

Studi Analisis Instrument Penilaian: Menguji Reliabilitas, Taraf Kesukaran, Daya Pembeda Secara Empirik dan Rasional

Ratna Sari^{1*}, Julita Evawani², Zuhraini³, Ansor⁴

^{1,2,3,4} IAIN Takengon, Aceh, Indonesia

e-mail: ratnasaritandep@gmail.com

*Corresponding Author

ARTICLE INFO

Article history:

Received: Mei 3, 2026

Revised: Juni 12, 2026

Accepted: Juni 16, 2026

Kata Kunci:

Analisis Empiric; Analisis Rasional; Daya Pembeda; Reliabilitas Tes; Taraf Kesukaran

Keywords:

Empirical Analysis; Rational Analysis; Difficulty Level; Test Reliability; Discriminatory Power

ABSTRACT

Dalam dunia pendidikan tentunya tak luput dari yang namanya penilaian, untuk mengetahui tingkat keberhasilan pembelajaran yang disampaikan kepada peserta didik, sebelum melakukan penilaian tentunya memerlukan beberapa analisis agar penilaian tersebut tidak dianggap sebagai anomali deviasi. Penelitian ini bertujuan untuk menganalisis instrumen penilaian yang di gunakan dalam proses pembelajaran guna mengetahui tingkat kelayakan, kevalidan, dan keefektifannya dalam mengukur hasil belajar peserta didik. Instrumen penilaian merupakan komponen penting dalam evaluasi pembelajaran karena berfungsi untuk mengukur pencapaian kompetensi secara objektif dan sistematis. Metode yang digunakan dalam penelitian ini adalah pendekatan deskriptif dengan melakukan kajian terhadap bentuk, isi, serta kesesuaian instrumen dengan indikator pembelajaran. Hasil analisis menunjukkan bahwa instrumen penilaian yang baik harus memenuhi kriteria valid, reliabel, objektif, serta sesuai dengan tujuan dan indikator pembelajaran yang telah ditetapkan. Dengan demikian, analisis instrumen penilaian sangat penting dilakukan agar proses penilaian dapat memberikan gambaran yang akurat mengenai kemampuan peserta didik.

In the world of education, of course, there is no escape from the assessment, to know the level of learning success that is conveyed to the students, before conducting an assessment, of course, it requires several analysis so that the assessment is not considered as a deviation anomaly. This research aims to analyze the assessment instruments used in the learning process to determine the level of eligibility, validity, and effectiveness in measuring the learning outcomes of learners. Assessment instruments are an important component in learning evaluation because they function to measure the achievement of competence objectively and systematically. The method used in this research is a descriptive approach by studying the form, content, and suitability of the instrument with learning indicators. The results of the analysis show that a good assessment instrument must meet the criteria that are valid, reliable, objective, and in accordance with the goals and learning indicators that have been set. Thus, the analysis of assessment instruments is very important to be carried out so that the assessment process can provide an accurate picture of the student's ability.

This is an open access article under the [CC-BY-SA](#) license



Corresponding Author:

Ratna Sari,

Pendidikan Agama Islam, Institut Agama Islam Negeri Takengon

Jl. Yos Sudarso/A. Dimot No. 10 Takengon, Aceh Tengah, Aceh, Indonesia.

Email: ratnasaritandep@gmail.com

PENDAHULUAN

Analisis data adalah salah satu tahap penelitian yang dilakukan setelah seluruh informasi yang diperlukan untuk menyelesaikan permasalahan penelitian telah tersedia secara lengkap. Ketepatan dan ketelitian dalam pemilihan alat analisis sangat menentukan akurasi kesimpulan yang dihasilkan, sehingga kegiatan analisis data tidak boleh diabaikan dalam proses penelitian. Kesalahan dalam memilih alat analisis dapat menimbulkan dampak fatal pada kesimpulan yang diperoleh, yang selanjutnya akan merusak pemanfaatan dan standar hasil penelitian tersebut. Oleh karena itu, pemahaman dan pengetahuan mendalam tentang berbagai teknik analisis sangat dibutuhkan oleh seorang peneliti agar hasil penelitiannya memberikan manfaat nyata dalam pemecahan masalah sekaligus dapat dipertanggungjawabkan secara ilmiah. (Muhson, 2021)

