

## RELIABILITAS ALAT UKUR: JENIS-JENIS, CARA PENGUKURAN, DAN FAKTOR-FAKTOR YANG MEMPENGARUHINYA DALAM PEMBELAJARAN

Fityatul Muharromah<sup>1\*</sup>, Farid Al Fani<sup>2</sup>, Ahmad Rosyid Ridho<sup>3</sup>

<sup>1,2,3</sup>Institut Islam Mamba'ul Ulum, Surakarta

\* Corresponding Email: [pelangifita@gmail.com](mailto:pelangifita@gmail.com)

### ABSTRAK

Penelitian ini bertujuan untuk mendeskripsikan konsep reliabilitas alat ukur dalam evaluasi pembelajaran serta jenis, cara pengukuran, dan faktor-faktor yang memengaruhinya. Metode penelitian menggunakan studi pustaka dengan menganalisis literatur primer dan sekunder yang relevan dalam sepuluh tahun terakhir. Hasil kajian menunjukkan bahwa reliabilitas merupakan aspek kunci yang menentukan konsistensi dan keajegan suatu instrumen, yang dapat diuji melalui berbagai metode seperti test-retest, bentuk paralel, konsistensi internal, dan belah dua. Faktor yang memengaruhi reliabilitas meliputi jumlah butir soal, kejelasan instruksi, kondisi peserta, waktu pelaksanaan, serta kesesuaian bentuk tes. Penelitian ini menegaskan bahwa reliabilitas instrumen sangat penting untuk menjamin akurasi penilaian, mendukung validitas, dan meningkatkan kualitas pengambilan keputusan dalam pembelajaran.

**Kata kunci:** reliabilitas, evaluasi pembelajaran, instrumen penilaian, konsistensi internal, studi pustaka.

### ABSTRACT

*This study aims to describe the concept of measurement instrument reliability in learning evaluation, including its types, measurement methods, and influencing factors. The research method employed a literature review, analyzing relevant primary and secondary literature from the past ten years. The study's results indicate that reliability is a key aspect determining the consistency and reliability of an instrument, which can be tested through various methods such as test-retest, parallel forms, internal consistency, and split-half tests. Factors influencing reliability include the number of items, clarity of instructions, participant conditions, time of administration, and the appropriateness of the test format. This study confirms that instrument reliability is crucial for ensuring assessment accuracy, supporting validity, and improving the quality of decision-making in learning.*

**Keywords:** *reliability, learning evaluation, assessment instruments, internal consistency, literature review.*

### PENDAHULUAN

Dalam dunia pendidikan, pengukuran merupakan salah satu aspek yang sangat krusial dalam memastikan efektivitas proses pembelajaran. Melalui proses pengukuran, pendidik dan peneliti dapat mengetahui sejauh mana tujuan pembelajaran telah tercapai serta menilai kompetensi peserta didik secara objektif. Pengukuran bukan sekadar proses teknis, tetapi merupakan landasan penting dalam pengambilan keputusan pendidikan, mulai dari evaluasi hasil belajar, perbaikan proses pembelajaran, hingga penentuan

kebijakan pendidikan. Dalam konteks ini, reliabilitas alat ukur menjadi elemen kunci yang tidak dapat diabaikan. Sebuah instrumen pengukuran yang tidak reliabel dapat menyebabkan hasil evaluasi yang bias, tidak akurat, dan menyesatkan.

Reliabilitas pada dasarnya mengacu pada sejauh mana suatu alat ukur mampu memberikan hasil yang konsisten dan stabil ketika digunakan dalam kondisi yang sama atau serupa. Ketika suatu tes atau instrumen evaluasi memiliki reliabilitas tinggi, maka skor atau hasil yang diperoleh peserta didik akan relatif sama jika tes tersebut diulang dalam waktu yang berbeda namun dalam kondisi yang setara. Sebaliknya, jika reliabilitas suatu alat ukur rendah, maka hasil yang diperoleh menjadi fluktuatif, sehingga sulit untuk menarik kesimpulan yang valid tentang kemampuan atau pemahaman peserta didik. Oleh karena itu, reliabilitas tidak hanya sekadar aspek teknis, melainkan juga menyangkut keakuratan dan kredibilitas suatu proses evaluasi pembelajaran.

Pentingnya reliabilitas alat ukur tidak terbatas pada instrumen tes akademik seperti ujian pilihan ganda, uraian, atau tes kinerja saja. Dalam konteks pendidikan yang lebih luas, banyak instrumen non-tes yang digunakan untuk mengukur aspek sikap, minat, motivasi, perilaku, dan kompetensi sosial peserta didik. Instrumen tersebut dapat berupa angket, lembar observasi, skala penilaian, wawancara terstruktur, hingga portofolio. Dalam semua bentuk instrumen tersebut, reliabilitas tetap menjadi faktor utama yang harus diperhatikan. Instrumen yang tidak reliabel akan menghasilkan data yang tidak konsisten, sehingga dapat menurunkan kualitas kesimpulan penelitian maupun efektivitas evaluasi pembelajaran.

