

ANALISIS KUALITAS BUTIR SOAL: KAJIAN VALIDITAS, RELIABILITAS, DAN KEPRAKTISAN DALAM EVALUASI PEMBELAJARAN

Debora Try Emi Pakpahan¹, Ida Bagus Putrayasa²
debora@student.undiksha.ac.id¹, ib.putrayasa@undiksha.ac.id²
Universitas Pendidikan Ganesha

ABSTRAK

Analisis butir soal merupakan salah satu langkah penting dalam menjamin kualitas instrumen evaluasi pembelajaran. Instrumen yang berkualitas harus memenuhi aspek validitas, reliabilitas, dan kepraktisan agar mampu menghasilkan data yang akurat, konsisten, dan mudah digunakan. Artikel ini bertujuan untuk mendeskripsikan konsep analisis butir soal serta mengkaji validitas, reliabilitas, dan kepraktisan sebagai indikator kualitas instrumen evaluasi pembelajaran. Penelitian ini menggunakan metode studi kepustakaan (library research) dengan memanfaatkan berbagai sumber berupa buku dan artikel ilmiah yang relevan. Hasil kajian menunjukkan bahwa analisis butir soal berfungsi untuk menilai kualitas instrumen serta menentukan kelayakan penggunaannya dalam proses evaluasi. Validitas berperan dalam memastikan ketepatan pengukuran, reliabilitas menunjukkan konsistensi hasil pengukuran, sedangkan kepraktisan berkaitan dengan kemudahan penggunaan instrumen dalam pelaksanaan evaluasi. Ketiga aspek tersebut saling melengkapi dalam menghasilkan instrumen evaluasi yang berkualitas. Oleh karena itu, pendidik perlu melakukan analisis instrumen secara sistematis sebelum menggunakan tes sebagai alat evaluasi pembelajaran.

Kata Kunci: Analisis Butir Soal, Validitas, Reliabilitas, Kepraktisan, Evaluasi Pembelajaran.

ABSTRACT

Item analysis is an important step in ensuring the quality of educational assessment instruments. A high-quality instrument must fulfill the aspects of validity, reliability, and practicality in order to produce accurate, consistent, and applicable data. This article aims to describe the concept of item analysis and examine validity, reliability, and practicality as indicators of assessment instrument quality. The study employed a library research method by reviewing relevant books and scientific journal articles. The findings indicate that item analysis functions as a procedure for evaluating instrument quality and determining its feasibility for use in educational assessment. Validity ensures the accuracy of measurement, reliability reflects the consistency of measurement results, while practicality relates to the ease of instrument implementation. These three aspects complement one another in producing high-quality assessment instruments. Therefore, educators need to conduct systematic instrument analysis before using tests as tools for educational assessment.

Keywords: Item Analysis, Validity Reliability, Practicality, Educational Assessment.

PENDAHULUAN

Evaluasi merupakan bagian yang tidak dapat dipisahkan dari proses pembelajaran. Melalui evaluasi, pendidik dapat mengetahui tingkat pencapaian kompetensi peserta didik serta memperoleh informasi yang diperlukan untuk memperbaiki proses pembelajaran. Agar hasil evaluasi dapat dipercaya, diperlukan instrumen yang memiliki kualitas baik. Salah satu cara untuk mengetahui kualitas instrumen adalah melalui analisis butir soal. Analisis butir soal merupakan kegiatan yang dilakukan untuk mengkaji kualitas setiap butir soal sehingga dapat diketahui apakah suatu soal layak digunakan, perlu direvisi, atau harus dibuang.

