

Kajian Validitas dan Reliabilitas Tes dalam Evaluasi Pembelajaran

Nadiva Syahira Daulay¹, Nur Fadilah Harahap², Fitrah Hayati³, Sonia Pertiwi⁴,
Masitoh⁴, Ahmad Husein⁵

Pendidikan Guru Madrasah Ibtidaiyah, Sekolah Tinggi Agama Islam Negeri Mandailing Natal^{1,2,3,4,5}

*Email Korespondensi: nadivasyahirad@gmail.com

Sejarah Artikel:

Diterima 24-06-2026
Disetujui 28-06-2026
Diterbitkan 30-06-2026

ABSTRACT

This study aims to examine the concepts of validity and reliability as two fundamental aspects that determine the quality of research instruments. Validity refers to the extent to which an instrument measures what it is intended to measure, whereas reliability indicates the consistency of the measurement results obtained. The study employed a library research method by reviewing various relevant scientific sources, including books, journal articles, and academic documents discussing validity and reliability in research instruments. The findings reveal that validity consists of several types, namely content validity, construct validity, criterion validity, face validity, and empirical validity. Meanwhile, reliability includes test-retest reliability, parallel forms reliability, internal consistency reliability, inter-rater reliability, and intra-rater reliability. Each type of validity and reliability possesses distinct characteristics and testing procedures depending on the objectives and forms of the instruments used. Therefore, validity and reliability testing should be conducted systematically to ensure that research instruments produce accurate, consistent, and trustworthy data. This study is expected to serve as a reference for researchers and educational practitioners in understanding the importance of validity and reliability as the foundation for developing high-quality research instruments.

Keywords: validity, reliability, research instruments, educational evaluation, measurement

ABSTRAK

Penelitian ini bertujuan untuk mengkaji konsep validitas dan reliabilitas sebagai dua aspek utama yang menentukan kualitas instrumen penelitian. Validitas berkaitan dengan kemampuan instrumen dalam mengukur apa yang seharusnya diukur, sedangkan reliabilitas menunjukkan tingkat konsistensi hasil pengukuran yang diperoleh. Metode yang digunakan dalam penelitian ini adalah studi kepustakaan (library research) dengan mengkaji berbagai sumber ilmiah yang relevan, seperti buku, artikel jurnal, dan dokumen akademik yang membahas validitas dan reliabilitas instrumen penelitian. Hasil kajian menunjukkan bahwa validitas terdiri atas beberapa jenis, yaitu validitas isi, validitas konstruk, validitas kriteria, validitas muka, dan validitas empiris. Sementara itu, reliabilitas meliputi reliabilitas tes ulang, reliabilitas bentuk paralel, reliabilitas konsistensi internal, reliabilitas antarpemilai, dan reliabilitas intra-pemilai. Setiap jenis validitas dan reliabilitas memiliki karakteristik serta prosedur pengujian yang berbeda sesuai dengan tujuan dan bentuk instrumen yang digunakan. Oleh karena itu, pengujian validitas dan reliabilitas perlu dilakukan secara sistematis agar instrumen penelitian mampu menghasilkan data yang akurat, konsisten, dan dapat dipercaya. Kajian ini diharapkan dapat menjadi referensi bagi peneliti dan praktisi pendidikan dalam memahami pentingnya validitas dan reliabilitas sebagai dasar pengembangan instrumen penelitian yang berkualitas.

Kata kunci: validitas, reliabilitas, instrumen penelitian, evaluasi pendidikan, pengukuran.

