

ANALISIS VALIDITAS DAN RELIABILITAS BUTIR SOAL DALAM EVALUASI PEMBELAJARAN

Fanazal Risqianto¹, Ma'mun Hanif², Angga Tri Ajie³, Heida Rahma Eksanti⁴,
Naila Zulfa Nabila⁵

Fakultas Tarbiyah dan Ilmu Keguruan, Universitas Islam Negeri K. H.
Abdurrahman Wahid Pekalongan

[1fanazal.risqianto24095@mhs.uingusdur.ac.id](mailto:fanazal.risqianto24095@mhs.uingusdur.ac.id), [2mamun.hanif63@gmail.com](mailto:mamun.hanif63@gmail.com),
[3angga.tri.ajie24086@mhs.uingusdur.ac.id](mailto:angga.tri.ajie24086@mhs.uingusdur.ac.id), [4heida.rahma.eksanti24984@mhs.uingusdur.ac.id](mailto:heida.rahma.eksanti24984@mhs.uingusdur.ac.id),
[5naila.zulfa.nabila24105@mhs.uingusdur.ac.id](mailto:naila.zulfa.nabila24105@mhs.uingusdur.ac.id)

ABSTRACT

Learning assessment plays a crucial role in determining the level of students' competency achievement and serves as the basis for educational decision-making. The quality of assessment results is greatly influenced by the quality of the instruments used; therefore, aspects of validity, reliability, and item analysis must be considered in the development of assessment instruments. This study aims to examine the concepts of validity, reliability, item analysis techniques, the relationship between validity and reliability, and the importance of item quality in learning assessment. The study employs a literature review (library research) method with a qualitative descriptive approach. Data were obtained through a review of various scientific sources, including journals, books, and articles relevant to the research topic. The results of the study indicate that validity and reliability are the two primary aspects that determine the quality of an assessment instrument. Furthermore, item analysis which encompasses validity, reliability, difficulty level, discriminative power, and the effectiveness of distractors can be used to objectively assess the quality of each test item. Consequently, instruments that meet these criteria are capable of producing data that is more accurate, consistent, and reliable. In conclusion, the quality of test items plays a crucial role in supporting the implementation of effective, objective, and education quality improvement oriented learning evaluations.

Keywords: learning assessment, validity, reliability, item analysis.

ABSTRAK

Evaluasi pembelajaran memiliki peran penting dalam menentukan tingkat pencapaian kompetensi peserta didik serta sebagai dasar dalam pengambilan keputusan pendidikan. Kualitas hasil evaluasi sangat dipengaruhi oleh kualitas instrumen yang digunakan, sehingga aspek validitas, reliabilitas, dan analisis butir soal perlu diperhatikan dalam penyusunan instrumen penilaian. Penelitian ini bertujuan untuk mengkaji konsep validitas, reliabilitas, teknik analisis butir soal, hubungan validitas dan reliabilitas, serta pentingnya kualitas butir soal dalam

evaluasi pembelajaran. Penelitian menggunakan metode studi kepustakaan (library research) dengan pendekatan deskriptif kualitatif. Data diperoleh melalui kajian berbagai sumber ilmiah berupa jurnal, buku, dan artikel yang relevan dengan topik penelitian. Hasil kajian menunjukkan bahwa validitas dan reliabilitas merupakan dua aspek utama yang menentukan kualitas instrumen evaluasi. Selain itu, analisis butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh dapat digunakan untuk menilai kualitas setiap butir soal secara objektif. Dengan demikian, instrumen yang memenuhi kriteria tersebut mampu menghasilkan data yang lebih akurat, konsisten, dan dapat dipercaya. Kesimpulannya, kualitas butir soal berperan penting dalam mendukung pelaksanaan evaluasi pembelajaran yang efektif, objektif, dan berorientasi pada peningkatan mutu pendidikan.

Kata Kunci: evaluasi pembelajaran, validitas, reliabilitas, analisis butir soal.

A. Pendahuluan

Evaluasi pembelajaran merupakan bagian yang tidak dapat dipisahkan dari proses pendidikan karena berfungsi untuk mengetahui tingkat ketercapaian tujuan pembelajaran yang telah ditetapkan. Melalui kegiatan evaluasi, guru dapat memperoleh informasi mengenai kemampuan, pemahaman, serta perkembangan belajar peserta didik. Hasil evaluasi juga dapat digunakan sebagai dasar untuk memperbaiki strategi pembelajaran, menentukan tindak lanjut, dan meningkatkan kualitas proses pendidikan secara keseluruhan. Oleh karena itu, instrumen yang digunakan dalam evaluasi harus mampu menghasilkan data yang akurat dan dapat dipercaya agar keputusan yang diambil

berdasarkan hasil evaluasi benar-benar sesuai dengan kondisi peserta didik. Sunaryati et al., (2024) menjelaskan bahwa kualitas evaluasi pembelajaran sangat dipengaruhi oleh kualitas instrumen yang digunakan karena instrumen menjadi alat utama dalam memperoleh informasi mengenai hasil belajar siswa.

Dalam praktiknya, masih ditemukan berbagai permasalahan terkait kualitas instrumen penilaian yang digunakan di sekolah. Tidak sedikit soal yang disusun tanpa melalui proses pengujian yang memadai sehingga belum diketahui tingkat validitas, reliabilitas, maupun karakteristik butir soalnya. Kondisi tersebut dapat menyebabkan hasil evaluasi kurang mampu menggambarkan kemampuan peserta

didik secara objektif. Menurut Buhori, (2024), kualitas suatu instrumen evaluasi perlu ditinjau dari beberapa aspek penting, seperti validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, karena aspek-aspek tersebut menentukan ketepatan hasil pengukuran yang diperoleh. Instrumen yang tidak memenuhi karakteristik tersebut berpotensi menghasilkan informasi yang kurang akurat dan dapat memengaruhi kualitas pengambilan keputusan dalam pembelajaran.