Menurut KBBI, instrumen adalah alat penelitian, seperti rangkaian tes atau sejenisnya, yang dimaksudkan untuk mengumpulkan data guna dianalisis lebih lanjut. Arikunto menjelaskan bahwa instrumen merupakan alat pendukung yang dipilih dan dimanfaatkan oleh peneliti untuk mengumpulkan data sehingga prosesnya lebih terstruktur dan sederhana. Suharsimi mendefinisikan alat atau instrumen evaluasi sebagai sesuatu yang membantu seseorang melaksanakan tugas atau meraih tujuan secara lebih efektif dan efisien. Kemudian, Sudjiono menyatakan bahwa penilaian merupakan proses pengambilan keputusan mengenai suatu objek dengan mempertimbangkan kriteria seperti baik atau buruk, sehat atau sakit, cerdas atau kurang cerdas, dan sebagainya. (Maulana, 2022)

Sebuah instrumen tidak boleh digunakan secara langsung, melainkan wajib diuji validitasnya lebih dulu. Proses tersebut diperlukan agar instrumen yang dipakai untuk mengukur sesuatu terbukti sah dan pantas. Karenanya, pengujian validitas harus dilaksanakan sebelum penelitian dimulai. Suatu instrumen atau tes dikatakan valid jika tes itu benar-benar mengukur apa yang dimaksudkan. Validitas berasal dari kata *validity*, yang merujuk pada tingkat ketepatan dan akurasi alat ukur dalam menjalankan peran pengukurannya (Azwar 1986). Di samping itu, validitas juga merupakan indikator bahwa variabel yang diukur selaras dengan variabel yang hendak diteliti peneliti; validitas berkaitan dengan apakah suatu peubah mengukur apa yang seharusnya diukur. Validitas dalam penelitian mencerminkan seberapa tepat alat ukur penelitian merepresentasikan isi yang sebenarnya diukur. Uji validitas adalah metode untuk menunjukkan sejauh mana alat ukur dapat dipercaya dalam mengukur hal yang dimaksud. Ghazali menyatakan bahwa uji validitas berfungsi untuk menilai apakah suatu kuesioner sah atau valid. Kuesioner dianggap valid jika pertanyaan-pertanyaannya mampu menggambarkan dengan tepat apa yang ingin diukur oleh kuesioner tersebut.

Evaluasi dalam dunia pendidikan memiliki peran penting karena digunakan untuk mengambil keputusan dalam merancang kebijakan yang mendorong terbentuknya sumber daya manusia yang mampu bersaing dalam era globalisasi saat ini. Karena itu, evaluasi atau penilaian merupakan bagian yang tidak terpisahkan dari seluruh proses pendidikan di setiap lembaga. Secara umum, ada dua tujuan utama dilakukannya evaluasi, yaitu untuk mengetahui atau membuktikan sejauh mana pemahaman peserta didik setelah melalui proses pembelajaran, serta untuk menggunakan hasil evaluasi sebagai dasar dalam mencari tahu kelemahan dan meningkatkan kualitas proses pembelajaran.

Kegiatan penilaian siswa adalah bagian yang sangat penting serta tidak terpisahkan dari proses belajar mengajar di sekolah. Untuk mengetahui sejauh mana peserta didik mencapai hasil belajar yang telah ditentukan, penilaian hasil belajar sangat diperlukan. Wahidmurni menyatakan bahwa salah satu peran utama bagi pendidik dalam menilai proses belajar siswa adalah untuk memberikan umpan balik yang membantu mereka dalam menilai efektivitas dan efisiensi proses pembelajaran yang telah dilakukan. Menurut Miller, penilaian hasil belajar siswa melibatkan berbagai langkah untuk mendapatkan informasi mengenai pembelajaran siswa dan mengambil keputusan terkait kinerja atau hasil yang diperoleh. Kegiatan penilaian hasil belajar ini merupakan tugas guru yang berhubungan dengan pengambilan keputusan mengenai pencapaian kompetensi siswa selama proses belajar. Penilaian ini harus memenuhi beberapa prinsip yang diungkapkan oleh Anderson, termasuk (1) memiliki makna (*meaningfulness*), (2) transparansi atau keterbukaan (*expicitness*), (3) keadilan (*fairness*). Data penilaian yang dikumpulkan oleh guru harus mengikuti prosedur dan alat ukur yang relevan dengan kompetensi yang diharapkan dari siswa atau indikator yang telah ditetapkan. Alat penilaian hasil belajar dapat memanfaatkan teknologi modern, sehingga proses penilaiannya menjadi lebih efisien dengan menggunakan mesin (komputasi), dapat mencakup semua tugas rutin (otomatisasi), serta memungkinkan komunikasi dari lokasi dan waktu mana saja. Teknologi Informasi dan Komunikasi (TIK) memberikan kontribusi yang sangat besar terhadap pencapaian tujuan pendidikan. (Destiana et al., 2020)

