

Analisis Soal Sumatif Mata Pelajaran Pendidikan Jasmani Olahraga dan Kesehatan Kelas VIII Di Sekolah Menengah Pertama Negeri 25 Kota Padang

¹Rahmi Eka Putri, ²Khairuddin, ³Sefri Hardiansyah, ⁴Despita Antoni

¹program Studi Pendidikan Jasmani Kesehatan Dan Rekreasi Departemen Pendidikan

Olahraga Fakultas Ilmu Keolahragaan Universitas Negeri Padang

*Corresponding Author e-mail: khairuddin@fik.unp.ac.id

Received: July 2026; Revised: July 2026; Published: August 2026

Abstrak

Penelitian ini dilatarbelakangi oleh belum diketahuinya kualitas soal sumatif mata pelajaran Pendidikan Jasmani, Olahraga, dan Kesehatan (PJOK) kelas VIII di SMP Negeri 25 Kota Padang berdasarkan karakteristik butir soal. Penelitian ini bertujuan menganalisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal sumatif PJOK. Penelitian menggunakan pendekatan deskriptif kuantitatif dengan sampel 53 peserta didik yang dipilih melalui *proportional stratified random sampling* dari populasi 272 peserta didik kelas VIII. Data dikumpulkan melalui dokumentasi berupa perangkat soal, kisi-kisi, dan respons jawaban peserta didik. Analisis dilakukan terhadap 20 butir soal pilihan ganda menggunakan korelasi *point biserial* untuk validitas, *Cronbach's Alpha* untuk reliabilitas, indeks tingkat kesukaran, indeks daya pembeda, dan efektivitas pengecoh. Hasil penelitian menunjukkan bahwa seluruh 20 butir soal (100%) dinyatakan valid dan reliabilitas tes berada pada kategori tinggi dengan *Cronbach's Alpha* sebesar 0,784. Tingkat kesukaran didominasi kategori mudah sebanyak 17 butir (85%) dan sedang sebanyak 3 butir (15%), tanpa butir berkategori sukar. Daya pembeda terdiri atas 1 butir kategori bagus sekali, 3 butir lumayan bagus, 8 butir belum memuaskan, dan 8 butir jelek. Dari 60 pengecoh, 50 pengecoh (83,33%) efektif dan 10 pengecoh (16,67%) tidak efektif. Disimpulkan bahwa soal sumatif memiliki validitas dan reliabilitas yang baik, tetapi kualitas keseluruhan belum optimal karena tingkat kesukaran kurang proporsional dan sebagian besar butir memiliki daya pembeda rendah, sehingga diperlukan revisi dan analisis butir secara berkala.

Kata kunci: Analisis Butir Soal; Asesmen Sumatif; Pjok; Kualitas Soal; Evaluasi Pembelajaran

Analysis of Summative Assessment Questions for Physical Education, Sports, and Health (Grade VIII) at SMP Negeri 25 Padang City

Abstract

This study was motivated by the lack of empirical information regarding the quality of summative assessment items in Physical Education, Sports, and Health (PJOK) for eighth-grade students at SMP Negeri 25 Kota Padang. The study aimed to analyze item validity, test reliability, difficulty level, discrimination power, and distractor effectiveness. A quantitative descriptive approach was employed with a sample of 53 students selected through proportional stratified random sampling from a population of 272 eighth-grade students. Data were collected through documentation, including the test items, test blueprint, and students' response sheets. Twenty multiple-choice items were analyzed using point-biserial correlation for validity, Cronbach's Alpha for reliability, the difficulty index, the discrimination index, and distractor effectiveness analysis. The results showed that all 20 items (100%) were valid, while the test reliability was high with a Cronbach's Alpha coefficient of 0.784. Regarding difficulty, 17 items (85%) were categorized as easy and 3 items (15%) as moderate, with no difficult items. Discrimination analysis identified 1 excellent item, 3 reasonably good items, 8 unsatisfactory items, and 8 poor items. Of the 60 distractors analyzed, 50 (83.33%) were effective and 10 (16.67%) were ineffective. It can be concluded that the test demonstrated good validity and reliability; however, its overall quality was not yet optimal because the difficulty distribution was unbalanced and most items had low discrimination power. Therefore, periodic item analysis and revision are required to improve the quality of future summative assessments.

Keywords: item analysis; summative assessment; physical education; test quality; learning evaluation

How to Cite: Putri, R. E., Khairuddin, K., Hardiansyah, S. ., & Antoni, D. . (2026). Analisis Soal Sumatif Mata Pelajaran Pendidikan Jasmani Olahraga dan Kesehatan Kelas VIII Di Sekolah Menengah Pertama Negeri 25 Kota Padang. *Journal of Authentic Research*, 5(3), 6385-6396. <https://doi.org/10.36312/jar.v5i3.6620>



<https://doi.org/10.36312/jar.v5i3.6620>

Copyright© 2026, Putri

This is an open-access article under the CC-BY-SA License.



