EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

ANALISIS KARAKTERISTIK BUTIR SOAL ASESMEN PELATIHAN *EFFECTIVE RECRUITMENT* BERBASIS TEORI TES KLASIK

Fatichatur Rohmah[1], Citra Fitri Kholidiya[2], Bachtiar Sjaiful Bachri[3], Hirnanda Dimas Pradana[4]
Program Studi Teknologi Pendidikan, Fakultas Ilmu Pendidikan, Universitas Negeri Surabaya[1,2,3,4]
e-mail: fatichatur.23183@mhs.unesa.ac.id

Diterima: *4/10/2026*; Direvisi: *11/10/2026*; Diterbitkan: *23/10/2026*

ABSTRAK

Kualitas instrumen asesmen kognitif menentukan sejauh mana hasil tes dapat merepresentasikan penguasaan kompetensi peserta pelatihan secara memadai. Oleh karena itu, evaluasi karakteristik butir diperlukan untuk memastikan bahwa instrumen yang digunakan menghasilkan informasi penilaian yang dapat dipertanggungjawabkan. Penelitian ini bertujuan mengevaluasi kualitas psikometrik instrumen asesmen kognitif pada Program Pelatihan *Effective Recruitment* di PT Edukarir Global Nusantara. Penelitian menggunakan pendekatan kuantitatif deskriptif-evaluatif dengan kerangka *Classical Test Theory* (CTT). Sampel ditentukan melalui teknik *total sampling*, yang mencakup 40 peserta pelatihan. Data diperoleh dari dokumentasi hasil *pre-test* berupa 10 butir soal pilihan ganda dengan empat opsi jawaban. Analisis meliputi validitas empiris, reliabilitas, tingkat kesukaran, daya pembeda, dan fungsi pengecoh. Hasil menunjukkan bahwa hanya 50% butir memenuhi kriteria validitas empiris, sedangkan koefisien reliabilitas KR-20 sebesar 0,324 tergolong rendah. Sebanyak 90% butir berada dalam kategori mudah dan 80% memiliki daya pembeda yang buruk. Temuan tersebut memperlihatkan bahwa instrumen belum optimal dalam membedakan tingkat penguasaan peserta. Dengan demikian, peninjauan substansi soal, penyempurnaan opsi jawaban, dan pengujian ulang diperlukan sebelum instrumen digunakan sebagai dasar evaluasi kompetensi peserta pelatihan.

Kata Kunci: *Asesmen Kognitif, Classical Test Theory, Daya Pembeda, Effective Recruitment, Kualitas Psikometrik*

ABSTRACT

The quality of a cognitive assessment instrument determines the extent to which test results accurately represent participants' mastery of training competencies. Therefore, evaluating item characteristics is necessary to ensure that the instrument produces accountable assessment information. This study aimed to evaluate the psychometric quality of the cognitive assessment instrument used in the *Effective Recruitment* Training Program at PT Edukarir Global Nusantara. The study employed a descriptive-evaluative quantitative approach based on *Classical Test Theory* (CTT). The sample was determined using *total sampling*, comprising 40 training participants. Data were obtained from documented *pre-test* results consisting of 10 multiple-choice items, each with four answer options. The analysis covered empirical validity, reliability, item difficulty, discrimination power, and distractor functioning. The results showed that only 50% of the items met the empirical validity criteria, while the KR-20 reliability coefficient was 0.324, indicating low internal consistency. Furthermore, 90% of the items were classified as easy, and 80% demonstrated poor discrimination power. These findings indicate

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

that the instrument is not yet optimal in distinguishing differences in participants' levels of mastery. Therefore, item content should be reviewed, answer options refined, and the instrument retested before it is used as a basis for evaluating training participants' competencies.

Keywords: *Cognitive Assessment, Classical Test Theory, Discrimination Index, Effective Recruitment, Psychometric Quality*

PENDAHULUAN

Pengembangan kompetensi sumber daya manusia melalui pelatihan tidak cukup dinilai dari terselenggaranya kegiatan atau tersampaikannya materi, tetapi juga perlu ditinjau dari kemampuan peserta memahami dan menerapkan pengetahuan yang diperoleh. Dalam organisasi, evaluasi pelatihan memerlukan instrumen yang mampu menggambarkan penguasaan kompetensi secara terukur agar kebutuhan pengembangan profesional dapat diidentifikasi secara lebih tepat. O'Neill dan Cleary (2023) membahas pengukuran transfer pelatihan melalui tujuan yang dirumuskan dalam lima komponen, sehingga evaluasi perlu mempertimbangkan keterkaitan antara sasaran pelatihan dan bukti pencapaian peserta. Perspektif tersebut relevan bagi program *Effective Recruitment* di PT Edukarir Global Nusantara, yang ditujukan kepada praktisi *Human Resources* (HR) dan manajer lini (*user*) untuk memperkuat pemahaman mengenai proses rekrutmen. Dengan demikian, asesmen awal perlu dirancang agar mampu memberikan informasi yang dapat dipercaya mengenai penguasaan pengetahuan peserta sebelum pelatihan dilaksanakan.