Namun, pada realitasnya di lapangan, kegiatan analisis butir soal sering kali diabaikan oleh pendidik. Banyak tes hasil belajar disusun secara tergesa-gesa tanpa melalui uji validitas dan reliabilitas yang sistematis, sehingga data pencapaian belajar yang dihasilkan menjadi kurang akurat dan bias. Terlebih lagi dalam pembelajaran Bahasa Indonesia, instrumen evaluasi dituntut untuk mampu mengukur berbagai keterampilan berbahasa

secara tepat dan komprehensif, meliputi keterampilan menyimak, berbicara, membaca, memirsa, dan menulis. Karakteristik materi Bahasa Indonesia yang kaya akan teks dan konteks memerlukan kecermatan tinggi dalam penyusunan soal agar tidak menimbulkan takwil ganda bagi peserta didik. Ketidaksesuaian antara kompetensi atau tujuan pembelajaran dengan instrumen yang digunakan dapat menyebabkan hasil evaluasi gagal menggambarkan kemampuan berbahasa peserta didik yang sebenarnya.

Instrumen yang berkualitas tidak hanya ditentukan oleh isi soal, tetapi juga oleh validitas, reliabilitas, dan kepraktisan instrumen tersebut. Validitas menunjukkan ketepatan instrumen dalam mengukur kompetensi yang hendak diukur. Reliabilitas menunjukkan konsistensi hasil pengukuran, sedangkan kepraktisan berkaitan dengan kemudahan penggunaan instrumen dalam pelaksanaan evaluasi. Ketiga aspek tersebut menjadi indikator penting dalam menentukan kualitas suatu instrumen evaluasi pembelajaran.

Berdasarkan pentingnya ketiga aspek tersebut, artikel ini bertujuan untuk mengkaji analisis butir soal serta menjelaskan validitas, reliabilitas, dan kepraktisan sebagai indikator kualitas instrumen evaluasi pembelajaran, khususnya dalam konteks pendidikan bahasa.

METODE PENELITIAN

Penelitian ini menggunakan metode studi kepustakaan (library research). Data diperoleh melalui kajian terhadap berbagai sumber pustaka yang relevan dengan topik analisis butir soal, validitas, reliabilitas, dan kepraktisan instrumen evaluasi. Sumber data terdiri atas buku-buku evaluasi pendidikan dan artikel ilmiah yang berkaitan dengan kualitas instrumen pembelajaran.

Teknik pengumpulan data dilakukan secara sistematis melalui tiga tahapan utama yang saling berkesinambungan, yaitu identifikasi, klasifikasi, dan telaah pustaka. Pada tahap identifikasi, pelacakan sumber pustaka dilakukan secara digital pada database Google Scholar menggunakan kombinasi kata kunci seperti analisis butir soal, validitas instrumen, reliabilitas tes, dan evaluasi pembelajaran Bahasa Indonesia. Selanjutnya, pada tahap klasifikasi dilakukan pemilahan literatur berdasarkan kriteria inklusi, yakni kesesuaian substansi teori dengan topik bahasan serta kebaruan publikasi untuk menjamin aktualitas data. Proses pengumpulan data ini diakhiri dengan tahap telaah mendalam untuk membaca, mencatat, dan membandingkan konsep-konsep dari berbagai ahli mengenai indikator kualitas instrumen.

Setelah data dari berbagai literatur terkumpul, tahap berikutnya adalah melakukan analisis data menggunakan teknik analisis isi (content analysis) dan analisis deskriptif kualitatif. Langkah ini dilaksanakan dengan cara mereduksi data, mengorganisasikan informasi ke dalam tema-tema teoretis yang relevan, serta menarik kesimpulan yang mendalam. Melalui pendekatan deskriptif kualitatif ini, peneliti dapat menyajikan pemahaman yang komprehensif dan utuh mengenai konsep analisis butir soal serta pemenuhan aspek validitas, reliabilitas, dan kepraktisan sebagai indikator kualitas instrumen evaluasi pembelajaran.