PENDAHULUAN

Evaluasi merupakan bagian integral dari proses pembelajaran karena memberikan informasi mengenai ketercapaian tujuan, tingkat penguasaan kompetensi, dan dasar pengambilan keputusan untuk tindak lanjut pembelajaran. Dalam implementasi Kurikulum Merdeka, asesmen sumatif digunakan untuk menilai pencapaian kompetensi peserta didik setelah menyelesaikan suatu lingkup materi atau pada akhir periode pembelajaran, sehingga kualitas instrumen yang digunakan menentukan ketepatan informasi hasil belajar. Hal ini juga berlaku pada mata pelajaran Pendidikan Jasmani, Olahraga, dan Kesehatan (PJOK) yang tidak hanya diarahkan pada kebugaran dan keterampilan gerak, tetapi juga pada pengembangan aspek pengetahuan, sikap, dan karakter peserta didik. Oleh karena itu, soal sumatif PJOK perlu disusun sebagai instrumen yang mampu mengukur kemampuan peserta didik secara objektif, akurat, dan konsisten. Salah satu bentuk instrumen yang banyak digunakan ialah soal pilihan ganda karena memungkinkan cakupan materi yang luas dan penskoran yang relatif objektif. Meskipun demikian, kualitas soal pilihan ganda tidak cukup dinilai dari kesesuaian isi semata, tetapi perlu dibuktikan melalui karakteristik empiris setiap butir. Analisis butir memungkinkan peneliti atau guru menilai validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh, kemudian menggunakan hasilnya sebagai dasar untuk mempertahankan, memperbaiki, atau membuang butir soal. Dengan demikian, analisis butir merupakan langkah penting dalam membangun perangkat asesmen dan bank soal yang lebih berkualitas serta mendukung keputusan pembelajaran yang dapat dipertanggungjawabkan (DeCarlo, 2023; Rashwan et al., 2024).

Permasalahan utama dalam praktik asesmen adalah masih digunakannya butir soal tanpa pengujian empiris yang memadai, sehingga instrumen berpotensi memiliki tingkat kesukaran yang tidak proporsional, daya pembeda rendah, atau pengecoh yang tidak berfungsi. Kondisi tersebut juga ditemukan pada konteks penelitian ini. Berdasarkan wawancara dan pengamatan awal di SMP Negeri 25 Kota Padang, soal objektif PJOK yang digunakan belum dianalisis secara menyeluruh untuk memperoleh informasi mengenai validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecohnya. Padahal, hasil asesmen sumatif digunakan sebagai dasar untuk menggambarkan pencapaian kompetensi peserta didik dan menentukan tindak lanjut pembelajaran. Pengalaman di berbagai konteks menunjukkan bahwa analisis pascaujian dapat memberikan dasar yang kuat untuk pengendalian mutu instrumen. Rashwan et al. (2024) di Mesir menunjukkan bahwa analisis kuantitatif terhadap kesukaran, daya pembeda, efektivitas pengecoh, dan reliabilitas dapat digunakan untuk menentukan butir yang layak dipertahankan, direvisi, atau dikeluarkan dari bank soal. Rezigalla et al. (2024) di Arab Saudi juga menemukan bahwa efektivitas pengecoh berkaitan dengan indeks kesukaran dan daya pembeda, sehingga perbaikan pengecoh berpotensi meningkatkan kualitas butir. Dalam konteks Indonesia, Logayah et al. (2024) menegaskan bahwa analisis kesukaran, daya pembeda, dan efektivitas pengecoh penting untuk memperbaiki soal yang akan digunakan kembali. Temuan-temuan tersebut memperlihatkan bahwa

analisis butir bukan sekadar prosedur statistik, tetapi merupakan mekanisme penjaminan mutu asesmen yang relevan lintas bidang dan jenjang pendidikan.

Meskipun analisis butir telah banyak diterapkan, masih terdapat kesenjangan yang penting untuk dikaji pada konteks asesmen PJOK di sekolah menengah pertama. Penelitian mutakhir umumnya dilakukan pada tes di bidang pendidikan kedokteran, bahasa, atau kompetisi akademik, sedangkan bukti empiris mengenai karakteristik soal sumatif PJOK pada jenjang SMP, khususnya dalam implementasi Kurikulum Merdeka, masih terbatas. Selain itu, sebagian penelitian menitikberatkan pada beberapa parameter, seperti tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, tanpa mengintegrasikannya dengan validitas butir dan reliabilitas tes dalam satu evaluasi yang utuh. Padahal, keputusan mengenai kualitas sebuah instrumen tidak seharusnya ditentukan hanya berdasarkan satu karakteristik. Konteks SMP Negeri 25 Kota Padang juga memberikan alasan empiris yang kuat karena sekolah tersebut belum melakukan analisis butir secara sistematis terhadap soal PJOK sejak penerapan Kurikulum Merdeka, sehingga kualitas soal yang digunakan dalam asesmen sumatif belum diketahui secara terukur. Berdasarkan kesenjangan tersebut, kebaruan penelitian ini terletak pada analisis komprehensif terhadap lima karakteristik utama kualitas soal, yaitu validitas empiris, reliabilitas perangkat tes, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh pada soal sumatif PJOK kelas VIII. Analisis secara simultan terhadap kelima aspek tersebut diharapkan memberikan profil kualitas soal yang lebih lengkap dibandingkan evaluasi parsial, sekaligus menghasilkan dasar empiris untuk menentukan butir yang dapat dipertahankan, perlu direvisi, atau sebaiknya tidak digunakan kembali dalam penyusunan bank soal PJOK.

Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis kualitas soal sumatif mata pelajaran Pendidikan Jasmani, Olahraga, dan Kesehatan kelas VIII di SMP Negeri 25 Kota Padang secara empiris dan komprehensif. Ruang lingkup penelitian dibatasi pada soal pilihan ganda yang digunakan dalam asesmen sumatif dan dianalisis melalui lima indikator utama kualitas instrumen. Pertama, validitas butir digunakan untuk menilai ketepatan setiap butir dalam mengukur kemampuan yang menjadi sasaran melalui korelasi *point biserial*. Kedua, reliabilitas digunakan untuk mengetahui konsistensi internal perangkat tes sebagai satu kesatuan. Ketiga, tingkat kesukaran dianalisis berdasarkan proporsi peserta didik yang menjawab benar untuk mengelompokkan soal ke dalam kategori mudah, sedang, atau sukar. Keempat, daya pembeda digunakan untuk mengetahui kemampuan butir dalam membedakan peserta didik kelompok berkemampuan tinggi dan rendah. Kelima, efektivitas pengecoh digunakan untuk mengetahui sejauh mana alternatif jawaban salah dapat berfungsi secara logis dalam menarik peserta didik yang belum menguasai materi. Penelitian menggunakan data respons peserta didik terhadap 20 butir soal sumatif dengan sampel 53 peserta didik kelas VIII dari populasi 272 peserta didik. Melalui analisis kelima indikator tersebut, penelitian diharapkan menghasilkan gambaran objektif mengenai kekuatan dan kelemahan perangkat soal PJOK serta memberikan dasar bagi guru untuk melakukan perbaikan butir, meningkatkan

keseimbangan karakteristik soal, dan mengembangkan bank soal yang lebih valid, reliabel, serta mampu mendukung evaluasi pembelajaran secara berkelanjutan.