Dalam praktiknya, reliabilitas sering kali diukur dengan berbagai pendekatan, misalnya dengan metode *test-retest*, reliabilitas bentuk paralel, konsistensi internal, serta metode belah dua (*split-half*). Masing-masing pendekatan tersebut memiliki kelebihan dan kelemahan, namun secara umum semuanya bertujuan untuk mengukur tingkat keajegan hasil pengukuran. Misalnya, dalam metode *test-retest*, reliabilitas diukur dengan cara memberikan tes yang sama kepada kelompok peserta didik dalam dua waktu yang berbeda dan kemudian membandingkan hasilnya. Jika hasilnya konsisten, maka reliabilitasnya dapat dikatakan tinggi. Pendekatan ini menunjukkan bahwa reliabilitas tidak hanya berbicara tentang alat ukur itu sendiri, tetapi juga berkaitan dengan stabilitas respon peserta didik terhadap pengukuran.

Dalam konteks pembelajaran di sekolah, reliabilitas alat ukur sangat penting karena berkaitan langsung dengan keadilan dan akurasi penilaian. Misalnya, jika sebuah tes tidak reliabel, maka peserta didik yang memiliki kemampuan sama dapat memperoleh skor yang sangat berbeda. Hal ini tidak hanya merugikan peserta didik secara individual, tetapi juga dapat mempengaruhi persepsi terhadap kualitas proses pembelajaran itu sendiri. Sebaliknya, instrumen dengan reliabilitas tinggi dapat memberikan gambaran yang lebih objektif mengenai pencapaian belajar peserta didik, sehingga guru dapat merancang strategi pembelajaran dan intervensi yang lebih tepat sasaran.

Selain itu, reliabilitas juga memiliki kaitan erat dengan validitas alat ukur. Instrumen yang reliabel belum tentu valid, tetapi instrumen yang valid harus reliabel. Dengan kata lain, reliabilitas merupakan syarat dasar bagi tercapainya validitas. Oleh karena itu, dalam proses pengembangan instrumen evaluasi pendidikan, pengujian reliabilitas menjadi tahap penting yang tidak boleh dilewatkan. Pengujian ini dapat

dilakukan melalui analisis statistik, seperti koefisien reliabilitas (misalnya Cronbach's Alpha) yang memberikan gambaran tentang konsistensi internal butir-butir tes atau instrumen.

Dalam penelitian pendidikan, reliabilitas juga berperan penting dalam menjamin kredibilitas temuan penelitian. Data yang dikumpulkan menggunakan instrumen yang tidak reliabel akan menghasilkan kesimpulan yang lemah dan sulit dipertanggungjawabkan secara ilmiah. Sebaliknya, instrumen yang reliabel memungkinkan peneliti untuk menginterpretasikan hasil penelitian dengan lebih meyakinkan. Hal ini sangat relevan dalam era pendidikan modern yang menuntut transparansi, akuntabilitas, dan kualitas pembelajaran yang tinggi.

Dengan demikian, reliabilitas alat ukur bukan hanya sekadar konsep teknis dalam evaluasi pembelajaran, melainkan menjadi fondasi penting dalam memastikan kualitas proses dan hasil pendidikan. Melalui pemahaman dan penerapan reliabilitas yang baik, para pendidik, peneliti, dan pengambil kebijakan dapat mengambil keputusan yang lebih tepat, objektif, dan berdampak positif terhadap peningkatan mutu pendidikan.

Salah satu bagian penting dalam proses pengukuran adalah reliabilitas alat ukur. Reliabilitas artinya tingkat keajegan atau konsistensi alat ukur dalam menghasilkan data atau informasi. Alat ukur yang reliabel akan memberikan hasil yang stabil dan dapat dipercaya ketika digunakan dalam kondisi yang sama di waktu yang berbeda. Jika alat ukur tidak reliabel, hasil yang diperoleh bisa bervariasi dan tidak konsisten, sehingga berpotensi menyebabkan kesimpulan yang salah. Hal ini sangat berpengaruh terhadap proses evaluasi pembelajaran karena keputusan yang diambil bisa terganggu.

## **METODE PENELITIAN**

Pendekatan ini dipilih karena penelitian fokus pada pengkajian mendalam mengenai konsep reliabilitas alat ukur, jenis-jenisnya, cara pengukurannya, serta faktor-faktor yang memengaruhinya dalam pembelajaran, berdasarkan teori dan hasil penelitian sebelumnya, tanpa mengumpulkan data langsung di lapangan. Sumber data berasal dari literatur primer dan sekunder. Literatur primer mencakup buku teks, artikel jurnal baik nasional maupun internasional, serta hasil penelitian ilmiah yang relevan. Literatur sekunder berupa laporan penelitian, skripsi, tesis, disertasi, dan sumber daring yang terpercaya, seperti Google Scholar, ResearchGate, dan Garuda Kemdikbud. Literatur yang digunakan dibatasi pada publikasi dalam 10 tahun terakhir untuk menjaga relevansi kajian.

