

## Pentingnya Validitas Tes dalam Meningkatkan Kualitas Evaluasi

Annisa Maryam<sup>1</sup>, Nuraisa Tambak<sup>2</sup>, Husnul Khotimah<sup>3</sup>, Nabila<sup>4</sup>, Niah Amanda Nur Anisah<sup>5</sup>

Sekolah Tinggi Agama Islam Negeri Mandailing Natal<sup>1,2,3,4,5</sup>

\*Email

annisamaryam2005@gmail.com, aisyahnuraisyah552@gmail.com,  
khusnulhotimah240999@gmail.com, nabila.boukna2005@gmail.com, sariminda874@gmail.com,  
anisah25236@gmail.com

Diterima: 07-602-26 | Disetujui: 13-06-2026 | Diterbitkan: 15-06-2026

### ABSTRACT

*Test validity is a crucial aspect of educational evaluation because it determines the extent to which an assessment instrument measures the competencies it is intended to assess. This article aims to examine the importance of test validity in improving the quality of educational evaluation through a literature review approach. Data were collected from various scholarly sources, including books and journal articles, and analyzed descriptively. The findings indicate that test validity plays a significant role in ensuring the accuracy, fairness, and trustworthiness of assessment results. Three main types of validity are identified: content validity, construct validity, and criterion-related validity, each contributing to the alignment of assessment instruments with measurement objectives. Furthermore, the development of valid test instruments requires a systematic process, including the formulation of objectives, preparation of test blueprints, item writing, expert review, pilot testing, and empirical analysis. Instruments with high validity provide evaluation results that can serve as a basis for objective, accurate, and accountable educational decision-making. Therefore, test validity is a fundamental element in enhancing the quality of teaching, learning, and educational evaluation.*

**Keywords:** test validity, educational evaluation, assessment instruments, evaluation quality.

### ABSTRAK

Validitas tes merupakan aspek penting dalam evaluasi pendidikan karena menentukan sejauh mana instrumen penilaian mampu mengukur kompetensi yang seharusnya diukur. Artikel ini bertujuan mengkaji pentingnya validitas tes dalam meningkatkan kualitas evaluasi pendidikan melalui pendekatan studi kepustakaan. Data diperoleh dari berbagai sumber ilmiah, seperti buku dan artikel yang relevan, kemudian dianalisis secara deskriptif. Hasil kajian menunjukkan bahwa validitas tes berperan dalam memastikan keakuratan, keadilan, dan keterpercayaan hasil penilaian. Terdapat tiga jenis validitas yang utama, yaitu validitas isi, validitas konstruk, dan validitas kriteria, yang masing-masing memiliki fungsi dalam menjamin kesesuaian instrumen dengan tujuan pengukuran. Selain itu, penyusunan instrumen yang valid memerlukan tahapan yang sistematis, mulai dari perumusan tujuan, penyusunan kisi-kisi, penulisan butir soal, telaah ahli, hingga uji coba dan analisis empiris. Dengan instrumen yang memiliki validitas tinggi, hasil evaluasi dapat digunakan sebagai dasar pengambilan keputusan pendidikan yang lebih objektif, akurat, dan dapat dipertanggungjawabkan. Oleh karena itu, validitas tes menjadi unsur fundamental dalam upaya meningkatkan kualitas pembelajaran dan evaluasi pendidikan.

**Kata kunci:** validitas tes, evaluasi pendidikan, instrumen penilaian, kualitas evaluasi.

Bagaimana Cara Sitasi Artikel ini:

Maryam, A., Tambak, N. ., Khotimah, H., Nabila, N., & Anisah , N. A. . N. . (2026). Pentingnya Validitas Tes dalam Meningkatkan Kualitas Evaluasi. Educational Journal, 1(4), 1889-1895.  
<https://doi.org/10.63822/r5n6a130>

## PENDAHULUAN

Validitas berasal dari kata *valid* yang berarti tepat, benar, dan sah. Dalam konteks pengukuran dan evaluasi pendidikan, validitas merujuk pada sejauh mana suatu instrumen tes benar-benar mengukur apa yang seharusnya diukur.