Bagaimana Cara Sitasi Artikel ini:

Daulay, N. S., Harahap, N. F. ., Hayati, F. ., Pertiwi, S. ., Masitoh, M., & Husein, A. . (2026). Kajian Validitas dan Reliabilitas Tes dalam Evaluasi Pembelajaran. Jejak Digital: Jurnal Ilmiah Multidisiplin, 2(4), 6956-6965. <https://doi.org/10.63822/pdzmh775>

PENDAHULUAN

Sebagai bagian penting dalam sistem pendidikan, evaluasi digunakan untuk menilai sejauh mana tujuan pembelajaran telah tercapai oleh peserta didik. Kegiatan ini mencakup pengukuran berbagai aspek kemampuan, baik yang berhubungan dengan pengetahuan, sikap, maupun keterampilan (Febriani et al., 2026). Dalam praktiknya, tes menjadi salah satu alat evaluasi yang paling banyak digunakan untuk memperoleh informasi mengenai capaian dan perkembangan belajar peserta didik. Namun, informasi yang dihasilkan dari suatu tes hanya akan memiliki nilai guna apabila instrumen tersebut mampu memberikan hasil yang akurat dan dapat dipercaya. Oleh sebab itu, kualitas tes perlu diperhatikan secara serius, tidak hanya dari segi penyusunan soal, tetapi juga dari kemampuannya dalam mengukur kompetensi yang menjadi sasaran pengukuran serta menghasilkan skor yang konsisten.

Kualitas suatu instrumen evaluasi tidak terlepas dari validitas dan reliabilitas yang dimilikinya. Validitas berhubungan dengan kemampuan tes dalam mengukur aspek yang memang menjadi tujuan pengukuran. Dengan kata lain, tes yang valid mampu memberikan gambaran yang tepat mengenai kompetensi yang ingin diketahui. Ketika suatu instrumen tidak memiliki tingkat validitas yang memadai, hasil yang diperoleh berisiko menghasilkan informasi yang kurang akurat dan tidak sepenuhnya mencerminkan kemampuan peserta didik yang sesungguhnya.

Berbeda dengan validitas yang menekankan pada ketepatan pengukuran, reliabilitas berkaitan dengan tingkat kestabilan dan konsistensi hasil yang diperoleh dari suatu instrumen. Sebuah tes dikatakan reliabel apabila mampu menghasilkan skor yang relatif tetap ketika digunakan pada kondisi yang setara, baik dalam pelaksanaan yang berbeda maupun pada kelompok peserta yang memiliki karakteristik yang sebanding. Dengan demikian, reliabilitas menjadi indikator penting untuk memastikan bahwa hasil pengukuran tidak dipengaruhi oleh faktor-faktor yang bersifat kebetulan atau tidak relevan dengan tujuan pengukuran.

Pada kenyataannya, masih terdapat berbagai instrumen tes yang belum memenuhi kriteria validitas dan reliabilitas secara optimal. Kondisi ini dapat dipengaruhi oleh sejumlah faktor, antara lain keterbatasan pengetahuan pendidik dalam mengembangkan instrumen evaluasi, terbatasnya waktu yang tersedia untuk penyusunan dan pengujian instrumen, serta minimnya kemampuan dalam melakukan analisis butir soal. Akibatnya, data yang diperoleh dari proses evaluasi belum sepenuhnya mampu memberikan gambaran yang akurat mengenai hasil belajar peserta didik. Keadaan tersebut berpotensi mengurangi ketepatan penggunaan hasil evaluasi sebagai dasar untuk menilai keberhasilan pembelajaran maupun sebagai acuan dalam menyusun program tindak lanjut pembelajaran.

Dari uraian tersebut dapat dipahami bahwa validitas dan reliabilitas merupakan unsur yang sangat penting dalam penyusunan maupun penggunaan tes sebagai instrumen evaluasi. Kedua aspek tersebut menjadi dasar untuk memastikan bahwa hasil pengukuran yang diperoleh benar-benar akurat dan dapat dipercaya. Apabila suatu tes tidak memenuhi kriteria validitas dan reliabilitas, maka informasi yang dihasilkan berisiko tidak mencerminkan kemampuan peserta didik secara tepat (Sibarani, 2024). Oleh karena itu, pemahaman yang baik mengenai kedua konsep tersebut sangat diperlukan oleh pendidik sebagai bekal dalam menyusun instrumen evaluasi yang berkualitas guna mendukung peningkatan mutu pembelajaran dan pendidikan.