HASIL DAN PEMBAHASAN

Analisis Butir Soal

a. Pengertian Analisis Butir Soal

Menurut Arikunto, “analisis soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun”. Nana Sudjana menambahkan bahwa, Analisis butir soal atau analisis item adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai. Analisis soal juga dilakukan untuk mengetahui

berfungsi tidaknya sebuah soal. Analisis pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif (qualitatif control) dan analisis kuantitatif (quantitatif control). Analisis kualitatif sering pula dinamakan sebagai validitas logis (logical validity) yang dilakukan sebelum soal digunakan untuk melihat berfungsi tidaknya sebuah soal. Analisis soal secara kuantitatif sering pula dinamakan sebagai validitas empiris (empirical validity) yang dilakukan untuk melihat lebih berfungsi tidaknya sebuah soal, setelah itu diujicobakan kepada sampel yang representatif. Dari beberapa pendapat para ahli diatas, penulis menyimpulkan bahwa analisis butir soal adalah sebuah proses yang harus ditempuh dan dilaksanakan oleh seorang guru untuk mengetahui sejauh mana kualitas butir soal yang diberikan dan dengan adanya kegiatan analisis ini bisa membantu para guru terkait butir soal yang bagus dan yang layak untuk dipertahankan dan butir soal yang harus dibuang.b.Fungsi analisis butir soal

b. Fungsi Analisis Butir Soal

Arikunto (2006:205) dalam jurnal Analisis Butir Soal Pilihan Ganda Dari Aspek Kebahasaan mengatakan bahwa fungsi analisis soal yaitu: membantu kita dalam mengidentifikasi butir-butir soal yang jelek, memperoleh informasi yang akan dapat digunakan untuk menyempurnakan soal-soal untuk kepentinganlanjut, memperoleh gambaran secara selintas tentang keadaan yang kita susun.

c. Tujuan Analisis Butir Soal

Tujuan analisis butir soal menurut Thorndik dan Hagn (1977) sebagaimana dikutip purwanto (2001:118)dalam jurnal Analisis Butir Soal Pilihan Ganda dari Aspek Kebahasaanadalah sebagai berikut:1)Jawaban-jawaban soal itu merupakan informasi diagnostik untuk meneliti pelajaran dari kelas itu dan kegagalan-kegagalan belajarnya, serta selanjutnya untuk membimbing belajar yang lebih baik.2)Jawaban-jawaban terhadap soal yang terpisah dan perbaikan soal-soal didasarkan atas jawaban-jawaban itu merupakan basis penyiapan tes-tes yang lebih baik untuk tahun berikutnya.

Mulyasa (2022) menegaskan bahwa analisis instrumen merupakan bagian penting dalam proses evaluasi pembelajaran karena mampu memberikan informasi mengenai kualitas alat ukur yang digunakan. Melalui analisis yang sistematis, pendidik dapat mengidentifikasi kelemahan instrumen, melakukan perbaikan terhadap butir soal yang kurang efektif, serta meningkatkan kualitas evaluasi pada pelaksanaan berikutnya. Oleh karena itu, analisis butir soal tidak hanya berfungsi sebagai kegiatan penilaian, tetapi juga sebagai upaya peningkatan mutu pembelajaran secara berkelanjutan.

Validitas

Validitas adalah sejauh mana tes itu mengukur apa yang dimaksudkan untuk diukur. Ada dua unsur penting dalam validitas ini. Pertama, validitas menunjukkan suatu derajat, ada yang sempurna, ada yang sedang, dan ada pula yang rendah. Kedua validitas selalu dihubungkan dengan suatu putusan atau tujuan yang spesifik. Sebagaimana pendapat R.L. Thorndike dan H.P. Hagen dalam Arifin (2010) bahwa “*validity is always in relation to a specific decision or use*”. Sementara itu, Gronlund dalam Arifin (2010) mengemukakan ada tiga faktor yang mempengaruhi validitas hasil tes, yaitu “faktor instrumen evaluasi, faktor administrasi evaluasi dan penskoran, dan faktor dari jawaban peserta didik.”