Menurut Arikunto (2013), "validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan atau kesahihan sesuatu instrumen." Ia menjelaskan bahwa sebuah instrumen dikatakan valid apabila mampu mengukur apa yang seharusnya diukur dan dapat mengungkap data dari variabel yang diteliti secara tepat dan akurat. Semakin tinggi nilai validitas sebuah instrumen, maka semakin kecil pula tingkat kesalahan atau penyimpangan yang terjadi dalam proses pengukuran. Dalam konteks evaluasi pendidikan, tes yang valid akan memberikan gambaran yang benar-benar mencerminkan kemampuan peserta didik sesuai dengan indikator dan tujuan pembelajaran yang telah ditetapkan, sehingga keputusan yang diambil berdasarkan hasil tes tersebut dapat bersifat tepat sasaran dan adil bagi seluruh peserta didik.

Menurut Allen dan Yen (1979), menegaskan bahwa "a test is valid if it measures what it is supposed to measure." menyatakan ini secara sederhana namun tegas menggaris bawahi bahwa inti dari validitas adalah kesesuaian antara apa yang diukur oleh tes dengan apa yang memang seharusnya diukur berdasarkan tujuan evaluasi yang telah dirancang. Apabila sebuah tes mengukur hal-hal yang tidak relevan atau menyimpang dari tujuan pembelajaran yang ditetapkan, maka tes tersebut tidak dapat dikategorikan sebagai alat ukur yang valid, seberapa canggih atau sistematis cara penyusunannya. Dengan demikian, validitas tes merupakan prasyarat fundamental yang harus selalu dipastikan terpenuhi sebelum sebuah tes digunakan dalam konteks evaluasi pendidikan yang sesungguhnya.

Berdasarkan para ahli di atas dapat disimpulkan bahwa validitas adalah tingkat ketepatan suatu instrumen tes dalam mengukur apa yang seharusnya diukur, sehingga hasil yang diperoleh benar-benar mencerminkan kemampuan peserta didik secara akurat dan dapat dipertanggungjawabkan.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan jenis studi kepustakaan, yaitu dengan mengumpulkan data yang kemudian dianalisis dari berbagai sumber seperti artikel dan buku. Penelitian ini tidak hanya merangkum isi bacaan, tetapi juga memaparkan materi dari beragam sumber sehingga berbagai gagasan dapat dipadukan dan menghasilkan kesimpulan terhadap permasalahan yang dikaji.

## HASIL PEMBAHASAN

### Depenisi Validitas Tes

Validitas tes adalah ukuran yang menunjukkan kualitas suatu alat penilaian dalam pendidikan. Secara sederhana, validitas tes berarti kemampuan tes untuk mengukur apa yang seharusnya diukur dengan tepat. Tes yang valid tidak hanya menghasilkan nilai, tetapi juga dapat menggambarkan kemampuan peserta didik yang sebenarnya sesuai dengan tujuan pembelajaran.

validitas tes juga dapat dipahami sebagai jaminan kualitas dari sebuah instrumen evaluasi. Apabila sebuah tes tidak valid, maka hasil yang diperoleh tidak dapat dijadikan acuan yang dapat dipercaya dalam menilai kemampuan peserta didik. Ibarat sebuah timbangan yang tidak akurat, tes yang tidak valid akan

menghasilkan informasi yang menyesatkan dan merugikan peserta didik maupun pendidik dalam proses pengambilan keputusan pendidikan.

Dengan demikian, validitas tes bukan hanya sekadar persyaratan teknis dalam penyusunan soal, melainkan merupakan tanggung jawab moral seorang pendidik terhadap peserta didiknya. Pendidik yang menggunakan tes yang valid berarti telah memberikan penilaian yang adil, objektif, dan dapat dipertanggungjawabkan kepada seluruh peserta didik tanpa terkecuali.

Arikunto (2013), menyatakan bahwa "validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan atau kesahihan sesuatu instrumen. Suatu instrumen yang valid atau sah mempunyai validitas tinggi. Sebaliknya, instrumen yang kurang valid berarti memiliki validitas rendah."

Gronlund dan Linn (1990,) Validitas mengacu pada kesesuaian, kebermaknaan, dan kegunaan inferensi spesifik yang dibuat berdasarkan skor tes, di mana validitas bukan sifat mutlak tes itu sendiri melainkan bergantung pada tujuan penggunaannya.