METODE PENELITIAN

Penulisan artikel ini menggunakan pendekatan studi kepustakaan dengan mengkaji berbagai sumber ilmiah yang berkaitan dengan tema yang dibahas, meliputi buku, jurnal, dan publikasi akademik lainnya. Proses kajian dilakukan melalui identifikasi, penelaahan, serta analisis terhadap berbagai konsep dan temuan penelitian yang berkaitan dengan validitas dan realibilitas tes. Informasi yang diperoleh dari berbagai sumber kemudian disintesis dan disusun secara sistematis sehingga membentuk suatu uraian yang utuh dan mudah dipahami. Melalui kajian ini, diharapkan pembaca memperoleh pemahaman konseptual yang lebih mendalam mengenai analisis butir soal serta dapat memanfaatkan hasil pembahasan sebagai referensi dalam pengembangan dan pelaksanaan evaluasi pembelajaran.

HASIL DAN PEMBAHASAN

1. Validitas Tes

Sebagai instrumen yang digunakan dalam proses evaluasi, tes harus memiliki tingkat validitas yang memadai agar hasil pengukuran yang diperoleh dapat dipercaya. Validitas menjadi aspek yang sangat penting karena berkaitan dengan ketepatan instrumen dalam mengukur kemampuan atau konstruk yang hendak dinilai. Dengan kata lain, kualitas suatu tes sangat dipengaruhi oleh kemampuannya dalam menghasilkan informasi yang sesuai dengan tujuan pengukuran yang telah ditentukan (Putri et al., 2023).

Validitas suatu tes tercermin dari tingkat kesesuaian antara tujuan evaluasi, materi yang diukur, dan hasil pengukuran yang diperoleh. Sebuah instrumen hanya dapat dikatakan valid apabila mampu memberikan informasi yang benar-benar merepresentasikan kompetensi yang menjadi fokus penilaian. Oleh karena itu, validitas mengharuskan adanya keterhubungan yang kuat antara aspek yang hendak diukur dengan alat ukur yang digunakan dalam proses evaluasi (Dewy et al., 2025).

Menurut azwar (2017) Validitas berasal dari istilah *validity* yang merujuk pada tingkat ketepatan suatu instrumen dalam melaksanakan fungsi pengukuran. Suatu pengukuran dianggap memiliki validitas yang tinggi apabila data yang dihasilkan mampu menggambarkan variabel yang diukur secara akurat sesuai dengan tujuan pengukuran yang telah ditentukan. Dengan demikian, validitas menunjukkan sejauh mana hasil yang diperoleh benar-benar merefleksikan aspek yang menjadi sasaran pengukuran. Sebaliknya, instrumen dengan validitas rendah cenderung menghasilkan data yang tidak sesuai dengan tujuan pengukuran sehingga kurang mampu memberikan gambaran yang tepat mengenai variabel yang diukur.

Berdasarkan berbagai definisi yang dikemukakan para ahli, validitas dapat dimaknai sebagai tingkat ketepatan instrumen dalam mengukur suatu variabel sesuai dengan tujuan pengukuran yang telah ditetapkan. Instrumen yang memiliki validitas tinggi akan menghasilkan data yang mampu menggambarkan secara akurat aspek yang menjadi sasaran pengukuran. Oleh karena itu, validitas menunjukkan sejauh mana hasil yang diperoleh benar-benar mencerminkan kompetensi atau karakteristik yang hendak diukur. Sebaliknya, apabila data yang dihasilkan tidak sesuai dengan tujuan pengukuran, maka instrumen tersebut memiliki validitas yang rendah sehingga informasi yang diperoleh kurang dapat dipercaya dalam merepresentasikan kondisi yang sebenarnya.

Untuk memberikan gambaran yang lebih sederhana, validitas dapat dianalogikan dengan penggunaan termometer untuk mengukur suhu tubuh. Ketika seseorang ingin mengetahui suhu tubuhnya, alat yang tepat digunakan adalah termometer karena memang dirancang untuk mengukur suhu. Jika termometer menunjukkan suhu tubuh yang sesuai dengan kondisi sebenarnya, maka alat tersebut dapat dikatakan valid. Namun, apabila seseorang menggunakan jam tangan untuk mengetahui suhu tubuh, hasil

yang diperoleh tentu tidak akan sesuai karena alat tersebut tidak memiliki fungsi untuk mengukur suhu. Analogi ini menunjukkan bahwa validitas berkaitan dengan kesesuaian antara alat yang digunakan dan tujuan pengukuran yang ingin dicapai.

