
KUALITAS BUTIR SOAL SEBAGAI FONDASI EVALUASI PEMBELAJARAN: SEBUAH KAJIAN TEORETIS TENTANG PENGECOH, KESUKARAN, DAN DAYA PEMBEDA

Mhd Ali Wardana¹; Masyitoh Hasibuan²; Nur Asyifa³; Aisyah Amini⁴; Yusro Layla⁵

¹²³⁴⁵Sekolah Tinggi Agama Islam Negeri Mandailing Natal, Indonesia

¹muhammadaliwardana782@gmail.com, ²masyitohhasibuan63@gmail.com, ³asyipanur375@gmail.com,

⁴aisyahamini160706@gmail.com, ⁵rangkutyusrohlaila@gmail.com.

Corresponding Author: Masyitoh Hasibuan- masyitohhasibuan63@gmail.com

Abstract

This research focuses on examining the quality of question items in learning evaluation tools based on three main criteria: the effectiveness of the trickster, the level of difficulty, and the ability to distinguish. A literature review approach was used for this study, by exploring a variety of relevant academic books and articles related to educational assessment. The analysis is conducted in a structured manner, combining theoretical perspectives and insights from previous studies to gain an in-depth understanding of the attributes of high-quality question items. The findings of the study show that the effectiveness of the question items is greatly influenced by how well the wrong answer choices are designed, the distribution of difficulty levels, and the ability to distinguish between students with different abilities. Effective question items usually have a balanced level of difficulty, strong discriminating abilities, and incorrect answer choices that fit the intended purpose. Conversely, question items that are too low or too high in difficulty, along with incorrect answer choices that fail to fulfill their function, will compromise the overall reliability and validity of the assessment tool. These results highlight the important role of question item analysis in the educational assessment process

Keywords: Question Item Analysis; Deceivers; Difficulty Level; Differentiating Power; Learning Evaluation

Abstrak

Penelitian ini berfokus pada pemeriksaan kualitas item soal dalam alat evaluasi pembelajaran berdasarkan tiga kriteria utama: efektivitas penipu, tingkat kesulitan, dan kemampuan membedakan. Pendekatan tinjauan literatur digunakan untuk penelitian ini, dengan mengeksplorasi berbagai buku dan artikel akademik yang relevan terkait penilaian pendidikan. Analisis dilakukan secara terstruktur, menggabungkan perspektif teoretis dan wawasan dari penelitian sebelumnya untuk mendapatkan pemahaman mendalam tentang atribut item pertanyaan berkualitas tinggi.

Temuan penelitian menunjukkan bahwa efektivitas item soal sangat dipengaruhi oleh seberapa baik pilihan jawaban yang salah dirancang, distribusi tingkat kesulitan, dan kemampuan untuk membedakan antara siswa dengan kemampuan yang berbeda. Item pertanyaan yang efektif biasanya memiliki tingkat kesulitan yang seimbang, kemampuan pembeda yang kuat, dan pilihan jawaban yang salah yang sesuai dengan tujuan yang dimaksudkan. Sebaliknya, item pertanyaan yang kesulitannya terlalu rendah atau terlalu tinggi, bersama dengan pilihan jawaban yang salah yang gagal memenuhi fungsinya, akan membahayakan keandalan dan validitas alat penilaian secara keseluruhan. Hasil ini menyoroti peran penting analisis item soal dalam proses penilaian pendidikan.

Kata kunci: Analisis Item Pertanyaan; Pengecoh; Tingkat Kesulitan; Kekuatan Pembeda; Evaluasi Pembelajaran

PENDAHULUAN

Penilaian pembelajaran dapat diartikan sebagai bagian penting dari kerangka pendidikan, yang bertujuan untuk menilai seberapa baik tujuan pembelajaran telah tercapai. Penilaian tidak hanya digunakan sebagai sarana untuk mengukur prestasi siswa, melainkan menjadi landasan untuk

membuat keputusan tentang bagaimana meningkatkan pengalaman pendidikan, yang mencakup pendekatan, teknik, dan sumber daya yang digunakan. Menurut Arikunto (2018) Evaluasi pembelajaran melibatkan pengumpulan data mengenai prestasi akademik siswa, yang kemudian digunakan untuk menilai tingkat keberhasilan pendidikan. Sesuai dengan hal tersebut Sudijono (2017) Menekankan bahwa penilaian memainkan peran penting dalam memberikan wawasan bagi guru dan murid dalam meningkatkan efektivitas pengajaran dan pengalaman pendidikan.

Dalam bidang pendidikan, instrumen penilaian yang umum digunakan adalah ujian. Ujian berfungsi sebagai mekanisme atau metode yang dirancang untuk mengukur atau mengevaluasi kemampuan siswa menggunakan serangkaian pertanyaan atau aktivitas yang membutuhkan respons. Seperti yang dinyatakan oleh Sudijono (2017), ujian adalah teknik atau metode untuk mengukur dan mengevaluasi hasil pendidikan, yang terwujud sebagai tugas atau pertanyaan yang menghasilkan skor yang mencerminkan kinerja atau pencapaian siswa. Ujian digunakan untuk menilai sejauh mana siswa telah memahami isi, kemampuan kognitif mereka, dan kemajuan pendidikan mereka setelah terlibat dalam pengalaman belajar.