Messick (1989), Validitas merupakan penilaian terpadu yang mempertimbangkan bukti empiris dan landasan teoretis untuk mendukung ketepatan interpretasi dan tindakan yang diambil berdasarkan skor tes, termasuk mempertimbangkan dampak sosial dari penggunaan hasil tes tersebut.

Berdasarkan pendapat ketiga ahli tersebut, dapat disimpulkan bahwa inti dari validitas tes adalah ketepatan tes dalam mengukur apa yang seharusnya diukur. Arikunto menegaskan bahwa semakin valid sebuah instrumen, maka hasilnya semakin dapat dipercaya. Sebaliknya, tes yang tidak valid akan menghasilkan informasi yang kurang akurat. Gronlund dan Linn menjelaskan bahwa validitas tidak melekat pada tes itu sendiri, tetapi bergantung pada tujuan dan cara penggunaannya. Artinya, sebuah tes bisa saja valid dalam satu situasi, tetapi belum tentu valid dalam situasi lain. Sementara itu, Messick menekankan bahwa validitas tidak hanya berkaitan dengan aspek teknis, tetapi juga tanggung jawab pendidik terhadap dampak penggunaan hasil tes bagi peserta didik. Dengan demikian, validitas tes mencerminkan kejujuran dan tanggung jawab guru dalam menilai kemampuan siswa secara adil.

Validitas tes berarti ketepatan tes dalam mengukur kemampuan yang seharusnya diukur. Jika tes memiliki validitas yang baik, maka hasilnya dapat dipercaya. Sebaliknya, tes yang tidak valid bisa memberikan informasi yang keliru. Validitas juga tidak melekat pada tes itu sendiri, tetapi bergantung pada tujuan dan cara tes tersebut digunakan. Selain itu, validitas bukan hanya masalah teknis, tetapi juga berkaitan dengan tanggung jawab guru karena hasil tes dapat memengaruhi masa depan peserta didik. Oleh karena itu, validitas tes mencerminkan kejujuran dan tanggung jawab pendidik dalam menilai siswa secara adil.

## **Jenis-Jenis Tes Validitas**

### **1. Validitas Isi**

Menurut Gronlund (1985) menyatakan bahwa validitas isi ditentukan oleh seberapa baik komponen sampel dari tes merepresentasikan domain konten yang ingin diukur. Proses penilaian dilakukan secara logis dan sistematis oleh para ahli materi.

Popham (1981) mengemukakan bahwa validitas isi merupakan pertimbangan ahli terhadap kesesuaian butir tes dengan domain yang direpresentasikan. Penilaian dilakukan dengan menggunakan kisi-kisi soal sebagai acuan.

Dalam perhitungan validitas isi, metode yang digunakan umumnya mengacu pada formula tertentu seperti Indeks Validitas Isi (Content Validity Index/CVI) atau rumus Aiken's V, yang melibatkan penilaian

para ahli terhadap tingkat relevansi setiap butir instrumen. Adapun rumus yang digunakan untuk menghitung validitas isi adalah sebagai berikut:

$$V = \Sigma S / [n(C-1)]$$

$$S = R - Lo$$

V = Indeks persetujuan pakar tentang butir soal

R = nilai yang diberikan pakar

Lo = Skor terendah

C = Skor tertinggi

N = Jumlah pakar

Menunjukkan sejauh mana butir soal merepresentasikan materi pembelajaran. Validitas ini memastikan tidak ada materi penting yang terlewat atau berlebihan.

## 2. Validitas Konstruk (Construct Validity)

Validitas konstruk merupakan tingkat ketepatan suatu instrumen dalam mengukur konsep atau kemampuan teoretis yang menjadi sasaran pengukuran. Instrumen dikatakan memiliki validitas konstruk yang baik apabila setiap butir soal benar-benar mencerminkan aspek kemampuan yang hendak diukur, seperti kemampuan berpikir kritis, pemahaman konsep, atau sikap tertentu. Dengan demikian, validitas konstruk memastikan bahwa tes tidak sekadar mengukur pengetahuan umum, tetapi secara spesifik menilai konstruk yang telah dirumuskan dalam tujuan pembelajaran.