Cara mengumpulkan data dilakukan melalui penelusuran dan dokumentasi literatur. Peneliti mencari, memilih, dan mengelompokkan sumber sesuai dengan fokus penelitian. Analisis data menggunakan analisis isi (content analysis), yaitu menelaah dan mengelompokkan isinya ke dalam beberapa tema utama, yaitu: (1) pengertian reliabilitas, (2) jenis-jenis reliabilitas, (3) metode pengukuran, dan (4) faktor-faktor yang memengaruhi reliabilitas alat ukur.

Untuk meningkatkan validitas hasil, dilakukan triangulasi sumber, yaitu membandingkan berbagai literatur agar informasi yang diperoleh konsisten. Seluruh literatur dicatat dan dikutip secara sistematis berdasarkan aturan penulisan ilmiah. Pendekatan ini diharapkan mampu memberikan pemahaman teoretis yang kuat tentang reliabilitas alat ukur dan penerapannya dalam evaluasi pembelajaran.

## PEMBAHASAN

### 1. Konsep Dasar Reliabilitas Alat Ukur

Dalam bidang pengukuran dan evaluasi pendidikan, reliabilitas merupakan salah satu komponen fundamental yang menentukan kualitas suatu instrumen penilaian. Reliabilitas mengacu pada tingkat konsistensi atau keajegan hasil pengukuran yang diperoleh ketika suatu alat ukur digunakan berulang kali dalam kondisi yang relatif sama (Robert L. Linn & Norman E. Gronlund, 2000). Dengan kata lain, semakin tinggi reliabilitas suatu alat ukur, semakin besar kemungkinan bahwa hasil yang diperoleh mencerminkan kemampuan atau kondisi peserta didik secara stabil dan dapat dipercaya. Dalam konteks pembelajaran, hal ini sangat penting karena hasil pengukuran akan menjadi dasar pengambilan keputusan pendidikan, seperti penentuan keberhasilan belajar, kelulusan, maupun intervensi pembelajaran.

Reliabilitas sering kali dikaitkan dengan keakuratan hasil evaluasi. Instrumen yang memiliki reliabilitas tinggi akan menghasilkan skor yang konsisten meskipun digunakan oleh penguji yang berbeda atau pada waktu yang berbeda, selama kondisi pengukuran relatif sama (J. P. Guilford & Benjamin Fruchter, 1978). Sebaliknya, jika reliabilitas rendah, maka perbedaan skor dapat terjadi bukan karena perubahan kemampuan peserta didik, tetapi karena kelemahan instrumen itu sendiri. Oleh karena itu, reliabilitas berperan penting dalam menjamin objektivitas hasil pengukuran dan mencegah kesalahan interpretasi terhadap prestasi belajar siswa.

Menurut Lee J. Cronbach (1951), reliabilitas dapat dipahami sebagai proporsi varians skor yang disebabkan oleh perbedaan kemampuan aktual peserta didik terhadap total varians skor. Dengan demikian, reliabilitas berhubungan erat dengan sejauh mana hasil pengukuran bebas dari kesalahan acak (*random error*). Instrumen yang reliabel akan meminimalkan kesalahan pengukuran dan menghasilkan data yang lebih stabil. Konsep ini sangat relevan dalam pembelajaran karena penilaian yang tidak reliabel dapat berimplikasi pada ketidakadilan bagi peserta didik dan kesalahan kebijakan pendidikan.

Selain itu, reliabilitas tidak dapat dilepaskan dari konsep validitas. Menurut Robert L. Thorndike dan Elizabeth Hagen (1977), reliabilitas merupakan prasyarat bagi validitas. Instrumen tidak mungkin valid apabila tidak reliabel, meskipun instrumen yang reliabel belum tentu valid. Artinya, reliabilitas lebih menekankan pada konsistensi hasil pengukuran, sedangkan validitas menekankan ketepatan pengukuran terhadap konstruk yang diukur. Dalam konteks pembelajaran, tes yang reliabel akan membantu memastikan bahwa perbedaan skor siswa benar-benar mencerminkan perbedaan dalam penguasaan materi pelajaran, bukan karena faktor lain seperti kondisi emosional, lingkungan ujian, atau instrumen yang kurang baik.