Tes tersedia dalam berbagai jenis, termasuk esai tertulis, penilaian verbal, dan evaluasi objektif. Jenis penilaian objektif yang umum digunakan adalah format pertanyaan pilihan ganda. Gaya pengujian ini dianggap bermanfaat karena kemampuannya untuk mencakup berbagai macam konten, tingkat imparialitas penilaian yang tinggi, dan kemudahannya dalam mengevaluasi prestasi siswa bagi pendidik. Linn dan Gronlund (2000) menyatakan bahwa pertanyaan pilihan ganda yang dirancang dengan tepat dapat mengukur berbagai tingkatan keterampilan kognitif jika sesuai dengan pedoman penulisan tes yang telah ditetapkan. (Juni, 2023).

Dalam pendidikan modern, evaluasi tidak lagi dianggap hanya sebagai kegiatan yang dilakukan pada akhir pembelajaran, tetapi telah menjadi bagian yang tidak terpisahkan dari seluruh rangkaian proses pembelajaran. Pandangan ini sejalan dengan pendapat Mardapi (2012) Pelaksanaan penilaian harus dilakukan secara terstruktur dan berkelanjutan guna memperoleh informasi yang akurat dan terpercaya. Oleh sebab itu, evaluasi tidak hanya digunakan untuk mengetahui tingkat keberhasilan belajar siswa, tetapi juga menjadi acuan bagi pendidik dalam merancang perbaikan dan pengembangan pembelajaran. Mengenali kemampuan dan tantangan peserta didik terkait dengan isi mata pelajaran. Dalam penerapannya secara praktis, salah satu jenis penilaian yang umum adalah ujian tertulis, terutama yang terdiri dari soal pilihan ganda. Metode pengujian ini sering digunakan karena memberikan evaluasi yang lebih objektif, memfasilitasi penilaian yang adil, menyederhanakan proses pemberian nilai, dan memungkinkan berbagai macam topik untuk dibahas dalam jangka waktu yang terbatas. (Linn & Gronlund, 2000).

Selain itu, menurut Brookhart & Nitko (2011) Soal pilihan ganda yang dirancang dengan baik dapat menilai berbagai keterampilan kognitif, mulai dari pengetahuan dasar hingga penalaran analitis, selama mengikuti pedoman pengujian yang telah ditetapkan. Namun demikian, efektivitas hasil penilaian sangat terkait dengan kualitas soal tes yang digunakan. Soal yang dirancang secara tidak efektif dapat menyebabkan evaluasi yang salah dan menyesatkan tentang kemampuan siswa. Misalnya, soal yang terlalu sederhana mungkin gagal membedakan peserta didik yang memiliki kemampuan tinggi dengan peserta didik yang memiliki kemampuan rendah. Sebaliknya, soal yang terlalu sulit dapat menyebabkan siswa kehilangan motivasi dan belum tentu mencerminkan tingkat penguasaan mereka terhadap materi secara tepat. Oleh karena itu, analisis terhadap butir soal perlu dilakukan untuk memastikan bahwa instrumen penilaian yang digunakan memiliki kualitas yang baik, valid, serta reliabel dalam mengukur hasil belajar. (Salsabilla et al., 2024).

Analisis butir soal adalah kegiatan yang dilakukan untuk mengevaluasi kualitas setiap butir soal yang dikerjakan siswa dari kajian teori yang ada (Febriani et al., 2026). Tugas ini dirancang untuk menilai apakah suatu pertanyaan tes memenuhi standar instrumen pengukuran yang efektif atau apakah diperlukan penyesuaian lebih lanjut. Menganalisis suatu item melibatkan beberapa faktor penting: tingkat kesulitan, kemampuan untuk membedakan antara peserta tes, dan efektivitas pilihan yang menyesatkan. Ketiga faktor ini sangat penting untuk memutuskan apakah

suatu pertanyaan tes sesuai untuk diimplementasikan, perlu dimodifikasi, atau harus dibuang dari alat penilaian (Mahendra, 2019).

Tingkat kesulitan adalah penilaian terhadap kemudahan atau tantangan suatu pertanyaan, yang ditentukan oleh proporsi murid yang menjawab dengan benar. Pertanyaan yang efektif biasanya memiliki tingkat kesulitan yang seimbang, sehingga memberikan wawasan terbaik tentang kemampuan siswa. Daya diskriminasi mengacu pada seberapa baik suatu pertanyaan dapat membedakan antara siswa dengan berbagai tingkat kemampuan. Pertanyaan yang menunjukkan daya diskriminasi yang kuat dapat secara jelas dan tepat menyoroti perbedaan kemampuan siswa. (Fatimah & Alfath, 2019).

Kemudian, efektivitas pengecoh berkaitan dengan sejauh mana alternatif jawaban yang salah mampu menarik perhatian peserta didik yang kurang memahami materi. Pengecoh yang baik harus berfungsi secara optimal agar soal tidak menjadi terlalu mudah ditebak (Haladyna et al., 2013). Analisis butir soal merupakan hal penting dalam membantu guru mengetahui kualitas soal yang dibuat, sehingga soal dapat diperbaiki dan digunakan sebagai alat evaluasi yang lebih baik dan lebih tepat dalam mengukur kemampuan siswa.