Secara umum, validitas dikelompokkan ke dalam tiga jenis utama, yaitu validitas isi (*content validity*), validitas konstruk (*construct validity*), dan validitas berdasarkan kriteria (*criterion-related validity*) (Daulay et al., 2026). Adapun penjelasan nya sebagai berikut:

a. Validitas isi (*content validity*)

Validitas isi (*content validity*) mengacu pada tingkat keterwakilan isi instrumen terhadap seluruh aspek, indikator, atau ruang lingkup materi yang menjadi sasaran pengukuran. Suatu instrumen dinilai memiliki validitas isi yang baik apabila setiap butir pertanyaan atau soal yang disusun mampu mencerminkan tujuan pengukuran, materi yang diujikan, serta indikator kompetensi yang telah ditentukan sebelumnya secara tepat dan proporsional.

Validitas isi dapat dianalogikan seperti ujian Matematika yang dirancang untuk mengukur kemampuan siswa pada materi pecahan. Apabila soal-soal yang diberikan seluruhnya berkaitan dengan konsep pecahan sesuai indikator pembelajaran, maka ujian tersebut memiliki validitas isi yang baik. Namun, jika sebagian besar soal justru membahas geometri atau statistika, maka validitas isinya rendah karena isi tes tidak mewakili materi yang seharusnya diukur. Dengan kata lain, validitas isi berkaitan dengan kesesuaian antara isi instrumen dan materi yang menjadi tujuan pengukuran.

Validitas isi suatu instrumen dapat dianalisis menggunakan koefisien Aiken's V yang diperoleh berdasarkan penilaian para ahli (*expert judgment*) terhadap relevansi setiap butir instrumen. Adapun rumusnya dan contoh perhitungannya sebagai berikut:

$$V = \frac{\sum s}{n(c - 1)}$$

Keterangan:

V = Koefisien validitas Aiken

s = Skor yang telah dikoreksi

r = Skor yang diberikan oleh validator

l_o = Skor terendah pada skala penilaian

n = Jumlah validator

c = Jumlah kategori penilaian

Sebagai contoh, suatu instrumen terdiri atas lima butir soal yang dinilai oleh tiga orang ahli menggunakan skala penilaian 1–5, dengan skor 1 menunjukkan sangat tidak relevan dan skor 5 menunjukkan sangat relevan. Berikut ditunjukkan langkah perhitungan untuk butir soal nomor 1.

Tabel 1. Contoh Hasil Perhitungan Validitas Isi Butir Soal Menggunakan Indeks Aiken's V

No. Butir	Ahli 1	Ahli 2	Ahli 3	Σs	Aiken's V	Kategori
1	5	4	5	11	0,92	Sangat Valid
2	4	4	5	10	0,83	Sangat Valid
3	5	5	4	11	0,92	Sangat Valid
4	4	3	4	8	0,67	Valid
5	5	4	4	10	0,83	Sangat Valid

Adapun langkah-langkah nya yaitu:

- 1) Menentukan skor yang diberikan oleh validator

Validator	Skor (r)
Ahli 1	5
Ahli 2	4
Ahli 3	5

- 2) Menentukan parameter perhitungan

Komponen	Simbol	Nilai
Skor terendah	(l _o)	1
Skor tertinggi	-	5
Jumlah kategori penilaian	(c)	5
Jumlah validator	(n)	3

- 3) Menghitung Nilai $S = r - L_o$

Validator	Skor (r)	Skor Terendah ((l _o))	Perhitungan	Nilai s
Ahli 1	5	1	5-1	4
Ahli 2	4	1	4-1	3
Ahli 3	5	1	5-1	4
Jumlah				11

- 4) Menghitung nilai Aiken's V

Komponen	Nilai
(\sum s)	11
(n)	3
(c)	5
(n(c-1))	$3 \times (5-1) = 12$
Nilai Aiken's V	$11 \div 12 = 0,92$

