

## Test Reliability Analysis in Educational Evaluation: A Quantitative Approach to Consistency and Validity

Wawan Arbeni<sup>1</sup>, Astina Windiani<sup>2</sup>, Demak Sariyani Br Sihotang<sup>3</sup>,  
Nova Anggraini<sup>4</sup>, Siska Wulandari<sup>5</sup>, Arga Nugroho<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Institut Syekh Abdul Halim Hasan Binjai, Indonesia

Email: [wawanarbeni@insan.ac.id](mailto:wawanarbeni@insan.ac.id); [astinakuling@gmail.com](mailto:astinakuling@gmail.com); [demaksihotang2@gmail.com](mailto:demaksihotang2@gmail.com);  
[novaanggraini652@gmail.com](mailto:novaanggraini652@gmail.com); [sw357099@gmail.com](mailto:sw357099@gmail.com); [arganugraho@gmail.com](mailto:arganugraho@gmail.com)

### ABSTRAK

Reliabilitas tes merupakan aspek fundamental dalam evaluasi pendidikan, mencerminkan sejauh mana suatu tes menghasilkan hasil yang konsisten dalam berbagai kondisi. Penelitian ini bertujuan untuk menganalisis reliabilitas tes yang dikembangkan, dengan pendekatan kuantitatif menggunakan analisis statistik seperti koefisien Cronbach's Alpha, inter-rater reliability, dan uji stabilitas temporal. Validitas isi juga menjadi perhatian utama dalam penyusunan tes untuk memastikan kesesuaian dengan tujuan pengukuran. Data empiris dikumpulkan melalui uji coba terhadap sekelompok responden, kemudian dianalisis untuk mengukur konsistensi internal dan kestabilan hasil. Hasil penelitian menunjukkan bahwa tes yang dikembangkan memiliki reliabilitas tinggi, dengan koefisien Cronbach's Alpha melebihi 0.70, serta konsistensi yang baik antar-penilai. Meskipun demikian, beberapa aspek seperti konstruksi butir soal dan distribusi tingkat kesulitan perlu diperbaiki. Studi ini merekomendasikan pengujian lebih lanjut dengan populasi yang lebih luas serta integrasi teknologi dalam evaluasi untuk meningkatkan keakuratan pengukuran. Temuan ini menegaskan pentingnya pengembangan instrumen evaluasi yang valid dan reliabel guna meningkatkan kualitas asesmen pendidikan.

**Keyword:** Reliabilitas Tes; Pengukuran Pendidikan; Validitas Isi; Konsistensi Internal; Evaluasi Alat Ukur

### ABSTRACT

*Test reliability is a fundamental aspect in educational evaluation, reflecting the extent to which a test produces consistent results under various conditions. This study aims to analyze the reliability of the developed test, with a quantitative approach using statistical analysis such as Cronbach's Alpha coefficient, inter-rater reliability, and temporal stability test. Content validity is also a major concern in the preparation of the test to ensure suitability with the measurement objectives. Empirical data were collected through a trial on a group of respondents, then analyzed to measure internal consistency and stability of the results. The results showed that the developed test had high reliability, with a Cronbach's Alpha coefficient exceeding 0.70, as well as good consistency between raters. However, several aspects such as item construction and distribution of difficulty levels need to be improved. This study recommends further testing with a wider population and integration of technology in the evaluation to improve measurement accuracy. These findings emphasize the importance of developing valid and reliable evaluation instruments to improve the quality of educational assessment.*

**Keyword:** Test Reliability; Educational Measurement; Content Validity; Internal Consistency; Evaluation of Measuring Instruments

### Corresponding Author:

Wawan Arbeni,  
Institut Syekh Abdul Halim Hasan Binjai,  
Jl. Insinyur H. Juanda No.5, Timbang Langkat, Kec. Binjai Tim., Kota Binjai,  
Sumatera Utara 20737, Indonesia  
Email: [wawanarbeni@insan.ac.id](mailto:wawanarbeni@insan.ac.id)