Menurut Azwar (1995) Melalui evaluasi butir soal, pendidik dapat mengidentifikasi kekurangan dalam cara pertanyaan dirancang dan mendorong peningkatan berkelanjutan. Selain itu, temuan dari penilaian butir soal dapat membantu dalam pemilihan dan pengarsipan pertanyaan berkualitas tinggi dalam basis data pertanyaan, menjamin bahwa pertanyaan yang digunakan telah melalui pengujian, sesuai untuk menilai keterampilan siswa, dan menghasilkan hasil yang konsisten dan dapat diandalkan. Dengan demikian, dapat disimpulkan bahwa evaluasi butir soal merupakan komponen penting dalam menjaga efektivitas alat penilaian pendidikan. Tanpa analisis menyeluruh, hasil penilaian berisiko menjadi bias dan mungkin tidak secara akurat mewakili kemampuan siswa yang sebenarnya. Oleh karena itu, sangat penting bagi pendidik dan mereka yang terlibat dalam pendidikan untuk memiliki pemahaman yang kuat tentang metode analisis butir soal dan menerapkannya secara teratur dalam proses pembelajaran evaluatif.

Berdasarkan uraian tersebut, sebagian besar penelitian mengenai kualitas butir soal masih berfokus pada analisis masing-masing indikator secara terpisah, seperti efektivitas pengecoh, tingkat kesukaran, atau daya pembeda. Padahal, ketiga komponen tersebut memiliki keterkaitan yang saling memengaruhi dalam menentukan kualitas instrumen evaluasi secara keseluruhan. Oleh karena itu, diperlukan suatu kajian yang tidak hanya mendeskripsikan setiap indikator, tetapi juga mensintesis hubungan konseptual di antara ketiganya.

Artikel ini merupakan kajian konseptual berbasis studi pustaka yang bertujuan untuk mensintesis berbagai teori dan hasil penelitian mengenai efektivitas pengecoh, tingkat kesukaran, dan daya pembeda dalam analisis butir soal. Melalui sintesis tersebut, artikel ini berupaya membangun model konseptual yang menjelaskan keterkaitan ketiga komponen sebagai fondasi kualitas instrumen evaluasi pembelajaran. Dengan demikian, kontribusi artikel tidak hanya terletak pada pemaparan teori, tetapi juga pada pengembangan pemahaman yang lebih komprehensif mengenai hubungan antarindikator kualitas butir soal.

METODE

Penelitian ini menggunakan metode studi pustaka (*library research*) dengan menelaah berbagai sumber literatur yang relevan, seperti buku, jurnal ilmiah, artikel penelitian, dan dokumen akademik yang berkaitan dengan efektivitas pengalih perhatian, tingkat kesulitan, dan daya pembeda dalam analisis butir soal. Data dikumpulkan melalui kegiatan membaca, mengidentifikasi, mencatat, dan menyeleksi informasi yang sesuai dengan fokus penelitian. Selanjutnya, data dianalisis menggunakan pendekatan deskriptif kualitatif dengan cara mengelompokkan, membandingkan, menginterpretasikan, dan mensintesis berbagai temuan dari literatur untuk memperoleh pemahaman yang komprehensif mengenai karakteristik butir soal yang berkualitas dalam evaluasi pembelajaran.

Penelitian ini tidak menggunakan data hasil tes peserta didik secara langsung. Seluruh pembahasan dan contoh perhitungan yang disajikan dalam artikel digunakan sebagai ilustrasi konseptual untuk menjelaskan penerapan teori analisis butir soal berdasarkan berbagai sumber literatur yang relevan. Oleh karena itu, temuan yang dihasilkan bersifat konseptual dan merupakan hasil sintesis teoritis dari berbagai penelitian terdahulu, bukan hasil analisis empiris terhadap data lapangan.

HASIL DAN PEMBAHASAN

Hasil

1. Efektivitas Pengecoh dalam Analisis Butir Soal

Hasil studi literatur menunjukkan bahwa efektivitas pengecoh masih menjadi salah satu aspek yang sering menjadi kelemahan dalam penyusunan instrumen evaluasi pembelajaran. Banyak penelitian melaporkan bahwa terdapat opsi jawaban yang tidak dipilih peserta didik atau hanya dipilih oleh sebagian kecil peserta tes sehingga tidak berfungsi sebagai pengecoh yang baik. Kondisi tersebut mengindikasikan bahwa alternatif jawaban yang disediakan kurang mampu menarik perhatian peserta didik yang belum menguasai materi. Padahal, pengecoh yang efektif seharusnya dipilih oleh minimal 5% peserta didik agar dapat berfungsi sebagai alat pembeda antara peserta didik yang memahami materi dan yang belum memahaminya (Dewi et al., 2019). Temuan ini sejalan dengan pendapat Sumanti dan Susanti (2019) yang menyatakan bahwa pengecoh harus disusun secara logis dan berkaitan dengan konsep yang diukur agar mampu meningkatkan kualitas instrumen evaluasi.