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan tujuan untuk menganalisis kualitas instrumen penilaian dalam pembelajaran. Pendekatan ini dipilih karena mampu menggambarkan secara sistematis karakteristik instrumen berdasarkan aspek rasional dan empirik. Sumber data dalam penelitian ini berasal dari 8 artikel ilmiah yang relevan, termasuk jurnal nasional dan publikasi akademik yang membahas analisis instrumen penilaian pendidikan. Data tersebut digunakan sebagai dasar untuk mengkaji konsep reliabilitas, tingkat kesukaran, dan daya pembeda butir soal.

Teknik pengumpulan data dilakukan melalui studi dokumentasi (*document analysis*) dengan menelaah isi literatur secara mendalam. Proses analisis data dilakukan melalui tiga tahap, yaitu reduksi data, penyajian data, dan penarikan kesimpulan. Pada tahap reduksi, data yang tidak relevan diseleksi. Selanjutnya, data disusun berdasarkan kategori analisis instrumen, seperti analisis rasional dan empirik. Keabsahan data diperkuat melalui teknik triangulasi sumber, yaitu membandingkan temuan dari beberapa artikel untuk memperoleh konsistensi konsep. Hasil analisis kemudian diinterpretasikan secara naratif untuk menghasilkan pemahaman yang komprehensif mengenai kualitas instrumen penilaian.

HASIL DAN PEMBAHASAN

Analisis Rasional Instrumen Penilaian

Analisis rasional merupakan tahap awal dalam pengembangan instrumen penilaian yang berfungsi memastikan kesesuaian antara tujuan pembelajaran, indikator kompetensi, dan butir soal yang dikembangkan. Tahap ini menjadi dasar konseptual sebelum instrumen diuji secara empiris. Hasil kajian menunjukkan bahwa kesesuaian antara indikator dan butir soal menentukan validitas isi instrumen. Ketidakesesuaian dapat menyebabkan kesalahan pengukuran sistematis yang berdampak pada ketidaktepatan interpretasi hasil belajar peserta didik. Hal ini sejalan dengan temuan penelitian sebelumnya yang menegaskan bahwa validitas instrumen sangat ditentukan oleh keterkaitan antara indikator dan representasi butir soal yang diukur (Saputri et al., 2023).

Selain itu, analisis rasional juga berkaitan dengan prinsip konstruksi instrumen yang menuntut keselarasan antara tingkat kognitif dan capaian pembelajaran. Instrumen yang tidak mempertimbangkan aspek ini cenderung menghasilkan bias pengukuran. Penelitian terdahulu juga menunjukkan bahwa kualitas instrumen asesmen sangat bergantung pada analisis awal sebelum pengujian statistik dilakukan (Son, 2019). Dengan demikian, analisis rasional berfungsi sebagai filter awal untuk memastikan instrumen layak diuji secara empirik.

Analisis Empirik Instrumen Penilaian

Analisis empirik dilakukan berdasarkan data hasil uji coba instrumen untuk menilai kualitas butir soal secara objektif. Parameter utama yang digunakan meliputi tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil analisis menunjukkan bahwa instrumen yang baik harus memiliki distribusi jawaban yang seimbang antara kelompok siswa berkemampuan tinggi dan rendah. Instrumen yang terlalu mudah atau terlalu sulit tidak mampu memberikan informasi diagnostik yang optimal.

