

MENGENAL TIGA PILAR KUALITAS SOAL: DAYA BEDA, TINGKAT KESUKARAN, DAN PENGECOH

Wafda Sabilla¹, Ma'mun Hanif², Indah Ayu Febriana³, Slamet Gunawan⁴,
Kharis Barik Lana⁵

^{1,2,3,4,5}Pendidikan Agama Islam,
Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan,
wafda.sabilla2409@mhs.uingusdur.ac.id¹
, ma'mun.hanif@uingusdur.ac.id²
, indah.ayu.febriana.24113@mhs.uingusdur.ac.id³,
, slamet.gunawan24093@mhs.uingusdur.ac.id⁴
, kharis.barik.lana24108@mhs.uingusdur.ac.id⁵

ABSTRACT

Item quality is an important factor in determining the validity and reliability of educational evaluation instruments. This study aims to examine three main pillars in item quality analysis, namely discriminating power, difficulty level, and distractor effectiveness, in the context of improving the quality of multiple-choice evaluation instruments in learning activities. This study uses a literature study method by collecting and synthesizing various scientific references related to the topic studied including books, national journals, and articles obtained through Google Scholar. The findings of the study revealed that discriminating power has a primary function as a tool to differentiate between students with high abilities and students with low abilities, with an ideal index value that should reach ≥ 0.40 . The difficulty level describes the degree of ease or difficulty of a question item, with an ideal index in the medium range of 0.31–0.70. Meanwhile, distractor effectiveness refers to the ability of incorrect answer alternatives to influence students who have not yet mastered the material to choose it. A distractor is considered to function well and is declared effective if it is chosen by at least 5% of the total participants who take the test. These three components are interrelated and complementary in determining the overall quality of a test item. This research's contribution lies in providing a comprehensive theoretical framework for educators to analyze and improve the quality of evaluation instruments, ensuring that the tests used in the assessment process accurately, validly, and reliably measure student learning outcomes.

Keywords: item analysis, discriminating power, difficulty level, distractor effectiveness, evaluation instrument

ABSTRAK

Kualitas butir soal merupakan faktor penting dalam menentukan validitas dan reliabilitas instrumen evaluasi pendidikan. Penelitian ini bertujuan untuk mengkaji tiga pilar utama dalam analisis kualitas butir soal, yaitu daya beda, tingkat kesukaran, dan efektivitas pengecoh, dalam konteks peningkatan mutu instrumen evaluasi berbentuk pilihan ganda pada kegiatan pembelajaran. Penelitian ini menggunakan metode studi literatur dengan mengumpulkan dan mensintesis berbagai referensi ilmiah yang berkaitan dengan topik yang dikaji meliputi buku, jurnal nasional, dan artikel yang diperoleh melalui Google Scholar. Temuan dari hasil kajian mengungkapkan bahwa daya pembeda memiliki fungsi utama sebagai alat untuk membedakan antara peserta didik yang memiliki kemampuan tinggi dan peserta didik yang berkemampuan rendah, dengan nilai indeks ideal yang seharusnya mencapai $\geq 0,40$. Tingkat kesukaran menggambarkan derajat kemudahan atau kesulitan suatu butir soal, dengan indeks ideal pada rentang sedang yaitu 0,31–0,70. Sementara itu, efektivitas pengecoh merujuk pada kemampuan alternatif jawaban yang salah dalam memengaruhi peserta didik yang belum menguasai materi untuk memilihnya. Sebuah pengecoh dinilai berfungsi dengan baik dan dinyatakan efektif apabila dipilih oleh paling sedikit 5% dari total keseluruhan peserta yang mengikuti tes. Ketiga komponen tersebut saling berkaitan dan saling melengkapi dalam menentukan kualitas suatu butir soal secara menyeluruh. Kontribusi penelitian ini terletak pada tersedianya kerangka teoritis yang komprehensif bagi pendidik dalam menganalisis dan meningkatkan kualitas instrumen evaluasi, sehingga tes yang diterapkan dalam proses penilaian benar-benar dapat mengukur capaian hasil belajar peserta didik secara tepat, valid, dan reliabel.