- 5) Interpretasi Hasil Validitas Isi

Nilai Aiken's V	Kategori
0,80 – 1,00	Sangat Valid
0,60 – 0,79	Valid
0,40 – 0,59	Cukup Valid
< 0,40	Kurang Valid

Dalam pengujian validitas isi melalui penilaian para ahli (*expert judgment*), suatu butir instrumen dinyatakan valid apabila memperoleh nilai koefisien Aiken's V sebesar $\geq 0,30$.

b. Validitas konstruk (*construct validity*)

Suatu instrumen dikatakan memiliki validitas konstruk apabila butir-butir yang terdapat di dalamnya mampu mencerminkan konsep atau konstruk yang menjadi fokus pengukuran. Validitas ini menekankan kesesuaian antara landasan teoritis suatu variabel dan indikator yang

digunakan dalam instrumen. Untuk membuktikan validitas konstruk, diperlukan analisis teoritis yang sistematis terhadap variabel yang diteliti, mulai dari mendefinisikan konstruk secara konseptual, mengidentifikasi dimensi dan indikatornya, hingga menerjemahkannya ke dalam bentuk butir instrumen yang dapat digunakan untuk proses pengukuran (Ananda & Fhadli, 2018).

Validitas konstruk (*construct validity*) merupakan aspek penting dalam pengembangan instrumen yang menunjukkan kemampuan alat ukur untuk mengukur konstruk atau konsep teoretis secara tepat sesuai dengan karakteristik yang hendak diungkap. Validitas konstruk memastikan bahwa instrumen benar-benar merepresentasikan konstruk yang menjadi sasaran pengukuran sehingga hasil yang diperoleh dapat menggambarkan konsep yang diukur secara akurat (Umar & Nisa, 2020).

Validitas konstruk dapat dianalogikan seperti seseorang yang ingin mengukur kemampuan berenang. Kemampuan berenang tidak cukup dinilai hanya dari kemampuan menggerakkan tangan, tetapi juga harus mencakup kemampuan menggerakkan kaki, mengatur pernapasan, dan menjaga keseimbangan di air. Jika seluruh aspek tersebut diukur, maka instrumen tersebut memiliki validitas konstruk yang baik karena benar-benar merepresentasikan konstruk "kemampuan berenang" secara utuh.

Berdasarkan berbagai pendapat ahli, validitas konstruk (*construct validity*) dapat dipahami sebagai tingkat ketepatan suatu instrumen dalam mengukur konstruk atau konsep teoretis yang menjadi sasaran pengukuran. Validitas ini menunjukkan kesesuaian antara konstruk yang dirumuskan secara konseptual dengan indikator serta butir-butir yang digunakan dalam instrumen.

c. Validitas kriteria (*criterion-related validity*).

Validitas empiris atau validitas kriterium (*criterion-related validity*) merujuk pada tingkat keterkaitan antara skor yang diperoleh dari suatu tes dengan kriteria tertentu yang digunakan sebagai acuan pembandingan. Kriteria tersebut dapat berupa hasil pengukuran dari instrumen lain yang telah terbukti valid maupun indikator kinerja tertentu yang terdapat dalam situasi nyata. Dalam pelaksanaannya, validitas empiris terbagi menjadi dua bentuk, yakni validitas konkuren dan validitas prediktif.

Validitas konkuren (*concurrent validity*) menggambarkan derajat kesesuaian antara hasil suatu tes dan hasil pengukuran lain yang dilakukan pada periode waktu yang sama. Sebagai contoh, skor yang diperoleh pada tes matematika dapat dibandingkan dengan nilai rapor pada mata pelajaran matematika. Adapun validitas prediktif (*predictive validity*) menunjukkan kemampuan suatu tes dalam memperkirakan atau memprediksi capaian maupun performa individu pada masa mendatang. (Wardah, 2019).