Temuan ini sejalan dengan penelitian Rahmawati et al. (2025) yang menunjukkan bahwa analisis empirik pada butir soal mampu mengidentifikasi kualitas item berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda secara simultan. Selain itu, penelitian lain juga menegaskan bahwa analisis butir soal berbasis data empirik mampu meningkatkan kualitas evaluasi pembelajaran secara signifikan (Iskandar, 2018). Oleh karena itu, analisis empirik menjadi tahapan penting untuk memastikan instrumen tidak hanya valid secara teori, tetapi juga berfungsi secara nyata di lapangan.

Reliabilitas Instrumen Penilaian

Reliabilitas menunjukkan tingkat konsistensi instrumen dalam mengukur konstruk yang sama secara berulang pada kondisi yang stabil. Instrumen yang reliabel akan menghasilkan skor yang konsisten meskipun dilakukan pengukuran berulang. Hasil kajian menunjukkan bahwa reliabilitas dipengaruhi oleh kualitas item, jumlah soal, serta konsistensi penskoran. Instrumen dengan reliabilitas tinggi menunjukkan bahwa variasi skor lebih banyak disebabkan oleh perbedaan kemampuan siswa, bukan oleh kesalahan pengukuran.

Penelitian terdahulu juga menunjukkan bahwa instrumen dengan nilai reliabilitas di atas 0,80 termasuk kategori sangat baik dan layak digunakan dalam evaluasi formal (Elizanoviyanti et al., 2025). Hal ini memperkuat bahwa reliabilitas merupakan indikator utama dalam menentukan kelayakan instrumen penilaian.

Uji reliabilitas dilaksanakan setelah uji kelayakan rampung. Sebuah alat ukur dianggap dapat diandalkan jika hasil pengukuran yang diberikan konsisten. Uji keandalan mengukur sejauh mana hasil pengukuran dapat dipercaya. Ketika aspek-aspek yang dinilai dalam subjek tetap stabil, hasil pengukuran dapat dianggap dapat diandalkan jika memberikan hasil yang serupa melalui berbagai evaluasi pada kelompok orang yang sama. (Putri Juliani & Erita, 2023). Mengetahui tingkat ketepatan (presisi) dan konsistensi skor tes adalah tujuan utama dari penghitungan reliabilitas skor tes. (Nasir, 2015) Untuk mengetahui tinggi rendahnya koefisien reliabilitas dapat dilakukan dengan cara sebagai berikut:

Tabel 1. Reliabilitas Tes

No	Reliabilitas tes	Kategori instrument tes
1	$0,00 < r \leq 0,20$	Sangat rendah
2	$0,20 < r \leq 0,40$	Rendah
3	$0,40 < r \leq 0,60$	Sedang
4	$0,60 < r \leq 0,80$	Tinggi
5	$0,80 < r \leq 1,00$	Sangat tinggi

Tarap Kesukaran Butir Soal

Tingkat kesukaran merupakan indikator yang menunjukkan proporsi peserta didik yang mampu menjawab benar suatu butir soal. Indeks kesukaran (P) digunakan untuk mengklasifikasikan tingkat kesulitan soal. Hasil analisis menunjukkan bahwa tingkat kesukaran ideal berada pada kategori sedang karena mampu memberikan informasi paling optimal mengenai kemampuan siswa. Soal yang terlalu mudah tidak mampu membedakan kemampuan siswa, sedangkan soal yang terlalu sulit dapat menurunkan motivasi belajar. Penelitian sebelumnya juga menunjukkan bahwa distribusi tingkat kesukaran yang tidak proporsional dapat mengurangi kualitas evaluasi secara keseluruhan (Prastika, 2021). Oleh karena itu, keseimbangan tingkat kesukaran menjadi faktor penting dalam penyusunan instrumen.