Tes pilihan ganda merupakan salah satu instrumen yang dapat digunakan untuk memetakan penguasaan pengetahuan melalui respons peserta terhadap sejumlah alternatif jawaban. Kemudahan pelaksanaan dan penskorannya tidak dengan sendirinya menjamin bahwa skor yang dihasilkan telah menggambarkan kemampuan peserta secara akurat. Ramadani dan Handayani (2024) membahas instrumen penilaian hasil pembelajaran kognitif melalui tes objektif, sedangkan Nurjanah et al. (2024) mengkaji kualitas psikometrik tes pilihan ganda berdasarkan pendekatan *Classical Test Theory* (CTT) dengan bantuan AnBuso, Iteman, dan R. Kedua kajian tersebut menjadi landasan untuk memandang kualitas instrumen sebagai persoalan empiris yang perlu diperiksa, bukan sekadar diasumsikan dari bentuk soal atau penilaian penyusunnya. Dalam pelatihan profesional, pemeriksaan tersebut diperlukan agar skor asesmen dapat digunakan untuk mengidentifikasi kebutuhan belajar dan mendukung pengambilan keputusan terkait pengembangan kompetensi.

Pendekatan CTT menyediakan kerangka untuk menelaah karakteristik butir melalui validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Kelima parameter tersebut memberikan informasi yang saling melengkapi mengenai fungsi instrumen, sehingga kualitas tes tidak cukup ditentukan hanya oleh satu indikator. Nurcahya et al. (2026) menganalisis butir soal asesmen sumatif mata pelajaran ekonomi, sedangkan Sabat et al. (2026) mengkaji kualitas soal pilihan ganda materi statistika menggunakan pendekatan CTT. Kedua penelitian tersebut memberikan pembanding empiris mengenai penggunaan analisis butir dalam konteks pendidikan formal, tetapi karakteristik hasilnya tidak dapat langsung digeneralisasikan pada asesmen di lingkungan perusahaan. Perbedaan sasaran kompetensi, karakteristik peserta, dan tujuan penggunaan tes menjadikan pemeriksaan instrumen pelatihan korporat sebagai kebutuhan tersendiri.

Kebutuhan untuk menyesuaikan instrumen dengan kompetensi yang hendak diukur juga terlihat pada pengembangan asesmen vokasional dan evaluasi kinerja. Rozi et al. (2026)

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

membahas pengembangan instrumen asesmen kognitif pada kompetensi sistem *air conditioning* kendaraan ringan, sementara Rasyid et al. (2025) mengkaji validasi instrumen evaluasi kinerja menggunakan metode *Balanced Scorecard* di lingkungan perusahaan. Hasanah dan Sahlan (2026) juga menelaah validitas dan reliabilitas asesmen berbasis proyek dalam pembelajaran *hybrid*. Walaupun konteks dan bentuk instrumen pada penelitian tersebut berbeda, ketiganya memberikan perspektif bahwa ketepatan pengukuran perlu disesuaikan dengan tujuan asesmen dan karakteristik kompetensi yang dinilai. Oleh sebab itu, instrumen kognitif untuk pelatihan rekrutmen perlu dievaluasi berdasarkan respons peserta yang sesungguhnya agar kelayakannya tidak hanya bergantung pada pertimbangan konseptual penyusun.

Salah satu persoalan yang dapat mengurangi kemampuan tes dalam memetakan kompetensi adalah dominasi butir yang terlalu mudah sehingga sebagian besar peserta memberikan jawaban benar. Keadaan tersebut berpotensi menimbulkan *ceiling effect*, yaitu kondisi ketika skor terkonsentrasi pada batas atas pengukuran dan informasi untuk membedakan kemampuan peserta menjadi terbatas. Rezigalla et al. (2024) mengkaji dampak efisiensi distraktor terhadap indeks kesukaran dan daya pembeda soal pilihan ganda, sehingga fungsi pengecoh perlu diperhatikan sebagai bagian dari kualitas pengukuran. Sementara itu, Kambara (2026) menelaah karakteristik psikometrik skala pembelajaran eksperiensial di tempat kerja pada ahli gizi terdaftar berbasis rumah sakit di Jepang, yang memberikan perspektif mengenai evaluasi instrumen dalam konteks profesional meskipun konstruk yang diukur berbeda. Astuti dan Hanifa (2026), melalui kajian edukasi *personal hygiene* menstruasi berbasis *leaflet*, menyediakan konteks tambahan mengenai evaluasi pengetahuan melalui intervensi edukatif, tetapi tidak dapat dijadikan bukti langsung mengenai karakteristik psikometrik tes pilihan ganda. Perbedaan konteks tersebut menegaskan perlunya memilih bukti penelitian terdahulu secara kritis dan menguji karakteristik instrumen sesuai dengan tujuan pengukurannya.

Pada program *Effective Recruitment* di PT Edukarir Global Nusantara, persoalan kualitas instrumen perlu diperiksa karena asesmen awal dimaksudkan untuk menggambarkan penguasaan pengetahuan peserta sebelum mereka menerima materi pelatihan. Pemeriksaan pendahuluan memperlihatkan pemusatan skor pada rentang tinggi, sehingga dugaan *ceiling effect* perlu ditelaah melalui analisis empiris setiap butir soal. Jika sebagian besar butir terlalu mudah atau pilihan pengecohnya tidak berfungsi, hasil *pre-test* berpotensi kurang informatif untuk mengidentifikasi kesenjangan kompetensi peserta. Penelitian terdahulu telah membahas analisis butir dalam pendidikan formal, pengembangan asesmen vokasional, evaluasi kinerja, dan pengujian instrumen pada konteks profesional, tetapi karakteristik empiris tes kognitif khusus untuk pelatihan rekrutmen korporat masih perlu dikaji lebih lanjut. Kesenjangan penelitian ini terletak pada kebutuhan untuk memeriksa secara terpadu validitas butir, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor dalam satu instrumen asesmen pelatihan rekrutmen.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengevaluasi karakteristik psikometrik instrumen asesmen kognitif pada program pelatihan *Effective Recruitment* di PT Edukarir Global Nusantara menggunakan pendekatan CTT. Analisis difokuskan pada lima parameter, yaitu validitas butir, konsistensi internal, tingkat kesukaran, daya pembeda, dan efektivitas distraktor, untuk mengidentifikasi aspek instrumen yang telah berfungsi dan aspek yang memerlukan perbaikan. Kontribusi penelitian ini terletak pada penerapan analisis terpadu terhadap instrumen asesmen kognitif dalam konteks pelatihan rekrutmen korporat, sehingga