Upaya untuk menghasilkan instrumen yang berkualitas dapat dilakukan melalui analisis butir soal. Analisis ini bertujuan untuk mengetahui karakteristik setiap soal sehingga dapat ditentukan apakah suatu butir layak digunakan, perlu direvisi, atau harus diganti. Selain membantu meningkatkan kualitas instrumen, analisis butir soal juga dapat memberikan gambaran mengenai tingkat kesukaran, daya pembeda, dan efektivitas soal dalam mengukur kemampuan peserta didik. Penelitian yang dilakukan oleh Ikhsaniyah et al., (2024) menunjukkan bahwa analisis butir soal menjadi langkah penting dalam evaluasi pembelajaran karena mampu

mengidentifikasi kelemahan dan kelebihan setiap item soal secara lebih objektif. Dengan demikian, kualitas tes yang digunakan dalam proses penilaian dapat terus ditingkatkan sesuai dengan kebutuhan pembelajaran.

Berdasarkan pentingnya kualitas instrumen evaluasi, pembahasan mengenai validitas, reliabilitas, dan teknik analisis butir soal menjadi hal yang perlu dikaji secara lebih mendalam. Pemahaman yang baik mengenai ketiga aspek tersebut akan membantu pendidik dalam menyusun instrumen yang mampu mengukur kemampuan peserta didik secara tepat dan konsisten. Selain itu, hubungan antara validitas dan reliabilitas serta pentingnya kualitas butir soal dalam evaluasi pembelajaran juga perlu dipahami agar proses penilaian dapat memberikan informasi yang akurat, objektif, dan bermanfaat bagi peningkatan mutu pembelajaran. Oleh karena itu, artikel ini membahas konsep validitas, reliabilitas, teknik analisis butir soal, hubungan validitas dan reliabilitas, serta pentingnya kualitas butir soal dalam evaluasi pembelajaran sebagai dasar untuk

menghasilkan instrumen penilaian yang berkualitas.

B. Metode Penelitian

Penelitian ini menggunakan metode penelitian kepustakaan (library research) dengan pendekatan deskriptif kualitatif. Metode penelitian kepustakaan dilakukan dengan cara mengumpulkan, mengkaji, dan menganalisis berbagai sumber literatur yang relevan dengan topik penelitian. Sumber data yang digunakan dalam penelitian ini berupa artikel jurnal ilmiah, buku, prosiding, dan sumber akademik lainnya yang membahas mengenai validitas, reliabilitas, analisis butir soal, serta kualitas instrumen evaluasi pembelajaran. Teknik pengumpulan data dilakukan melalui studi dokumentasi dengan menelusuri berbagai referensi yang diperoleh dari jurnal nasional maupun sumber ilmiah lainnya yang berkaitan dengan tema penelitian. Literatur yang digunakan dipilih berdasarkan relevansi topik, kredibilitas sumber, serta keterbaruan publikasi agar informasi yang diperoleh sesuai dengan perkembangan kajian evaluasi pembelajaran.

C. Hasil Penelitian dan Pembahasan

1. Konsep dan Jenis Validitas Butir Soal

Validitas menjadi salah satu aspek dasar yang wajib dimiliki oleh instrumen penilaian karena berkaitan dengan tingkat ketepatan instrumen dalam mengukur konstruk atau kemampuan yang menjadi sasaran pengukuran. Instrumen yang memiliki validitas tinggi mampu memberikan data yang mencerminkan keadaan sebenarnya, sehingga hasil pengukuran dapat dijadikan dasar yang kuat dalam pengambilan keputusan. Dalam konteks evaluasi pembelajaran, validitas merupakan persyaratan penting yang harus dipastikan terlebih dahulu sebelum instrumen diterapkan kepada peserta didik (Utami et al., 2023). Oleh karena itu, proses penyusunan, pengembangan, dan pengujian butir soal perlu dilakukan dengan cermat agar setiap item yang digunakan benar-benar sesuai dengan kompetensi dan tujuan pembelajaran yang ingin diukur.

Sebagaimana yang disampaikan oleh (Saputri et al., 2023), bahwa dalam evaluasi pembelajaran, validitas berperan penting karena menentukan tingkat ketepatan suatu instrumen dalam mengukur kemampuan atau aspek yang menjadi tujuan pengukuran. Hasil yang diperoleh dari instrumen yang valid akan menggambarkan kondisi peserta didik secara lebih akurat, sehingga dapat digunakan sebagai dasar yang terpercaya dalam proses pengambilan keputusan pendidikan. Oleh karena itu, sebelum instrumen penilaian diterapkan, kualitas validitasnya perlu dibuktikan terlebih dahulu melalui proses pengembangan dan pengujian yang tepat. Setiap butir soal harus disusun berdasarkan indikator dan tujuan pembelajaran yang telah ditetapkan agar data yang dihasilkan benar-benar mencerminkan kompetensi yang ingin diukur.

Pengujian validitas butir soal bertujuan untuk menilai sejauh mana setiap item mampu merepresentasikan kompetensi yang ingin diukur dalam suatu tes.

Suatu butir dapat dikatakan valid apabila memberikan kontribusi yang sejalan dengan hasil tes secara keseluruhan, sehingga mampu mengidentifikasi perbedaan tingkat kemampuan peserta didik secara tepat. Informasi yang diperoleh dari analisis tersebut menjadi bahan pertimbangan dalam menentukan kualitas setiap item, baik untuk dipertahankan, diperbaiki, maupun dieliminasi. Dengan demikian, proses evaluasi instrumen dapat menghasilkan perangkat tes yang lebih akurat, efektif, dan sesuai dengan tujuan pengukuran yang telah ditetapkan (Hermanto et al., 2023).