Soal yang baik tidak terlalu mudah maupun terlalu sulit. Soal yang terlalu mudah tidak mendorong siswa untuk berusaha lebih keras dalam memahaminya. Sebaliknya, soal yang terlalu sulit akan membuat siswa kehilangan semangat dan enggan mencoba lagi karena melebihi kemampuan mereka. Indeks kesulitan, yang juga disebut "indeks kesulitan", mengindikasikan tingkat sulit atau mudahnya suatu hal. Indeks kesulitan berkisar antara 0,00 hingga 1,0 dan menunjukkan derajat kesulitan soal. Soal dengan indeks kesulitan 0,0 berarti soal terlalu sulit, sedangkan soal dengan indeks 1,0 berarti soal terlalu mudah.



Simbol P (huruf kapital besar), merupakan singkatan dari kata "proporsi", dipakai untuk mengilustrasikan indeks kesulitan ini dalam proses evaluasi. Maka, soal dengan $P = 0,70$ lebih mudah daripada soal dengan $P = 0,20$, serta soal dengan $P = 0,30$ lebih sulit daripada soal dengan $P = 0,80$..(Solichin, 2017)

Tingkat kesulitan dihitung sebagai rasio antara jumlah peserta yang menjawab soal dengan benar dan jumlah peserta yang gagal. Tingkat kesulitan, menurut teori skor-murni klasik, merupakan parameter yang mengindikasikan seberapa sulit bagi sekelompok peserta ujian untuk memberikan jawaban benar pada suatu butir soal. Secara teoritis, tingkat kesulitan didefinisikan sebagai probabilitas empiris untuk berhasil pada butir soal tertentu bagi peserta ujian tertentu..(Suryani, 2017)

Daya Pembeda Butir Soal

Daya pembeda merupakan kemampuan butir soal untuk membedakan peserta didik berkemampuan tinggi dan rendah. Indeks ini menjadi salah satu indikator utama dalam evaluasi kualitas soal. Hasil analisis menunjukkan bahwa butir soal dengan daya pembeda tinggi mampu mengidentifikasi perbedaan kemampuan siswa secara lebih akurat. Sebaliknya, butir soal dengan nilai daya pembeda rendah atau negatif menunjukkan adanya kelemahan dalam konstruksi soal. Temuan ini sejalan dengan penelitian Nurhalimah et al. (2022) yang menyatakan bahwa daya pembeda memiliki hubungan erat dengan kualitas validitas item dan tingkat kesukaran. Artinya, ketiga indikator tersebut saling berkaitan dalam menentukan kualitas instrumen penilaian.

Daya pembeda soal merujuk pada kemampuan suatu soal untuk membedakan antara siswa berpengetahuan tinggi dengan siswa berpengetahuan rendah. Angka yang mengindikasikan tingkat daya pembeda ini disebut indeks diskriminasi, yang disingkat D (huruf kapital besar). Indeks diskriminasi memiliki rentang nilai dari 0,00 hingga 1,00. Meskipun demikian, berbeda dengan indeks kesukaran yang tidak memiliki nilai negatif, indeks diskriminasi dapat memiliki nilai negatif. Nilai negatif pada indeks diskriminasi digunakan ketika sebuah soal menunjukkan hasil yang tidak sesuai dengan harapan dari peserta tes.

Angka yang menunjukkan tingkat daya beda ini merupakan Indeks Diskriminasi yang berkisar antara 0,00 hingga 1,00. Namun, pada indeks diskriminasi, terdapat nilai negatif (-) yaitu -1,0 sampai dengan 0,0 hingga 1,0 (semakin jauh ke kanan nilai soal semakin baik, sementara semakin jauh ke kiri nilai soal menunjukkan kualitas yang semakin buruk; karena semakin ke kanan, siswa yang cerdas mengalami kesulitan menjawab dengan tepat, sedangkan siswa yang kurang pintar di sisi kiri dapat menjawab dengan sembarangan).

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

D = Indek Diskriminasi (Daya Beda)

JA = Banyaknya Peserta Kelompok Atas

JB = Banyaknya Peserta Kelompok Bawah

BA = Banyaknya Peserta Kelompok Atas Yang Menjawab Soal Dengan Benar

- BB = Abnyaknya Peserta Kelompok Bawah Yang Menjawab Soal Dengan Benar
- PA = Proporsi peserta Kelompok Atas Yang Menjawab Benar
- PB = Proporsi Peserta Kelompok Bawah Yang Menjawab Benar
- D = 0,00-----0,20 → Jelek (*poor*)
- D = 0,21-----0,40 → cukup (*satisfactory*)
- D = 0,41-----0,70 → baik (*good*)
- D = 0,71-----1,00 → baik sekali (*excellent*)
- D = Negatif, Semuanya → tidak baik, jadi semua butir soal yang mempunyai nilai D negative sebaiknya di buang.