- **Faktor instrumen evaluasi** Dalam mengembangkan instrumen evaluasi, seorang evaluator harus memperhatikan hal hal yang mempengaruhi validitas instrumen dan berkaitan dengan prosedur penyusunan instrumen, seperti silabus kisi kisi soal petunjuk mengerjakan soal dan pengisian lembar soal dan pengisian lembar jawaban, kunci jawaban, penggunaan kalimat efektif, bentuk alternatif jawaban, tingkat kesukaran, daya pembeda dan sebagainya.

- **Faktor administrasi evaluasi dan penskoran** Dalam administrasi evaluasi dan penskoran, banyak sekali penyimpangan atau kekeliruan, seperti alokasi waktu untuk pengerjaan soal yang tidak proporsional, memberikan bantuan kesalahan penskoran, termasuk kondisi fisik dan psikis peserta didik yang kurang menguntungkan.
- **Faktor jawaban dari peserta didik** Dalam praktiknya, faktor jawaban peserta didik justru lebih banyak berpengaruh dari pada faktor sebelumnya. Faktor ini meliputi kecenderungan peserta didik untuk menjawab secara tepat, tapi tidak tepat, keinginan melakukan coba-coba, dan penggunaan gaya bahasa tertentu dalam menjawab soal bentuk uraian.

Beberapa jenis validitas adalah sebagai berikut.

a. Validitas tampilan

Validitas tampilan lebih berhubungan dengan bagaimana tanggapan dari pihak awam mengenai alat ukur tersebut. Bila kita mengacu ke konsep validitas yang telah dibahas sebelumnya, validitas tipe ini kurang tepat dikategorikan ke dalam tipe validitas karena makna dari validitas tampilan tidak terkait dengan kemampuan mengukur dari suatu alat ukur. Ada kemungkinan validitas tipe ini tidak terlalu ilmiah dan hanya berdasarkan kebiasaan yang ada, misalnya format penyusunan pilihan-pilihan dalam soal pilihan ganda.

Seiring dengan perkembangan teori belajar bahasa asing, validitas tampilan mengalami perkembangan dari waktu ke waktu. Misalnya, pada awalnya orang awam beranggapan bahwa penggunaan bahasa itu secara lisan sehingga alat ukurnya pun selalu dilaksanakan secara lisan. Waktu itu sementara orang mungkin beranggapan bahwa kemampuan berkomunikasi dalam bahasa Arab yang diukur secara tertulis dianggap tidak valid. Namun, akhir-akhir ini tes untuk mengukur kemampuan bahasa Arab dalam bentuk tulisan sudah mulai populer.

b. Validitas Isi

Validitas isi digunakan untuk mengukur sejauh mana tes mencerminkan apa yang akan diukur dari kemampuan siswa sehubungan dengan tujuan pembelajaran yang ingin dicapai. Validitas isi mengukur lebih pada ranah kognitif siswa seperti yang tercantum dalam kurikulum. Penilaian harus dicocokkan dengan Tujuan Instruksional Khusus (ITK). Karena itu isi tes hendaknya sesuai dengan pokok-pokok bahasan yang diberikan kepada siswa. Cara penilaian validitas isi dapat dilakukan oleh beberapa judges kalau bias 3 atau 5 orang, usahakan jangan meminta judges dalam jumlah genap agar tidak sulit menarik kesimpulan jika ada dua pendapat yang berlainan bias ada penengah.

c. Validitas prediktif

Validitas ini menggambarkan sejauh mana hasil tes dari suatu alat ukur mempunyai korelasi dengan suatu keberhasilan di masa mendatang. Dengan kata lain, suatu alat ukur yang mempunyai validitas prediktif dapat digunakan untuk memprediksi apakah seseorang akan lebih berhasil atau kurang berhasil dalam belajar sesuatu. Dalam pembelajaran bahasa arab suatu keberhasilan seseorang dapat diprediksi oleh suatu alat ukur walaupun belum terjadi proses belajar. Sebagai contoh, keberhasilan belajar bahasa arab mahasiswa dapat diprediksi dengan penggunaan strategi belajar.