Validitas kriteria dapat dianalogikan seperti penggunaan sebuah termometer untuk mengukur suhu tubuh seseorang. Untuk memastikan bahwa termometer tersebut benar-benar akurat, hasil pengukurannya dibandingkan dengan termometer standar yang telah terbukti ketepatannya. Apabila kedua alat menunjukkan hasil yang hampir sama, maka termometer tersebut dapat dianggap valid. Dalam konteks pengukuran pendidikan, prinsip yang sama berlaku pada validitas kriteria, yaitu sejauh mana hasil suatu tes memiliki kesesuaian dengan kriteria atau ukuran pembandingan yang telah dipercaya keabsahannya.

Misalnya, hasil tes matematika dibandingkan dengan nilai rapor matematika yang diperoleh pada waktu yang sama untuk melihat kesesuaiannya (validitas konkuren), atau hasil tes digunakan untuk memperkirakan prestasi belajar siswa di masa mendatang (validitas prediktif).

Dengan demikian, validitas kriteria menunjukkan kemampuan suatu instrumen untuk menghasilkan skor yang sejalan dengan ukuran pembanding atau mampu meramalkan kinerja yang akan datang.

2. Realibilitas tes

Menurut Yusup (2018), reliabilitas merupakan tingkat konsistensi suatu instrumen dalam memberikan hasil pengukuran. Instrumen yang reliabel akan menghasilkan data yang relatif sama ketika digunakan berulang kali untuk mengukur objek yang sama dalam kondisi yang serupa. Sejalan dengan itu, Saidah et al. (2026) menyatakan bahwa reliabilitas menunjukkan sejauh mana instrumen mampu memberikan hasil yang tetap dan dapat dipercaya apabila digunakan berulang kali.

Sementara itu, Subhaktiyasa (2024) menjelaskan bahwa reliabilitas merupakan aspek penting yang menentukan kualitas instrumen karena berkaitan dengan konsistensi hasil pengukuran yang dihasilkan. Berdasarkan ketiga pendapat tersebut, reliabilitas dapat disimpulkan sebagai tingkat konsistensi, kestabilan, dan keterpercayaan suatu instrumen dalam menghasilkan data yang relatif sama ketika digunakan berulang kali pada kondisi yang serupa.

Konsep reliabilitas dapat dianalogikan seperti sebuah timbangan badan yang digunakan setiap hari. Jika seseorang dengan berat badan yang sama menimbang dirinya beberapa kali dan timbangan selalu menunjukkan angka yang relatif sama, maka timbangan tersebut dapat dikatakan reliabel. Sebaliknya, apabila hasil pengukuran selalu berubah-ubah meskipun kondisi yang diukur tidak berubah, maka timbangan tersebut tidak reliabel. Dalam penelitian, instrumen yang reliabel berfungsi seperti timbangan yang konsisten, yaitu mampu memberikan hasil pengukuran yang stabil dan dapat dipercaya.

Berdasarkan ketiga pendapat tersebut, reliabilitas dapat disimpulkan sebagai tingkat konsistensi, kestabilan, dan keterpercayaan suatu instrumen dalam menghasilkan data ketika digunakan berulang kali pada kondisi yang sama. Semakin konsisten hasil yang diperoleh, semakin tinggi tingkat reliabilitas instrumen tersebut, sehingga data yang dihasilkan dapat dipercaya sebagai dasar pengambilan kesimpulan penelitian.

Dalam evaluasi pendidikan, reliabilitas terdiri atas beberapa jenis yang digunakan untuk mengkaji tingkat konsistensi suatu instrumen pengukuran. Setiap jenis reliabilitas memiliki karakteristik dan pendekatan yang berbeda dalam menilai sejauh mana suatu tes mampu menghasilkan data yang stabil dan dapat dipercaya. Pemahaman mengenai berbagai jenis reliabilitas tersebut menjadi penting karena dapat membantu pendidik maupun peneliti dalam menentukan metode pengujian yang paling sesuai dengan tujuan pengukuran, karakteristik instrumen yang digunakan, serta kondisi pelaksanaan evaluasi. Adapun jenis-jenis reliabilitas tersebut meliputi beberapa bentuk sebagai berikut (Marjuki et al., 2018).

a. Reliabilitas Tes Ulang (Test-Retest Reliability)