Kata Kunci: analisis butir soal, daya pembeda, tingkat kesukaran, efektivitas pengecoh, instrumen evaluasi

A. Pendahuluan

Evaluasi merupakan komponen krusial dalam kegiatan belajar mengajar, karena melalui proses evaluasi dapat diperoleh gambaran mengenai sejauh mana tujuan pembelajaran yang telah dirancang dapat tercapai (Muhadi, 2025). Tes atau soal ujian menjadi instrumen evaluasi yang paling lazim digunakan untuk mengukur kemampuan peserta didik. Akan tetapi, pada kenyataannya tidak semua soal yang dibuat oleh pendidik maupun tim penyusun tes memenuhi standar kualitas yang diharapkan. Berbagai kendala sering ditemukan dalam penyusunan soal, misalnya tingkat kesulitan yang tidak proporsional, baik terlalu mudah maupun terlalu sulit, serta ketidakmampuan soal dalam membedakan antara peserta didik dengan kemampuan tinggi dan mereka yang berkemampuan lebih rendah.. Kondisi semacam ini berpotensi mengurangi keakuratan dan keandalan hasil penilaian, sehingga diperlukan langkah-langkah sistematis untuk memastikan setiap butir soal benar-benar memenuhi kriteria kualitas yang baik.

Analisis butir soal merupakan tahapan strategis yang perlu

dilakukan guna meningkatkan kualitas instrumen evaluasi dalam pembelajaran (Basri et al, 2025). Melalui proses analisis ini, setiap butir soal dapat ditelaah secara mendalam dari berbagai dimensi, di antaranya daya pembeda, indeks kesukaran, dan keberfungsian distraktor. Daya pembeda merujuk pada kemampuan suatu butir soal dalam mengidentifikasi perbedaan antara peserta didik yang telah memahami materi dengan baik dan mereka yang belum menguasainya. Sementara itu, tingkat kesukaran menunjukkan seberapa mudah atau sulit suatu soal, apakah tergolong mudah, sedang, atau sulit. Adapun distraktor atau pilihan pengecoh pada soal berbentuk pilihan ganda berperan sebagai alat uji terhadap ketelitian dan kedalaman pemahaman siswa, sehingga perancangan dan penyusunannya perlu dilakukan secara cermat agar setiap pengecoh dapat menjalankan fungsinya secara efektif dan optimal.

Manfaat analisis butir soal tidak hanya terbatas pada penyempurnaan kualitas soal yang sudah pernah digunakan, melainkan juga berfungsi sebagai bahan refleksi bagi pendidik dalam merancang dan menyusun soal di masa yang akan datang. Dengan

memahami keunggulan dan kelemahan yang terdapat pada setiap butir soal, guru dapat melakukan penyesuaian dan perbaikan yang diperlukan, sehingga instrumen penilaian yang digunakan benar-benar mampu mengukur kompetensi siswa secara objektif, valid, dan akurat. Di samping itu, hasil dari proses analisis ini juga dapat dijadikan dasar pertimbangan dalam pengambilan keputusan terkait status kelayakan sebuah soal, yakni apakah soal tersebut layak untuk tetap dipertahankan dan digunakan kembali, perlu dilakukan revisi dan perbaikan, atau sebaiknya dieliminasi karena tidak memenuhi standar kualitas yang ditetapkan.

B. Metode Penelitian

Penelitian ini menerapkan metode studi literatur atau kajian pustaka sebagai pendekatan utama dalam pelaksanaannya. Pemilihan metode ini didasarkan pada tujuan penelitian yang diarahkan untuk mempelajari, menelaah, dan mengintegrasikan berbagai teori, konsep, serta temuan-temuan dari penelitian terdahulu yang memiliki relevansi dengan topik analisis kualitas butir soal.

Penelitian ini tidak menggunakan responden secara langsung karena seluruh data diperoleh melalui telaah pustaka. Dengan demikian, populasi penelitian adalah seluruh literatur ilmiah yang berkaitan dengan topik, sedangkan sampel penelitian adalah literatur yang memenuhi kriteria kelayakan dan relevansi jurnal.

kejelasan metodologis, dan berhubungan langsung dengan fokus kajian. Data penelitian diambil dari berbagai sumber pustaka, seperti jurnal ilmiah yang membahas kualitas butir soal seperti daya beda, Tingkat kesukaran, dan pengecoh.

Pengumpulan data dilakukan dengan mencari literatur menggunakan kata kunci yang sesuai, memilih sumber yang terpercaya, dan mengelompokkan informasi berdasarkan tema utama. Pencarian literatur dilakukan dengan mengakses buku, Google Scholar, dan jurnal-jurnal nasional yang relevan dengan topik dengan menggunakan kata kunci kualitas butir soal.