Menurut Widodo (2006), validitas konstruk berkaitan dengan sejauh mana hasil pengukuran dapat diinterpretasikan sebagai representasi dari konsep teoretis yang menjadi dasar penyusunan instrumen. Artinya, setiap indikator dan butir soal harus disusun berdasarkan landasan teori yang jelas agar instrumen mampu mengukur kemampuan yang dimaksud secara tepat. Proses pengujian validitas konstruk umumnya dilakukan melalui analisis teoritis, telaah ahli, serta analisis statistik.

Sejalan dengan itu, Jahja (2020), menjelaskan bahwa validitas konstruk menuntut adanya kesesuaian antara teori, indikator, dan data empiris yang diperoleh dari hasil pengukuran. Instrumen yang memiliki validitas konstruk yang baik akan menghasilkan data yang konsisten dengan konsep yang diukur, sehingga interpretasi hasil tes menjadi lebih dapat dipercaya.

Oleh karena itu, validitas konstruk merupakan aspek penting dalam penyusunan instrumen evaluasi agar hasil penilaian benar-benar menggambarkan kemampuan seseorang secara akurat.

## 3. Validitas kriteria

Validitas kriteria adalah validitas yang melihat keterkaitan antara hasil skor yang diperoleh peserta dalam suatu tes dengan tolok ukur kinerja nyata dari luar tes itu sendiri yang disebut sebagai kriteria. Jenis validitas ini berfokus pada seberapa erat hubungan antara perolehan skor tes dengan variabel pembanding lain yang dianggap relevan dan sudah teruji.

Pengujian validitas kriteria dapat dilakukan menggunakan uji korelasi Pearson. Nilai korelasi yang mendekati +1 menunjukkan hubungan positif yang kuat, nilai yang mendekati -1 menunjukkan hubungan negatif yang kuat, sedangkan nilai yang mendekati 0 menunjukkan tidak adanya hubungan yang signifikan antara kedua variabel (Subhaktiyasa, 2024).

Validitas kriteria merupakan tingkat kesesuaian antara prediktor (skor hasil instrumen) dengan kriteria (skor acuan). Validitas ini bertujuan menilai apakah alat ukur prediktor memiliki kualitas yang baik sehingga dapat digunakan untuk berbagai keperluan yang sejenis. Validitas kriteria terdiri atas

validitas serentak, yaitu ketika kriteria diukur pada waktu yang sama dengan prediktor, serta validitas prediktif, yaitu ketika kriteria diukur setelah prediktor (Arafah, 2017).

Berdasarkan pendapat kedua ahli tersebut, dapat dipahami bahwa validitas kriteria menekankan pada seberapa kuat hubungan antara skor hasil instrumen dengan skor acuan yang dijadikan pembandingan. Pengujiannya dapat dilakukan melalui korelasi Pearson untuk melihat kekuatan hubungan antar variabel. Selain itu, validitas kriteria mencakup dua bentuk, yaitu validitas serentak dan validitas prediktif, yang menunjukkan bahwa instrumen tidak hanya harus sesuai dengan kondisi saat ini, tetapi juga mampu memprediksi hasil di masa mendatang.

### **Langkah Langkah Penyusunan Validitas Instrumen Tes**

Penyusunan instrumen tes yang valid diawali dengan penetapan tujuan penilaian yang jelas serta penentuan kompetensi yang akan diukur. Tahap ini dilanjutkan dengan penyusunan kisi-kisi sebagai pedoman dalam menulis butir soal agar setiap soal memiliki keterkaitan langsung dengan indikator dan tujuan pembelajaran. Kisi-kisi berfungsi sebagai peta yang memastikan bahwa materi yang diujikan mencakup seluruh ruang lingkup kompetensi secara proporsional.

Setelah kisi-kisi disusun, langkah berikutnya adalah penulisan butir soal yang dilanjutkan dengan telaah ahli (expert judgment). Telaah ini bertujuan untuk menilai kesesuaian materi, kejelasan bahasa, serta ketepatan bentuk soal sebelum instrumen digunakan. Tahap ini sangat penting karena dapat mengidentifikasi kelemahan instrumen sejak awal sehingga perbaikan dapat dilakukan sebelum tes diujikan kepada peserta didik (Armedi et al. 2025).