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

pembahasan kualitas soal dikaitkan langsung dengan kebutuhan pemetaan kompetensi awal praktisi. Hasil penelitian diharapkan menjadi dasar empiris bagi perusahaan untuk merevisi butir yang belum memadai, mengembangkan pengecoh yang lebih representatif terhadap kemungkinan kesalahan pemahaman peserta, dan memperbaiki prosedur evaluasi pelatihan. Selain itu, temuan penelitian dapat menjadi titik awal bagi pengujian ulang instrumen pada kelompok profesional yang lebih beragam serta kajian lanjutan mengenai pendekatan pengukuran yang sesuai dengan karakteristik asesmen kompetensi kerja.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk mengevaluasi kualitas psikometrik instrumen asesmen kognitif pada pelatihan *Effective Recruitment* di PT Edukarir Global Nusantara. Evaluasi dilakukan dengan kerangka *Classical Test Theory* (CTT), yang digunakan untuk mengidentifikasi karakteristik butir soal berdasarkan respons peserta terhadap instrumen yang diujikan. Data penelitian bersumber dari hasil pengerjaan *pre-test* yang terdiri atas 10 butir soal pilihan ganda dan diikuti oleh 40 peserta pelatihan. Seluruh respons peserta menjadi dasar penghitungan parameter kualitas butir dan konsistensi internal instrumen. Penelitian ini berfokus pada evaluasi karakteristik instrumen yang digunakan, bukan pada pengujian hubungan kausal antara pelaksanaan pelatihan dan peningkatan kompetensi peserta.

Analisis data dilaksanakan melalui lima tahap yang saling melengkapi, yaitu pengujian validitas empiris, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Validitas butir diperiksa menggunakan korelasi *point-biserial* dengan membandingkan koefisien hasil perhitungan terhadap nilai kritis yang ditetapkan pada taraf signifikansi 5% untuk jumlah responden sebanyak 40 orang. Konsistensi internal instrumen dianalisis menggunakan koefisien Kuder–Richardson 20 (KR-20), sesuai dengan karakteristik skor dikotomis yang membedakan jawaban benar dan salah. Tingkat kesukaran dihitung berdasarkan proporsi peserta yang menjawab setiap butir dengan benar, sedangkan daya pembeda dievaluasi melalui perbandingan hasil kelompok atas dan kelompok bawah yang masing-masing mencakup 27% peserta. Selanjutnya, efektivitas distraktor diperiksa dengan mengidentifikasi distribusi pilihan pada setiap opsi jawaban untuk mengetahui pola pemilihan pengecoh oleh peserta.

Hasil perhitungan setiap parameter disajikan secara deskriptif dalam bentuk tabel agar karakteristik masing-masing butir dapat dibandingkan secara sistematis. Interpretasi validitas mengacu pada perbandingan koefisien korelasi dengan nilai kritis yang digunakan, sedangkan reliabilitas ditafsirkan berdasarkan kriteria kategorisasi yang ditetapkan dalam penelitian. Indeks tingkat kesukaran dan daya pembeda digunakan untuk mengelompokkan butir sesuai dengan kriteria evaluasi yang diterapkan, sementara analisis distraktor diarahkan pada frekuensi serta proporsi pemilihan setiap opsi jawaban. Apabila ditemukan butir tanpa variasi respons atau hasil perhitungan yang tidak terdefinisi secara normal, kondisi tersebut perlu dicatat dan diverifikasi sebelum ditarik kesimpulan mengenai kualitas butir. Melalui prosedur tersebut, evaluasi diharapkan menghasilkan gambaran empiris mengenai kekuatan dan keterbatasan instrumen serta menyediakan dasar untuk memperbaiki butir soal pada pelaksanaan asesmen berikutnya.

.

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

HASIL DAN PEMBAHASAN

Hasil

Evaluasi kualitas instrumen asesmen kognitif pada pelatihan Effective Recruitment dilakukan menggunakan pendekatan Classical Test Theory (CTT). Analisis mencakup lima aspek, yaitu validitas butir, reliabilitas instrumen, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Data diperoleh dari respons 40 peserta terhadap 10 butir soal pre-test. Hasil analisis setiap parameter disajikan pada subbagian berikut.

1. Validitas Butir Soal

Validitas empiris dianalisis untuk mengetahui kemampuan setiap butir dalam membedakan peserta berdasarkan skor total tes. Pengujian menggunakan korelasi point-biserial (r_pbi) dengan nilai pembanding r_tabel = 0,312 pada taraf signifikansi 5% untuk 40 responden. Butir dinyatakan valid apabila koefisien korelasinya memenuhi kriteria yang digunakan dalam pengujian. Hasil perhitungan untuk seluruh butir disajikan pada Tabel 1.