Artikel-artikel yang telah dihimpun tersebut diseleksi secara bertahap. Tahap awal dilakukan dengan memastikan judul dan abstrak

artikel-artikel tersebut sesuai dengan fokus penelitian.

Kemudian, artikel yang dinilai relevan, ditelaah lebih lanjut dengan membaca secara keseluruhan untuk melihat isi dan metodologinya. Artikel yang telah terpilih kemudian dikelompokkan berdasarkan tema-tema utama.

Data yang terkumpul dianalisis menggunakan cara deskriptif-kualitatif. Analisis ini meliputi penggabungan berbagai konsep dari sumber yang berbeda, membandingkan hasil penelitian sebelumnya untuk melihat persamaan atau perbedaan, serta menafsirkan makna data.

C. Hasil Penelitian dan Pembahasan

1. Analisis Kualitas Instrumen Tes Berbasis Butir Soal

Analisis soal adalah proses penelaahan terhadap setiap butir pertanyaan dalam suatu tes untuk menilai dan menentukan kualitasnya. Melalui kegiatan ini, setiap item soal dikaji secara mendalam sehingga dapat dihasilkan seperangkat pertanyaan yang layak dan berkualitas tinggi untuk digunakan sebagai alat evaluasi

(Ivtari., & Aminatuz , 2024). Di sisi lain, (Arikunto, 2018) menyatakan bahwa analisis butir soal pada dasarnya merupakan sebuah prosedur yang bersifat sistematis dan terstruktur, yang bertujuan untuk memberikan gambaran informasi secara khusus dan rinci berkaitan dengan setiap butir soal yang telah disusun oleh pendidik maupun tim pengembang tes. Studi oleh Afryaningsih et al (2022) juga menyebutkan bahwa Analisis butir soal pada hakikatnya merupakan upaya untuk mengidentifikasi dan memilah butir soal yang berkualitas baik dari yang kurang baik, sekaligus menelusuri alasan yang mendasari penilaian tersebut. Untuk meningkatkan kualitas soal tertulis, pendidik harus menganalisis kualitas butir-butir tes.

(Solichin, 2017) menyatakan bahwa analisis kualitas butir adalah proses yang digunakan untuk menentukan pertanyaan mana yang baik, dan kurang baik, serta untuk mengumpulkan saran perbaikan. Untuk membuat penilaian pada setiap

ujian, tugas ini melibatkan pengumpulan, merangkum, dan menerapkan data dari respons siswa.

Analisis kualitas butir soal merupakan suatu kegiatan yang dilakukan untuk menelaah serta mengidentifikasi setiap soal tes guna mengetahui tingkat mutunya, sebagaimana dijelaskan oleh para ahli. Melalui proses ini, setiap soal dapat diperbaiki dan disempurnakan berdasarkan hasil penelaahan dan identifikasi yang telah dilakukan.

Kegiatan analisis butir soal menjadi penting karena didasari oleh beberapa alasan, di antaranya sebagai berikut:

- a) Mengidentifikasi keunggulan dan kelemahan butir tes, sehingga dapat dilakukan seleksi dan revisi butir soal.
- b) Untuk memberikan gambaran tentang spesifikasi butir soal secara lengkap, sehingga penyusun soal dapat lebih mudah merancang perangkat tes yang sesuai dengan kebutuhan ujian, baik dari

segi bidang maupun tingkatannya.

- c) Untuk segera mengidentifikasi berbagai permasalahan yang terdapat dalam setiap butir soal, seperti soal yang memiliki makna ganda, penempatan kunci jawaban yang keliru, tingkat kesulitan soal yang terlalu tinggi atau terlalu rendah, maupun soal yang kurang mampu membedakan kemampuan peserta didik. Dengan mendeteksi permasalahan tersebut sedini mungkin, pembuat soal dapat segera mengambil langkah yang tepat sebelum digunakan sebagai alat untuk menentukan nilai siswa, apakah terdapat pertanyaan yang mengandung masalah atau ada yang perlu diperbaiki
- d) Untuk mengevaluasi soal ujian yang akan dikumpulkan
- e) Untuk menyempurnakan kualitas butir soal melalui tiga aspek analisis utama, yakni indeks tingkat

kesukaran, daya pembeda, serta efektivitas pengecoh soal.