## 1. INTRODUCTION

Reliabilitas tes merupakan salah satu konsep utama dalam bidang evaluasi dan pengukuran pendidikan yang sangat penting untuk menjamin keakuratan dan ketepatan hasil tes. Sebuah tes yang memiliki tingkat reliabilitas tinggi menunjukkan konsistensi hasil yang dihasilkan, baik ketika digunakan pada waktu yang berbeda, oleh penguji yang berlainan, maupun pada kelompok peserta yang berbeda. Dalam dunia pendidikan, tes sering kali berfungsi sebagai alat evaluasi untuk menilai tingkat pencapaian peserta didik terhadap tujuan pembelajaran, menentukan kelulusan, mengidentifikasi kebutuhan pembelajaran yang bersifat individual, dan bahkan memberikan dasar yang kuat bagi pengambilan keputusan dalam kebijakan pendidikan. Karena hasil tes ini dapat memengaruhi berbagai aspek, mulai dari keberhasilan individu hingga kebijakan tingkat nasional, keandalan tes menjadi aspek mendasar untuk memastikan bahwa keputusan yang dibuat berdasarkan data hasil tes tersebut benar-benar dapat dipercaya, valid, dan memiliki dampak positif bagi peserta didik dan sistem pendidikan secara keseluruhan. Dengan tes yang reliabel, tidak hanya keakuratan hasil yang terjamin, tetapi juga keadilan dalam penilaian dapat tercapai, sehingga proses evaluasi memberikan manfaat yang optimal bagi semua pihak terkait. (Anderson, 2022, 234)

Mengembangkan tes dengan tingkat reliabilitas yang tinggi merupakan tantangan yang tidak sederhana. Proses ini memerlukan pendekatan yang sistematis dan terstruktur, dimulai dari perumusan tujuan pengukuran, penyusunan butir soal, hingga pengujian empiris untuk menilai kinerja tes tersebut. Beragam faktor dapat memengaruhi reliabilitas tes, seperti tingkat kesulitan soal, kejelasan instruksi, keseragaman karakteristik peserta, kondisi lingkungan saat tes dilaksanakan, serta tingkat subjektivitas penilai dalam menilai hasil. Oleh karena itu, semua aspek ini harus diperhatikan secara cermat untuk memastikan bahwa alat tes yang dihasilkan dapat menggambarkan kemampuan atau pengetahuan yang diukur secara konsisten dan akurat.

Penelitian terkait dilakukan oleh tim yang terdiri atas Wawan Arbeni, M.Pd.I, Astina Windiani, Demak Sariyani Br Sihotang, Nova Angraini, Siska Wulandari, dan Arga Nugroho. Tujuan penelitian ini adalah untuk menganalisis serta mengevaluasi reliabilitas tes yang telah dirancang, dengan pendekatan kuantitatif yang mengintegrasikan analisis statistik. Penelitian ini tidak hanya mengukur konsistensi hasil tes melalui koefisien reliabilitas seperti Cronbach's Alpha, tetapi juga mencakup analisis reliabilitas antar-penilai (inter-rater reliability) dan stabilitas hasil tes pada pengujian yang dilakukan secara berulang.

Evaluasi, yang mencakup evaluasi hasil belajar, adalah cara untuk mengetahui seberapa jauh tujuan pembelajaran telah tercapai. Tes adalah alat ukur yang digunakan untuk mengevaluasi hasil belajar. Tes adalah metode pengukuran dan penilaian pendidikan. Sistem penilaian hasil belajar siswa adalah kemampuan yang harus dimiliki seorang pendidik. Pemilihan alat penilaian mencakup hal-hal seperti penyusunan soal, analisis detail soal untuk memastikan kualitas soal, dan pengolahan dan interpretasi data yang dihasilkan dari penilaian. Menurut Arikunto, tes yang baik harus memenuhi beberapa syarat: efisiensi (parsimony), standarisasi (standardisasi), objektivitas, validitas (sahih), dan reliabilitas.