Tahap akhir dalam penyusunan instrumen tes adalah pelaksanaan uji coba serta analisis butir soal secara empiris. Uji coba dilakukan dengan memberikan instrumen kepada kelompok responden yang memiliki karakteristik serupa dengan peserta didik sasaran. Melalui proses ini, peneliti atau guru memperoleh data nyata mengenai bagaimana setiap butir soal berfungsi dalam mengukur kemampuan peserta didik.

Data hasil uji coba kemudian dianalisis untuk mengetahui kualitas setiap butir soal, meliputi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Analisis validitas bertujuan memastikan ketepatan soal dalam mengukur kompetensi, sedangkan reliabilitas menunjukkan konsistensi hasil pengukuran. Tingkat kesukaran digunakan untuk melihat apakah soal terlalu mudah atau terlalu sulit, sementara daya pembeda berfungsi mengidentifikasi kemampuan soal dalam membedakan peserta didik yang berkemampuan tinggi dan rendah.

Hasil analisis tersebut menjadi dasar penting dalam melakukan revisi atau seleksi butir soal yang layak digunakan. Soal yang tidak memenuhi kriteria dapat diperbaiki atau diganti, sedangkan soal yang memenuhi standar dipertahankan dalam instrumen akhir. Dengan demikian, instrumen tes yang dihasilkan benar-benar mampu memberikan informasi penilaian yang akurat, objektif, dan dapat dipercaya (Muthi et al. 2025).

Berdasarkan pembahasan yang telah diuraikan, dapat disimpulkan bahwa validitas instrumen tes merupakan aspek yang sangat penting dalam memastikan bahwa alat ukur benar-benar mampu mengukur kemampuan yang seharusnya diukur. Proses penyusunan instrumen tidak dapat dilakukan secara sembarangan, melainkan harus melalui tahapan sistematis mulai dari perumusan tujuan, penyusunan kisi-kisi, penulisan butir soal, hingga proses uji coba. Setiap tahapan memiliki peran penting untuk menjaga kesesuaian antara tujuan pembelajaran dengan indikator yang diukur, sehingga instrumen yang dihasilkan memiliki dasar teoritis dan praktis yang kuat.

## KESIMPULAN

Validitas tes merupakan unsur utama dalam menjamin kualitas evaluasi pendidikan karena menentukan sejauh mana instrumen mampu mengukur kemampuan peserta didik secara tepat. Tes yang valid tidak hanya memberikan skor, tetapi juga menghadirkan gambaran nyata mengenai pencapaian belajar sehingga dapat menjadi dasar pengambilan keputusan pendidikan yang adil dan akurat. Keberadaan validitas isi, konstruk, dan kriteria menunjukkan bahwa proses pengukuran harus mempertimbangkan kesesuaian materi, landasan teori, serta keterkaitan dengan kinerja nyata.

Penyusunan instrumen tes yang valid memerlukan proses yang sistematis dan berkelanjutan, mulai dari perumusan tujuan hingga analisis hasil uji coba. Melalui tahapan tersebut, setiap butir soal dapat diseleksi dan diperbaiki sehingga instrumen yang digunakan benar-benar layak. Dengan instrumen yang valid, proses penilaian akan lebih objektif, dapat dipercaya, serta memberikan dampak positif bagi peningkatan kualitas pembelajaran dan evaluasi pendidika.

## DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Anastasi, A. (1988). *Psychological Testing* (6th ed.). New York: Macmillan Publishing Company.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik (Edisi Revisi)*. Jakarta: Rineka Cipta.
- Azwar, S. (2012). *Validitas dan Reliabilitas*. Yogyakarta: Pustaka Pelajar.
- Coaley, K. (2010). *An Introduction to Psychological Assessment and Psychometrics*. London: Sage Publications.
- Devon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A Psychometric Toolbox for Testing Validity and Reliability. *Journal of Nursing Scholarship*, 39(2), 155–164.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Evaluation in Teaching* (6th ed.). New York: Macmillan Publishing Company.
- Messick, S. (1989). Validity. Dalam R. L. Linn (Ed.), *Educational Measurement* (3rd ed., hlm. 13–103). New York: American Council on Education & Macmillan.