- f) Untuk memperoleh data yang cukup mengenai karakteristik masing-masing butir soal, sehingga memungkinkan penyusunan beberapa paket tes yang memiliki tingkat kesetaraan dan kesebandingan. Pengembangan perangkat soal seperti ini sangat berguna, terutama saat diperlukan pelaksanaan ujian ulang atau ketika ingin membandingkan kemampuan beberapa kelompok peserta tes yang diuji pada waktu yang berbeda (Qadir, Huda, 2024).

Berdasarkan uraian di atas, menunjukkan bahwa mengevaluasi kualitas butir soal adalah tugas penting yang harus dilakukan oleh setiap pendidik. Agar para pendidik dapat terus mengembangkan dan meningkatkan kualitas butir soal yang akan digunakan dalam kegiatan penilaian, latihan ini dapat memberikan mereka

gambaran yang jelas tentang berbagai kekurangan dan kelemahan yang terdeteksi dalam butir-butir soal tersebut.

2. Indeks Daya Pembeda Butir Soal

Daya pembeda pada suatu butir soal merujuk pada kemampuan soal tersebut untuk membedakan antara peserta didik yang mampu menjawab dengan benar karena memiliki penguasaan tinggi, dan peserta didik dengan kemampuan lebih rendah yang cenderung tidak dapat menjawab dengan tepat. (Lestari & Yudhanegara, 2017). Dalam pengertian lain, daya pembeda diartikan sebagai kemampuan soal dalam memisahkan kelompok siswa berkemampuan tinggi dengan kelompok peserta didik dengan tingkat kemampuan rendah (Arikunto, 2018). Dengan demikian, analisis daya pembeda pada dasarnya merupakan kegiatan mengkaji setiap butir soal berdasarkan kemampuannya dalam membedakan peserta didik yang berada pada kategori kemampuan rendah dan

kategori kemampuan tinggi. Pendapat yang sejalan menyatakan bahwa daya pembeda adalah kemampuan suatu butir soal untuk mengidentifikasi perbedaan antara peserta tes yang berkemampuan tinggi dan mereka yang memiliki kemampuan lebih rendah (Setiyawan, 2020).

Secara praktis, analisis daya pembeda dilakukan dengan menelaah setiap butir soal berdasarkan kemampuannya dalam mengelompokkan peserta didik ke dalam kategori berkemampuan rendah dan berkemampuan tinggi, yang dilihat dari persentase capaian hasil yang diperoleh. Dalam aplikasi Anates, informasi mengenai daya pembeda setiap butir soal dapat dilihat pada kolom DP persen di tabel daya pembeda. Kegiatan analisis daya pembeda memberikan sejumlah manfaat penting, antara lain membantu meningkatkan kualitas setiap butir soal berdasarkan data empiris yang diperoleh, serta memberikan gambaran

mengenai kemampuan masing-masing soal dalam mengidentifikasi perbedaan tingkat pemahaman peserta didik terhadap materi yang telah dipelajari. Untuk menghitung daya pembeda, digunakan rumus yang merujuk pada pendapat Kusaeri dan Supranto. (Magdalena et al., 2021).

$$D = \frac{BA}{JA} - \frac{BB}{JB} = PA - PB$$

$$PA = \frac{BA}{JA}, PB = \frac{BB}{JB}$$

Keterangan:

D : Daya Beda

BA : Banyak peserta tes kelompok atas yang menjawab dengan benar

BB : Banyak peserta tes kelompok bawah yang menjawab dengan benar

JA : Jumlah peserta tes kelompok atas

JB : Jumlah peserta tes kelompok bawah

PA : Butir tes kelompok atas yang menjawab benar

PB : Butir tes kelompok bawah yang menjawab benar

Indeks daya pembeda berada pada rentang $-1,00$ hingga $+1,00$, di mana semakin tinggi nilai positif yang diperoleh

menunjukkan kualitas butir soal yang semakin baik. Suatu soal dinyatakan berkualitas apabila memiliki indeks daya pembeda $\geq 0,40$. Adapun Kriteria Klasifikasi Indeks Daya Beda :

- a) 0,70-1,00: Baik Sekali
- b) 0,40-0,69: Baik
- c) 0,30-0,39: Cukup (diterima/direvisi)
- d) 0,20-0,29: Kurang Baik (direvisi)
- e) < 0,19: Jelek (dibuang/direvisi)
- f) Negatif (-): (dibuang)

Dengan demikian, dapat dipahami bahwa daya pembeda digunakan untuk menilai kemampuan suatu butir soal dalam mengidentifikasi perbedaan antara peserta didik yang memiliki kemampuan tinggi dan mereka yang berkemampuan lebih rendah. Dengan mempelajari daya beda, guru dapat menilai kualitas soal apakah sudah efektif digunakan dalam evaluasi, karena soal yang baik seharusnya lebih banyak dijawab benar oleh siswa yang pandai dibandingkan siswa yang kurang mampu. Selain itu, hasil analisis daya beda dapat

digunakan sebagai dasar untuk memperbaiki, mengganti, atau membuang soal yang kurang baik agar tes yang dibuat menjadi lebih valid dan sesuai dengan tujuan pembelajaran.