Kemampuan alat ukur untuk memprediksi dapat dibuktikan setelah terjadi suatu proses belajar. Waktu yang digunakan untuk membuktikan kemampuan memprediksi bisa dalam waktu yang singkat dan bisa terbukti dalam waktu yang lama hingga 5 tahun. Misalnya, masuk perguruan tinggi islam di luar negeri atau beberapa perguruan tinggi islam di indonesia mesyaratkan kemampuan berbahasa arab dengan nilai tertentu

yang diperoleh dari TOAFL (Test Of Arabic as a Foreign Language/ ikhtibar al-lughah al-'arabiyyah ka al-lughah al-ajnabiyyah). Asumsi yang mendasari kebijakan ini adalah bahwa TOAFL bukan hanya mengukur kemampuan berbahasa arab saja namun alat ini juga dikembangkan untuk mengukur kemampuan berpikir, yang sangat diperlukan untuk belajar di perguruan tinggi.

Bila ternyata terdapat korelasi yang baik antara nilai dari tes bakat kebahasaan yang diukur sebelum terjadi poses belajar dengan nilai pengetahuan dan keterampilan dari bahasa arab yang dipelajari setelah dilakukan tes kebahasaan, alat ukur bakat kebahasaan tersebut mempunyai validitas prediktif yang lazim digunakan untuk alat analisis korelasi dan regresi. Jenis validitas ini ditetapkan dengan cara mengkorelasikan skor tes yang sedang disusun dengan kriteria yang menyangkut hasil karya atau prestasinya di masa yang akan datang.

d. Validitas konstruk

Validitas konstruk atau construct validity diperlukan alat ukur yang mempunyai beberapa indikator dalam mengukur konstruk. Jika ada alat ukur yang mempunyai beberapa aspek dan setiap aspek diukur dengan beberapa indikator, indikator yang sejenis harus berasosiasi positif satu dengan lainnya. Sebaliknya, indikator-indikator tersebut harus berasosiasi negatif dengan indikator lainnya jika indikator tersebut mengukur aspek yang berbeda atau berlawanan.

Penggunaan analisis faktor sering juga berfungsi sebagai cara untuk mendapatkan aspek validitas konstruk. Walaupun validitas konstruk sering dikaitkan dengan alat ukur yang terkait dengan psikolinguistik, misalnya persepsi, sikap kebahasaan, motivasi dalam belajar bahasa asing, validitas konstruk sering pula digunakan untuk alat tes pengetahuan dan keterampilan kebahasaan. Jika alat ukur tersebut hanya mengukur satu aspek saja, misalnya kosakata, validitas konstruknya dapat diukur dengan mengevaluasi dengan semua butir soalnya. Jika semua butir soalnya mengukur penguasaan kosakata, alat ukur tersebut sudah memenuhi aspek validitas konstruk.

Contoh, seorang peneliti mengembangkan suatu alat untuk mengukur gaya belajar bahasa arab berdasarkan teori bahwa belajar bahasa dapat dikelompokkan ke dalam gaya mandiri dan kerja kelompok. Bila alat ukur tersebut mempunyai 10 indikator untuk gaya belajar mandiri, dan 10 indikator untuk kerja kelompok, sepuluh indikator yang pertama harus saling berasosiasi atau berkorelasi positif, dan sepuluh indikator kedua juga harus saling berkorelasi positif.

Karena kedua kelompok tersebut mengukur aspek yang berlawanan, kedua kelompok indikator tersebut harus berkorelasi negatif. Jika ada indikator yang mengukur aspek yang sama namun berkorelasi negatif, indikator tersebut harus dibuang dari kuesioner yang dipakai. Jika dianalisis dengan analisis faktor, indikator-indikator tersebut akan dikelompokkan dengan sendirinya secara statistik sehingga masing-masing kelompok akan berkorelasi tinggi.