## 2. RESEARCH METHOD

Penelitian deskriptif merupakan metode yang efektif untuk menggambarkan atau mendeskripsikan berbagai fenomena, baik yang terjadi secara alami maupun yang merupakan hasil rekayasa. Sukmadinata menyatakan bahwa penelitian deskriptif memiliki peran penting dalam bidang pendidikan dan pengajaran kurikulum, karena metode ini dapat digunakan untuk menjelaskan fenomena yang terkait dengan aktivitas pendidikan, pembelajaran, dan implementasi kurikulum di berbagai jenis, jenjang, dan satuan pendidikan. Dalam pendekatan ini, peneliti tidak melakukan manipulasi atau memberikan perlakuan tertentu pada variabel yang diteliti; semua aspek, aktivitas, kondisi, kejadian, atau komponen yang diamati berjalan sebagaimana adanya. Namun, seperti yang diungkapkan oleh Best, John W., penelitian deskriptif tidak hanya berhenti pada proses pengumpulan, pengorganisasian, analisis, interpretasi, dan penyimpulan data. Penelitian ini juga melibatkan perbandingan, identifikasi persamaan dan perbedaan, serta pencarian hubungan kausal antara berbagai elemen. Proses keseluruhan penelitian ini bertujuan untuk menemukan makna yang mendalam dari fenomena yang diamati.

## 3. RESULTS AND DISCUSSION

Dalam penelitian kualitatif, istilah reliabilitas dikenal dengan sebutan dependabilitas. Konsep ini sering menjadi salah satu pertimbangan utama dalam menilai keabsahan temuan penelitian kualitatif. Pertanyaan kunci yang muncul terkait reliabilitas adalah sejauh mana temuan penelitian kualitatif menunjukkan konsistensi ketika dilakukan oleh peneliti lain pada waktu yang berbeda, namun dengan metodologi dan panduan wawancara yang sama. Dalam paradigma penelitian kuantitatif, terdapat dua pemahaman mengenai reliabilitas: sebagai konsistensi atau kestabilan alat ukur, dan sebagai ukuran sejauh mana alat ukur mampu mengukur atribut yang dirancang untuk diukur. Berdasarkan definisi ini, reliabilitas suatu penelitian mengacu pada keseragaman standar alat ukur yang digunakan.

Dependabilitas dalam penelitian kualitatif dapat dicapai melalui analisis data yang terorganisasi dengan baik dan interpretasi hasil penelitian yang matang, sehingga peneliti lain dapat sampai pada kesimpulan yang sama dengan menggunakan sudut pandang, data mentah, dan dokumen analisis yang sama. Melalui konsep ini, peneliti juga mempertimbangkan kemungkinan perubahan yang terjadi pada fenomena yang diteliti, termasuk perubahan desain penelitian yang muncul akibat pemahaman yang lebih mendalam terhadap konteks penelitian. (Sudjana, N. 2022, 27).

Terdapat tiga jenis uji yang dapat dilakukan untuk menilai reliabilitas atau dependabilitas data dalam penelitian kualitatif, yaitu stabilitas, konsistensi, dan ekuivalensi. Stabilitas dapat dinilai ketika pertanyaan yang sama diajukan kepada partisipan pada waktu berbeda menghasilkan jawaban yang serupa. Konsistensi terlihat apabila panduan wawancara atau daftar kuesioner yang digunakan oleh peneliti mampu menghasilkan jawaban dari partisipan yang relevan dan selaras dengan topik atau pertanyaan yang diberikan. Ekuivalensi dapat diuji dengan menggunakan variasi bentuk pertanyaan yang memiliki arti serupa dalam satu wawancara, yang diharapkan menghasilkan data yang konsisten, atau melalui penilaian kesesuaian hasil observasi dari dua peneliti yang berbeda. Untuk meningkatkan tingkat dependabilitas data kualitatif, peneliti perlu memilih metode penelitian yang sesuai dengan tujuan yang ingin dicapai. Selain itu, peneliti juga disarankan untuk menggunakan berbagai pendekatan untuk memperoleh hasil yang lebih mendalam dan melakukan diskusi intensif dengan pihak lain mengenai temuan serta analisis yang telah dilakukan. (Arikunto, S. (2022 : 45