Lebih lanjut, menurut Anastasia Anne dan Susana Urbina (1997), reliabilitas dapat diklasifikasikan menjadi beberapa bentuk, antara lain: (1) reliabilitas *test-retest* yang mengukur kestabilan hasil dari waktu ke waktu, (2) reliabilitas bentuk paralel yang menguji konsistensi antar versi tes, dan (3) reliabilitas konsistensi internal yang mengukur homogenitas butir-butir dalam satu tes. Dalam dunia pendidikan, terutama di sekolah, reliabilitas konsistensi internal sering menjadi pilihan karena relatif mudah dilakukan menggunakan koefisien statistik seperti Cronbach's Alpha.

Dalam praktik pembelajaran, reliabilitas alat ukur memiliki peran strategis. Misalnya, ketika guru menyusun ujian tengah semester, instrumen yang reliabel akan memberikan gambaran yang konsisten mengenai kemampuan akademik siswa. Jika tes tidak reliabel, maka siswa dengan kemampuan setara dapat memperoleh nilai yang sangat berbeda, yang pada akhirnya dapat menimbulkan ketidakadilan dalam proses penilaian. Oleh karena itu, setiap guru atau peneliti pendidikan perlu memperhatikan aspek reliabilitas sejak tahap perencanaan instrumen, bukan hanya saat pelaksanaan.

Dengan demikian, konsep dasar reliabilitas alat ukur dalam pembelajaran menekankan pentingnya keajegan, stabilitas, dan konsistensi hasil pengukuran. Instrumen dengan reliabilitas tinggi akan meningkatkan kepercayaan terhadap hasil evaluasi, memperkuat dasar pengambilan keputusan pendidikan, dan menjamin proses penilaian yang lebih objektif dan adil bagi seluruh peserta didik.

## 2. Jenis-Jenis Reliabilitas

Reliabilitas merupakan aspek fundamental dalam pengukuran dan evaluasi pendidikan. Salah satu cara memahami reliabilitas secara lebih mendalam adalah dengan mengenali berbagai jenis-jenis reliabilitas yang dapat digunakan untuk menilai konsistensi instrumen. Setiap jenis reliabilitas memiliki fokus pengukuran dan prosedur pengujian yang berbeda, namun semuanya bertujuan untuk memastikan bahwa alat ukur memberikan hasil yang stabil dan dapat dipercaya. Dalam konteks pembelajaran, pemilihan jenis reliabilitas harus disesuaikan dengan karakteristik instrumen dan tujuan penilaian (Robert L. Linn & Norman E. Gronlund, 2000).

### a. Reliabilitas Test-Retest (Stability Reliability)

Reliabilitas *test-retest* mengukur kestabilan hasil pengukuran dari waktu ke waktu. Caranya adalah dengan memberikan tes yang sama kepada kelompok peserta didik dalam dua kesempatan yang berbeda dengan interval waktu tertentu, kemudian menghitung koefisien korelasi antara kedua hasil tes tersebut (Anastasia Anne & Susana Urbina, 1997). Jika koefisien korelasi tinggi, berarti instrumen tersebut memiliki reliabilitas yang baik. Jenis reliabilitas ini sangat cocok digunakan pada instrumen yang mengukur kemampuan yang relatif stabil, seperti kemampuan kognitif umum. Instrumen ini digunakan jika pendidik ingin mengetahui kestabilan hasil tes dari waktu ke waktu, misalnya untuk mengukur konsistensi pemahaman siswa terhadap suatu materi pelajaran dalam periode tertentu. Jenis tes yang cocok adalah tes hasil belajar (*achievement test*), tes sikap (*attitude test*), atau tes kemampuan dasar yang tidak berubah secara cepat. Contohnya adalah guru memberikan tes pemahaman konsep matematika pada awal minggu dan mengulang tes yang sama dua minggu kemudian untuk melihat apakah hasilnya konsisten. Namun, kelemahannya adalah rentan terhadap efek belajar atau perubahan kondisi peserta didik selama jeda waktu.

### b. Reliabilitas Bentuk Paralel (Equivalent Forms Reliability)

Jenis reliabilitas ini mengukur konsistensi hasil pengukuran antara dua bentuk tes yang setara dalam isi, tingkat kesulitan, dan tujuan pengukuran (J. P. Guilford & Benjamin Fruchter, 1978). Misalnya, guru menyusun dua versi ujian tengah semester dengan soal berbeda tetapi memiliki tingkat kesulitan yang sama. Kedua tes diberikan kepada kelompok siswa yang sama atau setara, kemudian hasilnya

dikorelasikan. Jika koefisien korelasinya tinggi, reliabilitas bentuk paralel dianggap baik. Jenis reabilitas ini digunakan saat peneliti ingin menghindari efek ingatan (*memory effect*) dan mengukur konsistensi antar bentuk tes yang setara. Jenis tes yang cocok Tes pilihan ganda atau bentuk objektif yang dapat dibuat dalam dua versi setara (misalnya ujian tryout dan ujian akhir). Salah satu contoh penggunaannya diantaranya dua bentuk tes Bahasa Indonesia dengan tingkat kesulitan yang sama diberikan pada waktu yang berdekatan untuk melihat kesamaan hasil. Jenis reliabilitas ini cocok untuk ujian berskala besar seperti ujian semester atau ujian seleksi masuk