3. Indeks Tingkat Kesukaran Butir Soal

Tingkat kesukaran merupakan ukuran yang digunakan untuk menunjukkan sejauh mana suatu butir soal tergolong mudah atau sulit bagi peserta didik. Apabila seluruh peserta dapat menjawab dengan benar, maka soal tersebut dapat dikategorikan terlalu mudah. Sebaliknya, jika tidak ada peserta yang mampu menjawab dengan tepat, maka soal tersebut termasuk dalam kategori terlalu sulit. Dengan demikian, Tingkat kesukaran menggambarkan derajat kemudahan atau kesulitan suatu soal ketika diberikan kepada sekelompok peserta didik (Sofiani et al, 2023). Secara umum, tingkat kesukaran digunakan sebagai indikator untuk menilai derajat kesulitan suatu butir soal, dan membantu dalam mengelompokkan soal ke

dalam kategori mudah, sedang, dan sukar. Nilai tingkat kesukaran suatu soal dinyatakan dalam bentuk indeks yang berada pada rentang 0,00 hingga 1,00. Nilai 0,00 menunjukkan bahwa soal tersebut sangat sulit karena tidak ada peserta didik yang dapat menjawab dengan benar, sedangkan nilai 1,00 menandakan bahwa soal sangat mudah karena seluruh peserta didik mampu menjawabnya dengan tepat. Melalui rentang nilai ini, tingkat kesukaran setiap butir soal dapat dinilai secara lebih objektif (Iskandar & Rizal, 2018).

Butir soal yang berkualitas adalah soal yang memiliki tingkat kesukaran yang seimbang, tidak terlalu mudah dan tidak terlalu sulit. Soal yang terlalu mudah umumnya kurang mampu merangsang peserta didik untuk berusaha lebih dalam menyelesaikan masalah, sedangkan soal yang terlalu sulit berpotensi menurunkan motivasi karena dianggap berada di luar kemampuan mereka. Oleh karena itu, diperlukan ukuran

yang dapat menunjukkan tingkat kesulitan suatu soal, yang disebut dengan indeks kesukaran (difficulty index). Dalam evaluasi pembelajaran, indeks kesukaran ini biasanya dilambangkan dengan huruf P (huruf besar), yang merupakan singkatan dari kata “proporsi” (Hera et al, 2023)

Rumus Mencari P :

$$P = \frac{B}{JS}$$

P : Indeks Kesukaran

B : Banyak Siswa Yang

Menjawan Soal Dengan Betul

JS : Jumlah Seluruh Siswa

Peserta Tes

Indeks Kesukaran	Kategori Tingkat Kesukaran
0-0.30	Sukar
0,31-0,70	Sedang
0,71-1,00	Mudah

Tabel 1 (Kurniasi et al., 2020).

Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa nilai indeks kesukaran (p) pada suatu butir soal hanya berlaku untuk kelompok peserta yang diuji, sehingga nilainya dapat berubah apabila soal tersebut diberikan kepada kelompok yang berbeda. Selain itu, indeks kesukaran yang

diperoleh dari perhitungan ini mencerminkan tingkat kesukaran soal secara keseluruhan pada tingkat kelompok, bukan pada individu. Oleh karena itu, tingkat kesukaran untuk masing-masing peserta tidak dapat ditentukan hanya dengan melihat nilai proporsi jawaban benar (p) pada suatu butir soal (Erfan et al 2020).