2. Tingkat Kesukaran Butir Soal

Berdasarkan hasil sintesis literatur, tingkat kesukaran merupakan salah satu indikator utama dalam menentukan kualitas suatu butir soal. Berbagai penelitian menunjukkan bahwa soal dengan tingkat kesukaran sedang cenderung memberikan informasi yang lebih optimal mengenai kemampuan peserta didik dibandingkan soal yang terlalu mudah maupun terlalu sulit. Sebagai ilustrasi konseptual yang sering digunakan dalam analisis butir soal, nilai indeks kesukaran sebesar 0,70 termasuk dalam kategori sedang. Nilai tersebut digunakan untuk menjelaskan interpretasi indeks kesukaran dan bukan merupakan hasil pengukuran empiris dalam penelitian ini. Temuan dari berbagai literatur menunjukkan bahwa soal dengan kategori sedang memiliki kemampuan yang lebih baik dalam menggambarkan variasi kemampuan peserta didik secara objektif dan reliabel. Temuan ini sejalan dengan Son (2019) yang menjelaskan bahwa tingkat kesukaran sedang memberikan informasi yang lebih representatif mengenai kemampuan peserta didik. Selain itu, Sudijono (2017) menegaskan bahwa komposisi soal yang baik sebaiknya didominasi oleh soal berkategori sedang karena mampu menghasilkan pengukuran yang lebih reliabel dan objektif.

Berdasarkan hasil sintesis literatur, tingkat kesukaran merupakan salah satu indikator utama dalam menentukan kualitas suatu butir soal. Berbagai penelitian menunjukkan bahwa soal dengan tingkat kesukaran sedang cenderung memberikan informasi yang lebih optimal mengenai kemampuan peserta didik dibandingkan soal yang terlalu mudah maupun terlalu sulit. Sebagai ilustrasi konseptual yang sering digunakan dalam analisis butir soal, nilai indeks kesukaran sebesar 0,70 termasuk dalam kategori sedang. Nilai tersebut digunakan untuk menjelaskan interpretasi indeks kesukaran dan bukan merupakan hasil pengukuran empiris dalam penelitian ini. Temuan dari berbagai literatur menunjukkan bahwa soal dengan kategori sedang memiliki kemampuan yang lebih baik dalam menggambarkan variasi kemampuan peserta didik secara objektif dan reliabel.

3. Daya Pembeda Butir Soal

Hasil sintesis berbagai literatur menunjukkan bahwa daya pembeda merupakan indikator yang sangat penting dalam menentukan efektivitas suatu instrumen evaluasi. Daya pembeda menggambarkan kemampuan suatu butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Sebagai ilustrasi konseptual yang banyak digunakan dalam pembahasan analisis butir soal, nilai daya pembeda sebesar 0,40 termasuk dalam kategori sangat baik. Nilai tersebut digunakan sebagai contoh interpretasi indeks daya pembeda dan bukan berasal

dari hasil pengumpulan data empiris pada penelitian ini. Literatur menunjukkan bahwa semakin tinggi nilai daya pembeda suatu soal, semakin baik kemampuan soal tersebut dalam mengidentifikasi perbedaan kompetensi peserta didik. (Nurhalimah et al., 2022). Kondisi ini menunjukkan bahwa proses penyusunan soal perlu memperhatikan kemampuan butir soal dalam mengidentifikasi perbedaan kompetensi peserta didik. Temuan tersebut diperkuat oleh penelitian Salsabilla et al. (2024) yang menyatakan bahwa daya pembeda yang rendah dapat menurunkan kualitas instrumen dan menghasilkan informasi yang kurang akurat mengenai kemampuan peserta didik.

Hasil sintesis berbagai literatur menunjukkan bahwa daya pembeda merupakan indikator yang sangat penting dalam menentukan efektivitas suatu instrumen evaluasi. Daya pembeda menggambarkan kemampuan suatu butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Sebagai ilustrasi konseptual yang banyak digunakan dalam pembahasan analisis butir soal, nilai daya pembeda sebesar 0,40 termasuk dalam kategori sangat baik. Nilai tersebut digunakan sebagai contoh interpretasi indeks daya pembeda dan bukan berasal dari hasil pengumpulan data empiris pada penelitian ini. Literatur menunjukkan bahwa semakin tinggi nilai daya pembeda suatu soal, semakin baik kemampuan soal tersebut dalam mengidentifikasi perbedaan kompetensi peserta didik.

Secara keseluruhan, hasil studi literatur menunjukkan bahwa kualitas instrumen evaluasi pembelajaran sangat dipengaruhi oleh efektivitas pengecoh, tingkat kesukaran, dan daya pembeda soal. Ketiga komponen tersebut saling berkaitan dalam menentukan kemampuan instrumen untuk mengukur hasil belajar peserta didik secara objektif dan akurat. Masih ditemukannya pengecoh yang tidak berfungsi, dominasi soal mudah, serta rendahnya daya pembeda pada sebagian butir soal menunjukkan bahwa kualitas instrumen evaluasi belum sepenuhnya optimal. Temuan ini sejalan dengan hasil penelitian Magdalena et al. (2021) dan Trianingsih (2023) yang menyatakan bahwa analisis butir soal perlu dilakukan secara berkelanjutan agar kualitas evaluasi pembelajaran dapat terus ditingkatkan. Oleh karena itu, guru dan penyusun instrumen perlu melakukan analisis butir soal sebelum dan sesudah tes digunakan sehingga instrumen yang dihasilkan benar-benar mampu mengukur capaian pembelajaran peserta didik secara valid dan reliabel.