METODE

Desain Penelitian

Penelitian ini menggunakan pendekatan deskriptif kuantitatif yang bertujuan memberikan gambaran empiris mengenai kualitas soal sumatif mata pelajaran Pendidikan Jasmani, Olahraga, dan Kesehatan (PJOK) kelas VIII di SMP Negeri 25 Kota Padang. Pendekatan deskriptif kuantitatif digunakan karena penelitian tidak memberikan perlakuan terhadap peserta didik, tetapi menganalisis karakteristik perangkat tes berdasarkan respons peserta didik terhadap soal yang telah digunakan dalam asesmen. Analisis diarahkan pada lima karakteristik utama kualitas butir soal, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Objek yang dianalisis berupa 20 butir soal pilihan ganda pada asesmen sumatif mata pelajaran PJOK. Penelitian dilaksanakan di UPT SMP Negeri 25 Kota Padang dengan menggunakan dokumen soal sumatif semester ganjil Tahun Pelajaran 2025/2026 sebagai sumber utama data penelitian. Melalui desain tersebut, setiap butir dianalisis secara kuantitatif untuk memperoleh informasi mengenai tingkat kelayakannya sebagai bagian dari perangkat evaluasi pembelajaran. Hasil analisis selanjutnya digunakan untuk memberikan gambaran mengenai butir yang dapat dipertahankan, butir yang membutuhkan perbaikan, serta karakteristik perangkat soal secara keseluruhan. Pendekatan ini sesuai dengan tujuan analisis butir, yaitu memperoleh informasi empiris mengenai kualitas soal sehingga instrumen evaluasi yang digunakan guru dapat memberikan informasi hasil belajar secara lebih tepat (Sanusi & Aziez, 2021).

Populasi dan Sampel

Populasi penelitian terdiri atas seluruh peserta didik kelas VIII SMP Negeri 25 Kota Padang sebanyak 272 siswa yang tersebar pada delapan kelas, yaitu VIII.1 sampai dengan VIII.8. Jumlah peserta didik pada masing-masing kelas berkisar antara 33 dan 34 siswa. Sampel penelitian ditentukan menggunakan teknik *proportional stratified random sampling*. Teknik tersebut digunakan agar setiap kelas sebagai strata memperoleh keterwakilan dalam sampel sesuai dengan proporsi jumlah peserta didiknya. Penentuan sampel dilakukan dengan mengambil sekitar 20% peserta didik laki-laki dan perempuan pada masing-masing kelas. Berdasarkan prosedur tersebut diperoleh sampel sebanyak 53 peserta didik. Distribusi sampel meliputi 7 peserta didik dari kelas VIII.1, masing-masing 6 peserta didik dari kelas VIII.2, VIII.3, dan VIII.4, serta masing-masing 7 peserta didik dari kelas VIII.5, VIII.6, VIII.7, dan VIII.8. Pemilihan sampel secara proporsional dimaksudkan agar data respons yang digunakan dalam analisis tidak hanya bersumber dari satu kelas, tetapi menggambarkan peserta didik kelas VIII dari seluruh kelas yang terdapat di sekolah penelitian. Seluruh respons dari 53 peserta didik tersebut selanjutnya digunakan sebagai dasar perhitungan validitas empiris, reliabilitas, dan tingkat kesukaran, sedangkan analisis daya pembeda dilakukan dengan membandingkan peserta didik kelompok tinggi dan kelompok rendah berdasarkan skor hasil tes.

Instrumen dan Prosedur Pengumpulan Data

Data penelitian dikumpulkan dengan teknik dokumentasi. Dokumen yang menjadi sumber data terdiri atas perangkat soal sumatif PJOK kelas VIII, respons atau lembar jawaban peserta didik, kisi-kisi soal, serta dokumen indikator pembelajaran yang digunakan dalam penyusunan soal. Perangkat tes yang dianalisis terdiri atas 20 butir soal berbentuk pilihan ganda dengan empat alternatif jawaban pada setiap butir. Dengan demikian, selain satu jawaban benar, setiap butir memiliki tiga alternatif jawaban yang berfungsi sebagai pengecoh. Dalam penelitian ini tidak digunakan instrumen angket atau tes tambahan karena perangkat soal sumatif itu sendiri merupakan objek utama yang dievaluasi kualitas empirisnya. Prosedur penelitian dimulai dengan pengumpulan perangkat soal dan respons jawaban peserta didik. Selanjutnya, data jawaban disusun untuk memungkinkan analisis setiap butir berdasarkan lima aspek yang telah ditentukan. Validitas butir dianalisis berdasarkan hubungan antara skor butir dan skor total, reliabilitas digunakan untuk menilai konsistensi perangkat soal secara keseluruhan, tingkat kesukaran ditentukan dari proporsi peserta didik yang menjawab benar, daya pembeda diperoleh melalui perbandingan kelompok berkemampuan tinggi dan rendah, sedangkan efektivitas pengecoh dianalisis berdasarkan distribusi peserta didik yang memilih alternatif jawaban salah. Rangkaian prosedur tersebut digunakan untuk memperoleh gambaran yang utuh mengenai kualitas perangkat soal sumatif PJOK.