4. Efektivitas Pengecoh Dalam Butir Soal

Analisis kualitas butir soal merupakan bagian yang tidak terpisahkan dari kegiatan evaluasi pembelajaran, terutama pada instrumen berbentuk pilihan ganda. Salah satu komponen penting dalam analisis tersebut adalah efektivitas pengecoh (distractor) merujuk pada fungsi pilihan jawaban yang tidak benar dalam suatu butir soal, namun disusun sedemikian rupa agar tetap terlihat masuk akal dan mampu menarik peserta didik yang belum sepenuhnya memahami materi. Fungsi pengecoh tidak sekadar melengkapi variasi opsi jawaban, melainkan juga

memiliki peran sebagai alat diagnostik dalam mengungkap tingkat pemahaman sekaligus kesalahan konsep yang dialami peserta didik. Oleh karena itu, keberfungsian pengecoh menjadi salah satu indikator penting dalam menilai kualitas suatu butir soal secara menyeluruh (Arikunto, 2018).

Efektivitas pengecoh merupakan kemampuan pilihan jawaban yang tidak benar dalam mengarahkan peserta didik yang belum memiliki pemahaman yang memadai untuk memilih alternatif jawaban tersebut. Pengecoh dikatakan berfungsi dengan baik apabila mampu mengalihkan perhatian peserta tes sehingga mereka tidak memilih jawaban yang benar. Analisis efektivitas pengecoh (distractor) merupakan penelaahan terhadap pilihan jawaban yang memiliki kemiripan dengan jawaban benar, sehingga berpotensi menjebak peserta didik dalam menjawab butir soal yang diujikan. Selain itu, suatu pengecoh dinilai dinyatakan efektif apabila dipilih oleh

sekurang-kurangnya sekitar 5% dari jumlah keseluruhan peserta tes, yang menunjukkan bahwa opsi tersebut cukup menarik dan berfungsi sebagaimana mestinya dalam mengukur kemampuan peserta didik (Uno, H. B., & Koni, 2018).

Dalam kajian evaluasi pendidikan, pengecoh yang baik adalah pengecoh yang mampu berfungsi secara efektif, yaitu dipilih oleh peserta didik yang berkemampuan rendah, namun tidak menarik bagi peserta didik yang berkemampuan tinggi. Efektivitas pengecoh dapat dilihat dari distribusi pemilihan jawaban oleh peserta tes. Apabila suatu pengecoh tidak dipilih sama sekali atau hanya dipilih oleh sebagian kecil peserta, maka pengecoh tersebut dikategorikan tidak berfungsi. Sebaliknya, pengecoh yang dipilih oleh proporsi peserta yang cukup, khususnya dari kelompok bawah, menunjukkan bahwa pengecoh tersebut bekerja dengan baik. Hal ini menunjukkan bahwa pengecoh berperan dalam meningkatkan daya pembeda butir soal,

sehingga mampu mengidentifikasi perbedaan antara peserta didik yang telah memahami materi dan yang belum memahaminya (Sudijono, 2015).

Analisis pengecoh umumnya dilakukan dengan pendekatan kuantitatif melalui penghitungan frekuensi pemilihan setiap alternatif jawaban. Dalam pelaksanaan tes, peserta tes umumnya dibagi menjadi dua kelompok, yaitu kelompok berkemampuan tinggi dan kelompok berkemampuan rendah, yang ditentukan berdasarkan skor total yang diperoleh. Pengecoh dikatakan efektif apabila lebih sering dipilih oleh kelompok berkemampuan rendah dibandingkan oleh kelompok berkemampuan tinggi. Selain itu, secara umum suatu pengecoh dinilai berfungsi dengan baik apabila dipilih oleh setidaknya sekitar 5% dari jumlah seluruh peserta tes.. Jika tidak memenuhi kriteria tersebut, maka pengecoh perlu direvisi atau diganti karena tidak memberikan kontribusi yang

signifikan terhadap kualitas butir soal (Daryanto, 2008).

Kriteria pengecoh yang baik tidak hanya dilihat dari aspek statistik, tetapi juga dari aspek konstruksi soal. Pengecoh harus disusun secara homogen dan sejajar dengan kunci jawaban, baik dari segi panjang kalimat, bentuk bahasa, maupun tingkat kesulitan. Selain itu, pengecoh sebaiknya didasarkan pada kesalahan umum atau miskonsepsi yang sering dialami peserta didik, sehingga memiliki daya tarik yang kuat bagi mereka yang belum memahami materi. Pengecoh yang terlalu mudah dikenali sebagai jawaban salah akan cenderung tidak dipilih, sehingga menjadi tidak efektif dan menurunkan kualitas soal secara keseluruhan (Wulandari, A., & Pramusinto, 2020).