Tabel 1. Hasil Uji Validitas Butir Soal Pre-test

Butir	Konstruk Pengukuran	(r_pbi)	Keputusan
Q1	Tujuan utama proses rekrutmen	Tidak terhitung (0,000)	Tidak valid
Q2	Peran User dalam rekrutmen	0,051	Tidak valid
Q3	Alur umum tahapan rekrutmen	0,470	Valid
Q4	Tujuan esensial wawancara kerja	0,236	Tidak valid
Q5	Karakteristik wawancara behavioral	0,799	Valid
Q6	Karakteristik wawancara situational	0,470	Valid
Q7	Pemanfaatan teknik probing	0,421	Valid
Q8	Etika profesional dalam wawancara	0,051	Tidak valid
Q9	Tujuan technical assessment	0,529	Valid
Q10	Urgensi kolaborasi HR dan User	0,051	Tidak valid

Berdasarkan Tabel 1, sebanyak 5 dari 10 butir soal (50%) memenuhi kriteria validitas empiris, yaitu Q3, Q5, Q6, Q7, dan Q9. Lima butir lainnya belum memenuhi kriteria yang ditetapkan. Pada Q1, seluruh peserta menjawab benar sehingga tidak terdapat variasi skor jawaban. Kondisi tersebut menyebabkan koefisien korelasi point-biserial tidak dapat dihitung secara normal; angka 0,000 perlu dipahami sebagai penanda keluaran perhitungan, bukan koefisien korelasi yang menunjukkan tidak adanya hubungan secara substantif.

2. Reliabilitas Instrumen

Setelah validitas setiap butir diperiksa, analisis dilanjutkan dengan pengujian reliabilitas untuk mengetahui konsistensi internal instrumen. Karena respons setiap butir dikodekan secara dikotomis, yaitu benar dan salah, koefisien reliabilitas dihitung menggunakan rumus Kuder–Richardson 20 (KR-20). Pengujian dilakukan terhadap keseluruhan 10 butir soal dengan data dari 40 peserta. Ringkasan hasil pengujian disajikan pada Tabel 2.

Tabel 2. Hasil Uji Reliabilitas Instrumen

Parameter	Hasil
Jumlah peserta	40
Jumlah butir soal	10
Metode pengujian	KR-20

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Koefisien reliabilitas	0,324
Kategori berdasarkan kriteria penelitian	Sangat rendah

Tabel 2 menunjukkan bahwa koefisien reliabilitas instrumen sebesar 0,324. Berdasarkan kategori yang digunakan dalam penelitian ini, nilai tersebut termasuk dalam kategori sangat rendah. Temuan ini memperlihatkan bahwa konsistensi internal instrumen masih terbatas. Oleh karena itu, hasil pengukuran perlu ditafsirkan secara hati-hati, khususnya apabila digunakan untuk menggambarkan tingkat penguasaan kompetensi peserta secara individual.

3. Tingkat Kesukaran (*Difficulty Index*)

Analisis tingkat kesukaran dilakukan untuk mengetahui proporsi peserta yang menjawab setiap butir dengan benar. Indeks kesukaran (P) dihitung dengan membandingkan jumlah jawaban benar dengan jumlah seluruh peserta. Nilai indeks yang semakin mendekati 1 menunjukkan bahwa semakin banyak peserta menjawab butir tersebut dengan benar. Distribusi tingkat kesukaran pada 10 butir soal disajikan dalam Tabel 3.

Tabel 3. Distribusi Indeks Tingkat Kesukaran Butir Soal

Butir	Jawaban Benar	Indeks (P)	Proporsi	Kategori
Q1	40	1,000	100,0%	Mudah
Q2	39	0,975	97,5%	Mudah
Q3	38	0,950	95,0%	Mudah
Q4	39	0,975	97,5%	Mudah
Q5	23	0,575	57,5%	Sedang
Q6	38	0,950	95,0%	Mudah
Q7	39	0,975	97,5%	Mudah
Q8	39	0,975	97,5%	Mudah
Q9	37	0,925	92,5%	Mudah
Q10	39	0,975	97,5%	Mudah

Berdasarkan Tabel 3, sembilan butir soal (90%) berada dalam kategori mudah, sedangkan satu butir, yaitu Q5, termasuk kategori sedang. Proporsi jawaban benar pada sembilan butir tersebut berkisar antara 92,5% hingga 100%. Sebaliknya, Q5 memiliki indeks kesukaran sebesar 0,575, dengan 57,5% peserta menjawab benar. Dengan demikian, distribusi indeks kesukaran menunjukkan bahwa instrumen didominasi oleh butir yang dapat dijawab dengan benar oleh hampir seluruh peserta.

4. Daya Pembeda (*Discrimination Index*)

Daya pembeda dianalisis untuk mengetahui kemampuan setiap butir dalam membedakan peserta dengan skor tinggi dan peserta dengan skor rendah. Perhitungan menggunakan metode kelompok atas dan kelompok bawah sebesar 27%, dengan masing-masing kelompok terdiri atas 11 peserta. Indeks daya pembeda (D) diperoleh dari selisih proporsi jawaban benar kelompok atas dan kelompok bawah. Hasil perhitungan setiap butir disajikan pada Tabel 4.

Tabel 4. Indeks Daya Pembeda Butir Soal

Butir	BA	BB	Indeks (D)	Kategori

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Q1	11	11	0,00	Jelek
Q2	11	11	0,00	Jelek
Q3	11	9	0,18	Jelek
Q4	11	10	0,09	Jelek
Q5	11	0	1,00	Sangat baik
Q6	11	9	0,18	Jelek
Q7	11	10	0,09	Jelek
Q8	11	11	0,00	Jelek
Q9	11	8	0,27	Cukup
Q10	11	11	0,00	Jelek

Tabel 4 memperlihatkan bahwa delapan butir soal (80%) memiliki daya pembeda dalam kategori jelek, dengan indeks (D) antara 0,00 dan 0,18. Q9 memperoleh indeks 0,27 dan termasuk kategori cukup, sedangkan Q5 memiliki indeks tertinggi, yaitu 1,00, dengan seluruh peserta kelompok atas menjawab benar dan tidak ada peserta kelompok bawah yang menjawab benar. Hasil ini memperlihatkan adanya perbedaan kemampuan diskriminasi yang mencolok antara Q5 dan sebagian besar butir lainnya.