Kelebihan metode ini adalah mengurangi efek hafalan dari tes pertama, namun kelemahannya adalah membutuhkan waktu dan tenaga lebih untuk membuat dua tes yang benar-benar sepadan.

c. Reliabilitas Konsistensi Internal (Internal Consistency Reliability)

Reliabilitas konsistensi internal mengukur homogenitas butir-butir soal dalam satu tes, yakni sejauh mana setiap butir soal mengukur konstruk yang sama. Salah satu teknik yang umum digunakan untuk mengukur jenis reliabilitas ini adalah koefisien Cronbach's Alpha, yang banyak digunakan dalam penelitian dan penilaian pendidikan (Lee J. Cronbach, 1951). Semakin tinggi nilai alpha (mendekati 1), semakin baik konsistensi internalnya. Jenis reliabilitas ini sering digunakan pada instrumen seperti angket sikap, skala psikologis, dan tes prestasi akademik. Jenis reliabilitas ini digunakan jika ingin mengetahui sejauh mana butir-butir dalam satu tes mengukur konstruk yang sama secara konsisten. Jenis tes yang cocok antara lain tes pilihan ganda, tes uraian, angket skala Likert, tes kompetensi tunggal. Contoh penggunaan antara lain untuk tes pemahaman membaca dengan 40 soal dianalisis menggunakan koefisien alpha untuk melihat konsistensi antar soal. Sangat umum digunakan dalam penelitian pendidikan dan psikologi, terutama pada satu waktu tes saja.

d. Reliabilitas Belah Dua (Split-Half Reliability)

Metode ini mengukur reliabilitas dengan cara membagi butir-butir soal dalam satu tes menjadi dua bagian yang setara, kemudian menghitung korelasi antara skor kedua bagian tersebut. Koefisien reliabilitas kemudian disesuaikan menggunakan rumus Spearman-Brown (Charles Spearman & William Brown, 1910). Kelebihan metode ini adalah tidak memerlukan dua kali pengukuran, namun kelemahannya adalah hasilnya dapat bervariasi tergantung cara pembagian butir tes. Instrumen ini digunakan

saat hasil tes bergantung pada penilaian manusia (subjektif), sehingga konsistensi antar penilai menjadi penting.. Jenis tes yang cocok antara lain tes kinerja (*performance test*), tes praktik, penilaian presentasi, penilaian esai. Salah satu contoh penggunaan diantaranya jika ada dua guru memberikan skor terhadap presentasi siswa dan kemudian hasilnya dibandingkan untuk mengukur konsistensi penilaian. Cocok untuk pembelajaran berbasis proyek dan penilaian autentik.

e. Inter-Rater Reliability (Reliabilitas Antar Penilai)

Jenis reliabilitas ini digunakan ketika instrumen melibatkan penilaian subyektif dari beberapa penilai, seperti penilaian kinerja, observasi kelas, atau tes praktik.

Reliabilitas antar penilai mengukur tingkat kesepakatan antar penilai terhadap skor yang diberikan pada peserta didik. Semakin tinggi tingkat kesepakatan, semakin reliabel instrumen tersebut (Robert L. Thorndike & Elizabeth Hagen, 1977). Penggunaan rubrik penilaian yang jelas dapat membantu meningkatkan reliabilitas antar penilai.

Dalam praktik pembelajaran, pemilihan jenis reliabilitas sangat bergantung pada jenis instrumen dan tujuan penilaian. Misalnya, tes objektif seperti pilihan ganda lebih cocok diuji dengan reliabilitas konsistensi internal atau belah dua, sedangkan penilaian kinerja atau portofolio lebih cocok diuji menggunakan reliabilitas antar penilai. Dengan memahami berbagai jenis reliabilitas ini, pendidik dapat memastikan bahwa instrumen penilaian yang digunakan benar-benar menghasilkan data yang stabil, konsisten, dan dapat dipercaya sebagai dasar pengambilan keputusan pendidikan.

### 3. Cara Mengukur Reliabilitas

Reliabilitas instrumen merupakan tingkat konsistensi hasil pengukuran suatu alat ukur ketika digunakan dalam kondisi yang sama. Dalam konteks pembelajaran, reliabilitas sangat penting untuk memastikan bahwa skor tes mencerminkan kemampuan atau atribut yang sebenarnya dan bukan akibat faktor kebetulan atau kesalahan pengukuran (Psychological Testing and Assessment, Cohen & Swerdlik, 2018).