Pembahasan

1. Pengecoh

Pengecoh merupakan alternatif jawaban yang salah namun dirancang sedemikian rupa sehingga menyerupai jawaban yang benar. Keberadaan pengecoh dalam soal pilihan ganda memiliki peran penting dalam meningkatkan kualitas instrumen evaluasi, karena berfungsi untuk menguji ketelitian serta pemahaman peserta didik terhadap materi yang diujikan. Pengecoh yang disusun dengan baik tidak boleh terlalu jelas kesalahannya, melainkan harus logis dan masih berkaitan dengan konsep yang diukur, sehingga mampu membuat peserta didik terkecoh yang kurang memahami materi secara mendalam (Sumanti & Susanti, 2019).

Pengecoh yang baik adalah alternatif jawaban yang dapat menarik perhatian peserta didik yang belum menguasai materi yang diujikan, khususnya mereka yang memiliki tingkat penguasaan materi yang rendah. Hal ini ditandai dengan adanya distribusi pemilihan jawaban pada setiap opsi yang relatif merata, terutama pada kelompok peserta didik berkemampuan rendah. Sebaliknya, pengecoh yang tidak berfungsi biasanya tidak dipilih sama sekali atau justru lebih banyak dipilih oleh peserta didik yang berkemampuan tinggi, sehingga mengurangi kualitas soal. Oleh karena itu, penyusunan pengecoh perlu dilakukan secara cermat dan berdasarkan analisis yang matang agar dapat berfungsi optimal dalam (Riinawati, 2021).

Berdasarkan beberapa pendapat tersebut, dapat dipahami bahwa pengecoh adalah pilihan jawaban salah yang dibuat menyerupai jawaban benar agar dapat mengukur pemahaman peserta didik. Pengecoh yang baik harus masuk akal, berkaitan dengan materi yang diujikan, dan mampu menarik peserta didik yang belum memahami materi dengan baik. Sebaliknya, jika pengecoh tidak dipilih atau sangat sedikit dipilih oleh peserta didik, maka pengecoh tersebut dianggap kurang

efektif. Oleh karena itu, penyusunan pengecoh perlu dilakukan dengan cermat karena berpengaruh terhadap kualitas butir soal dan ketepatan hasil evaluasi pembelajaran.

Cara Menghitung Efektivitas Pengecoh Pengecoh dikatakan berfungsi apabila dipilih minimal 5% peserta didik. Rumus: Persentase Pengecoh = (Jumlah peserta yang memilih pengecoh / Jumlah seluruh peserta) \times 100%.

Contoh:

Suatu soal dikerjakan oleh 40 peserta didik dengan hasil sebagai berikut:

Tabel 1. Distribusi Pemilihan Opsi Jawaban untuk Analisis Efektivitas Pengecoh

Opsi Jawaban	Jumlah Peserta
A (Jawaban Benar)	24
B	8
C	6
D	2
Total	40

Tabel 2. Persentase Pemilihan Opsi Pengecoh pada Butir Soal

Pengecoh	Jumlah Peserta	Perhitungan	Persentase
B	8	$(8/40) \times 100\%$	20%
C	6	$(6/40) \times 100\%$	15%
D	2	$(2/40) \times 100\%$	5%

Berdasarkan temuan ini, pilihan pengecoh B dan C terbukti efektif karena dipilih oleh lebih dari 5% peserta. Selain itu, pilihan pengecoh D juga dianggap efektif karena dipilih oleh tepat 5% peserta.

2. Tingkat Kesukaran

Tingkat kesulitan adalah metrik dalam evaluasi pertanyaan yang mencerminkan menggambarkan derajat kesulitan sebuah soal yang dirasakan oleh peserta didik. Tingkat kesulitan suatu pertanyaan dinilai berdasarkan proporsi siswa yang dapat memberikan jawaban yang benar relatif terhadap jumlah keseluruhan peserta ujian. Ketika sebagian besar siswa berhasil menjawab dengan akurat, pertanyaan tersebut dianggap lebih mudah; namun, jika hanya sebagian kecil yang mencapai jawaban yang benar, pertanyaan tersebut dianggap sulit. Tingkat kesulitan ideal biasanya berada dalam kisaran menengah, karena dapat memberikan wawasan yang paling efektif dalam menilai kompetensi siswa secara adil. (Son, 2019).

Dalam menilai pembelajaran, tingkat kesulitan sangat penting karena memengaruhi efektivitas alat pengujian. Soal yang terlalu mudah tidak membedakan antara berbagai kemampuan siswa, sedangkan soal yang terlalu sulit dapat menyebabkan penurunan motivasi siswa dan tidak secara akurat menunjukkan kemampuan mereka yang sebenarnya. Oleh karena itu, ketika mengembangkan soal, pendidik harus merancang soal yang mempertahankan tingkat kesulitan yang adil, idealnya termasuk lebih banyak soal dalam kisaran sedang. Pendekatan ini memastikan bahwa hasil penilaian dapat lebih andal dan akurat menunjukkan kemampuan siswa (Dewi et al., 2019).