Penghitungan reliabilitas tes ulang dilakukan dengan menerapkan instrumen yang sama kepada kelompok responden yang sama pada dua kesempatan yang berbeda. Skor yang diperoleh dari pelaksanaan pertama dan kedua selanjutnya dianalisis melalui korelasi menggunakan koefisien korelasi, seperti Pearson Product Moment. Tingginya nilai korelasi yang dihasilkan menunjukkan semakin baik tingkat reliabilitas instrumen tersebut (Latipah et al., 2026).

b. Reliabilitas Bentuk Paralel (Parallel Forms Reliability)

Pengukuran reliabilitas bentuk paralel dilakukan dengan memberikan dua versi tes yang setara kepada kelompok responden yang sama. Setelah itu, skor yang diperoleh dari kedua bentuk tes tersebut dikorelasikan untuk mengetahui tingkat kesetaraannya. Apabila nilai korelasi yang diperoleh tinggi, maka kedua bentuk tes tersebut dinyatakan memiliki reliabilitas yang baik.

c. Reliabilitas Konsistensi Internal (Internal Consistency Reliability)

Perhitungan reliabilitas konsistensi internal didasarkan pada keterkaitan antarbutir soal dalam satu kali pelaksanaan pengukuran. Beberapa teknik yang lazim digunakan meliputi Split-Half, Kuder-Richardson (KR-20 dan KR-21), serta Cronbach's Alpha. Instrumen dinilai reliabel apabila menghasilkan koefisien reliabilitas yang tinggi dan memperlihatkan konsistensi yang baik antarbutir soal.

d. Reliabilitas Antarpenilai (Inter-Rater Reliability)

Reliabilitas antarpenilai ditentukan dengan membandingkan hasil penilaian yang diberikan oleh dua orang penilai atau lebih terhadap objek yang sama. Tingkat kesesuaian antarpenilai kemudian dianalisis menggunakan koefisien korelasi atau indeks kesepakatan, seperti Cohen's Kappa maupun Intraclass Correlation Coefficient (ICC). Semakin besar tingkat kesepakatan yang dicapai, semakin tinggi pula reliabilitas antarpenilai yang dimiliki.

e. Reliabilitas Intra-Penilai (Intra-Rater Reliability)

Perhitungan reliabilitas intra-penilai dilakukan dengan membandingkan hasil penilaian yang diberikan oleh penilai yang sama terhadap objek yang sama pada waktu yang berbeda. Selanjutnya, hasil penilaian tersebut dikorelasikan untuk mengidentifikasi tingkat konsistensinya. Koefisien korelasi yang tinggi menunjukkan bahwa penilai memiliki tingkat konsistensi yang baik dalam memberikan penilaian.

Taeb1.2 krtiteria umum koefisian realibilitas

Koefisien Reliabilitas	Interpretasi
0,80 – 1,00	Sangat tinggi
0,60 – 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat rendah

Catatan: semakin mendekati angka 1,00, semakin baik reliabilitas instrumen yang digunakan.

KESIMPULAN

Validitas dan reliabilitas merupakan dua komponen fundamental yang menentukan kualitas suatu instrumen penelitian. Validitas berhubungan dengan kemampuan instrumen dalam mengukur aspek yang seharusnya diukur, sedangkan reliabilitas berkaitan dengan tingkat konsistensi hasil pengukuran yang dihasilkan. Kedua aspek tersebut saling melengkapi dalam memastikan bahwa data yang diperoleh memiliki tingkat ketepatan dan keterpercayaan yang memadai untuk digunakan dalam penelitian.

Kajian ini menunjukkan bahwa validitas memiliki beberapa jenis, yaitu validitas isi, validitas konstruk, validitas kriteria, validitas muka, dan validitas empiris. Sementara itu, reliabilitas terdiri atas reliabilitas tes ulang, reliabilitas bentuk paralel, reliabilitas konsistensi internal, reliabilitas antarpenilai, dan

reliabilitas intra-penilai. Masing-masing jenis validitas dan reliabilitas memiliki karakteristik, fungsi, serta prosedur pengujian yang berbeda sesuai dengan tujuan pengukuran dan bentuk instrumen yang digunakan.