Kajian mengenai validitas instrumen dalam evaluasi pembelajaran umumnya mencakup tiga bentuk utama, yaitu validitas isi, validitas konstruk, dan validitas kriteria. Validitas isi berkaitan dengan sejauh mana materi yang diujikan telah mencerminkan kompetensi serta indikator pembelajaran yang ditetapkan. Sementara itu, validitas konstruk menilai kesesuaian instrumen dengan konsep teoretis atau karakteristik

yang menjadi dasar pengukuran. Adapun validitas kriteria menekankan pada tingkat keterkaitan antara hasil tes dan ukuran lain yang dijadikan standar pembanding. Masing-masing jenis validitas memiliki fungsi yang berbeda, tetapi secara bersama-sama memberikan gambaran mengenai ketepatan dan kualitas instrumen yang digunakan dalam kegiatan penilaian (Ikhsaniyah et al., 2024a).

(I Gusti Ngurah Puger, 2021) menjelaskan bahwa penilaian validitas isi pada instrumen hasil belajar sering dilakukan dengan melibatkan sejumlah pakar yang memiliki keahlian sesuai bidang yang diukur. Dalam proses tersebut, digunakan dua indikator utama, yaitu Content Validity Ratio (CVR) dan Content Validity Index (CVI). Nilai CVR berfungsi untuk menyeleksi butir soal yang layak dipertahankan dalam penyusunan instrumen, di mana item dengan koefisien di atas nol umumnya dianggap memenuhi kriteria kelayakan. Setelah proses seleksi dilakukan, nilai CVI dihitung untuk menggambarkan tingkat kesepakatan para ahli terhadap

kualitas keseluruhan item yang digunakan. Dengan demikian, CVI tidak dapat ditentukan sebelum penetapan item yang lolos berdasarkan hasil CVR. Meskipun kedua indeks tersebut saling berkaitan dalam proses validasi, tingkat validitas isi suatu tes pada akhirnya lebih banyak direpresentasikan oleh nilai CVI karena indeks ini mencerminkan kualitas instrumen secara lebih menyeluruh.

Validitas konstruk menunjukkan tingkat ketepatan suatu instrumen dalam mengukur konsep, karakteristik, atau kemampuan yang menjadi fokus pengukuran. Dalam pengembangan butir soal, aspek ini mengharuskan setiap item disusun berdasarkan landasan teori yang relevan, indikator yang telah ditetapkan, serta kemampuan berpikir yang ingin dievaluasi. Penilaian terhadap validitas konstruk tidak hanya meninjau kesesuaian isi soal, tetapi juga mencakup kualitas penyajian, penggunaan bahasa, tingkat proses kognitif yang dituntut, dan keterhubungannya dengan kompetensi yang menjadi

target pembelajaran. Melalui pengujian tersebut, dapat dipastikan bahwa instrumen yang digunakan benar-benar merefleksikan dan mengukur konstruk yang dimaksud secara tepat (Sihotang et al., 2025).

2. Konsep dan Teknik Reliabilitas Butir Soal

Reliabilitas merupakan ukuran yang digunakan untuk mengetahui tingkat kestabilan dan konsistensi hasil yang diperoleh dari suatu instrumen pengukuran. Suatu instrumen dikatakan memiliki reliabilitas yang baik apabila mampu menghasilkan skor yang relatif serupa ketika diterapkan pada kondisi yang sebanding. Dalam evaluasi dan analisis instrumen asesmen, reliabilitas menjadi aspek penting karena mencerminkan tingkat kepercayaan terhadap data yang dihasilkan. Nilai koefisien reliabilitas yang tinggi menunjukkan bahwa hasil pengukuran cenderung lebih konsisten dan tidak banyak dipengaruhi oleh kesalahan pengukuran maupun faktor-faktor lain yang bersifat acak. Dengan demikian, reliabilitas menjadi salah

satu indikator utama dalam menentukan kualitas suatu instrumen penilaian (Saputri et al., 2023).

(Utami et al., 2023) menyampaikan bahwa tujuan utama pengujian reliabilitas adalah memastikan bahwa suatu instrumen mampu memberikan hasil yang dapat dipercaya ketika digunakan sebagai alat pengukuran. Reliabilitas mencerminkan sejauh mana instrumen mampu menghasilkan data yang konsisten apabila diterapkan berulang kali pada subjek atau kondisi yang sama. Tingkat reliabilitas yang tinggi menunjukkan bahwa skor yang diperoleh relatif stabil dari waktu ke waktu, sehingga hasil pengukuran memiliki tingkat kepercayaan yang lebih baik dan dapat dijadikan landasan yang kuat dalam proses penentuan maupun pengambilan keputusan. Dalam teori tes klasik, Coefficient of Internal Consistency (CIC) yang tersedia pada Moodle memiliki fungsi yang sepadan dengan koefisien Alpha Cronbach. Indeks ini digunakan untuk menilai tingkat konsistensi internal suatu instrumen dengan menganalisis

keterkaitan antarbutir soal yang menyusunnya. Nilai Alpha Cronbach yang semakin tinggi menunjukkan bahwa setiap item dalam tes memiliki keselarasan yang baik dalam mengukur konstruk yang sama. Oleh karena itu, koefisien yang tinggi umumnya mengindikasikan bahwa instrumen memiliki reliabilitas yang lebih baik, sehingga hasil pengukuran yang diperoleh cenderung lebih stabil dan dapat dipercaya (Suwanda, 2024).