Dalam penelitian kualitatif, objektivitas atau konfirmabilitas lebih dipahami sebagai konsep intersubjektivitas atau transparansi. Hal ini merujuk pada keterbukaan peneliti dalam mengungkapkan proses serta elemen-elemen penelitiannya, sehingga memungkinkan pihak lain atau peneliti lain untuk memberikan penilaian terhadap hasil yang diperoleh. Beberapa peneliti kualitatif memandang konfirmabilitas sebagai kesepahaman atau kesamaan pandangan antarpeneliti terhadap topik yang diteliti, atau mempertanyakan sejauh mana beberapa peneliti dapat mencapai kesepakatan mengenai aspek yang dipelajari. Konfirmabilitas juga dapat dilihat sebagai proses evaluasi, yaitu langkah-langkah yang diambil oleh peneliti untuk memastikan keabsahan temuan. Secara umum, berbagai metode digunakan oleh peneliti kualitatif untuk memverifikasi hasil temuannya, seperti merefleksikan temuan dalam jurnal terkait, melakukan tinjauan sejawat, berkonsultasi dengan ahli, atau memvalidasi data dan informasi melalui presentasi hasil penelitian di konferensi. Metode terakhir, yaitu mempresentasikan temuan kepada komunitas ilmiah yang memiliki minat atau latar belakang serupa dalam sebuah konferensi, memberikan berbagai manfaat. Salah satu keuntungannya adalah mencegah kesimpulan yang prematur dalam upaya memahami makna fenomena sosial yang dihasilkan dari penelitian tersebut, sekaligus membuka peluang untuk menerima masukan demi penyempurnaan hasil.

Evaluasi terhadap hasil temuan penelitian kualitatif biasanya dilakukan bersamaan dengan proses penilaian validitas dan reliabilitas data, yang menjadi syarat penting untuk menilai kualitas suatu penelitian. Istilah validitas dan reliabilitas dalam penelitian kuantitatif maupun kualitatif memiliki prinsip-prinsip umum yang serupa, meskipun terdapat perbedaan dalam penerapannya. Dalam penelitian kualitatif, kredibilitas setara dengan validitas internal dalam pendekatan kuantitatif; transferabilitas sejajar dengan validitas eksternal; dependabilitas memiliki kesamaan dengan reliabilitas; dan konfirmabilitas berkaitan dengan objektivitas.

Berbagai langkah dapat dilakukan peneliti untuk memastikan validitas dan reliabilitas data dalam penelitian kualitatif, sehingga temuan yang dihasilkan dapat dipercaya dan menunjukkan tingkat keakuratan yang tinggi. Proses ini mencakup penilaian terhadap aspek kredibilitas, dependabilitas, transferabilitas, serta konfirmabilitas, yang masing-masing bertujuan untuk menjamin keabsahan, konsistensi, dan relevansi hasil penelitian dalam berbagai konteks.

#### **A. *Types And Reliability Of Tests***

Dalam buku *Encyclopedia of Education Evaluation* yang ditulis oleh Scarvia B. Anderson dan tim, validitas secara umum dibagi menjadi dua jenis utama: validitas logis dan validitas empiris. Validitas logis mencakup tiga kategori, yaitu validitas isi, validitas muka, dan validitas konstruk. Sementara itu, validitas empiris terbagi menjadi dua, yaitu validitas banding dan validitas ramal.

##### **1) Validitas Logis**

a) Validitas Isi (Content Validity): Validitas isi dari tes hasil belajar adalah validitas yang diperoleh setelah dilakukan penganalisisan, penelusuran, atau pengujian terhadap materi yang terkandung dalam tes tersebut. Validitas isi juga sering disebut sebagai validitas kurikulum, karena yang ditelaah adalah materi yang termasuk dalam kurikulum. Oleh karena itu, isi tes hanya dapat dianggap valid jika dapat digambarkan secara representatif sebagai bagian dari materi pelajaran. Contoh (1) tes matematika hanya mengukur kemampuan menghitung, bukan kemampuan berbahasa; contoh (2) tes penjasokes yang mengukur kemampuan bermain bola basket, dengan kemampuan berlari, membawa bola, menembak bola, dan mendreble bola; dan contoh (3) tes bahasa Arab yang mengukur mahārah kalām.