Dalam proses analisis kualitas tes dan butir soal, terdapat manfaat dari daya pembeda butir soal yang kami ambil dari pandangan Karjono Natar berikut ini: 1). Untuk meningkatkan kualitas setiap butir soal berdasarkan data nyata yang tersedia. Melalui indeks daya pembeda, dapat diketahui apakah butir soal tersebut berkualitas baik, perlu direvisi, atau harus dibuang. 2) Untuk mengetahui seberapa efektif setiap butir soal dalam membedakan kemampuan siswa, yakni siswa yang sudah menguasai atau yang belum menguasai materi yang diberikan guru. Jika suatu butir soal gagal membedakan kedua kelompok kemampuan siswa tersebut, maka butir soal itu kemungkinan mencerminkan masalah seperti berikut: (a) kunci jawaban butir soal tersebut tidak tepat; (b) butir soal memiliki dua atau lebih jawaban benar; (c) kompetensi yang diukur kurang jelas; (d) pilihan jawaban tidak efektif; (e) materi yang diujikan terlalu sulit, sehingga banyak siswa hanya menebak; (f) mayoritas siswa yang menguasai materi yang ditanyakan merasa ada kesalahan dalam butir soalnya (Solichin, 2017).

Keterkaitan Antar Parameter Instrumen

Analisis kualitas instrumen penilaian tidak dapat dilakukan secara parsial, melainkan harus dipahami sebagai suatu sistem yang saling terintegrasi antara reliabilitas, tingkat kesukaran, dan daya pembeda. Ketiga parameter tersebut membentuk satu kesatuan konstruksi evaluasi yang menentukan apakah suatu instrumen benar-benar mampu mengukur kemampuan peserta didik secara akurat, konsisten, dan diskriminatif.

Reliabilitas pada dasarnya merepresentasikan stabilitas pengukuran. Namun, stabilitas ini sangat dipengaruhi oleh karakteristik butir soal yang membentuk instrumen. Instrumen yang didominasi oleh butir soal yang terlalu mudah atau terlalu sulit cenderung menghasilkan variansi skor yang rendah, sehingga menurunkan koefisien reliabilitas. Sebaliknya, instrumen dengan distribusi tingkat kesukaran yang seimbang akan menghasilkan variasi skor yang lebih baik, yang pada akhirnya meningkatkan konsistensi internal instrumen.

Selain itu, daya pembeda memiliki peran langsung dalam menentukan kualitas informasi yang dihasilkan oleh setiap butir soal. Butir soal dengan daya pembeda tinggi mampu memisahkan kelompok peserta didik berdasarkan tingkat kemampuan secara lebih akurat. Namun, daya pembeda ini tidak akan optimal apabila tingkat kesukaran tidak berada pada

rentang yang proporsional. Soal yang terlalu mudah atau terlalu sulit cenderung memiliki daya pembeda yang rendah karena tidak mampu menciptakan variasi respons antar peserta didik.

Dengan demikian, tingkat kesukaran dapat dipahami sebagai variabel yang memengaruhi distribusi respons, sedangkan daya pembeda berfungsi sebagai indikator sensitivitas instrumen dalam membedakan kemampuan. Kedua aspek ini secara langsung berkontribusi terhadap reliabilitas sebagai indikator konsistensi keseluruhan instrumen. Hubungan ini menunjukkan bahwa reliabilitas bukan hanya hasil dari jumlah item, tetapi juga kualitas psikometrik setiap butir soal.