5. Efektivitas Distraktor

Analisis efektivitas distraktor dilakukan untuk mengetahui sejauh mana opsi jawaban yang salah berfungsi sebagai pengecoh bagi peserta. Analisis ini difokuskan pada Q5 tentang *Behavioral Event Interview* (BEI), yang memiliki tingkat kesukaran sedang dan daya pembeda sangat baik. Distribusi pilihan jawaban diperiksa berdasarkan jumlah dan proporsi peserta yang memilih setiap opsi. Hasil analisis pilihan jawaban pada Q5 disajikan pada Tabel 5.

Tabel 5. Analisis Opsi Jawaban Butir Q5

Opsi Jawaban	Status	Jumlah Pemilih	Proporsi	Indeks IP
A. Rencana kandidat masa depan	Distraktor	2	5,0%	35,3%
B. Pengetahuan teoretis kandidat	Distraktor	1	2,5%	17,6%
C. Pengalaman nyata masa lalu	Kunci jawaban	23	57,5%	—
D. Preferensi pribadi kandidat	Distraktor	14	35,0%	246,9%

Berdasarkan Tabel 5, opsi C merupakan jawaban yang paling banyak dipilih, yaitu oleh 23 peserta (57,5%). Opsi D menjadi distraktor yang paling banyak dipilih, yakni oleh 14 peserta (35,0%), sedangkan opsi A dan B masing-masing dipilih oleh 2 peserta (5,0%) dan 1 peserta (2,5%). Distribusi tersebut menunjukkan bahwa opsi pengecoh pada Q5 tidak memiliki tingkat pemilihan yang seragam. Opsi D cukup menarik bagi sebagian peserta, sementara opsi A dan B relatif jarang dipilih.

Pembahasan

Hasil evaluasi instrumen asesmen kognitif pada pelatihan *Effective Recruitment* memperlihatkan bahwa keberadaan tes pilihan ganda belum secara otomatis menjamin tersedianya informasi yang memadai mengenai penguasaan kompetensi peserta. Dari sepuluh butir yang dianalisis, hanya lima butir memenuhi kriteria validitas empiris, sedangkan koefisien reliabilitas KR-20 sebesar 0,324 menunjukkan konsistensi internal yang rendah menurut kriteria penelitian. Temuan tersebut menempatkan kualitas instrumen sebagai aspek yang perlu diperhatikan dalam evaluasi pelatihan, bukan sekadar sebagai pelengkap kegiatan

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

pembelajaran. Dalam konteks pengembangan sumber daya manusia, hasil tes idealnya dapat memberikan gambaran yang cukup akurat mengenai pengetahuan peserta agar keputusan tindak lanjut pelatihan tidak hanya bertumpu pada penyelesaian program. Perspektif ini sejalan dengan pembahasan Ni'mah et al. (2025) mengenai implementasi evaluasi pelatihan, yang relevan sebagai landasan untuk menempatkan pengukuran hasil belajar dalam proses evaluasi program secara lebih luas.

Permasalahan validitas butir dalam penelitian ini tampak pada banyaknya butir yang belum mampu menunjukkan hubungan empiris memadai dengan skor total. Kondisi tersebut mengisyaratkan perlunya peninjauan terhadap kesesuaian antara materi yang ditanyakan, kompetensi yang hendak diukur, konstruksi opsi jawaban, dan karakteristik peserta yang mengikuti asesmen. Andani dan Rizki (2023) membahas validitas instrumen dalam konteks penelitian pendidikan, sedangkan Pratiwi dan Sugiharto (2025) mengkaji validitas butir serta ketepatan target ukur melalui pemodelan *Rasch*. Kedua rujukan tersebut memberikan sudut pandang bahwa evaluasi instrumen tidak cukup dilakukan dengan memeriksa keberadaan soal, tetapi perlu mempertimbangkan bukti yang mendukung ketepatan pengukuran. Meskipun demikian, pendekatan *Rasch* berbeda dari CTT yang digunakan dalam penelitian ini, sehingga temuan dari kedua pendekatan tersebut tidak dapat diperlakukan sebagai hasil yang identik. Secara praktis, butir yang belum memenuhi kriteria validitas perlu diperiksa kembali substansi dan formulasi pertanyaannya sebelum diputuskan untuk dipertahankan, direvisi, atau diganti.

Distribusi tingkat kesukaran menunjukkan bahwa sembilan butir memiliki indeks antara 0,925 dan 1,000, sementara hanya Q5 berada pada kategori sedang dengan indeks 0,575. Dominasi butir yang sangat mudah membatasi variasi respons karena hampir seluruh peserta mampu memberikan jawaban benar pada sebagian besar soal. Pola ini sejalan dengan perhatian Fitryansyah dan Hilmiyati (2024) serta Wardana et al. (2026) terhadap keterkaitan tingkat kesukaran, daya pembeda, dan fungsi pengecoh dalam menentukan kualitas butir tes. Rivaldi et al. (2025), melalui kajian kualitas butir soal asesmen sumatif geografi, juga menjadi rujukan kontekstual untuk membahas evaluasi karakteristik butir dalam instrumen penilaian. Walaupun demikian, proporsi jawaban benar yang tinggi tidak dengan sendirinya membuktikan bahwa soal tidak berkualitas, sebab tingkat kesukaran yang sesuai tetap bergantung pada tujuan asesmen dan kompetensi yang ditargetkan. Dalam penelitian ini, persoalan utamanya bukan semata-mata keberadaan soal mudah, melainkan ketimpangan distribusi kesukaran yang menyebabkan instrumen kurang mampu menggambarkan perbedaan tingkat penguasaan peserta secara lebih terperinci.