Perhitungan reliabilitas dapat dilakukan menggunakan rumus Kuder-Richardson 20 (KR-20) pada instrumen yang berbentuk tes objektif dengan sistem penskoran dua kategori, seperti benar dan salah. Metode ini bertujuan untuk menilai konsistensi internal antaritem dalam suatu tes, sehingga dapat diketahui sejauh mana instrumen mampu menghasilkan data yang ajeg ketika digunakan sebagai alat ukur. Nilai koefisien KR-20 yang tinggi menunjukkan bahwa butir-butir soal memiliki keterkaitan yang baik dalam mengukur kemampuan yang sama. Dengan demikian, semakin besar nilai reliabilitas

yang diperoleh, semakin tinggi pula tingkat keyakinan terhadap ketepatan dan kestabilan hasil pengukuran yang dihasilkan oleh instrumen tersebut (Sunaryati Titin et al., 2024).

Sementara itu, (Gede Subhaktiyasa, 2024) turut menyampaikan tentang teknik Kuder-Richardson 21 (KR-21) yang merupakan salah satu metode yang digunakan untuk memperkirakan reliabilitas tes dengan sistem penskoran dikotomis, seperti jawaban benar dan salah. Metode ini dikembangkan sebagai bentuk penyederhanaan dari KR-20 karena proses perhitungannya tidak memerlukan informasi setiap butir soal secara rinci, melainkan cukup menggunakan jumlah item, nilai rata-rata tes, dan varians total skor. Penggunaan KR-21 lebih sesuai pada instrumen yang memiliki tingkat kesukaran antarbutir yang relatif seragam. Meskipun prosedurnya lebih sederhana, teknik ini tetap dapat memberikan gambaran mengenai tingkat konsistensi dan keajegan instrumen dalam menghasilkan

data pengukuran yang dapat dipercaya.

Teknik split-half atau metode belah dua digunakan untuk menilai reliabilitas internal suatu instrumen dengan cara memisahkan butir-butir soal ke dalam dua kelompok yang dianggap sebanding. Setelah pembagian dilakukan, hubungan antara skor kedua kelompok dianalisis melalui perhitungan korelasi. Pendekatan ini bertujuan untuk melihat sejauh mana seluruh bagian instrumen bekerja secara konsisten dalam mengukur konstruk atau variabel yang sama. Nilai korelasi yang tinggi menunjukkan bahwa kedua kelompok item memberikan hasil yang selaras, sehingga instrumen dapat dianggap memiliki tingkat reliabilitas yang memadai dan layak dimanfaatkan dalam kegiatan penelitian maupun pengukuran (Jago, 2019).

Berdasarkan pembahasan diatas, reliabilitas butir soal dapat disimpulkan sebagai suatu tingkat konsistensi suatu instrumen dalam menghasilkan data yang stabil dan dapat dipercaya. Pengujian reliabilitas penting dilakukan untuk memastikan bahwa instrumen

mampu mengukur kemampuan peserta didik secara konsisten dan bebas dari kesalahan pengukuran yang berlebihan. Reliabilitas dapat dianalisis lewat berbagai teknik, seperti Cronbach's Alpha, KR-20, KR-21, dan Split-Half, yang penggunaannya disesuaikan dengan karakteristik instrumen dan jenis data yang diperoleh. Oleh karena itu, instrumen yang memiliki koefisien reliabilitas tinggi dapat dianggap lebih layak digunakan dalam proses evaluasi pembelajaran karena dapat memberikan hasil pengukuran yang konsisten dan bisa dipertanggungjawabkan.

3. Teknik Analisis Butir Soal

Analisis butir soal merupakan salah satu tahapan penting dalam kegiatan evaluasi pembelajaran. Kegiatan ini dilakukan untuk mengetahui kualitas setiap butir soal yang digunakan dalam proses penilaian sehingga dapat dipastikan bahwa soal tersebut benar-benar mampu mengukur kemampuan peserta didik sesuai dengan tujuan pembelajaran yang telah ditetapkan. Melalui analisis butir soal, guru tidak hanya memperoleh informasi mengenai

hasil belajar siswa, tetapi juga dapat mengetahui kualitas instrumen yang digunakan dalam proses evaluasi. Menurut Suwanda, (2024), analisis butir soal dapat membantu mengidentifikasi karakteristik soal melalui pengujian validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Oleh karena itu, analisis butir soal menjadi bagian yang tidak dapat dipisahkan dari penyusunan instrumen evaluasi yang berkualitas.

Dalam dunia pendidikan, hasil evaluasi sering dijadikan dasar untuk mengambil berbagai keputusan, seperti menentukan tingkat pencapaian kompetensi siswa, melakukan perbaikan pembelajaran, maupun menyusun program tindak lanjut. Oleh sebab itu, instrumen yang digunakan harus memiliki kualitas yang baik agar data yang diperoleh benar-benar mencerminkan kemampuan peserta didik. Apabila soal yang digunakan kurang baik, maka hasil evaluasi yang diperoleh juga berpotensi kurang akurat. Rachmawati, (2025) menjelaskan bahwa analisis butir soal dapat memberikan informasi yang

objektif mengenai kualitas instrumen sehingga guru dapat melakukan revisi atau perbaikan terhadap soal yang masih memiliki kelemahan.

a) Validitas Soal

Validitas merupakan salah satu aspek utama yang harus diperhatikan dalam penyusunan instrumen evaluasi. Validitas menunjukkan sejauh mana suatu soal mampu mengukur kompetensi yang memang ingin diukur. Sebuah soal dapat dikatakan valid apabila isi, bentuk, dan tujuan pengukurannya sesuai dengan indikator pembelajaran yang telah ditetapkan sebelumnya. Dengan demikian, soal yang valid akan menghasilkan informasi yang sesuai dengan kemampuan yang dimiliki peserta didik.