Berdasarkan pendapat para ahli, dapat dipahami bahwa tingkat kesulitan menunjukkan seberapa mudah atau sulit suatu soal dikerjakan oleh peserta didik. Tingkat kesulitan dapat dilihat dari banyaknya peserta didik yang menjawab soal dengan benar. Soal yang terlalu mudah maupun terlalu sulit kurang baik digunakan karena tidak dapat menggambarkan kemampuan peserta didik secara optimal. Oleh karena itu, soal dengan tingkat kesulitan sedang lebih disarankan karena dapat memberikan hasil penilaian yang lebih tepat dan mampu menunjukkan kemampuan peserta didik secara lebih objektif. Adapun rumus dari tingkat kesukaran adalah:

$$P = (B / N)$$

Keterangan:

P = Indeks kesusahan Soal

B = jumlah peserta didik yang menjawab dengan benar

N = jumlah semua peserta didik

Tabel 3. Kriteria hasil Indeks Tingkat Kesulitan Butir Soal.

Nilai Indeks Kesulitan (P)	Kategori
$P < 0,30$	Sulit
$0,30 \leq P \leq 0,70$	Sedang
$P > 0,70$	Mudah

Contoh:

Sebuah soal dijawab dengan benar oleh 28 peserta didik dari 40 peserta.

Maka:

$$P = 28 / 40 = 0,70$$

Hasil tersebut menunjukkan bahwa soal termasuk kategori sedang

Tingkat kesukaran soal sangat berkaitan dengan tingkat kemampuan kognitif dalam Taksonomi Bloom. Soal pada tingkat C1 (mengingat) dan C2 (memahami) biasanya lebih mudah karena hanya menuntut kemampuan dasar peserta didik. Sementara itu, Soal-soal pada level C4, C5, dan C6 lebih sulit karena membutuhkan kemampuan berpikir yang lebih maju. (Lumbanraja, 2017). Sebagai contoh, soal yang hanya meyeruh peserta didik untuk menghafal definisi akan lebih mudah dibandingkan soal yang hanya meminta peserta didik menganalisis kasus. Oleh karena itu, semakin tinggi tingkat kognitif soal, maka biasanya tingkat kesukarannya juga semakin tinggi.

3. Daya Pembeda

Dalam analisis butir soal, daya pembeda menjadi indikator yang digunakan untuk mengukur sejauh mana suatu soal mampu mengelompokkan peserta didik berdasarkan tingkat kemampuannya. Butir soal yang memiliki daya pembeda tinggi umumnya dapat dijawab dengan benar oleh peserta didik yang memiliki kemampuan lebih baik, sedangkan peserta didik dengan kemampuan yang lebih rendah cenderung mengalami kesulitan dalam menjawabnya dengan tepat (Magdalena et al., 2021). Dengan demikian, daya pembeda berfungsi untuk menunjukkan tingkat efektivitas suatu soal dalam mengidentifikasi perbedaan kemampuan antar peserta didik.

Secara umum, daya pembeda diperoleh dengan membandingkan hasil jawaban kelompok peserta didik yang memiliki nilai tinggi (kelompok atas) dengan kelompok peserta didik yang memiliki nilai rendah (kelompok bawah). Semakin besar selisih jumlah jawaban benar antara kedua kelompok tersebut, semakin tinggi pula daya pembeda yang dimiliki oleh butir soal. Sebaliknya, apabila perbedaan skor jawaban benar diantara kelompok atas dengan kelompok rendah sangat kecil atau bahkan tidak ada, maka soal tersebut memiliki daya pembeda yang rendah sehingga kurang efektif digunakan sebagai alat evaluasi.

Berdasarkan tersebut, dapat dipahami bahwa daya pembeda adalah kemampuan suatu soal untuk membedakan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah. Soal yang baik biasanya lebih banyak dijawab benar oleh peserta didik yang memiliki kemampuan tinggi dibandingkan peserta didik yang memiliki kemampuan rendah. Semakin besar perbedaan hasil jawaban antara kedua kelompok tersebut, maka semakin baik daya pembeda soal. Oleh karena itu, daya pembeda penting untuk mengetahui apakah suatu soal sudah mampu mengukur kemampuan peserta didik dengan baik atau belum.

Rumus yang dapat digunakan untuk menghitung daya pembeda adalah sebagai berikut:

$$D = \frac{B_A - B_B}{J_A}$$

Keterangan:

D = indeks daya beda

BA= jumlah peserta kelompok atas yang menjawab benar.

BB= jumlah peserta kelompok bawah yang menjawab benar

JA= Jumlah peserta dalam kelompok atas

Nilai daya pembeda suatu butir soal dapat dikelompokkan ke dalam beberapa kategori seperti tabel di bawah ini:

Tabel 4. Tingkat Kemampuan Soal dalam Membedakan Peserta Didik.

Nilai Daya Pembeda (D)	Kategori
$D \geq 0,40$	Sangat Baik
$0,30 \leq D < 0,40$	Baik
$0,20 \leq D < 0,30$	Cukup
$D < 0,20$	Kurang Baik

Contoh:

dalam suatu tes yang diikuti oleh 40 peserta didik, terdapat 18 peserta didik dari kelompok atas yang menjawab benar, sedangkan dari kelompok bawah hanya 10 peserta didik yang berhasil memberikan jawaban yang tepat.