Menurut Anshari dkk. (2024), validitas merupakan indikator utama dalam menentukan kualitas suatu instrumen evaluasi karena berkaitan langsung dengan ketepatan pengukuran kompetensi peserta didik. Instrumen yang valid akan menghasilkan informasi yang sesuai dengan tujuan pembelajaran sehingga hasil evaluasi dapat digunakan sebagai dasar pengambilan keputusan pendidikan secara tepat. Oleh karena itu, proses pengujian validitas perlu dilakukan secara sistematis sebelum instrumen digunakan dalam kegiatan evaluasi pembelajaran.

Dalam pembelajaran Bahasa Indonesia, validitas memiliki peranan penting karena instrumen harus mampu mengukur berbagai keterampilan berbahasa secara tepat, seperti menyimak, berbicara, membaca, memirsa, dan menulis. Ketidaksesuaian antara tujuan

pembelajaran dan instrumen yang digunakan dapat menyebabkan hasil evaluasi tidak menggambarkan kemampuan peserta didik yang sebenarnya.

Reliabilitas

Reliabilitas berasal dari kata *reliability* berarti sejauh mana hasil suatu pengukuran dapat dipercaya. Reliabilitas merupakan derajat konsistensi suatu instrument. Suatu tes dapat dikatakan *reliable* apabila selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada kesempatan yang berbeda. Berikut ini merupakan empat factor yang mempengaruhi reliabilitas, yaitu:

a. Panjang tes (*Length of Test*)

Panjang tes berarti banyaknya soal tes. Ada kecenderungan, semakin panjang suatu tes akan lebih tinggi tingkat reliabilitas suatu tes, karena semakin banyak soal, maka akan semakin banyak sampel yang diukur dan proporsi jawaban yang benar semakin banyak, sehingga faktor tebakan (*guessing*) akan semakin rendah.

b. Sebaran skor (*spread of scores*)

Besarnya sebaran skor akan membuat tingkat reliabilitas menjadi lebih tinggi, karena koefisien reliabilitas yang lebih besar diperoleh ketika peserta didik tetap pada posisi yang relatif sama dalam satu kelompok pengujian ke pengujian berikutnya. Dengan kata lain, peluang selisih dari perubahan posisi dalam kelompok dapat memperbesar koefisien reliabilitas.

c. Tingkat kesukaran (*Difficulties Index*)

Dalam penilaian yang menggunakan pendekatan penilaian acuan norma, baik untuk soal yang mudah maupun sukar, cenderung menghasilkan tingkat reliabilitas yang rendah. Hal ini disebabkan antara hasil tes yang mudah dengan hasil tes yang sukar keduanya dalam satu sebaran skor yang terbatas. Untuk tes yang mudah, skor akan berada dibagian atas dan akhir dari skala penilaian. Bagi kedua tes (mudah dan sukar), perbedaan antar peserta didik kecil sekali dan cenderung tidak dapat dipercaya. Tingkat kesukaran soal yang ideal untuk meningkatkan koefisien reliabilitas adalah soal yang menghasilkan sebaran skor berbentuk genta atau kurva normal.

d. Obyektifitas

Obyektifitas disini menunjukkan skor tes kemampuan yang sama antara peserta didik yang satu dengan peserta didik lainnya. Peserta didik memperoleh hasil yang sama dalam mengerjakan suatu tes.

Suatu hasil pengukuran dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama, diperoleh hasil pengukuran yang relatif sama, selama aspek yang diukur dalam diri subyek memang belum berubah. Nur (1987: 47) menyatakan dalam jurnal *Validitas dan Reliabilitas Suatu Instrumen Penelitian* bahwa reliabilitas ukuran menyangkut seberapa jauh skor deviasi individu, atau skor-z, relatif konsisten apabila dilakukan pengulangan pengadministrasian dengan tes yang sama atau tes yang ekuivalen. Azwar (2003 : 176) menyatakan dalam jurnal *Validitas dan Reliabilitas Suatu Instrumen Penelitian* bahwa reliabilitas merupakan salah-satu ciri atau karakter utama instrumen pengukuran yang baik. Arifin (1991: 122) menyatakan dalam jurnal *Validitas dan Reliabilitas Suatu Instrumen Penelitian* bahwa suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda.