Menurut Nafs et al., (2023), validitas menjadi indikator penting karena berkaitan langsung dengan ketepatan hasil pengukuran. Jika suatu soal tidak sesuai dengan materi atau tujuan pembelajaran, maka hasil yang diperoleh tidak dapat menggambarkan kemampuan peserta didik secara tepat. Oleh karena itu, sebelum soal digunakan dalam kegiatan

evaluasi, perlu dilakukan pengujian validitas untuk memastikan bahwa setiap butir soal telah memenuhi kriteria yang diharapkan. Semakin tinggi tingkat validitas suatu soal, semakin baik pula kualitas informasi yang dihasilkan dari proses penilaian tersebut.

b) Reliabilitas Soal

Selain valid, instrumen evaluasi juga harus memiliki reliabilitas yang baik. Reliabilitas berkaitan dengan tingkat konsistensi suatu instrumen dalam menghasilkan data pengukuran. Sebuah tes dikatakan reliabel apabila memberikan hasil yang relatif sama ketika digunakan pada kondisi yang serupa. Dengan kata lain, reliabilitas menunjukkan tingkat kepercayaan terhadap hasil yang diperoleh dari suatu instrumen penilaian.

Reliabilitas yang tinggi menunjukkan bahwa instrumen mampu memberikan hasil yang stabil dan tidak mudah dipengaruhi oleh faktor-faktor di luar kemampuan peserta didik. Hal ini penting karena hasil evaluasi yang tidak konsisten dapat menyebabkan kesalahan dalam

pengambilan keputusan. Oleh sebab itu, pengujian reliabilitas perlu dilakukan untuk memastikan bahwa instrumen yang digunakan benar-benar dapat dipercaya sebagai alat ukur kemampuan siswa. Semakin tinggi nilai reliabilitas suatu tes, maka semakin besar pula tingkat kepercayaan terhadap hasil evaluasi yang diperoleh (Sarassanti, 2024).

c) Tingkat Kesukaran Soal

Tingkat kesukaran merupakan aspek yang digunakan untuk mengetahui tingkat kemudahan atau kesulitan suatu soal bagi peserta didik. Analisis tingkat kesukaran dilakukan dengan melihat banyaknya siswa yang mampu menjawab soal dengan benar. Melalui analisis ini, penyusun tes dapat mengetahui apakah suatu soal termasuk kategori mudah, sedang, atau sulit.

Instrumen yang baik seharusnya memiliki variasi tingkat kesukaran sehingga dapat mengukur kemampuan peserta didik secara lebih menyeluruh. Soal yang terlalu mudah cenderung tidak mampu menunjukkan perbedaan

kemampuan siswa karena hampir semua peserta didik dapat menjawabnya dengan benar. Sebaliknya, soal yang terlalu sulit juga kurang efektif karena sebagian besar siswa mengalami kesulitan dalam menjawabnya. Oleh karena itu, diperlukan keseimbangan antara soal mudah, sedang, dan sulit agar hasil evaluasi dapat memberikan gambaran yang lebih akurat mengenai tingkat penguasaan materi peserta didik (Muldia, 2025).

d) Daya Pembeda Soal

Daya pembeda merupakan kemampuan suatu butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dengan peserta didik yang memiliki kemampuan rendah. Aspek ini sangat penting karena dapat menunjukkan efektivitas suatu soal dalam mengukur kemampuan siswa secara objektif. Soal yang memiliki daya pembeda yang baik biasanya lebih banyak dijawab benar oleh siswa yang memiliki tingkat penguasaan materi tinggi dibandingkan siswa yang memiliki kemampuan rendah.

Daya pembeda yang tinggi menunjukkan bahwa suatu soal mampu berfungsi sesuai dengan tujuan evaluasi. Sebaliknya, apabila suatu soal dijawab benar atau salah secara merata oleh seluruh peserta didik, maka soal tersebut kurang efektif dalam membedakan tingkat kemampuan siswa. Oleh sebab itu, analisis daya pembeda perlu dilakukan untuk memastikan bahwa setiap butir soal benar-benar mampu mengidentifikasi perbedaan kemampuan yang dimiliki peserta didik.

e) Efektivitas Pengecoh (Distraktor)

Pada soal pilihan ganda, efektivitas pengecoh atau distraktor juga menjadi salah satu aspek yang perlu dianalisis. Distraktor merupakan pilihan jawaban yang salah tetapi dibuat sedemikian rupa sehingga tampak logis dan dapat menarik perhatian peserta didik yang belum memahami materi dengan baik. Keberadaan distraktor yang baik dapat meningkatkan kualitas soal karena membantu mengidentifikasi tingkat pemahaman siswa terhadap

materi yang diujikan (Wahyu et al., 2025).

Distraktor yang efektif akan dipilih oleh sejumlah peserta didik, khususnya kelompok siswa yang belum menguasai materi. Sebaliknya, apabila suatu pilihan jawaban tidak pernah dipilih oleh peserta didik, maka pilihan tersebut dianggap kurang berfungsi sebagai pengecoh. Oleh karena itu, analisis efektivitas distraktor penting dilakukan agar setiap alternatif jawaban memiliki fungsi yang optimal dalam mendukung kualitas instrumen evaluasi. Dengan adanya distraktor yang baik, soal pilihan ganda dapat memberikan informasi yang lebih akurat mengenai kemampuan peserta didik.