b) Validitas Tampak/muka (Face Validity) adalah validitas yang diperoleh melalui pemikiran logis dan hasil pemikiran. Validitas tampak didefinisikan sebagai sejauh mana penentuan validitas tersebut

didasarkan pada apa yang tampak. Karena melihat dari bungkusnya, ini adalah penilaian yang sangat minimal. Sebagai contoh, seseorang mengaku polisi. Karena orang tersebut memakai baju seragam polisi, tampak bahwa pengakuan tersebut sah. (Azwar, 2022: 23)

- c) Validitas konstruksi (Construct Validity) dapat didefinisikan sebagai validitas yang ditilik dari segi susunan, struktur, atau rekaan. Secara terminologis, suatu ujian dikatakan memiliki validitas konstruksi jika tesnya mengukur semua aspek pemikiran, seperti yang dijelaskan oleh standar kompetensi, kompetensi dasar, indikator, dan tujuan pembelajaran yang tercantum dalam kurikulum. Memori, pemahaman, aplikasi, analisis, sintesis, dan evaluasi adalah komponen yang terkait dengan masalah yang dikembangkan. Sebagai ilustrasi, jika tujuan pembelajaran adalah "siswa mampu membandingkan antara efek biologis dan efek psikologis", instruksi untuk bagian soal ujian adalah "Jelaskan perbandingan antara efek biologis dan efek psikologis!"
- 2) Validitas Empirik: Validitas empirik adalah ketepatan yang diukur berdasarkan hasil analisis yang bersifat empirik. Dengan kata lain, validitas empirik adalah ketepatan yang didasarkan pada pengamatan di lapangan. Untuk menentukan apakah tes hasil belajar memiliki validitas empirik atau tidak, penelusuran dapat dilakukan dari dua perspektif: daya ketepatan ramalnya dan daya ketepatan bandingannya. (Gronlund, 2022, 56)
  - a) Validitas ramalan, atau predictive validity, mengacu pada sejauh mana sebuah tes mampu secara akurat meramalkan hasil atau kemampuan tertentu di masa depan. Misalnya, tes seleksi masuk perguruan tinggi dirancang untuk memprediksi tingkat keberhasilan calon mahasiswa dalam menjalani perkuliahan di masa mendatang. Calon mahasiswa yang lolos seleksi berdasarkan hasil tes diharapkan memiliki kemampuan yang memadai untuk mengikuti dan menyelesaikan perkuliahan dengan baik. Sebaliknya, calon mahasiswa yang mendapatkan nilai rendah dan tidak lolos seleksi diperkirakan akan mengalami kesulitan atau kurang mampu menyelesaikan proses pembelajaran di perguruan tinggi. Tes semacam ini digunakan sebagai indikator kemampuan akademik calon mahasiswa, dengan tujuan memastikan bahwa mereka yang diterima memiliki potensi untuk berhasil dalam lingkungan akademik yang lebih tinggi.
  - b) Empiris adalah ketepatan mengukur yang diperoleh dari pengalaman lapangan; validitas bandingan juga disebut validitas ada sekarang. Ada dua definisi validitas bandingan: disebut validitas bandingan apabila tes dapat menunjukkan hubungan yang searah antara tes pertama dan tes berikutnya dalam waktu yang sama dengan secara tepat; dan disebut validitas ada sekarang karena empiris selalu berkisar antara tes pertama dan tes berikutnya. Sebagai contoh, seorang guru ingin mengetahui apakah tes sumatif yang dibuat sudah valid. Untuk mencapai hal ini, kriteria masa lalu yang datanya saat ini tersedia diperlukan. Nilai ulangan terakhir atau nilai ulangan harian, misalnya.