Analisis Data

Data dianalisis secara kuantitatif berdasarkan lima indikator kualitas soal. Pertama, validitas empiris setiap butir dihitung menggunakan korelasi *point biserial*. Nilai koefisien yang diperoleh dibandingkan dengan nilai (r_{tabel}) pada taraf signifikansi 5%. Dengan jumlah sampel 53 peserta didik, nilai (r_{tabel}) yang digunakan adalah 0,270; butir dinyatakan valid apabila ($r_{\text{hitung}} > r_{\text{tabel}}$) (Tussaadah, 2021). Kedua, reliabilitas perangkat tes ditentukan menggunakan koefisien *Cronbach's Alpha* dan diinterpretasikan ke dalam kategori amat rendah, rendah, cukup, tinggi, dan amat tinggi. Rentang 0,70–0,89 dikategorikan memiliki reliabilitas tinggi (Subhaktiyasa, 2024). Ketiga, tingkat kesukaran dihitung berdasarkan proporsi peserta didik yang menjawab benar. Indeks 0,00–0,30 dikategorikan sukar, 0,31–0,70 sedang, dan 0,71–1,00 mudah (Handayani, 2025). Keempat, analisis daya pembeda dilakukan menggunakan 14 peserta didik kelompok tinggi dan 14 peserta didik kelompok rendah. Indeks $>0,40$ dikategorikan bagus sekali; 0,30–0,39 lumayan bagus; 0,20–0,29 belum memuaskan; dan $<0,20$ jelek (Fitriani, 2021). Kelima, efektivitas pengecoh ditentukan berdasarkan proporsi peserta yang memilih setiap alternatif jawaban salah. Pengecoh dikategorikan efektif apabila dipilih sekurang-kurangnya 5% peserta tes (Sanusi & Aziez, 2021). Dengan 53 peserta didik, batas tersebut setara dengan minimal tiga peserta. Karena 20 soal memiliki tiga pengecoh per butir, secara keseluruhan terdapat 60 pengecoh yang dianalisis.

HASIL DAN PEMBAHASAN

Validitas Butir Soal

Hasil analisis validitas menunjukkan bahwa seluruh 20 butir soal sumatif PJOK memenuhi kriteria validitas empiris. Nilai (r_{hitung}) seluruh butir lebih besar dibandingkan (r_{tabel}) sebesar 0,270. Koefisien validitas terendah diperoleh pada butir S17 sebesar 0,271, sedangkan koefisien tertinggi diperoleh pada S18 sebesar 0,376. Dengan demikian, sebanyak 20 butir atau 100% soal dinyatakan valid. Temuan tersebut menunjukkan bahwa respons peserta didik terhadap setiap butir memiliki hubungan yang memadai dengan skor total tes berdasarkan kriteria yang digunakan dalam penelitian. Meskipun seluruh soal telah memenuhi batas validitas, nilai koefisien antarbutir menunjukkan variasi. Beberapa butir mempunyai nilai yang relatif dekat dengan batas minimum, sementara butir lainnya memperlihatkan koefisien yang lebih tinggi. Oleh karena itu, status valid tidak berarti bahwa semua butir memiliki kekuatan pengukuran yang identik. Analisis validitas tetap diperlukan pada penggunaan tes berikutnya karena karakteristik respons peserta didik dapat memberikan informasi mengenai konsistensi fungsi setiap butir. Dalam konteks pengembangan bank soal, hasil tersebut menunjukkan bahwa seluruh butir telah memenuhi kriteria validitas yang ditetapkan dalam penelitian sehingga secara empiris memiliki dasar untuk dipertahankan, dengan tetap mempertimbangkan karakteristik lainnya seperti tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.

Temuan tersebut mendukung pandangan Tussaadah (2021) yang menempatkan validitas sebagai indikator ketepatan suatu tes dalam mengukur kemampuan yang menjadi sasaran pengukuran. Amalia et al. (2021) juga menjelaskan bahwa analisis butir diperlukan untuk menentukan apakah soal yang disusun mempunyai kualitas yang memadai dalam mengukur hasil belajar. Apabila dibandingkan dengan penelitian Putri (2024), hasil penelitian ini menunjukkan karakteristik yang berbeda. Putri (2024) menemukan bahwa masih terdapat beberapa soal penilaian akhir Pendidikan Jasmani dengan validitas dan reliabilitas rendah, sedangkan seluruh butir pada penelitian ini memenuhi kriteria validitas empiris. Perbedaan tersebut terjadi pada konteks jenis tes dan lokasi penelitian yang berbeda sebagaimana diidentifikasi dalam kajian penelitian terdahulu. Temuan ini menegaskan bahwa kualitas empiris suatu soal tidak dapat digeneralisasikan hanya berdasarkan kesamaan mata pelajaran, tetapi perlu diperiksa pada perangkat tes dan kelompok peserta didik tempat soal tersebut digunakan. Dengan demikian, analisis validitas memberikan dasar awal untuk menentukan kelayakan butir, tetapi keputusan akhir mengenai penggunaan kembali soal perlu mempertimbangkan hasil analisis indikator lainnya. Butir yang valid namun mempunyai daya pembeda rendah atau tingkat kesukaran yang terlalu mudah tetap memerlukan penelaahan sebelum dimasukkan secara permanen ke dalam bank soal.