4. Hubungan Validitas dan Reliabilitas dalam Evaluasi Pembelajaran

Validitas dan reliabilitas merupakan dua aspek yang sangat penting dalam evaluasi pembelajaran karena keduanya menentukan kualitas suatu instrumen penilaian. Validitas berkaitan dengan ketepatan instrumen dalam mengukur

kompetensi, pengetahuan, keterampilan, atau sikap yang menjadi tujuan pembelajaran, sedangkan reliabilitas berkaitan dengan konsistensi hasil pengukuran yang diperoleh ketika instrumen digunakan dalam kondisi yang relatif sama. Kedua aspek ini tidak dapat dipisahkan karena sama-sama berperan dalam menghasilkan data penilaian yang akurat dan dapat dipercaya.

Hubungan antara validitas dan reliabilitas dapat dipahami melalui prinsip bahwa suatu instrumen yang valid harus memiliki reliabilitas yang baik. Instrumen yang tidak konsisten dalam memberikan hasil pengukuran akan sulit menggambarkan kemampuan peserta didik secara tepat. Dengan kata lain, jika hasil tes berubah-ubah secara signifikan meskipun kemampuan peserta didik relatif sama, maka ketepatan pengukuran menjadi diragukan. Oleh sebab itu, reliabilitas menjadi salah satu prasyarat penting bagi tercapainya validitas suatu instrumen (Matondang, 2009).

Namun demikian, reliabilitas yang tinggi tidak selalu menunjukkan bahwa suatu instrumen telah valid. Sebuah tes dapat memberikan hasil yang konsisten setiap kali digunakan, tetapi belum tentu mengukur kompetensi yang seharusnya diukur. Misalnya, guru ingin mengukur kemampuan berpikir kritis siswa, tetapi soal yang diberikan hanya menuntut kemampuan menghafal. Meskipun hasil tes tersebut konsisten dan reliabel, instrumen tersebut tetap tidak valid karena tidak sesuai dengan tujuan pengukuran. Hal ini menunjukkan bahwa validitas memiliki cakupan yang lebih luas dibandingkan reliabilitas karena tidak hanya menuntut konsistensi hasil, tetapi juga ketepatan sasaran pengukuran (Subhaktiyasa, 2024).

Dalam praktik evaluasi pembelajaran, validitas dan reliabilitas saling melengkapi untuk menjamin kualitas hasil penilaian. Instrumen yang valid dan reliabel akan menghasilkan informasi yang akurat mengenai tingkat pencapaian kompetensi peserta didik. Informasi tersebut kemudian

dapat digunakan oleh guru untuk menentukan nilai, memberikan umpan balik, memperbaiki proses pembelajaran, serta mengambil keputusan pendidikan secara objektif. Sebaliknya, apabila salah satu aspek tidak terpenuhi, maka hasil evaluasi berpotensi menimbulkan kesalahan dalam menilai kemampuan peserta didik dan mengurangi keakuratan keputusan yang diambil (Fernando et al., 2023).

Oleh karena itu, dalam penyusunan instrumen evaluasi pembelajaran, guru perlu memastikan bahwa setiap butir soal telah sesuai dengan indikator pembelajaran (valid) dan mampu menghasilkan skor yang stabil serta konsisten (reliabel). Dengan terpenuhinya kedua aspek tersebut, evaluasi pembelajaran dapat berfungsi secara optimal sebagai alat untuk mengukur keberhasilan proses pendidikan dan meningkatkan kualitas pembelajaran.

5. Pentingnya Kualitas Butir Soal dalam Evaluasi Pembelajaran

Evaluasi pembelajaran merupakan salah satu komponen penting dalam proses pendidikan. Melalui evaluasi, guru dapat

mengetahui tingkat pencapaian kompetensi peserta didik setelah mengikuti kegiatan pembelajaran. Hasil evaluasi tidak hanya digunakan untuk menentukan nilai, tetapi juga sebagai dasar dalam memperbaiki proses pembelajaran agar lebih efektif. Oleh karena itu, instrumen evaluasi yang digunakan harus memiliki kualitas yang baik, terutama pada aspek butir soal. Dalam evaluasi pembelajaran, kualitas butir soal sangat penting untuk memastikan bahwa tes yang diberikan dapat mengukur pemahaman dan penguasaan materi secara akurat. Butir soal yang baik harus dapat menguji berbagai tingkat kognitif, seperti pengetahuan, pemahaman, aplikasi, analisis, sintesis, dan evaluasi. Selain itu, butir soal juga harus dirancang dengan jelas dan mudah dipahami oleh siswa agar dapat menghasilkan data yang valid dan reliabel. Oleh karena itu, perlu adanya perhatian khusus dalam pembuatan butir soal agar dapat memberikan informasi yang berguna dalam proses evaluasi pembelajaran. Butir soal yang kurang berkualitas dapat menyebabkan tes yang diberikan tidak mampu mengukur secara tepat kemampuan siswa. Oleh karena itu, guru perlu memperhatikan setiap detail dalam merancang butir soal agar dapat mencapai tujuan evaluasi yang diinginkan (Farihan, 2024). Dengan begitu, hasil tes yang diperoleh akan memberikan informasi yang akurat tentang pemahaman dan penguasaan materi siswa, sehingga dapat digunakan untuk meningkatkan kualitas pembelajaran di masa depan.

