versi 4.02. *JURNAL RISET RUMPUN ILMU PENDIDIKAN*, *4*(1), 461–471.

Jimenez, S. (2023). The Performance Diagnostic Checklist–Human Services (1.1): An initial assessment of validity and reliability. *Journal of Applied Behavior Analysis*, *56*(4), 884–897. https://doi.org/10.1002/jaba.1017

Kim, Y. (2024). Development and Evaluation of Eating Traits Scale for Adult in Republic of Korea. *Research in Community and Public Health Nursing*, *35*(4), 339–350. https://doi.org/10.12799/rcphn.2024.00521

Leixiao, Z. (2024). Development and reliability and validity testing of a medication literacy scale for medical college students. *BMC Medical Education*, *24*(1). https://doi.org/10.1186/s12909-024-06222-3

Márquez-Herrera, R. M. (2022). Reliability and validity of a clinical competence test for dietitians caring patients with early chronic kidney disease. *Journal of Nutritional Science*, *11*. https://doi.org/10.1017/jns.2022.4

Menchetti, B. M., & Rusch, F. R. (1988). Reliability and validity of the vocational assessment and curriculum guide. *American Journal of Mental Retardation: AJMR*, *93*(3), 283–289.

Mia, Y. G., & Sulastri, S. (2023). Analisis kompetensi profesional guru. *Journal of Practice Learning and Educational Development*, *3*(1), 49–55.

Noor, N., Suherman, S., & Ganiadi, M. (2024). Ketercapaian Program Pelatihan Tata Boga Melalui Implementasi Evaluasi Model CIPP (Context, Input, Process, Product) di LPK Hesar Kota Cilegon. *Jurnal Ilmiah Profesi Pendidikan*. https://jipp.unram.ac.id/index.php/jipp/article/view/2139

Copyright © 2025 The Author(s)

Pan, L. (2023). Development of nutrition label use scale for patients of coronary heart disease and examination of its reliability and validity. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1168951

Pangestika, R. R. (2015). *Pendidikan profesi guru (PPG): Strategi pengembangan profesionalitas guru dan peningkatan mutu pendidikan Indonesia*.

Pasaribu, K. F., Saragih, V. G., & Renaldi, R. (2025). Pengembangan Buku Saku Interaktif sebagai Media Pembelajaran Etika Profesi pada Mahasiswa Pendidikan Tata Boga. *… RUMPUN ILMU PENDIDIKAN*. https://prin.or.id/index.php/JURRIPEN/article/view/5583

Puspitasari, S. (2025). Self-Assessment Terhadap Kemandirian Peserta Pelatihan Keterampilan Tata Boga di LPP Kelas IIA Jakarta. *Jurnal Kajian Praxis Pendidikan Masyarakat …*. https://journal.unj.ac.id/unj/index.php/penmasindo/article/view/48415

Sensus, M., Arifin, K., & Munir, A. (2022). Validitas Soal pada Asesmen Kompetensi Minimum Materi Ekologi SMA Kelas X. *Wahana-Bio: Jurnal Biologi Dan Pembelajarannya*, *14*(1), 1–10.

Subroto, J. G. (2019). Peningkatan kualitas pendidikan: Program pendidikan profesi guru prajabatan dalam pemenuhan kebutuhan guru profesional di Indonesia. *Artikel Ilmiah. Http://Download. Garuda. Kemdikbud. Go. Id/Article. Php*.

Sutadji, E., Susilo, H., Wibawa, A. P., Jabari, N. A. M., & Rohmad, S. N. (2021). Authentic assessment implementation in natural and social science. *Education Sciences*, *11*(9). https://doi.org/10.3390/educsci11090534

Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis kualitas butir soal tipe HOTS pada kompetensi sistem REM di sekolah menengah kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *17*(1), 95–104.

Winther, E., & Klotz, V. K. (2013). Measurement of vocational competences: an analysis of the structure and reliability of current assessment practices in economic domains. *Empirical Research in Vocational Education and Training*, *5*(1), 2.

Zhang, X. (2025). Construction of a Higher Vocational Culinary Education Quality Evaluation Model for University-Industry Integration Based on Adaptive BP and DDAE-SVR. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3603627

---Halaman ini sengaja dikosongkan---