Keterbatasan variasi kesukaran berkaitan dengan temuan daya pembeda, ketika delapan butir berada dalam kategori jelek, satu butir berkategori cukup, dan satu butir berkategori sangat baik. Q5 menunjukkan daya pembeda sebesar 1,00, sedangkan sebagian besar butir lain memiliki indeks antara 0,00 dan 0,18. Perbedaan tersebut menunjukkan bahwa setiap butir memberikan kontribusi yang tidak sama terhadap kemampuan instrumen membedakan peserta berdasarkan performa tes. Kajian Fitryansyah dan Hilmiyati (2024) serta Wardana et al. (2026) dapat digunakan untuk menjelaskan mengapa tingkat kesukaran perlu ditinjau bersama daya pembeda, bukan sebagai dua indikator yang berdiri sendiri. Sementara itu, Jumini et al. (2023) membahas analisis kualitas soal melalui CTT dan model *Rasch*, sehingga memberikan konteks perbandingan mengenai penggunaan pendekatan pengukuran dalam evaluasi butir. Dalam penerapannya, butir dengan daya pembeda rendah perlu ditelaah lebih lanjut untuk mengetahui apakah masalah berasal dari pertanyaan yang terlalu sederhana, opsi yang kurang berfungsi, ketidaktepatan kunci, atau ketidaksesuaian isi soal dengan kompetensi yang dituju.

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Analisis distraktor pada Q5 memperlihatkan bahwa opsi D dipilih oleh 35,0% peserta, sedangkan opsi A dan B masing-masing hanya dipilih oleh 5,0% dan 2,5% peserta. Distribusi tersebut menunjukkan bahwa daya tarik pengecoh berbeda antarpilihan, sehingga penelaahan terhadap fungsi setiap opsi diperlukan dalam proses perbaikan instrumen. Hendrita (2022) secara khusus membahas *item facility*, daya pembeda, dan efisiensi distraktor pada tes pilihan ganda, sedangkan Mustafa dan Hamid (2025) mengkaji analisis psikometrik soal pilihan ganda dalam konteks ujian bidang kedokteran gigi. Kedua penelitian tersebut relevan untuk menegaskan bahwa evaluasi soal pilihan ganda perlu mempertimbangkan pola respons terhadap setiap opsi, bukan hanya persentase jawaban benar. Namun, angka indeks IP pada tabel penelitian ini, terutama nilai 246,9% untuk opsi D, perlu diverifikasi berdasarkan definisi dan rumus yang digunakan sebelum ditafsirkan secara substantif. Oleh sebab itu, kesimpulan tentang efektivitas distraktor sebaiknya didasarkan pada frekuensi pilihan yang telah terverifikasi serta kriteria operasional yang dinyatakan secara jelas.

Temuan Q5 yang memiliki tingkat kesukaran sedang dan daya pembeda sangat baik memberikan petunjuk bahwa butir yang menuntut pemahaman lebih terarah dapat menghasilkan pola respons yang lebih bervariasi dibandingkan butir yang hampir seluruhnya dijawab benar. Akan tetapi, hasil tersebut tidak cukup untuk menyatakan bahwa seluruh soal dengan karakteristik serupa pasti memiliki kualitas yang lebih baik, karena setiap butir tetap perlu dinilai berdasarkan isi, kesesuaian kunci, dan keterwakilan kompetensi. Ruron dan Ramadhan (2025), melalui kajian pengembangan instrumen penilaian alternatif untuk mengukur berpikir kritis, dapat menjadi pembanding mengenai pentingnya kesesuaian antara rancangan instrumen dan kemampuan yang ditargetkan. Wulansari dan Sujatmiko (2024) juga membahas pengembangan media pembelajaran berbasis *web* untuk mengukur kompetensi pemrograman, sehingga relevan secara kontekstual dalam melihat pengukuran kompetensi pada lingkungan pembelajaran terapan. Meskipun kedua penelitian tersebut menggunakan konteks dan tujuan yang berbeda, keduanya dapat mendukung argumen bahwa instrumen perlu dirancang sesuai karakteristik kompetensi yang hendak diukur. Dalam pelatihan *Effective Recruitment*, pengembangan butir selanjutnya dapat diarahkan pada penerapan konsep melalui skenario rekrutmen, pemilihan tindakan wawancara, dan penilaian situasi kerja, sepanjang isi tersebut sesuai dengan tujuan pembelajaran yang ditetapkan.

Secara keseluruhan, hasil penelitian mengindikasikan bahwa instrumen *pre-test* pelatihan *Effective Recruitment* masih membutuhkan perbaikan sebelum digunakan sebagai satu-satunya dasar untuk menilai penguasaan pengetahuan peserta. Dominasi butir mudah, rendahnya daya pembeda pada sebagian besar soal, validitas empiris yang belum merata, dan koefisien reliabilitas yang rendah menunjukkan perlunya revisi terpadu terhadap konstruksi instrumen. Perbaikan dapat dimulai dengan memeriksa kembali kesesuaian butir dengan tujuan pelatihan, menyempurnakan opsi pengecoh, menambahkan variasi tingkat kesukaran, serta menguji ulang instrumen pada kelompok peserta berikutnya. Analisis lanjutan juga perlu memastikan ketepatan perhitungan koefisien validitas, reliabilitas, daya pembeda, dan indeks distraktor agar keputusan revisi didukung oleh data yang konsisten. Karena penelitian ini terbatas pada 40 peserta, sepuluh butir soal, dan satu konteks pelatihan perusahaan, hasilnya belum dapat digeneralisasikan secara langsung ke semua program pelatihan. Penelitian berikutnya dapat melibatkan sampel yang lebih beragam, menguji instrumen setelah revisi, serta membandingkan hasil CTT dengan pendekatan pengukuran lain untuk memperoleh bukti kualitas instrumen yang lebih komprehensif.