Pengaruh kualitas butir soal terhadap hasil evaluasi pembelajaran sangat signifikan. Selain itu, butir soal yang baik juga dapat memotivasi siswa untuk belajar dengan lebih serius dan fokus. Dengan adanya butir soal yang berkualitas, proses evaluasi pembelajaran dapat berjalan dengan lebih efektif dan efisien. Selain itu, guru juga perlu memperhatikan variasi jenis soal yang digunakan agar dapat mengukur berbagai aspek kemampuan siswa secara komprehensif. Dengan demikian, evaluasi pembelajaran dapat memberikan gambaran yang lebih lengkap tentang kemajuan belajar siswa. Butir soal yang terstruktur dengan baik juga dapat membantu guru dalam mengevaluasi pemahaman siswa terhadap materi pembelajaran. Dengan menggunakan variasi jenis soal seperti pilihan ganda, essay, dan ujian praktik, guru dapat memastikan bahwa semua aspek kemampuan siswa terukur dengan baik. Hal ini juga dapat membantu siswa dalam mengembangkan kemampuan analisis, sintesis, dan evaluasi dalam proses pembelajaran. Dengan demikian, evaluasi pembelajaran bukan hanya sebagai alat pengukur hasil belajar, tetapi juga sebagai sarana pengembangan kemampuan siswa secara menyeluruh (Phafiandita & Permadani, 2022).

Ada beberapa faktor yang dapat mempengaruhi kualitas butir soal dalam pembelajaran. Salah satunya adalah kejelasan dalam penyusunan soal, dimana butir soal

harus dirumuskan dengan jelas dan mudah dipahami oleh siswa. Selain itu, keberagaman jenis soal juga dapat mempengaruhi kualitas butir soal, karena dengan variasi soal, siswa dapat menguji kemampuan mereka secara menyeluruh. Faktor lain yang tidak kalah penting adalah tingkat kesulitan soal, dimana soal harus disesuaikan dengan tingkat kemampuan siswa agar dapat mengukur pemahaman mereka secara akurat. Oleh karena itu, guru harus memperhatikan dengan seksama setiap butir soal yang disusun untuk memastikan bahwa mereka memenuhi standar kualitas yang diinginkan. Selain itu, proses pembuatan soal juga harus dilakukan dengan teliti dan hati-hati agar tidak terjadi kesalahan atau kekurangan informasi yang dapat membingungkan siswa. Dengan demikian, kualitas butir soal akan menjadi faktor yang sangat penting dalam menentukan keberhasilan pembelajaran siswa (Muluki & Bundu, 2020). Sebagai contoh, dalam menyusun soal ujian matematika, guru harus mempertimbangkan tingkat kesulitan dari setiap soal agar sesuai dengan kemampuan siswa dan dapat mengukur pemahaman mereka secara akurat. Guru juga perlu teliti dalam mengecek setiap butir soal agar tidak terjadi kesalahan atau kekurangan informasi yang dapat membingungkan siswa, sehingga kualitas butir soal dapat terjaga dengan baik.

Strategi Meningkatkan Kualitas
Butir Soal dalam Evaluasi

Pembelajaran salah satu strategi yang dapat digunakan untuk meningkatkan kualitas butir soal dalam evaluasi pembelajaran adalah dengan melibatkan guru-guru dalam proses pengembangan soal. Dengan melibatkan para guru, maka dapat memastikan bahwa soal-soal yang disusun benar-benar sesuai dengan materi pembelajaran yang telah diajarkan. Selain itu, melakukan uji validitas dan reliabilitas soal secara berkala juga dapat membantu dalam meningkatkan kualitas butir soal (Ahmad et al., 2025). Dengan demikian, proses evaluasi pembelajaran dapat berjalan dengan lebih efektif dan akurat. Dengan melibatkan guru-guru dalam pengembangan soal, mereka dapat memberikan masukan dan saran untuk memperbaiki butir soal yang kurang jelas atau tidak sesuai dengan tujuan pembelajaran. Selain itu, uji validitas dan reliabilitas yang dilakukan secara berkala akan memastikan bahwa soal-soal yang digunakan benar-benar dapat mengukur pemahaman dan penguasaan siswa terhadap materi pembelajaran. Dengan adanya proses evaluasi yang efektif dan akurat, guru dapat mengetahui sejauh mana pencapaian siswa dalam pembelajaran dan dapat melakukan perbaikan yang diperlukan untuk meningkatkan kualitas pembelajaran di kelas.

Berdasarkan uraian tersebut, dapat disimpulkan bahwa kualitas butir soal memiliki peran yang sangat penting dalam evaluasi pembelajaran.

Butir soal yang memenuhi aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh akan menghasilkan informasi yang akurat mengenai kemampuan peserta didik.

Dengan demikian, hasil evaluasi dapat digunakan sebagai dasar yang tepat dalam pengambilan keputusan pendidikan dan perbaikan proses pembelajaran. Merangkum pentingnya kualitas butir soal dalam evaluasi pembelajaran dan bagaimana strategi yang dapat dilakukan untuk meningkatkannya. Dengan meningkatkan keakuratan dalam mengukur pemahaman siswa, memudahkan guru dalam mengevaluasi efektivitas pembelajaran, dan mendorong siswa untuk belajar secara aktif dan terarah, evaluasi pembelajaran dapat menjadi alat yang lebih efektif dalam meningkatkan kualitas pendidikan. Selain itu, dengan melakukan validasi dan revisi butir soal secara berkala, menggunakan variasi jenis soal, dan melibatkan siswa dalam proses pembuatan soal, kualitas evaluasi pembelajaran dapat terus ditingkatkan untuk mendukung proses pembelajaran yang lebih baik. Dengan demikian, evaluasi pembelajaran dapat membantu menciptakan lingkungan pembelajaran yang lebih efektif dan berorientasi pada hasil (Asrori, 2025). Selain itu, Dengan terus meningkatkan kualitas evaluasi pembelajaran, sekolah dan guru dapat memberikan pengalaman belajar yang lebih bermakna dan mendukung perkembangan siswa secara holistik.