Perhitungannya adalah:

$$D = (18 - 10) / (40/2)$$

$$D = 8 / 20$$

$$D = 0,40$$

Hasil analisis menunjukkan bahwa butir soal tersebut termasuk dalam kategori daya pembeda sangat baik, sebab dapat mengidentifikasi perbedaan kemampuan peserta didik secara efektif. Sebuah soal dapat dikatakan memiliki daya pembeda yang memadai jika nilai indeksnya termasuk dalam kategori cukup hingga sangat baik. Soal dengan daya pembeda yang baik berfungsi secara efektif dalam mengidentifikasi perbedaan kemampuan peserta didik, sehingga informasi yang diperoleh dari hasil tes menjadi lebih akurat dan dapat digunakan sebagai dasar evaluasi pembelajaran.

Soal dengan nilai daya pembeda di bawah 0,20 termasuk dalam kategori kurang baik karena kurang efektif dalam membedakan peserta didik berdasarkan tingkat kemampuannya. Apabila indeks daya pembeda bernilai negatif, maka dapat diartikan bahwa peserta didik berkemampuan rendah justru lebih banyak menjawab benar daripada peserta didik berkemampuan tinggi. Keadaan tersebut menandakan adanya masalah pada butir soal sehingga tidak layak digunakan sebagai alat evaluasi. Dengan demikian, soal yang memiliki daya pembeda rendah atau negatif sebaiknya diperbaiki secara menyeluruh atau dikeluarkan dari perangkat tes.

4. Hubungan Konseptual Efektivitas Pengecoh, Tingkat Kesukaran, dan Daya Pembeda

Berdasarkan sintesis berbagai teori dan hasil penelitian terdahulu, efektivitas pengecoh, tingkat kesukaran, dan daya pembeda merupakan tiga komponen yang saling berkaitan dalam menentukan kualitas suatu butir soal. Ketiga indikator tersebut tidak bekerja secara terpisah, melainkan membentuk suatu sistem yang secara bersama-sama memengaruhi validitas dan ketepatan instrumen evaluasi pembelajaran

Efektivitas pengecoh memiliki hubungan yang erat dengan tingkat kesukaran soal. Pengecoh yang dirancang secara logis dan sesuai dengan miskonsepsi peserta didik akan lebih mampu menarik peserta didik yang belum menguasai materi sehingga tingkat kesukaran soal menjadi lebih proporsional. Sebaliknya, pengecoh yang tidak berfungsi menyebabkan peserta didik lebih mudah menemukan jawaban yang benar sehingga soal menjadi terlalu mudah. Dengan demikian, kualitas pengecoh berkontribusi langsung terhadap terbentuknya tingkat kesukaran suatu butir soal.

Tingkat kesukaran selanjutnya berpengaruh terhadap daya pembeda. Berbagai penelitian menunjukkan bahwa soal dengan tingkat kesukaran sedang cenderung memiliki daya pembeda yang lebih tinggi dibandingkan soal yang terlalu mudah atau terlalu sulit. Pada soal yang terlalu

mudah, hampir seluruh peserta didik mampu menjawab dengan benar sehingga perbedaan kemampuan antarpeserta didik menjadi sulit diidentifikasi. Sebaliknya, pada soal yang terlalu sulit, sebagian besar peserta didik gagal menjawab dengan benar sehingga kemampuan soal dalam membedakan tingkat kompetensi peserta didik juga menjadi rendah.

Selain memengaruhi tingkat kesukaran, efektivitas pengecoh juga berkontribusi terhadap daya pembeda. Pengecoh yang efektif umumnya lebih banyak dipilih oleh peserta didik yang memiliki kemampuan rendah, sedangkan peserta didik yang memiliki kemampuan tinggi cenderung mampu mengidentifikasi jawaban yang benar. Kondisi ini menghasilkan perbedaan yang lebih jelas antara kelompok atas dan kelompok bawah sehingga daya pembeda menjadi lebih optimal.

Berdasarkan hubungan tersebut, dapat dirumuskan suatu model konseptual bahwa efektivitas pengecoh memengaruhi tingkat kesukaran soal, tingkat kesukaran memengaruhi daya pembeda, dan efektivitas pengecoh juga memberikan pengaruh tidak langsung terhadap daya pembeda. Semakin baik efektivitas pengecoh dan semakin proporsional tingkat kesukaran soal, maka semakin tinggi pula kemampuan soal dalam membedakan peserta didik berdasarkan tingkat kompetensinya. Oleh karena itu, analisis kualitas butir soal sebaiknya dilakukan secara terpadu dengan mempertimbangkan ketiga indikator tersebut secara bersamaan.

SIMPULAN

Berdasarkan hasil kajian literatur yang telah dilakukan, dapat disimpulkan bahwa kualitas butir soal merupakan fondasi penting dalam menghasilkan evaluasi pembelajaran yang valid, reliabel, dan objektif. Tiga komponen utama yang menentukan kualitas suatu butir soal, yaitu efektivitas pengecoh, tingkat kesukaran, dan daya pembeda, memiliki peran yang saling berkaitan dalam mengukur kemampuan peserta didik secara tepat. Pengecoh yang efektif mampu menarik peserta didik yang belum memahami materi, tingkat kesukaran yang proporsional dapat menggambarkan kemampuan peserta didik secara lebih akurat, sedangkan daya pembeda yang baik mampu membedakan peserta didik berdasarkan tingkat kompetensinya.