Hasil analisis terhadap pengecoh dapat dijadikan acuan dalam menentukan kualitas suatu butir soal. Jika seluruh pengecoh dalam soal tersebut bekerja secara efektif, maka soal dapat dinilai memiliki kualitas yang baik. Sebaliknya, apabila terdapat satu atau beberapa

pengecoh yang tidak berfungsi sebagaimana mestinya, maka butir soal tersebut perlu diperbaiki atau disempurnakan. Bahkan, jika sebagian besar pengecoh tidak berfungsi, maka soal sebaiknya diganti karena tidak mampu menjalankan fungsinya sebagai alat ukur yang baik. Dengan demikian, analisis pengecoh tidak hanya berperan dalam menilai kualitas soal, tetapi juga dalam meningkatkan kualitas instrumen evaluasi secara berkelanjutan (Rishan & Sulaiman, 2023).

Pengecoh dinyatakan berfungsi dengan baik apabila dipilih oleh sekurang-kurangnya sekitar 5% dari jumlah keseluruhan peserta tes. Indeks pengecoh selanjutnya dihitung menggunakan rumus:

$$IP = \frac{n}{N} \times 100\%$$

Keterangan :

IP = indeks pengecoh

n = jumlah siswa yang memilih opsi pengecoh (jawaban salah)

N = jumlah seluruh siswa yang ikut tes

Catatan: apabila seluruh peserta didik menjawab benar pada suatu butir soal sesuai dengan kunci jawaban, maka nilai IP = 0

yang menunjukkan bahwa butir soal tersebut kurang baik, sehingga pengecoh dinyatakan tidak berfungsi.

Secara keseluruhan, keberadaan pengecoh memiliki peran penting dalam menentukan mutu soal pilihan ganda. Pengecoh yang dirancang dengan baik mampu meningkatkan kemampuan soal dalam membedakan tingkat kemampuan peserta didik, menekan kemungkinan jawaban yang bersifat tebak-tebakan, serta memberikan gambaran yang lebih tepat mengenai kemampuan sebenarnya dari peserta didik. Oleh karena itu, penyusunan dan analisis pengecoh perlu dilakukan secara cermat dan sistematis agar instrumen evaluasi yang digunakan benar-benar mampu mengukur hasil belajar secara valid dan reliabel (Arikunto, 2018).

D. Kesimpulan

Memeriksa kualitas soal adalah hal yang sangat penting dalam pendidikan, terutama guna menjamin bahwa butir soal yang digunakan

dalam ujian memiliki mutu yang baik dan layak digunakan. Dengan melakukan pemeriksaan ini, guru dapat mengetahui mana soal yang sudah bagus dan mana yang masih perlu diperbaiki.

Terdapat tiga aspek pokok yang perlu diperhatikan dalam menilai kualitas suatu butir soal. Salah satunya adalah daya pembeda, yaitu kemampuan soal untuk mengidentifikasi perbedaan antara peserta didik yang berkemampuan tinggi dan mereka yang memiliki kemampuan lebih rendah. Soal yang bagus seharusnya lebih mudah dijawab oleh siswa yang pandai daripada siswa yang kurang mampu, dengan nilai minimal 0,40 agar soal dianggap baik. Kedua adalah tingkat kesukaran, yaitu ukuran seberapa mudah atau sulitnya sebuah soal. Soal yang ideal adalah soal dengan tingkat kesulitan yang seimbang, tidak terlalu mudah dan tidak terlalu sulit, sehingga mampu mendorong peserta didik untuk tetap berpikir dan berusaha dalam menemukan jawabannya. Soal dikatakan ideal jika nilai kesukaran berada di angka 0,31 sampai 0,70. Ketiga adalah pengecoh, yaitu pilihan jawaban yang salah tetapi dibuat semirip mungkin dengan

jawaban yang benar. Pengecoh yang baik adalah yang bisa membuat siswa yang belum paham materi terkecoh, dan harus dipilih oleh setidaknya 5% dari seluruh peserta ujian.

Ketiga hal tersebut saling berkaitan dan sama-sama penting dalam menilai apakah sebuah soal layak dipakai atau tidak. Dari hasil pemeriksaan ini, guru bisa memutuskan apakah soal tersebut bisa langsung digunakan, perlu diperbaiki, atau sebaiknya diganti dengan soal baru. Dengan kata lain, memeriksa kualitas soal bukan sekadar kegiatan tambahan, melainkan bagian penting dari proses ujian agar soal yang digunakan benar-benar mampu mengukur kemampuan siswa secara tepat dan adil.