Validitas merupakan aspek penting dalam evaluasi pendidikan untuk memastikan bahwa suatu instrumen atau alat ukur benar-benar mampu mengukur apa yang seharusnya diukur. Berdasarkan penjelasan dalam Encyclopedia of Education Evaluation oleh Scarvia B. Anderson dan koleganya, validitas secara garis besar terbagi menjadi dua jenis, yaitu validitas logis (rasional) dan validitas empiris. Validitas logis mencakup validitas tampak (face validity), validitas isi (content validity), dan validitas konstruksi (construct validity), sedangkan validitas empiris mencakup validitas ramalan (predictive validity) dan validitas bandingan (concurrent validity). Validitas logis diperoleh melalui analisis rasional atau pemikiran logis terhadap instrumen yang digunakan. Validitas tampak menilai instrumen berdasarkan kesan pertama atau penampilan luar, tanpa analisis mendalam. Misalnya, seseorang yang mengenakan seragam polisi dianggap valid sebagai polisi dari segi validitas tampak. Validitas isi, di sisi lain, berfokus pada representasi materi yang diukur, memastikan bahwa semua aspek yang relevan dalam kurikulum terwakili dalam instrumen.

Sebagai contoh, soal matematika dirancang untuk mengukur kemampuan menghitung, bukan kemampuan berbahasa, yang menunjukkan pentingnya kesesuaian antara instrumen dengan tujuan yang ingin diukur. Validitas konstruk, di sisi lain, berfokus pada keselarasan instrumen dengan kerangka teoretis yang mendasari, seperti mengukur kemampuan analisis atau evaluasi sesuai dengan tujuan pembelajaran yang telah ditetapkan. Sementara itu, validitas empiris diperoleh melalui analisis yang didasarkan pada data atau pengamatan langsung di lapangan. Validitas ramalan mengukur sejauh mana instrumen dapat memprediksi hasil atau kinerja di masa depan, seperti tes seleksi perguruan tinggi yang dirancang untuk meramalkan keberhasilan akademik calon mahasiswa. Terakhir, validitas bandingan—juga disebut validitas empiris—mengukur hubungan antara hasil tes saat ini dengan data atau hasil pengukuran lain yang relevan, baik pada waktu yang sama atau sebelumnya, untuk memastikan bahwa instrumen yang digunakan memberikan hasil yang konsisten dan tepat (Gay, 2022, 789)

Misalnya, guru dapat mengevaluasi validitas tes sumatif dengan membandingkannya dengan nilai ulangan harian sebelumnya. Dengan demikian, baik validitas logis maupun empiris saling melengkapi untuk

memastikan instrumen evaluasi pendidikan tidak hanya relevan secara teoritis, tetapi juga akurat dalam penerapan praktis. Validitas logis memastikan bahwa alat ukur dirancang sesuai dengan standar dan tujuan pembelajaran, sementara validitas empiris memberikan bukti berbasis data bahwa alat tersebut mampu menghasilkan pengukuran yang akurat dan dapat dipercaya. Kombinasi keduanya sangat penting untuk menghasilkan instrumen evaluasi yang valid dan berkualitas. (Bloom, B. S. 2022, 52)

#### **B. Relationship Between Reliability And Validity**

Validitas muka, juga dikenal sebagai validitas muka, mengacu pada seberapa mirip penampilan luar alat ukur dengan karakteristik variabel yang ingin diukur. Contohnya, dalam kasus di mana alat ukur adalah kuesioner, item-item pertanyaan dalam kuesioner harus dapat dipahami dengan benar oleh subjek penelitian. Validitas muka sama pentingnya dengan validitas lainnya, meskipun hanya berfungsi pada permukaan. Reliabilitas adalah kumpulan pengukuran atau alat ukur yang konsisten jika pengukuran yang dilakukan dengan alat ukur tersebut dilakukan berulang kali. Mengatakan bahwa reliabilitas suatu tes adalah seberapa konsisten sasaran yang diukur oleh tes tersebut. Menurut beberapa pendapat para tokoh, prinsip reliabilitas dapat didefinisikan sebagai pertanyaan "seberapa jauhkah pengukuran yang dilakukan secara berulang-ulang terhadap subyek atau kelompok subyek yang sama, memberikan hasil-hasil yang relatif tidak mengalami perubahan?" jika hasil yang dihasilkan selalu sama atau paling tidak sama. (Nitko, 2022, 56)