E. Kesimpulan

Selain validitas dan reliabilitas, kualitas instrumen juga dapat diketahui melalui analisis butir soal yang mencakup tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Analisis tersebut membantu guru dalam mengidentifikasi kelemahan maupun kelebihan setiap butir soal sehingga instrumen yang digunakan dapat diperbaiki dan disempurnakan. Melalui proses ini, soal yang digunakan dalam evaluasi tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai sarana untuk meningkatkan kualitas pembelajaran.

Hubungan antara validitas dan reliabilitas menunjukkan bahwa keduanya saling melengkapi dalam menghasilkan instrumen yang berkualitas. Instrumen yang valid harus didukung oleh reliabilitas yang baik agar hasil pengukurannya dapat dipercaya. Sebaliknya, reliabilitas yang tinggi belum tentu menjamin validitas instrumen apabila instrumen tersebut tidak mengukur kompetensi yang seharusnya diukur. Oleh sebab itu, pengujian kedua aspek ini perlu dilakukan secara bersamaan sebelum

instrumen digunakan dalam kegiatan evaluasi.

Pada akhirnya, kualitas butir soal memiliki peran yang sangat penting dalam menentukan ketepatan hasil evaluasi pembelajaran. Butir soal yang disusun dengan baik akan menghasilkan informasi yang lebih akurat mengenai kemampuan peserta didik, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan pendidikan, perbaikan proses pembelajaran, serta peningkatan mutu pendidikan secara berkelanjutan.

DAFTAR PUSTAKA

- Asrori, M. N. (2025). Intermediasi keuangan syariah: Sebuah eksplorasi tentang relasi antara bank, nasabah, dan regulasi, 3, 1830-1835.
- Buhori. (2024). Kualitas instrumen evaluasi pembelajaran, 3(1).
- Farihan, A. (2024). Implementasi guru dalam merancang butir kognitif di MTS An-Najiyah Duwi. Muaddib, 2(2), 422-429.
- Gede Subhaktiyasa, P. (2024). Evaluasi validitas dan reliabilitas instrumen penelitian kuantitatif: Sebuah studi pustaka. Journal of Education Research, 5(4), 5599-5609.
- Hermanto, F. Y., Anggraeni, A. D., & Sutirman. (2023). Analisis butir soal dengan classical test theory untuk mengukur kemampuan kognitif siswa. Efisiensi: Kajian Ilmu Administrasi, 18(1), 159-170.
- I Gusti Ngurah Puger. (2021). Pengujian validitas isi tes hasil belajar yang dinilai oleh subject matter expert (SME). Daiwi Widya Jurnal Pendidikan, 8(3), 1-15.
- Ikhsaniyah, S. N., Kurnia, A. D., Zuroida, M., Pratiwi, V., & Hakim, L. (2024). Analisis butir soal perpajakan PPh Pasal 21 menggunakan software Anates pada pendekatan teori tes klasik, 12(2), 77-88.
- Jago, N. (2019). Uji reliabilitas instrumen tingkat kegemaran membaca masyarakat umum di Provinsi Bali tahun 2018 dengan teknik belah dua (split half). Acarya Pustaka (Jurnal Ilmiah Perpustakaan dan Informasi), 6(2), 104-128.
- Muldia, H. (2025). Analisis butir soal berbasis level kognitif pada ujian akhir semester mata pelajaran PPKn kelas VII SMP Negeri 2 Pemangkat. Jurnal Pendidikan Kewarganegaraan, 9(2), 438-451.
- Muluki, A., & Bundu, P. (2020). Analisis kualitas butir tes semester ganjil mata pelajaran IPA kelas IV MI Radhiatul Adawiyah. 4(1), 86-96.
- Nafs, H., Sridana, N., Hikmah, N., & Soeprianto, H. (2023). Analisis kualitas butir soal ulangan akhir semester genap mata pelajaran matematika kelas VII

- SMPN 6 Mataram tahun ajaran. 8, 2324-2331.
- Phafiandita, A. N., & Permadani, A. (2022). Urgensi evaluasi pembelajaran di kelas. 3(2), 111-121.
- Rachmawati, D. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. 13(3), 273-284.
- Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). Analisis instrumen asesmen: Validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal. Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri, 9(5), 2986-2995.
- Sarassanti, Y. (2024). Tantangan dan peluang dalam mengoptimalkan peran artificial intelligence (AI) untuk menghasilkan karya ilmiah berkualitas menuju masa depan riset Indonesia maju 2045. Dalam Prosiding Seminar Nasional Penelitian dan Pengabdian kepada Masyarakat (SNPP) Tahun 2024 (hlm. 175-181).
- Sihotang, A. E., Hutagalung, S. H., Saragih, S., Akbar, R. F., & Hutagalung, S. M. (2025). Analisis validitas isi dan konstruk pada soal ujian akhir semester Bahasa Indonesia kelas VIII SMP Harapan 1 Medan. JKIP: Jurnal Kajian Ilmu Pendidikan, 6(2), 536-544.
- Sunaryati, T., Meilania, D. K., Lestari, F., & Aliifah, S. N. (2024). Analisis instrumen tes dan non tes dalam evaluasi pembelajaran. 8, 30461-30472.
- Suwanda, P. (2024). Analisis tes dan butir soal pada Moodle. Jurnal Evaluasi Pendidikan, 15(1), 33-42.
- Utami, Y., Rasmanna, P. M., & Khairunnisa. (2023). Uji validitas dan uji instrumen penilaian kerja dosen. SAINTEK (Jurnal Sains dan Teknologi), 4(2), 21-24.
- Wahyu, G., Pidekso, S., Pandu, G., & Ramadhan, S. (2025). Analysis of item characteristics of phase F accounting using software Anates. 84-95.