Reliabilitas Perangkat Tes

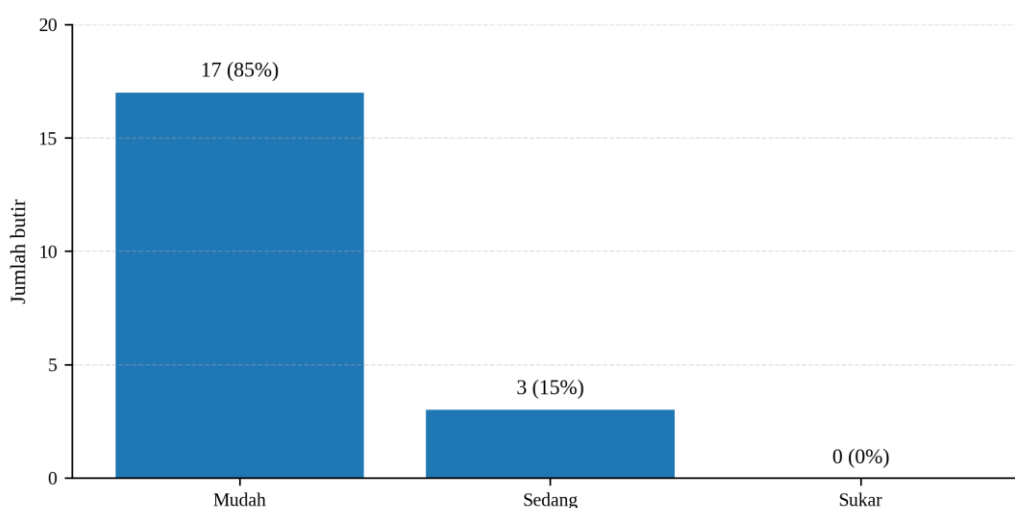
Hasil pengujian reliabilitas menunjukkan bahwa perangkat soal memperoleh nilai *Cronbach's Alpha* sebesar 0,784 untuk 20 butir soal. Berdasarkan kriteria yang digunakan dalam penelitian, nilai tersebut berada pada rentang 0,70–0,89 sehingga termasuk kategori tinggi. Hasil ini menunjukkan bahwa perangkat soal mempunyai konsistensi internal yang memadai sebagai satu kesatuan instrumen. Dengan kata lain, respons peserta didik terhadap keseluruhan butir memberikan pola pengukuran yang cukup konsisten. Reliabilitas yang tinggi menjadi karakteristik penting karena perangkat tes yang tidak konsisten akan menghasilkan informasi yang kurang dapat dipercaya dalam menggambarkan capaian peserta didik. Dalam penelitian ini, hasil reliabilitas memberikan indikasi positif terhadap perangkat soal sumatif PJOK, meskipun kualitas setiap butir tetap harus dipertimbangkan secara terpisah. Hal tersebut terlihat dari hasil analisis selanjutnya yang menunjukkan adanya variasi tingkat kesukaran dan daya pembeda. Oleh sebab itu, nilai reliabilitas perangkat yang tinggi tidak secara otomatis menunjukkan bahwa seluruh karakteristik setiap butir telah optimal. Reliabilitas lebih menggambarkan konsistensi perangkat secara keseluruhan, sedangkan evaluasi kualitas masing-masing butir memerlukan analisis validitas, kesukaran, daya pembeda, dan pengecoh secara bersama-sama.

Hasil tersebut sesuai dengan kriteria Subhaktiyasa (2024), yang menempatkan koefisien pada rentang 0,70–0,89 sebagai reliabilitas tinggi. Tussaadah et al. (2021) juga menjelaskan bahwa reliabilitas berkaitan dengan ketetapan suatu tes dalam menghasilkan pengukuran. Jika dibandingkan dengan penelitian Putri (2024), penelitian ini menunjukkan hasil yang lebih positif karena perangkat soal memiliki reliabilitas tinggi, sedangkan penelitian Putri melaporkan masih adanya karakteristik validitas dan reliabilitas yang rendah pada beberapa soal. Walaupun demikian, reliabilitas harus ditafsirkan bersama indikator lain. Hasil penelitian ini memperlihatkan bahwa tes yang reliabel masih dapat memiliki komposisi tingkat kesukaran yang kurang seimbang dan daya pembeda yang belum optimal pada sejumlah besar butir. Hal tersebut menunjukkan pentingnya analisis butir secara multidimensional dan mendukung Rahmawati (2024) yang menegaskan bahwa kualitas soal sumatif perlu dianalisis secara menyeluruh agar mampu memberikan gambaran kemampuan peserta didik secara optimal. Dengan demikian, reliabilitas tinggi merupakan kekuatan perangkat soal PJOK yang diteliti, tetapi proses perbaikan tetap diperlukan pada karakteristik individual butir agar konsistensi perangkat tersebut disertai dengan kemampuan soal dalam membedakan tingkat penguasaan peserta didik secara lebih baik.

Tingkat Kesukaran

Analisis tingkat kesukaran memperlihatkan bahwa distribusi soal belum proporsional. Dari 20 butir yang dianalisis, sebanyak 17 butir atau 85% berada pada kategori mudah, sedangkan hanya 3 butir atau 15% berada pada kategori sedang dan tidak terdapat soal berkategori sukar. Indeks kesukaran tertinggi terdapat pada S12 sebesar 0,868, sedangkan indeks terendah terdapat pada S11 dan S15 sebesar 0,660. Dominasi soal mudah menunjukkan bahwa sebagian besar peserta didik mampu memberikan jawaban benar terhadap sebagian besar butir yang digunakan. Kondisi

tersebut dapat menunjukkan bahwa materi yang ditanyakan telah dikuasai oleh banyak peserta didik. Namun, ditinjau dari fungsi perangkat evaluasi, komposisi yang terlalu didominasi oleh soal mudah menjadi kelemahan yang perlu diperhatikan. Ketika suatu soal dapat dijawab dengan benar oleh hampir seluruh peserta, ruang untuk mengidentifikasi variasi kemampuan peserta didik menjadi lebih terbatas. Kondisi tersebut tampak memiliki keterkaitan dengan temuan daya pembeda karena sebagian besar butir dalam penelitian juga berada pada kategori belum memuaskan dan jelek. Oleh karena itu, perbaikan perangkat soal tidak diarahkan untuk menghilangkan soal mudah, melainkan menghasilkan distribusi kesukaran yang lebih berimbang sehingga perangkat tes dapat menggambarkan variasi penguasaan peserta didik secara lebih komprehensif. Distribusi kategori tingkat kesukaran tersebut divisualisasikan pada Gambar 1.