Terdapat beberapa teknik statistik yang umum digunakan dalam mengukur reliabilitas, di antaranya: Korelasi Product Moment, Kuder Richardson (KR-20 dan KR-21), Alpha Cronbach, dan Split-Half. Masing-masing teknik memiliki karakteristik dan penggunaan yang berbeda sesuai dengan jenis tes dan data yang diperoleh.

#### a. Korelasi Product Moment

Metode Korelasi Product Moment digunakan untuk mengukur konsistensi hasil tes melalui hubungan linear antara dua set skor, misalnya pada metode test-retest atau bentuk paralel. Teknik ini umumnya menggunakan rumus Pearson Product Moment. Jika koefisien korelasi mendekati 1, maka reliabilitas tes dikatakan tinggi; sebaliknya jika mendekati 0, reliabilitasnya rendah. Metode ini cocok untuk tes objektif atau tes yang dilakukan dua kali pada kelompok yang sama dalam rentang waktu yang tidak terlalu jauh. Contoh penggunaannya yaitu pada saat menguji reliabilitas tes matematika yang diberikan dua kali dalam selang waktu dua minggu.

#### b. Kuder-Richardson (KR-20 dan KR-21)

Metode Kuder-Richardson digunakan untuk mengukur reliabilitas tes dengan item dikotomi (benar-salah, ya-tidak).

- KR-20 digunakan untuk tes dengan tingkat kesulitan butir yang bervariasi.
- KR-21 merupakan bentuk yang lebih sederhana dan digunakan jika semua butir dianggap memiliki tingkat kesulitan yang sama.

Kedua metode ini mengukur konsistensi internal instrumen. KR-20 lebih akurat, tetapi KR-21 lebih praktis. Koefisien reliabilitas yang tinggi menunjukkan bahwa butir-butir soal dalam tes tersebut homogen dan mengukur konstruk yang sama. Contoh penggunaan: Tes objektif pilihan ganda atau tes benar-salah dalam ujian akhir semester.

c. Alpha Cronbach

Alpha Cronbach digunakan untuk mengukur reliabilitas instrumen dengan skala lebih dari dua kategori jawaban, seperti skala Likert, rating scale, atau soal dengan bobot nilai tertentu. Teknik ini menghitung rata-rata korelasi antar butir soal untuk menentukan sejauh mana semua item dalam tes mengukur konstruk yang sama. Nilai  $\alpha$  (alpha) berkisar antara 0–1. Semakin mendekati 1, maka reliabilitasnya semakin tinggi. Secara umum, nilai  $\alpha \geq 0,70$  dianggap cukup baik untuk penelitian pendidikan. Metode ini sangat populer karena hanya membutuhkan satu kali pengambilan data dan efektif untuk mengukur konsistensi internal. Contoh penggunaannya adalah kuesioner persepsi siswa terhadap pembelajaran, angket motivasi belajar, atau instrument skala sikap.

d. Teknik Belah Dua (Split-Half)

Split-Half adalah metode pengukuran reliabilitas dengan cara membagi tes menjadi dua bagian yang setara, kemudian menghitung korelasi antara skor kedua belahan tersebut. Tes biasanya dibagi menjadi butir ganjil dan genap atau secara acak tetapi seimbang. Koefisien korelasi hasil perhitungan kemudian disesuaikan menggunakan rumus Spearman-Brown untuk memperkirakan reliabilitas tes penuh. Teknik ini praktis karena tidak membutuhkan dua kali pengukuran, namun pembagian tes harus benar-benar seimbang agar hasilnya akurat (Introduction to Classical and Modern Test Theory, Crocker & Algina, 2008). Contoh penggunaan adalah tes objektif dengan banyak butir soal seperti ujian pilihan ganda.

4. Faktor-Faktor yang Mempengaruhi Reliabilitas

Faktor yang mempengaruhi reliabilitas alat ukur antara lain:

a. Jumlah butir soal

Salah satu faktor utama yang memengaruhi reliabilitas adalah jumlah butir soal dalam suatu tes. Semakin banyak butir soal yang relevan dan berkualitas, maka reliabilitas tes cenderung meningkat. Hal ini karena dengan jumlah butir soal yang lebih banyak, peluang untuk mengukur kompetensi peserta didik secara lebih menyeluruh menjadi lebih besar. Instrumen dengan butir soal yang terlalu sedikit rentan menghasilkan hasil pengukuran yang tidak stabil dan kurang representatif terhadap kemampuan yang sebenarnya (Anastasi & Urbina, 1997; Purwanto, 2018).

b. Kejelasan instruksi

Instruksi atau petunjuk pengerjaan yang jelas dan mudah dipahami sangat penting untuk memastikan peserta tes mengerjakan instrumen sesuai tujuan pengukuran. Instruksi yang ambigu atau membingungkan dapat menimbulkan variasi yang tidak diinginkan dalam cara peserta memahami dan menjawab soal, sehingga menurunkan konsistensi hasil tes (Arikunto, 2013). Instruksi yang baik membantu mengurangi kesalahan non-sistematis yang dapat mengganggu reliabilitas.