Hasil kajian menunjukkan bahwa masih ditemukan berbagai kelemahan dalam penyusunan butir soal, seperti pengecoh yang tidak berfungsi, dominasi soal yang terlalu mudah atau terlalu sulit, serta daya pembeda yang rendah. Kondisi tersebut berpotensi menurunkan kualitas instrumen evaluasi dan menghasilkan informasi yang kurang akurat mengenai capaian belajar peserta didik. Oleh karena itu, analisis butir soal perlu dilakukan secara berkelanjutan baik sebelum maupun sesudah instrumen digunakan agar kualitas evaluasi pembelajaran dapat terus ditingkatkan.

Kajian ini menghasilkan sintesis konseptual yang menunjukkan bahwa efektivitas pengecoh, tingkat kesukaran, dan daya pembeda merupakan komponen yang saling berkaitan dalam menentukan kualitas butir soal. Efektivitas pengecoh berperan dalam membentuk tingkat kesukaran yang proporsional, sedangkan tingkat kesukaran yang ideal mendukung terciptanya daya pembeda yang optimal. Dengan demikian, kualitas instrumen evaluasi tidak dapat ditentukan hanya berdasarkan satu indikator, melainkan perlu dipahami melalui keterkaitan ketiga komponen tersebut secara menyeluruh.

Penelitian selanjutnya disarankan untuk melakukan analisis empiris terhadap butir soal menggunakan data hasil tes peserta didik secara langsung, sehingga diperoleh gambaran yang lebih komprehensif mengenai kualitas instrumen evaluasi dalam praktik pembelajaran. Selain itu, kajian dapat dikembangkan dengan memanfaatkan pendekatan analisis modern seperti Teori Respons Butir (Item Response Theory/IRT).

Penulis menyampaikan terima kasih kepada dosen pengampu mata kuliah Evaluasi Pembelajaran, pihak kampus STAIN Mandailing Natal, serta seluruh penulis dan peneliti yang karya-karyanya menjadi sumber rujukan dalam penyusunan artikel ini. Dukungan dan kontribusi berbagai pihak tersebut sangat membantu dalam penyelesaian penelitian ini.

REFERENSI

- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (3rd ed.). Bumi Aksara.
- Azwar, S. (1995). *Tes prestasi: Fungsi dan pengembangan pengukuran prestasi belajar*. Pustaka Pelajar.
- Brookhart, S. M., & Nitko, A. J. (2011). *Educational assessment of students* (6th ed.). Pearson Education.
- Dewi, N., Putra, I. K. A., & Astawa, I. N. T. (2019). Analisis kualitas butir soal pilihan ganda ditinjau dari tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. *Jurnal Pendidikan Indonesia*, 8(2), 120–129.
- Fatimah, L. U., & Alfath, K. (2019). Analisis kesukaran soal, daya pembeda, dan fungsi distraktor. *Al-Manar: Jurnal Komunikasi dan Pendidikan Islam*, 8(2), 37–64.
- Febriani, L., Safitri, R., Hamidah, C., Anisah, J., Ramadhana, S., Husein, A., & Al Fatih, F. (2026). MEMBANGUN INSTRUMEN EVALUASI YANG BERKUALITAS MELALUI ANALISIS TINGKAT KESUKARAN, DAYA BEDA, DAN EFEKTIVITAS PENGECOH SOAL. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 11(02), 237-251.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2013). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334. <https://doi.org/10.1207/S15324818AME15035>
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Prentice Hall.
- Lumbanraja, H. (2017). *Evaluasi hasil belajar dan analisis butir soal*. Deepublish.
- Magdalena, I., Fauzi, H. N., & Putri, R. (2021). Pentingnya analisis butir soal dalam meningkatkan kualitas evaluasi pembelajaran di sekolah dasar. *Jurnal Edukasi dan Sains*, 3(2), 245–257.
- Mahendra, I. W. E. (2019). Analisis kualitas butir soal pilihan ganda pada evaluasi pembelajaran. *Jurnal Pendidikan dan Pembelajaran*, 4(1), 55–63.
- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Nuha Medika.
- Nurhalimah, S., Rahmawati, D., & Hidayat, A. (2022). Analisis daya pembeda dan tingkat kesukaran butir soal sebagai upaya peningkatan kualitas evaluasi pembelajaran. *Jurnal Pendidikan Dasar*, 13(1), 45–56.
- Riinawati. (2021). *Evaluasi pendidikan: Teori dan praktik*. CV Budi Utama.
- Salsabilla, A., Rahman, M., & Fitriani, N. (2024). Analisis kualitas butir soal pilihan ganda berdasarkan tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. *Jurnal Pendidikan Indonesia*, 15(1), 1–12.
- Son, A. L. (2019). Instrumentasi kemampuan pemecahan masalah matematis: Analisis reliabilitas, validitas, tingkat kesukaran dan daya pembeda butir soal. *Gema Wiralodra*, 10(1), 41–52.
- Sudijono, A. (2017). *Pengantar evaluasi pendidikan*. PT RajaGrafindo Persada.
- Sumanti, S., & Susanti, E. (2019). *Evaluasi pendidikan*. Perdana Publishing.