Secara konseptual, keterkaitan ketiga parameter ini juga dapat dijelaskan melalui teori pengukuran klasik (Classical Test Theory/CTT), yang menyatakan bahwa skor observasi merupakan fungsi dari skor sebenarnya dan error pengukuran. Dalam konteks ini, tingkat kesukaran dan daya pembeda memengaruhi besarnya varians skor, sedangkan reliabilitas mencerminkan proporsi varians sejati terhadap varians total. Oleh karena itu, ketidakseimbangan pada salah satu parameter akan berdampak langsung pada penurunan kualitas instrumen secara keseluruhan.

Penelitian Iskandar (2018) memperkuat temuan ini dengan menegaskan bahwa instrumen penilaian yang baik harus memenuhi empat aspek utama secara simultan, yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Keempat aspek tersebut tidak dapat berdiri sendiri, tetapi harus dianalisis secara terpadu agar menghasilkan instrumen yang mampu memberikan gambaran objektif terhadap kemampuan peserta didik.

Dengan demikian, dapat disimpulkan bahwa kualitas instrumen penilaian sangat ditentukan oleh integrasi harmonis antara tingkat kesukaran yang proporsional, daya pembeda yang optimal, dan reliabilitas yang tinggi. Ketidakseimbangan pada salah satu komponen akan menurunkan kualitas pengukuran dan berpotensi menghasilkan interpretasi hasil belajar yang tidak akurat.

Implikasi Penelitian

Hasil penelitian ini memberikan implikasi yang signifikan terhadap pengembangan praktik evaluasi pembelajaran di lingkungan pendidikan. Temuan menunjukkan bahwa instrumen penilaian yang dianalisis melalui pendekatan rasional dan empirik mampu meningkatkan akurasi pengukuran hasil belajar peserta didik. Hal ini menegaskan bahwa proses evaluasi tidak hanya berfungsi sebagai alat ukur, tetapi juga sebagai dasar pengambilan keputusan pendidikan yang berbasis data (*data-driven decision making*).

Secara pedagogis, hasil analisis instrumen memberikan kontribusi langsung terhadap peningkatan kualitas pembelajaran di kelas. Guru dapat menggunakan informasi dari tingkat kesukaran, daya pembeda, dan reliabilitas untuk mengidentifikasi kelemahan materi pembelajaran serta karakteristik kemampuan siswa. Dengan demikian, evaluasi tidak hanya berakhir pada pemberian nilai, tetapi juga menjadi sumber umpan balik (*feedback*) yang sistematis untuk perbaikan proses pembelajaran berikutnya.

Selain itu, hasil penelitian ini menunjukkan bahwa analisis instrumen dapat digunakan sebagai dasar dalam pengambilan keputusan akademik yang lebih objektif, seperti penentuan kelulusan, pemberian remedial, serta pengayaan materi. Instrumen yang telah tervalidasi secara

rasional dan empirik memberikan tingkat kepercayaan yang lebih tinggi dalam interpretasi hasil belajar, sehingga mengurangi potensi bias subjektif dalam penilaian.

Dari perspektif profesionalisme pendidik, penerapan analisis instrumen juga berkontribusi dalam meningkatkan kompetensi guru dalam bidang asesmen pendidikan. Guru tidak hanya berperan sebagai penyusun soal, tetapi juga sebagai evaluator yang mampu membaca karakteristik butir soal secara kuantitatif dan kualitatif. Hal ini sejalan dengan tuntutan kompetensi abad ke-21 yang menekankan kemampuan analisis data dalam praktik pendidikan.

Lebih jauh, implikasi penelitian ini juga mengarah pada pentingnya budaya evaluasi berbasis bukti (evidence-based assessment) di lingkungan sekolah. Penggunaan instrumen yang dianalisis secara sistematis dapat mendorong peningkatan mutu pendidikan secara berkelanjutan, karena setiap keputusan pembelajaran didasarkan pada data yang valid dan reliabel. Dengan demikian, implikasi penelitian ini tidak hanya bersifat teknis dalam penyusunan instrumen, tetapi juga bersifat strategis dalam meningkatkan kualitas evaluasi, efektivitas pembelajaran, serta profesionalisme guru dalam sistem pendidikan.