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

KESIMPULAN

Evaluasi terhadap instrumen asesmen kognitif pada pelatihan *Effective Recruitment* menunjukkan bahwa kualitas pengukuran masih memerlukan penyempurnaan pada beberapa aspek psikometrik. Dari sepuluh butir yang dianalisis, hanya separuh memenuhi kriteria validitas empiris, sedangkan koefisien reliabilitas KR-20 sebesar 0,324 mengindikasikan konsistensi internal yang rendah berdasarkan kriteria penelitian. Dominasi butir berkategori mudah dan rendahnya daya pembeda pada sebagian besar soal juga memperlihatkan bahwa instrumen belum optimal dalam menggambarkan variasi penguasaan pengetahuan peserta. Di antara seluruh butir, Q5 menunjukkan karakteristik yang relatif lebih baik karena memiliki tingkat kesukaran sedang dan daya pembeda sangat tinggi, sehingga dapat dijadikan bahan telaah dalam pengembangan butir berikutnya, bukan langsung dianggap sebagai model yang pasti sesuai untuk semua kompetensi. Secara keseluruhan, temuan ini menegaskan bahwa evaluasi asesmen pelatihan perlu mempertimbangkan validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan fungsi distraktor secara terpadu agar hasil pengukuran lebih informatif.

Sebagai tindak lanjut, perbaikan instrumen perlu diarahkan pada peninjauan kesesuaian soal dengan tujuan pembelajaran, penyusunan opsi jawaban yang berfungsi, serta pengembangan butir dengan tingkat kesukaran yang lebih proporsional terhadap kompetensi peserta. Perhitungan yang tidak terdefinisi secara normal pada Q1 dan nilai indeks distraktor yang perlu diverifikasi juga harus diperiksa sebelum instrumen digunakan kembali untuk mengambil keputusan mengenai hasil asesmen. Interpretasi temuan penelitian ini perlu mempertimbangkan cakupan data yang terbatas pada 40 peserta, sepuluh butir soal, dan satu konteks pelatihan perusahaan, sehingga hasilnya tidak dapat langsung digeneralisasikan ke program pelatihan lain. Penelitian selanjutnya disarankan menguji kembali instrumen setelah revisi, melibatkan kelompok peserta yang lebih beragam, serta membandingkan hasil analisis *Classical Test Theory* (CTT) dengan pendekatan pengukuran lain apabila sesuai dengan tujuan penelitian. Langkah tersebut diharapkan menghasilkan instrumen yang lebih konsisten, mampu membedakan tingkat penguasaan peserta secara lebih memadai, dan mendukung evaluasi pelatihan berbasis bukti.

DAFTAR PUSTAKA

Andani, S., & Rizki, L. M. (2023). Validitas instrumen penelitian: Pengaruh pendekatan Ethno-RME terhadap kemampuan penalaran matematis peserta didik kelas VII di SMPN 2 Bangkinang Kota. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan, 2*(2), 7–11. https://jerkin.org/index.php/jerkin/article/view/190

Astuti, Y. W., & Hanifa, F. N. (2026). Edukasi personal hygiene menstruasi melalui program Body Myth Buster berbasis leaflet pada remaja SMP: Personal hygiene education during menstruation through a leaflet-based Body Myth Buster program for junior high school adolescents. *Jurnal Pengabdian Masyarakat Bunda Delima, 5*(2), 189–201. https://jpenmas.bundadelima.ac.id/index.php/jpmbd/article/view/164

Fitryansyah, M. A., & Hilmiyati, F. (2024). Peran tingkat kesukaran dan daya pembeda dalam analisis butir tes: Kajian literatur untuk pendidikan menengah. *Jurnal Riset Evaluasi Pendidikan, 1*(4), 252–262. https://www.etdci.org/journal/JREP/article/view/2250

Hasanah, I., & Sahlan, M. (2026). Evaluasi validitas dan reliabilitas asesmen berbasis proyek dalam pembelajaran hybrid. *Tut Wuri Handayani: Jurnal Keguruan dan Ilmu Pendidikan, 5*(2), 273–283. https://jurnal.risetilmiah.ac.id/index.php/jkip/article/view/1806

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

EDUCATIONAL : Jurnal Inovasi Pendidikan & Pengajaran
Vol. 6, No. 4, September-November 2026
e-ISSN : 2775-2593 | p-ISSN : 2775-2585
Online Journal System : https://jurnalp4i.com/index.php/educational

Hendrita, S. A. R. (2022). Analysing a multiple-choice test to find out the item facility (IF), the item discrimination (ID) and the distractor efficiency. *Inovasi Pendidikan, 9*(2). https://jurnal.umsb.ac.id/index.php/inovasipendidikan/article/view/3906

Jumini, S., Madnasri, S., Cahyono, E., & Parmin, P. (2023, June). Analisis kualitas butir soal pengukuran literasi sains melalui teori tes klasik dan Rasch model. In *Prosiding Seminar Nasional Pascasarjana*, *6*(1), 758–765. https://proceedings.unnes.ac.id/snpasca/article/view/2215