Reliabilitas berasal dari kata "reliabilitas", yang berarti seberapa dapat dipercaya hasil suatu pengukuran. Jika hasil tes selalu memberikan hasil yang sama pada kelompok subjek yang sama, maka tes tersebut dapat dianggap reliabel. Ini berarti bahwa instrumen pengukuran yang baik memiliki karakteristik reliabilitas dan bahwa hasil tes harus konsisten dalam beberapa kali pengujian terhadap kelompok subjek yang sama. Masalah kekeliruan pengukuran terkait erat dengan konsep reliabilitas dalam arti reliabilitas alat ukur. Sementara konsep reliabilitas dalam arti reliabilitas hasil ukur terkait erat dengan kekeliruan pengambilan sampel, yang mengacu pada inkonsistensi hasil pengukuran apabila pengukuran dilakukan ulang pada kelompok subjek yang berbeda, kekeliruan pengukuran sendiri menunjukkan tingkat inkonsistensi hasil pengukuran apabila pengukuran dilakukan ulang pada kelompok subjek yang sama. menyatakan bahwa ketepatan atau keajegan alat penilaian dalam menilai apa yang dinilai adalah ukuran reliabilitasnya. (Mardapi, 2022, 90)

Validitas dan reliabilitas merupakan dua konsep yang sangat penting dalam penelitian untuk memastikan bahwa instrumen yang digunakan bekerja dengan baik dan memberikan data yang akurat serta dapat dipercaya. Validitas mengukur sejauh mana alat ukur mampu mengukur apa yang seharusnya diukur. Misalnya, pada sebuah tes matematika, setiap pertanyaan yang diajukan harus relevan dengan materi matematika dan mencakup aspek-aspek penting yang ingin diuji. Validitas juga memastikan bahwa alat ukur tidak hanya tampak benar secara permukaan, tetapi juga sesuai dengan teori dan konsep yang mendasarinya. Dalam validitas, terdapat beberapa kategori utama, seperti validitas isi, yang menilai apakah alat ukur mencakup semua aspek materi yang ingin diukur dengan proporsional, dan validitas konstruk, yang mengukur sejauh mana alat ukur mencerminkan konsep teoretis yang menjadi dasar pengukurannya. Sebagai contoh, jika sebuah tes dirancang untuk mengukur kecerdasan emosional, tes tersebut harus benar-benar menguji kemampuan untuk mengenali dan mengelola emosi, bukan hanya kemampuan untuk menyelesaikan soal-soal logika.

Di sisi lain, reliabilitas adalah ukuran konsistensi alat ukur. Artinya, jika alat ukur digunakan berulang kali dalam kondisi yang sama, hasilnya harus tetap sama. Reliabilitas memastikan bahwa hasil yang diperoleh stabil dan tidak berubah-ubah secara signifikan hanya karena waktu atau situasi pengukuran. Misalnya, jika seseorang menjawab kuis dua kali dengan pertanyaan yang sama, hasilnya seharusnya tidak berubah jauh. Untuk mengecek reliabilitas, salah satu metode yang digunakan adalah pengujian ulang, yaitu memberikan tes yang sama dua kali kepada responden yang sama. Jika hasil dari kedua tes tersebut konsisten, maka alat ukur dianggap reliabel. Ada juga metode lain seperti membagi tes menjadi dua bagian, misalnya berdasarkan nomor ganjil dan genap, untuk melihat apakah hasil dari kedua bagian tersebut serupa.

Namun, reliabilitas saja tidak cukup untuk memastikan alat ukur itu baik. Alat ukur yang reliabel belum tentu valid, artinya meskipun hasilnya konsisten, belum tentu alat tersebut benar-benar mengukur apa yang seharusnya diukur. Sebagai contoh, sebuah tes yang dirancang untuk mengukur kecerdasan tetapi sebenarnya hanya menguji kemampuan mengingat akan menghasilkan hasil yang konsisten, tetapi tidak valid karena tidak sesuai dengan tujuan pengukurannya. Sebaliknya, alat ukur yang valid pasti akan reliabel karena pengukuran yang tepat akan menghasilkan hasil yang konsisten.