Memperhatikan pendapat diatas, maka reliabilitas merupakan koefisien yang menunjukan sejauh mana suatu instrumen/ alat pengukur dapat dipercaya, artinya apabila suatu instrumen digunakan berulang-ulang untuk mengukur sesuatu yang sama, maka hasilnya relatif stabil atau konsisten. Secara empiris tinggi rendahnya reliabilitas ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas,

besarnya koefisien reliabilitas berkisar antara 0 sampai dengan 1, dimana semakin tinggi angka reliabilitas berarti semakin konsisten hasil pengukuran, akan tetapi secara empiris koefisien reliabilitas yang mencapai angka 1 jarang dijumpai.

Menurut Allen dan Yen dalam jurnal Reliabilitas Instrumen Pendidikan bahwa koefisien reliabilitas besarnya ditentukan oleh satu dikurangi dengan perbandingan varians kesalahan pengukuran dan varians skor tampak. Atas dasar demikian semakin kecil varians kesalahan pengukuran, maka akan semakin tinggi koefisien reliabilitas, sehingga varians tampak atas dasar demikian semakin kecil varians kesalahan pengukuran, maka akan semakin tinggi koefisien reliabilitas, sehingga varians tampak dapat digunakan sebagai atribut koefisien reliabilitas yang sesungguhnya.

Rumus yang dapat digunakan untuk mencari reliabilitas menurut Arikunto (2011:87) adalah:

Spearman-Brown:

Rumus ini digunakan dalam menghitung besarnya reliabilitas berhubung dengan penambahan banyaknya butir soal.

Keterangan:

$$r_{11} = \frac{2 \cdot r_{1/2 \cdot 1/2}}{1 + r_{1/2 \cdot 1/2}}$$

Keterangan:

$r_{1/2 \cdot 1/2}$ = korelasi antara skor-skor setiap belahan tes.

r_{11} = koefisien reliabilitas yang sudah disesuaikan

Rumus Flanagan

Rumus yang penghitungannya menggunakan belah dua ganjil – genap.

$$r_{11} = 2 \left(1 - \frac{s_1^2 + s_2^2}{s_t^2} \right)$$

Keterangan:

r_{11} = reliabilitas tes

s_{21} = varians belahan pertama (1) yang dalam hal ini varians skor item ganjil

s_{22} = varians belahan kedua (2) yaitu varians skor item genap

s_{2t} = varians total yaitu varians skor total

Rumus Alpha:

Rumus ini digunakan untuk mencari soal uraian.

$r_{11} =$

$\frac{n}{n-1} \cdot 1 -$

$\frac{\sum \sigma_i^2}{\sigma_t^2}$

Keterangan:

r_{11} = reliabilitas yang dicari

$\sum \sigma_i^2$ = jumlah varians skor tiap-tiap item

σ_t^2 = varians total

Retnawati (2016) menjelaskan bahwa reliabilitas merupakan salah satu syarat utama instrumen yang berkualitas karena menunjukkan tingkat konsistensi hasil pengukuran. Instrumen yang memiliki reliabilitas tinggi akan menghasilkan skor yang relatif tetap

apabila digunakan pada kondisi yang setara. Dengan demikian, reliabilitas berfungsi untuk meningkatkan tingkat kepercayaan terhadap hasil evaluasi yang diperoleh.

Saputri dkk. (2023) menyatakan bahwa pengujian reliabilitas penting dilakukan untuk meminimalkan kesalahan pengukuran yang dapat memengaruhi hasil evaluasi. Semakin kecil kesalahan pengukuran yang terjadi, maka semakin tinggi tingkat reliabilitas instrumen tersebut. Oleh karena itu, reliabilitas menjadi salah satu indikator penting dalam menentukan kelayakan suatu instrumen evaluasi.