Kambara, C. (2026). Psychometric properties of a workplace experiential learning scale among hospital-based registered dietitians in Japan. *Journal of Human Nutrition and Dietetics, 39*(2), e70251. https://onlinelibrary.wiley.com/doi/abs/10.1111/jhn.70251

Mustafa, S., & Hamid, O. E. (2025). Psychometric item/question analysis of multiple-choice questions in fixed prosthodontics exam. *BMC Medical Education, 26*(1), 86. https://link.springer.com/article/10.1186/s12909-025-08429-4

Ni'mah, A., Gofur, A., & Effendie, R. (2025). Implementasi evaluasi pelatihan di Badan Pengembangan Sumber Daya Manusia Provinsi Kalimantan Tengah. *Concept: Journal of Social Humanities and Education, 4*(3), 129–142. https://journal-stiayappimakassar.ac.id/index.php/Concept/article/view/2216

Nurcahya, E., Sukidin, & Hudha, C. (2026). Analisis butir soal asesmen sumatif akhir tahun (ASAT) mata pelajaran ekonomi kelas X di SMAN Kalisat. *Jurnal Evaluasi Pendidikan, 17*(1), 89–102. https://journal.unj.ac.id/unj/index.php/jep/article/view/66119

Nurjanah, S., Iqbal, M., Zafrullah, Z., Mahmud, M. N., Seran, D. S. F., Suardi, I. K., & Arriza, L. (2024). Psychometric quality of multiple-choice tests under classical test theory (CTT): AnBuso, Iteman, and R. *Jurnal Penelitian dan Evaluasi Pendidikan, 28*(2). https://doi.org/10.21831/pep.v28i2.71542

O'Neill, A., & Cleary, Y. (2023). Measuring transfer of training using five-component objectives. *IEEE International Professional Communication Conference, 2023*, 1–8. https://doi.org/10.1109/ProComm57838.2023.00029

Pratiwi, E., & Sugiharto, B. (2025). Analisis validitas butir dan ketepatan target ukur instrumen biologi kelas XII: Aplikasi pemodelan Rasch. *Jurnal Pendidikan Biologi, 16*(2), 155–167. https://journal-fmipa.um.ac.id/index.php/jpb/article/view/53

Ramadani, E. N., & Handayani, D. F. (2024). Instrumen penilaian hasil pembelajaran kognitif pada tes objektif. *Jurnal Pendidikan Dan Ilmu Sosial (Jupendis)*, *2*(4), 86-96. https://doi.org/10.54066/jupendis.v2i4.2159

Rasyid, M., Manalu, D. N., & Hidayat, S. (2025). Validasi instrumen evaluasi kinerja di PT DUNKIN DONUT dengan menggunakan metode Balance ScoreCard. *Jurnal Manajemen dan Ekonomi Kreatif, 3*(1), 275–285. https://ukitoraja.id/index.php/jumek/article/view/538

Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., ... Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education, 24*(1), 445. https://link.springer.com/article/10.1186/s12909-024-05433-y

Rivaldi, H. R., Wahyudi, A., & Sari, R. A. N. I. (2025). Analisis kualitas butir soal asesmen sumatif geografi pada materi peta, penginderaan jauh, dan sistem informasi geografis.

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215

Jurnal Ilmu Pendidikan dan Pembelajaran, 4(1), 102–113. https://journal.edupartnerpublishing.co.id/index.php/JIPP/article/view/192

Rozi, A. N., Sukrawan, Y., & Noor, R. A. M. (2026). Pengembangan instrumen asesmen kognitif sistem air conditioning kendaraan ringan. *Jurnal Pendidikan Vokasi Otomotif, 8*(2), 89–100. https://jurnal.uny.ac.id/index.php/jpvo/article/view/97901

Ruron, A. T. T., & Ramadhan, F. (2025). Pengembangan instrumen penilaian alternatif berpikir kritis dalam pembelajaran Penjas untuk siswa SMP. *Jurnal Edukasi Citra Olahraga, 5*(1), 102–112. https://jurnal.citrabakti.ac.id/index.php/jor/article/view/5189

Sabat, A. O., Gusman, F., & Nenotaek, B. (2026). Analisis kualitas butir soal pilihan ganda materi statistika SMA berdasarkan pendekatan teori tes klasik. *Mandalika Mathematics and Educations Journal, 8*(3), 1904–1913. https://jurnalfkip.unram.ac.id/index.php/MANDALIKA/article/view/12863

Wardana, M. A., Hasibuan, M., Asyifa, N., Amini, A., & Layla, Y. (2026). Kualitas butir soal sebagai fondasi evaluasi pembelajaran: Sebuah kajian teoretis tentang pengecoh, kesukaran, dan daya pembeda. *Sasangga: Journal of Education and Learning, 4*(1), 1–10. https://sasangga.tfk.or.id/index.php/sasangga/article/view/55

Wulansari, A. D., & Sujatmiko, B. (2024). Pengembangan media pembelajaran berbasis web menggunakan PjBL untuk mengukur kompetensi pemrograman pada mata pelajaran pemrograman berbasis objek (studi kasus: SMKN 2 Surabaya). *IT-Edu: Jurnal Information Technology and Education, 9*(2), 9–17. https://ejournal.unesa.ac.id/index.php/it-edu/article/view/61338

Copyright (c) 2026 EDUCATIONAL: Jurnal Inovasi Pendidikan & Pengajaran
https://doi.org/10.51878/educational.v6i4.16215