c. Kondisi peserta tes

Faktor internal peserta, seperti kesehatan fisik, kondisi psikologis, tingkat kelelahan, dan motivasi belajar, sangat berpengaruh terhadap reliabilitas hasil tes. Peserta yang dalam kondisi tidak optimal cenderung memberikan jawaban yang tidak konsisten, sehingga menurunkan stabilitas hasil pengukuran. Oleh karena itu, kondisi peserta tes perlu diperhatikan untuk memastikan keandalan data (Sugiyono, 2019)



d. Waktu pelaksanaan

Waktu pelaksanaan tes yang kurang memadai, baik terlalu singkat atau terlalu lama, dapat memengaruhi fokus dan konsentrasi peserta didik. Waktu yang cukup memungkinkan peserta menyelesaikan soal dengan baik sesuai kemampuannya. Sebaliknya, waktu yang tidak sesuai dapat menimbulkan kesalahan dalam menjawab, yang akhirnya berdampak pada rendahnya reliabilitas instrumen (Arifin, 2017).

e. Kesesuaian bentuk tes dengan tujuan pengukuran

5. Implikasi Reliabilitas dalam Pembelajaran

Reliabilitas alat ukur memiliki implikasi yang sangat penting dalam proses pembelajaran. Dalam konteks evaluasi pendidikan, reliabilitas menunjukkan sejauh mana suatu instrumen penilaian mampu menghasilkan hasil yang konsisten dan stabil dari waktu ke waktu. Semakin tinggi reliabilitas suatu alat ukur, semakin besar kepercayaan pendidik terhadap hasil pengukuran yang diperoleh (Arikunto, 2013).

- a. Reliabilitas berimplikasi pada akurasi penilaian hasil belajar. Ketika instrumen penilaian memiliki reliabilitas tinggi, guru dapat lebih yakin bahwa perbedaan skor yang muncul benar-benar mencerminkan perbedaan kemampuan peserta didik, bukan akibat kesalahan pengukuran. Hal ini sangat penting untuk pengambilan keputusan yang adil dan tepat, seperti penentuan ketuntasan belajar, pemberian nilai, maupun rekomendasi tindak lanjut pembelajaran (Nitko & Brookhart, 2011).
- b. Reliabilitas mendukung validitas instrumen penilaian. Meskipun reliabilitas tidak menjamin validitas, sebuah tes yang tidak reliabel pasti tidak valid. Artinya, reliabilitas menjadi syarat dasar agar instrumen benar-benar dapat mengukur apa yang seharusnya diukur. Jika suatu alat ukur tidak konsisten, maka sulit memastikan bahwa hasilnya menggambarkan kompetensi sebenarnya dari peserta didik (Anastasi & Urbina, 1997). Dalam pembelajaran, hal ini dapat berdampak pada kesalahan interpretasi terhadap capaian belajar.
- c. Reliabilitas berimplikasi pada perencanaan pembelajaran yang efektif. Guru dan pendidik membutuhkan informasi yang akurat mengenai kemampuan siswa untuk menyusun rencana pembelajaran lanjutan. Instrumen yang reliabel memberikan data yang dapat diandalkan untuk menentukan strategi pembelajaran remedial, pengayaan, atau diferensiasi instruksional (Purwanto, 2018). Tanpa reliabilitas, perencanaan ini dapat menjadi tidak tepat sasaran.
- d. Reliabilitas mendukung objektivitas dalam evaluasi. Dalam pembelajaran, evaluasi yang objektif sangat penting untuk menjamin keadilan bagi seluruh peserta didik. Tes yang reliabel akan meminimalkan pengaruh faktor-faktor luar seperti kondisi lingkungan, kelelahan, atau perbedaan interpretasi instruksi, sehingga hasil evaluasi menjadi lebih konsisten (Arifin, 2017). Objektivitas ini juga penting dalam konteks asesmen berskala besar seperti ujian nasional atau asesmen sumatif.
- e. Reliabilitas juga berimplikasi pada peningkatan kualitas pembelajaran secara keseluruhan. Ketika guru menggunakan instrumen yang reliabel, mereka dapat memperoleh data evaluasi yang akurat, yang kemudian dapat digunakan untuk refleksi dan perbaikan praktik pembelajaran. Data yang konsisten membantu guru

mengidentifikasi area kekuatan dan kelemahan peserta didik serta efektivitas metode mengajar yang digunakan (Sugiyono, 2019).

Dengan demikian, reliabilitas bukan hanya sekadar aspek teknis dalam penyusunan instrumen penilaian, tetapi juga berdampak langsung terhadap mutu proses dan hasil pembelajaran. Oleh karena itu, guru, perancang tes, dan peneliti pendidikan perlu memberikan perhatian serius terhadap pengujian reliabilitas dalam setiap kegiatan evaluasi pembelajaran.