Seiring perkembangan teknologi, pengujian reliabilitas dapat dilakukan menggunakan berbagai perangkat lunak seperti SPSS, Microsoft Excel, maupun ANATES. Penggunaan teknologi tersebut dapat membantu pendidik memperoleh hasil analisis yang lebih cepat, akurat, dan efisien dibandingkan dengan perhitungan manual.

Kepraktisan

Kepraktisan mengandung arti kemudahan suatu tes, baik dalam mempersiapkan, menggunakan, mengolah dan menafsirkan, maupun mengadministrasikannya. Dimiyati dan Mudjiono dalam Arifin (2010) mengemukakan faktor-faktor yang mempengaruhi kepraktisan instrumen evaluasi meliputi:

- Kemudahan mengadministrasi
- Waktu yang disediakan untuk melancarkan evaluasi
- Kemudahan menskor
- Kemudahan interpretasi dan aplikasi
- Tersedianya bentuk instrumen evaluasi yang ekuivalen

Menurut Widoyoko (2022), instrumen evaluasi yang baik tidak hanya harus memenuhi aspek validitas dan reliabilitas, tetapi juga harus memiliki tingkat kepraktisan yang tinggi. Kepraktisan dapat dilihat dari kemudahan penggunaan instrumen, efisiensi waktu pelaksanaan, kemudahan penskoran, serta kemudahan dalam menginterpretasikan hasil evaluasi.

Priyadi (2024) menjelaskan bahwa pemanfaatan teknologi digital dalam evaluasi pembelajaran mampu meningkatkan kepraktisan instrumen melalui proses administrasi, pengolahan data, dan penyajian hasil evaluasi secara lebih efisien. Dengan adanya teknologi tersebut, pendidik dapat melaksanakan evaluasi dengan lebih efektif tanpa mengurangi kualitas hasil pengukuran.

Bagi guru Bahasa dan Sastra Indonesia, aspek kepraktisan menjadi penting karena proses penilaian sering melibatkan berbagai bentuk asesmen yang memerlukan waktu dan ketelitian tinggi. Oleh karena itu, instrumen yang praktis akan membantu guru melaksanakan evaluasi secara lebih optimal.

KESIMPULAN

Kualitas instrumen tes merupakan faktor penting dalam menghasilkan data pencapaian belajar peserta didik yang akurat, objektif, dan dapat dipertanggungjawabkan. Analisis butir soal berfungsi sebagai prosedur evaluatif yang membantu pendidik menentukan kualitas suatu instrumen serta kelayakannya untuk digunakan dalam proses evaluasi.

Validitas, reliabilitas, dan kepraktisan merupakan komponen utama yang menentukan kualitas instrumen evaluasi. Validitas memastikan bahwa instrumen mampu mengukur kompetensi yang ditargetkan, reliabilitas menjamin konsistensi hasil pengukuran, sedangkan kepraktisan mendukung kemudahan penggunaan instrumen dalam pelaksanaan evaluasi. Oleh karena itu, ketiga aspek tersebut perlu dipenuhi secara bersamaan agar instrumen evaluasi mampu mendukung proses pembelajaran yang efektif dan berkualitas.

DAFTAR PUSTAKA

- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *EDUKATIF: Jurnal Ilmu Pendidikan*, 6(1), 964–975.
- Arifin, Z. (2010). *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Bandung: Remaja Rosdakarya.
- Arikunto, S. (2011). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Muarrib Journal of Arabic Education*, 1(1), 34–44.
- Mulyasa, E. (2022). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Priyadi, S. (2024). Analisis Tes dan Butir Soal Pada Moodle. *Jurnal Evaluasi Pendidikan*, 15(1), 33–42.
- Purwanto. (2021). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama Publishing.
- 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 STKIP Subang*, 9(5).
- Sudjana, N. (2009). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Sugiyono. (2023). *Metode Penelitian Pendidikan*. Bandung: Alfabeta.
- Widoyoko, E. P. (2022). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.
- Yusuf, A. M. (2021). *Asesmen dan Evaluasi Pendidikan*. Jakarta: Kencana.