KESIMPULAN

Kualitas suatu instrumen penilaian sangat ditentukan oleh ketepatan dalam proses penyusunannya serta pengujiannya. Analisis rasional berperan penting dalam menilai kesesuaian isi instrumen dengan tujuan pembelajaran secara teoritis, sedangkan analisis empirik memberikan gambaran nyata tentang kinerja butir soal berdasarkan data hasil uji coba. Selain itu, reliabilitas tes menunjukkan tingkat konsistensi instrumen dalam mengukur kemampuan peserta didik, sehingga menjadi indikator penting dalam menentukan kepercayaan terhadap hasil penilaian. Tingkat kesukaran berfungsi untuk mengklasifikasikan butir soal agar tidak terlalu mudah maupun terlalu sulit, sehingga mampu mengukur kemampuan secara proporsional. Sementara itu, daya pembeda berperan dalam mengidentifikasi kemampuan soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Dengan demikian, kelima aspek tersebut saling berkaitan dan harus dianalisis secara komprehensif guna menghasilkan instrumen penilaian yang valid, reliabel, dan berkualitas, sehingga mampu memberikan gambaran yang akurat terhadap pencapaian hasil belajar peserta didik.

DAFTAR PUSTAKA

- Aridatul kholidah, Noviyanti, Santy ikaa, E. S. F. (2026). analisis rasional instrumen dalam pengembangan sistem evaluasi di tk IT al iffah bekasi. *Jurnal Pendidikan Sosial Dan Humaniora*, 5(2), 306–312.
- Destiana, D., Suchyadi, Y., & Anjaswuri, F. (2020). Pengembangan Instrumen Penilaian Untuk Meningkatkan Kualitas Pembelajaran Produktif di Sekolah Dasar [Development of Assessment Instruments to Improve the Quality of Productive Learning in Elementary Schools]. *Jurnal Pendidikan Dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, 3(2), 119–123. <https://journal.unpak.ac.id/index.php/JPPGuseda/article/view/2720>
- Elizanoviyanti, F., et al. (2025). Analisis tingkat validitas dan reliabilitas butir soal PAS Bahasa Indonesia. *Jurnal Literat*, 9(1), 45–60.

- Iskandar, A. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi. *Jurnal Penelitian Evaluasi Pendidikan*, 22(1), 1–15.
- Maulana, A. (2022). Analisis Validitas, Reliabilitas, dan Kelayakan Instrumen Penilaian Rasa Percaya Diri Siswa. *Jurnal Kualita Pendidikan*, 3(3), 133–139. <https://doi.org/10.51651/jkp.v3i3.331>
- Muhson, A. (2021). *TEKNIK ANALISIS KUANTITATIF*.
- Nasir, M. (2015). *Dalam Rangka Menghasilkan Soal Yang Baik Dan Bermutu*. 336–347.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan validitas item dengan daya pembeda dan tingkat kesukaran soal. *Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>
- Prastika, Y. D. (2021). Pengaruh validitas, reliabilitas dan tingkat kesukaran terhadap kualitas butir soal. *STKIP PGRI Bangkalan*.
- Putri Juliani, R., & Erita, S. (2023). Analisis Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Kritis dalam Konteks Sekolah Menengah. *JEID: Journal of Educational Integration and Development*, 3(3), 169–179. <https://doi.org/10.55868/jeid.v3i3.313>
- Rahmawati, N. S., et al. (2025). Analisis butir soal objektif penilaian sumatif akhir semester. *JIP Journal*, 8(2), 112–120.
- Saputri, H. A. S., Zulhijrah, N., Larasati, J., & Shaleh. (2023). Analisis instrumen asesmen: validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD*, 9(5), 2986–2995. <https://doi.org/10.36989/didaktik.v9i5.2268>
- Solichin, M. (2017). analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Manajemen & Pendidikan Islam*, 2, 192–213.
- Son, A. L. (2019). Instrumentasi kemampuan pemecahan masalah matematis. *Gema Wiralodra*, 10(1), 41–52.
- Suryani, Y. E. (2017). Pemetaan Kualitas Empirik Soal Ujian Akhir Semester Pada Mata Pembelajaran Bahasa Indonesia SMA Di Kabupaten Klaten. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 142–152. <https://journal.uny.ac.id/index.php/jpep/article/view/10725>