Oleh karena itu, pengujian validitas dan reliabilitas perlu dilakukan secara cermat sebelum instrumen digunakan dalam proses pengumpulan data. Instrumen yang memenuhi kedua kriteria tersebut akan menghasilkan data yang lebih akurat, konsisten, dan dapat dipertanggungjawabkan secara ilmiah. Dengan demikian, pemahaman yang baik mengenai konsep serta teknik pengujian validitas dan reliabilitas menjadi landasan penting bagi peneliti dalam menghasilkan penelitian yang berkualitas dan memiliki tingkat kepercayaan yang tinggi.

DAFTAR PUSTAKA

- Ananda,R.,& Fahdli,M. (2018) . Statistik pendidikan (Teori dan praktik Dalam Pendidikan). Medan : CV. Widya puspita
- Azwar,S (2017). Reabilitas dan Validitas edisi 4. Yogyakarta : Pustaka belajar
- Daulay, A., Ramadhani, S., Rafli, M., Zulpan, Z., Husein, A., & Wulandari, M. N. (2026). Pemahaman konsep validitas serta pengujiannya dalam penelitian. *Teaching and Learning Research Journal*, 2(1), 173-178.
- Dewy, M. S., Mulyana, D., Annisa, S., Zulkarnain, B., & Isnaini, M. (2025). ANALISIS UJICOBA INSTRUMEN TES BERBASIS HIGHER ORDER THINKING SKILLS (HOTS) PADA MATA KULIAH ELEKTRONIKA DASAR. *Jurnal Teknologi Pendidikan*, 18(2), 117–126.
- Febriani, L., Safitri, R., Hamidah, C., Anisah, J., Ramadhana, S., Husein, A., & Al Fatih, F. (2026). MEMBANGUN INSTRUMEN EVALUASI YANG BERKUALITAS MELALUI ANALISIS TINGKAT KESUKARAN, DAYA BEDA, DAN EFEKTIVITAS PENGECOH SOAL. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 11(02), 237-251.
- Latipah, N., Angriani, S., Alamsyah, R., & Husein, A. (2026). Memahami Tentang Realibilitas Serta Pengujiannya dalam Penelitian: Meningkatkan Konsistensi dan Validitas Hasil Pengukuran. *Teaching and Learning Research Journal*, 2(1), 195-200.
- Marjuki, Djemari Mardapi, B. K. (2018). Pengembangan Model Akreditasi Sekolah Menengah Atas/Madrasah Aliah (SMA/MA). *Jurnal Peneltian Dan Evaluasi Pendidikan*, 22(1), 105–117.
- Putri, N. A., Fauza, N., Bina, J., Km, W., & Panam, S. B. (2023). Validitas dan Reliabilitas Butir Soal Berbasis Kemampuan Berpikir Kritis. *Jurnal Pendidikan Fisika*, 12(1), 28–34.
- Saidah, N., Mawaddah, N., Hidayah, N., Zulpan, Z., Husein, A., & Arianto, A. (2026). Konsep reliabilitas dalam penelitian dan teknik pengujiannya. *Educational Journal*, 1(3), 775-784.
- Sibarani, S. M. (2024). Validitas dan Reliabilitas Tes Kemampuan Pemahaman Konsep Matematis di SMP Santo Petrus Medan. *Jurnal Fibonacci*, 05(1), 57–61.
- Subhaktiyasa, PG (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka. *Jurnal Penelitian Pendidikan* , 5 (4), 5599–5609. <https://doi.org/10.37985/jer.v5i4.1747>
- Umar, J., & Nisa, Y. F. (2020). Uji validitas konstruk dengan CFA dan pelaporannya. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia* , 9 (2), 1-11. <https://doi.org/10.15408/jp3i.v9i2.16964>
- Wardah, F. (2019). The validity and inter-rater reliability of project assessment in mathematics learning. *Jurnal Tadris Matematika*, 12(1), 1–13. <https://doi.org/10.20414/betajtm.v12i1.266>
- Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Tarbiyah : Jurnal Ilmiah Kependidikan* , 7 (1). <https://doi.org/10.18592/tarbiyah.v7i1.2100>