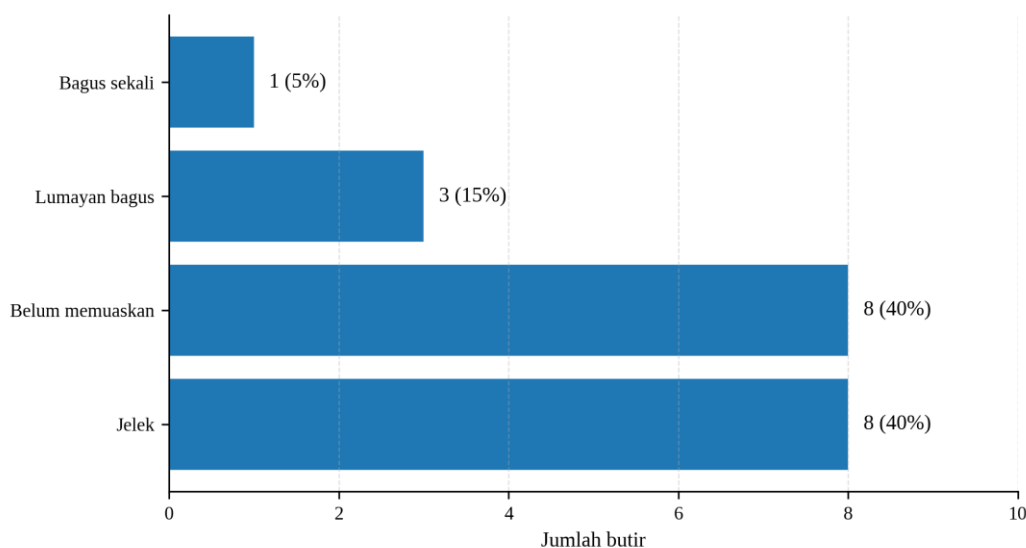


Gambar 1. Distribusi tingkat kesukaran soal sumatif PJOK.

Temuan tersebut sesuai dengan konsep tingkat kesukaran yang dikemukakan Handayani (2025), yaitu indeks kesukaran menunjukkan sejauh mana peserta didik mampu menjawab suatu butir dengan benar dan menjadi dasar pengelompokan soal ke dalam kategori mudah, sedang, dan sukar. Hasil penelitian ini berbeda dengan temuan Putri (2024) pada analisis soal penilaian akhir Pendidikan Jasmani, yang menunjukkan bahwa sebagian besar butir berada pada kategori sedang. Perbedaan tersebut menunjukkan bahwa meskipun kedua penelitian sama-sama mengkaji evaluasi dalam Pendidikan Jasmani, distribusi kesukaran instrumen dapat berbeda bergantung pada perangkat soal dan subjek yang dianalisis. Dari sudut pandang kualitas tes, temuan penelitian ini memperlihatkan pentingnya menyeimbangkan komposisi soal. Soal mudah tetap diperlukan untuk mengetahui penguasaan kompetensi dasar, tetapi perlu dipadukan dengan soal sedang dan beberapa soal yang lebih menantang. Keseimbangan tersebut penting karena perangkat yang terlalu mudah berpotensi menghasilkan skor yang terkonsentrasi pada kelompok nilai tinggi dan mengurangi kemampuan butir untuk membedakan tingkat penguasaan. Dengan demikian, hasil analisis tingkat kesukaran menjadi dasar penting bagi guru PJOK dalam merevisi komposisi perangkat soal sebelum digunakan pada asesmen berikutnya.

Daya Pembeda

Hasil analisis daya pembeda menunjukkan bahwa aspek ini merupakan salah satu kelemahan utama perangkat soal yang diteliti. Dari 20 butir soal, hanya 1 butir, yaitu S20, yang berada pada kategori bagus sekali dengan indeks 0,429. Sebanyak 3 butir, yaitu S10, S11, dan S12, berada pada kategori lumayan bagus dengan indeks masing-masing 0,357. Sebaliknya, 8 butir berada pada kategori belum memuaskan dengan indeks 0,214–0,286 dan 8 butir lainnya berada pada kategori jelek dengan indeks 0,000–0,143. Dengan demikian, 16 dari 20 butir masih berada pada dua kategori terendah. Hasil tersebut menunjukkan bahwa sebagian besar soal belum mampu menghasilkan perbedaan respons yang kuat antara kelompok peserta didik berkemampuan tinggi dan kelompok berkemampuan rendah. Keterbatasan daya pembeda tersebut dapat dikaitkan dengan distribusi tingkat kesukaran yang sangat didominasi soal mudah. Ketika peserta didik dari kelompok tinggi maupun rendah sama-sama dapat menjawab suatu butir dengan benar, selisih proporsi jawaban benar kedua kelompok menjadi kecil sehingga indeks daya pembeda juga rendah. Oleh karena itu, perbaikan daya pembeda perlu menjadi prioritas dalam revisi perangkat soal sumatif ini. Distribusi kategori daya pembeda tersebut disajikan pada Gambar 2.