DAFTAR PUSTAKA

- Afryaningsih, Y., Harjanto, A., & Fatmawati, R. A. (2022). Kedalaman Butir Soal Pilihan Ganda Penilaian Akhir Semester Ganjil Kelas V SD. *Edu Cendikia: Jurnal Ilmiah Kependidikan*, 2(02), 403–409. <https://doi.org/10.47709/educendikia.v2i02.1694>
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara. <https://doi.org/https://doi.org/10.66770/tlrj.v1i4.81>
- Basri, M., Sultan, S., Rapi, M., Baharman, B., & Sakaria, S. (2025). Meningkatkan Kompetensi Evaluasi Pembelajaran melalui Pelatihan Analisis Butir Soal bagi Mahasiswa Calon Guru. *Mallomo: Journal of Community Service*, 5(2), <https://doi.org/10.55678/mallomo.v5i2.2031>. <https://doi.org/https://doi.org/10.55678/mallomo.v5i2.2031>
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. <https://doi.org/https://doi.org/10.23887/ijerr.v3i1.24080>
- Hera Apriliana Saputri Zulhijrah, Nabila Joti Larasati, S. (2023). Analisis Instrumen Assesmen : Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Jurnal Ilmiah Pgsd Fkip Universitas Mandiri*, 2986–2995. <https://doi.org/https://doi.org/10.36989/didaktik.v9i5.2268>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23. <https://doi.org/10.21831/pep.v22i1.15609>
- Ivtari Savika, H., & Aminatuz Zuhriyah, I. (2024). Peran Analisis Butir Soal Terhadap Kualitas Soal, Kompetensi Guru, Dan Prestasi Belajar Peserta Didik Di Sekolah Dasar. *Pandu : Jurnal Pendidikan Anak Dan Pendidikan Umum*, 2(2), 43–51.

- <https://doi.org/https://doi.org/10.59966/pandu.v2i2.856>
- Kurniasi, E. R., Y, Y., & Karennisa, F. (2020). Analisis Soal Ulangan Harian Matematika Kelas IX SMP Negeri 1 Toboali. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, 12(1), 43–52. <https://doi.org/10.37640/jip.v12i1.276>
- Lestari, K. E., & Yudhanegara, M. R. (2017). Analisis Kemampuan Representasi Matematis Mahasiswa pada Mata Kuliah Geometri Transformasi Berdasarkan Latar Belakang Pendidikan Menengah. *Jurnal Matematika Integratif*, 13(1), 29. <https://doi.org/10.24198/jmi.v13.n1.11410.29-34>
- Muhadi. (2025). Evaluasi Perencanaan Desain Pembelajaran, Pelaksanaan Proses Kegiatan Pembelajaran, dan Evaluasi Instrumen Hasil Pembelajaran. *Edu Society: Jurnal Pendidikan, Ilmu Sosial, Dan Pengabdian Kepada Masyarakat*, 5(2), 160.
- Qadir, Huda, H. (2024). Analisis Butir Tes: Tingkat Kesukaran, Daya Pembeda, Efektivitas Pengecoh. *Cv.Eureka Media Aksara*, 3(3), 105–134.
- Rishan, M., & Sulaiman, S. (2023). Analisis Butir Soal Tipe Multiple Choice Ujian Akhir Semester Ganjil Kelas X Tahun Ajaran 2022/2023 Mata Pelajaran Pendidikan Agama Islam di SMA Negeri 1 Padang. *Islamika*, 5(3), 981–997. <https://doi.org/10.36088/islamika.v5i3.3517>
- Sofiani Putri Radja, Vidriana Oktoviana Bano, A. T. I. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda Dan Efektivitas Pengecoh Di Sman 1 Pandawai. *Jurnal Edusavana*, 1(1), 11–41. <https://doi.org/Universitas> <http://orcid.org/0000-0003-1156-6758>
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 195. <https://doi.org/https://doi.org/10.26594/dirasat.v2i2.879>
- Uno, H. B., & Koni, S. (2018). *Assessment Pembelajaran*. (PT Bumi (ed.)).
- Wulandari, A., & Pramusinto, H. (2020). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Otomatisasi Tata Kelola Sarana dan Prasarana Kelas XI Otomatisasi dan Tata Kelola Perkantoran. *Economic Education Analysis Journal*, 9(2), 366–378.