## SIMPULAN DAN SARAN

1. Reliabilitas merupakan komponen fundamental dalam proses evaluasi pembelajaran yang berperan penting dalam menjamin konsistensi dan keandalan hasil pengukuran. Instrumen penilaian yang memiliki reliabilitas tinggi akan menghasilkan skor yang stabil dan dapat dipercaya, sehingga keputusan pembelajaran yang diambil berdasarkan data tersebut menjadi lebih akurat dan objektif. Dalam konteks pendidikan, reliabilitas tidak hanya relevan untuk tes akademik, tetapi juga untuk berbagai instrumen non-tes seperti angket, observasi, maupun wawancara terstruktur.
2. Terdapat beberapa jenis reliabilitas yang dapat digunakan sesuai karakteristik tes dan tujuan pengukuran, antara lain reliabilitas tes ulang (test-retest), bentuk paralel (parallel forms), konsistensi internal, serta reliabilitas antar penilai (inter-rater). Masing-masing jenis memiliki waktu dan konteks penggunaan yang berbeda, namun semuanya bertujuan untuk memastikan konsistensi hasil pengukuran.
3. Metode pengukuran reliabilitas pun bervariasi, seperti Korelasi Product Moment, Kuder-Richardson (KR-20 dan KR-21), Alpha Cronbach, dan teknik belah dua (split-half). Pemilihan metode pengukuran harus disesuaikan dengan jenis instrumen dan karakteristik butir soal agar hasil perhitungan lebih akurat.
4. Reliabilitas juga dipengaruhi oleh beberapa faktor penting, antara lain jumlah butir soal, kejelasan instruksi, kondisi peserta tes, waktu pelaksanaan, dan kesesuaian bentuk tes dengan tujuan pengukuran. Faktor-faktor ini perlu diperhatikan dengan cermat oleh pendidik maupun penyusun instrumen agar reliabilitas dapat ditingkatkan.
5. Implikasinya terhadap pembelajaran sangat besar. Reliabilitas mendukung keakuratan penilaian, validitas instrumen, objektivitas evaluasi, serta efektivitas perencanaan pembelajaran. Oleh karena itu, reliabilitas bukan hanya sekadar aspek teknis dalam penyusunan tes, tetapi juga merupakan landasan penting bagi peningkatan mutu proses dan hasil pembelajaran secara keseluruhan.

## DAFTAR PUSTAKA

- Anastasi, A., & Urbina, S. (1997). *Psychological Testing*. Prentice Hall.
- Arifin, Z. (2017). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. (2020). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, S. (2019). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Bryman, A. (2016). *Social Research Methods*. Oxford University Press.
- Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment*. McGraw-Hill.

- Creswell, J. W. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods*. SAGE.
- Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. *Psychometrika*, 16(3), 297–334.
- DeVellis, R. F. (2017). *Scale Development: Theory and Applications*. SAGE.
- Field, A. (2018). *Discovering Statistics Using SPSS*. SAGE.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to Design and Evaluate Research in Education*. McGraw-Hill.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2018). *Educational Research*. Boston: Pearson.
- Guilford, J. P. (1965). *Fundamental Statistics in Psychology and Education*. McGraw-Hill.
- Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological Testing: Principles, Applications, and Issues*. Cengage.
- Kerlinger, F. N. (1986). *Foundations of Behavioral Research*. Holt, Rinehart and Winston.
- Kline, P. (2015). *A Handbook of Test Construction*. Routledge.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching*. Prentice Hall.
- Loewenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales*. Psychology Press.
- Messick, S. (1995). *Validity of psychological assessment*. *American Psychologist*.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of Students*. Pearson Education.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Polit, D. F., & Beck, C. T. (2017). *Nursing Research*. Wolters Kluwer.
- Purwanto, N. (2018). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
- Spearman, C., & Brown, W. (1910). *The Spearman-Brown Formula for Reliability*. *British Journal of Psychology*.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2019). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Tavakol, M., & Dennick, R. (2011). *Making sense of Cronbach's alpha*. *International Journal of Medical Education*.
- Thorndike, R. L., & Hagen, E. (1977). *Measurement and Evaluation in Psychology and Education*. Wiley.
- (Classroom Assessment: What Teachers Need to Know – W. James Popham, 2014).
- (Educational Measurement – Robert L. Linn & Norman E. Gronlund, 2000).
- (Introduction to Classical and Modern Test Theory – Linda Crocker & James Algina, 2008).
- (Measurement and Assessment in Education – Cecil R. Reynolds & Robert A. Livingston, 2012).
- (Psychological Testing and Assessment – Ronald Jay Cohen & Mark Swerdlik, 2018).