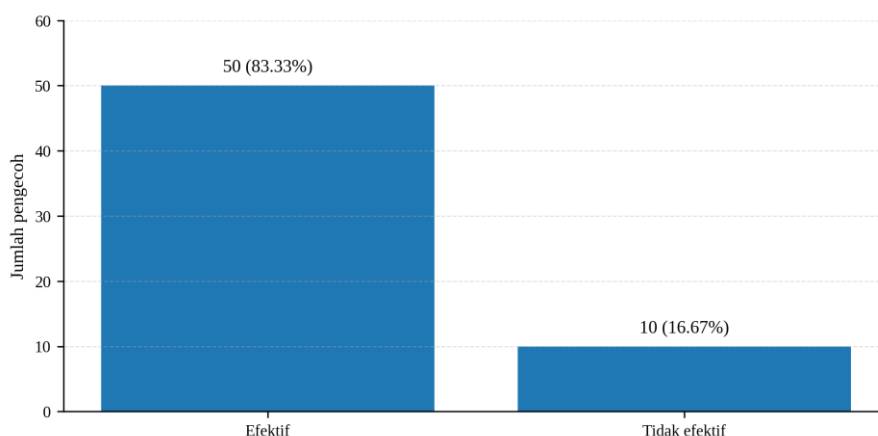
Gambar 2. Distribusi daya pembeda soal sumatif PJOK.

Temuan tersebut sesuai dengan Fitriani (2021), yang menyatakan bahwa daya pembeda menunjukkan kemampuan suatu butir untuk membedakan peserta yang telah menguasai materi dengan peserta yang kurang atau belum menguasainya. Dalam kriteria yang digunakan pada penelitian ini, butir dengan indeks kurang dari 0,20 berada pada kategori jelek dan perlu mendapat perhatian serius, sedangkan butir pada rentang 0,20–0,29 masih belum memuaskan dan perlu diperbaiki. Hasil penelitian ini juga sejalan dengan penelitian Andika (2025), yang menemukan bahwa tidak semua soal pilihan ganda memenuhi kriteria kualitas yang baik karena masih terdapat soal dengan daya pembeda rendah. Kesamaan tersebut memperlihatkan bahwa kemampuan diskriminatif menjadi salah satu persoalan penting dalam pengembangan soal pilihan ganda. Berdasarkan temuan penelitian ini, S20 dapat

dipertahankan karena memiliki daya pembeda paling baik, sedangkan S10, S11, dan S12 pada prinsipnya dapat digunakan dengan penelaahan atau perbaikan apabila diperlukan. Delapan butir kategori belum memuaskan perlu direvisi, sedangkan delapan butir kategori jelek perlu ditinjau secara lebih mendalam sebelum digunakan kembali. Penelaahan dapat diarahkan pada materi, konstruksi pertanyaan, kunci, dan alternatif jawaban sehingga butir mampu membedakan tingkat penguasaan peserta didik secara lebih tepat.

Efektivitas Pengecoh

Hasil analisis efektivitas pengecoh menunjukkan karakteristik yang relatif lebih baik dibandingkan tingkat kesukaran dan daya pembeda. Setiap butir mempunyai empat alternatif jawaban sehingga terdapat tiga pengecoh per butir dan secara keseluruhan sebanyak 60 pengecoh dianalisis. Berdasarkan batas efektivitas minimal 5% dari 53 peserta didik, suatu pengecoh harus dipilih sekurang-kurangnya oleh tiga peserta agar dikategorikan efektif. Hasil analisis menunjukkan bahwa sebanyak 50 pengecoh atau 83,33% telah berfungsi efektif, sedangkan 10 pengecoh atau 16,67% belum efektif. Sebagian besar alternatif jawaban salah dengan demikian mampu menarik pilihan peserta didik yang tidak menentukan jawaban benar. Meskipun hasil secara keseluruhan relatif baik, keberadaan 10 pengecoh yang tidak efektif menunjukkan bahwa kualitas alternatif jawaban pada beberapa butir masih perlu diperbaiki. Beberapa pengecoh hanya dipilih oleh satu atau dua peserta didik, bahkan terdapat pengecoh yang sama sekali tidak dipilih. Alternatif seperti ini memiliki fungsi yang terbatas sebagai pengalih perhatian karena peserta didik relatif mudah mengenalinya sebagai jawaban yang tidak tepat. Oleh sebab itu, penilaian kualitas soal pilihan ganda perlu mencakup tidak hanya kunci jawaban, tetapi juga kemampuan setiap alternatif salah dalam menjalankan fungsi sebagai pengecoh. Perbandingan pengecoh efektif dan tidak efektif divisualisasikan pada Gambar 3.



Gambar 3. Efektivitas pengecoh soal sumatif PJOK.

Hasil tersebut sejalan dengan Sanusi dan Aziez (2021), yang menjelaskan bahwa pengecoh yang baik akan dipilih oleh peserta didik yang menjawab salah secara relatif merata atau mendekati jumlah ideal. Sari dan Kurniawan (2024), sebagaimana digunakan dalam kajian teori penelitian, juga menyatakan bahwa pengecoh yang baik merupakan alternatif yang dipilih oleh sejumlah peserta didik, terutama peserta

dengan kemampuan lebih rendah. Temuan penelitian ini juga dapat dibandingkan dengan Andika (2025), yang menemukan masih adanya pengecoh tidak efektif pada analisis soal pilihan ganda. Kesamaannya terletak pada keberadaan alternatif jawaban yang belum menjalankan fungsi pengecoh secara optimal, meskipun dalam penelitian ini sebagian besar dari 60 pengecoh telah memenuhi kriteria efektivitas. Implikasi temuan tersebut adalah bahwa pengecoh yang telah efektif dapat dipertahankan, sedangkan 10 alternatif yang belum efektif perlu ditelaah dan direvisi sebelum digunakan kembali. Perbaikan diarahkan agar pilihan jawaban salah lebih logis, homogen dengan kunci, dan tetap masuk akal bagi peserta didik yang belum menguasai materi. Perbaikan tersebut diharapkan tidak hanya meningkatkan efektivitas pengecoh, tetapi juga berkontribusi terhadap peningkatan daya pembeda butir secara keseluruhan.