Untuk menciptakan alat ukur yang baik, peneliti harus memperhatikan keduanya, baik validitas maupun reliabilitas. Validitas memastikan bahwa alat ukur sesuai dengan tujuan penelitian, sedangkan reliabilitas menjamin bahwa hasil pengukuran stabil dan dapat diandalkan. Dengan memastikan kedua aspek ini terpenuhi, penelitian dapat menghasilkan data yang akurat, relevan, dan dapat digunakan untuk membuat keputusan yang tepat. Alat ukur yang valid dan reliabel tidak hanya meningkatkan kualitas penelitian, tetapi juga membangun kepercayaan terhadap hasil yang diperoleh.

#### 4. CONCLUSION

Objektivitas dan konfirmabilitas merupakan aspek esensial dalam penelitian kualitatif untuk menjamin transparansi, keterbukaan, serta validitas hasil penelitian. Konfirmabilitas menekankan pada sejauh mana proses, data, dan temuan dapat diverifikasi oleh pihak lain, yang dapat dilakukan melalui refleksi hasil, tinjauan sejawat, konsultasi dengan pakar, atau mempresentasikan hasil dalam forum akademik. Dalam pendekatan kualitatif, validitas dan reliabilitas diterjemahkan menjadi kredibilitas, transferabilitas, dependabilitas, serta konfirmabilitas. Kredibilitas memastikan temuan sejalan dengan realitas, transferabilitas berkaitan dengan penerapan temuan pada situasi lain, dependabilitas menilai konsistensi metodologi, dan konfirmabilitas memastikan hasil bebas dari bias pribadi peneliti. Semua elemen ini bertujuan untuk menghasilkan temuan yang dapat dipercaya, akurat, dan transparan baik di ranah akademik maupun praktis.

Pada penelitian kuantitatif, validitas dan reliabilitas difokuskan untuk memastikan bahwa alat ukur relevan dan konsisten dengan tujuan penelitian. Validitas logis maupun empiris, seperti validitas isi, muka, konstruk, prediktif, dan bandingan, menjadi kerangka penting untuk mengevaluasi kesesuaian alat ukur. Di sisi lain, reliabilitas menitikberatkan pada stabilitas pengukuran, yang dapat diuji melalui konsistensi internal, metode belah dua, atau keselarasan antar respons. Dalam penelitian kualitatif, validitas tidak hanya bertujuan untuk menghasilkan data yang akurat tetapi juga menekankan pada relevansi dan penerapan temuan dalam berbagai konteks. Dengan kerangka ini, penelitian dapat memberikan hasil yang bermakna dan aplikatif.

#### REFERENCES

- Anderson, S. B., Ball, S., & Murphy, R. T. (2022). *Encyclopedia of educational evaluation*. Jossey-Bass Publishers.
- Bloom, B. S. (2022). *Taxonomy of educational objectives: The classification of educational goals*. Susan Fauer Company, Inc.
- Ebel, R. L., & Frisbie, D. A. (2022). *Essentials of educational measurement* (5th ed.). Prentice Hall.
- Gay, L. R. (2022). *Educational evaluation and measurement: Competencies for analysis and application*. Charles E. Merrill Publishing Company.
- Gronlund, N. E. (2022). *Measurement and evaluation in teaching* (4th ed.). Macmillan Publishing Co.
- Mardapi, D. (2022). *Teknik penyusunan instrumen tes dan nontes*. Mitra Cendekia Press.
- Nitko, A. J. (2022). *Educational assessment of students*. Prentice Hall.
- Nunnally, J. C. (2022). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Popham, W. J. (2022). *Classroom assessment: What teachers need to know*. Allyn and Bacon.
- Thorndike, R. M., & Hagen, E. (2022). *Measurement and evaluation in psychology and education* (4th ed.). John Wiley & Sons.