Secara keseluruhan, kelima hasil analisis memperlihatkan bahwa kualitas perangkat soal sumatif PJOK kelas VIII SMP Negeri 25 Kota Padang belum sepenuhnya optimal. Kekuatan perangkat terletak pada seluruh butir yang memenuhi kriteria validitas, reliabilitas perangkat yang tinggi, serta sebagian besar pengecoh yang telah berfungsi efektif. Sebaliknya, kelemahan utama terdapat pada distribusi tingkat kesukaran yang didominasi soal mudah dan daya pembeda yang sebagian besar masih berada pada kategori belum memuaskan dan jelek. Temuan ini mendukung pandangan bahwa kualitas soal tidak dapat ditentukan hanya melalui satu parameter. Butir yang valid belum tentu mempunyai tingkat kesukaran dan daya pembeda yang baik, sementara perangkat yang reliabel masih dapat mengandung butir yang perlu diperbaiki. Oleh karena itu, keputusan untuk mempertahankan, merevisi, atau mengeluarkan suatu butir dari bank soal perlu mempertimbangkan seluruh karakteristik tersebut secara terpadu (Amalia et al., 2021; Rahmawati, 2024; Sanusi & Aziez, 2021). Pola pada Gambar 1-3 memperjelas bahwa kelemahan utama terletak pada dominasi soal mudah dan rendahnya daya pembeda, sementara efektivitas pengecoh relatif baik.

KESIMPULAN

Analisis terhadap soal sumatif PJOK kelas VIII SMP Negeri 25 Kota Padang menunjukkan bahwa perangkat tes telah memiliki kekuatan pada aspek validitas, konsistensi internal, dan fungsi sebagian besar pengecoh, tetapi kualitasnya secara keseluruhan belum optimal. Ketidakseimbangan terutama terlihat pada tingkat kesukaran yang sangat didominasi oleh soal mudah dan kemampuan sebagian besar butir yang masih rendah dalam membedakan peserta didik berkemampuan tinggi dan rendah. Temuan tersebut menunjukkan bahwa kelayakan perangkat evaluasi tidak cukup ditentukan hanya oleh validitas dan reliabilitas, tetapi harus mempertimbangkan karakteristik setiap butir secara terpadu. Oleh karena itu, perangkat soal memerlukan penyempurnaan terutama melalui penataan kembali komposisi tingkat kesukaran, revisi butir dengan daya pembeda rendah, dan perbaikan pengecoh yang belum berfungsi. Analisis butir secara berkala dapat digunakan sebagai dasar untuk mempertahankan soal yang berkualitas, memperbaiki

soal yang masih dapat dikembangkan, serta membangun bank soal PJOK yang lebih baik untuk asesmen berikutnya.

REFERENSI

- Amalia, N., dkk. (2021). Analisis kualitas soal ujian PJOK di sekolah menengah pertama. *Jurnal Pendidikan*, 15(2), 120–130.
- Andika. (2025). Analisis butir soal pilihan ganda menggunakan pendekatan klasik. *Jurnal Evaluasi Pendidikan*, 9(1), 45–55.
- DeCarlo, L. T. (2023). Classical item analysis from a signal detection perspective. *Journal of Educational Measurement*, 60(3), 520–547.
<https://doi.org/10.1111/jedm.12358>
- Fitriani, N. (2021). Analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal pelatihan kewaspadaan kegawatdaruratan maternal dan neonatal. *Paedagogia: Jurnal Kajian, Penelitian dan Pengembangan Kependidikan*, 12(2), 199–205. <https://doi.org/10.31764/paedagogia.v12i2.4956>
- Handayani, R. D., Prayoga, J. D., & Putri, N. M. (2025). Analisis butir soal evaluasi pada topik zat dan perubahannya berbasis program IteMan. *Natural Science Education Research*, 8(2), 149–160. <https://doi.org/10.21107/nser.v8i2.30485>
- Logayah, D. S., Ruhimat, M., Arrasyid, R., & Islami, M. R. F. (2024). Item analysis of National Geography Olympiad multiple-choice questions (MCQs) in Indonesia. *Cogent Social Sciences*, 10(1), 2354971.
<https://doi.org/10.1080/23311886.2024.2354971>
- Putri. (2024). Analisis kualitas soal PJOK berdasarkan taksonomi Bloom. *Jurnal Pendidikan Jasmani*, 6(2), 55–65.
- Rahmawati. (2024). Evaluasi hasil belajar siswa dalam pembelajaran PJOK. *Jurnal Pendidikan*, 12(2), 100–110.
- Rashwan, N. I., Aref, S. R., Nayel, O. A., & Rizk, M. H. (2024). Postexamination item analysis of undergraduate pediatric multiple-choice questions exam: Implications for developing a validated question bank. *BMC Medical Education*, 24, 168. <https://doi.org/10.1186/s12909-024-05153-3>
- Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & 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, 445. <https://doi.org/10.1186/s12909-024-05433-y>
- Sanusi, R. N. A., & Aziez, F. (2021). Analisis butir soal tes objektif dan subjektif untuk keterampilan membaca pemahaman pada kelas VII SMP N 3 Kalibagor. *Metafora: Jurnal Pembelajaran Bahasa dan Sastra*, 8(1), 99–109.
<https://doi.org/10.30595/mtf.v8i1.8501>
- Sari, & Kurniawan. (2024). Analisis daya pembeda soal pilihan ganda. *Jurnal Evaluasi Pendidikan*, 8(2), 66–75.
- Subhaktiyasa, P. G. (2024). Evaluasi validitas dan reliabilitas instrumen penelitian kuantitatif: Sebuah studi pustaka. *Journal of Education Research*, 5(4), 5599–5609. <https://doi.org/10.37985/jer.v5i4.1747>

Tussaadah, R. Z., Hendriana, H., & Yuliani, W. (2021). Validitas dan reliabilitas angket penyesuaian diri peserta didik SMA. *FOKUS: Kajian Bimbingan dan Konseling dalam Pendidikan*, 4(3), 213–218.
<https://doi.org/10.22460/fokus.v4i3.7250>