

# Pentingnya Validitas dan Reliabilitas dalam Pengembangan Tes Psikologi: Tinjauan Teoritis dan Praktis

Dita Dea Amanda\*<sup>1</sup>

<sup>1</sup> Universitas Negeri Surabaya  
\*e-mail: [24010014119@mhs.unesa.ac.id](mailto:24010014119@mhs.unesa.ac.id)<sup>1</sup>

## Abstrak

Validitas dan reliabilitas merupakan elemen kunci dalam pengembangan tes psikologi agar hasil evaluasi akurat dan konsisten. Artikel ini bertujuan menjelaskan konsep, penerapan, serta studi kasus terkait pengembangan Inventori Minat Karier (IMK) berbasis teori Holland. Metode yang digunakan adalah studi pustaka dan studi kasus deskriptif. Hasil menunjukkan validitas isi tinggi melalui CVI, validitas konstruk melalui CFA, serta reliabilitas memadai dengan nilai Cronbach's Alpha > 0,80. Simpulan menunjukkan pentingnya penguasaan konsep psikometri dan penerapan uji kualitas alat ukur dalam praktik bimbingan dan konseling.

**Kata kunci:** Bimbingan dan Konseling, Pengembangan Tes, Reliabilitas, Validitas

## Abstract

Validity and reliability are key elements in the development of psychological tests so that evaluation results are accurate and consistent. This article aims to explain the concept, application, and case studies related to the development of the Career Interest Inventory (IMK) based on Holland's theory. The methods used are literature study and descriptive case study. The results showed high content validity through CVI, construct validity through CFA, and adequate reliability with Cronbach's Alpha value > 0.80. The conclusion shows the importance of mastering psychometric concepts and the application of measuring instrument quality tests in guidance and counseling practice.

**Keywords:** Guidance and Counseling, Test Development, Reliability, Validity

## PENDAHULUAN

Alat pengukuran psikologi telah menjadi sarana utama dalam menilai berbagai aspek psikologis manusia yang tidak tampak secara langsung. Dalam ranah pendidikan dan layanan bimbingan serta konseling, penggunaan tes menjadi semakin krusial. Meskipun demikian, keberhasilan suatu tes sangat dipengaruhi oleh kualitas psikometriknya, terutama dalam hal validitas dan reliabilitas. Tanpa kedua unsur ini, hasil evaluasi bisa menyesatkan, berisiko merugikan klien, dan menurunkan kepercayaan terhadap layanan psikologi itu sendiri.

Ironisnya, dalam praktik sehari-hari, masih ada banyak instrumen psikologi yang diperoleh tanpa adanya uji validitas dan reliabilitas yang memadai. Hal ini memunculkan keraguan mengenai keabsahan hasil penilaian dan ketepatan data yang diperoleh dari alat tersebut. Karena itu, diskusi mengenai validitas dan reliabilitas tidak hanya menjadi isu akademis, tetapi juga berhubungan langsung dengan praktik profesional, khususnya di bidang bimbingan dan konseling.

Adapun rumusan masalahnya yakni sebagai berikut: (a) Apa konsep validitas dan reliabilitas dalam pengembangan tes psikologi? (b) Bagaimana pelaksanaannya dalam konteks bimbingan dan konseling? (c) Apa saja kelebihan dan kekurangan dalam pengembangan tes psikologi? Penyusunan artikel ini bertujuan untuk menguraikan konsep validitas dan reliabilitas dalam pengembangan tes psikologi, menjelaskan temuan-temuan terbaru yang relevan, serta menganalisis penerapan praktisnya dalam konteks pendidikan dan layanan konseling.

## KAJIAN TEORI

Validitas dan reliabilitas adalah dua prinsip dasar dalam teori pengukuran psikologi yang berasal dari pemikiran ilmiah mengenai bagaimana manusia bisa memahami dan menilai karakteristik tersembunyi, seperti kecerdasan, motivasi, sikap, atau kepribadian.

A. Konsep Validitas dalam Teori Psikometri

Secara umum, validitas mengacu pada seberapa sah suatu instrumen dalam mengukur hal yang seharusnya diukur (Taherdoost, 2016). Validitas bukan hanya tentang karakteristik dari tes itu sendiri, tetapi lebih kepada sejauh mana kesesuaian antara interpretasi skor dan cara penggunaan skor tersebut. Menurut American Educational Research Association (AERA), validitas didefinisikan sebagai tingkat di mana bukti dan teori mendukung interpretasi skor tes untuk tujuan tertentu (AERA, APA, dan NCME, 2014). Ada beberapa jenis validitas yang perlu dipahami dengan baik:

1. Validitas Isi (Content Validity)

Validitas isi berhubungan dengan sejauh mana konten dari alat ukur mencerminkan seluruh domain konsep yang diukur. Misalnya, untuk mengukur kecemasan sosial, sebuah tes harus memiliki item yang mencerminkan gejala kognitif, emosional, dan perilaku yang sesuai. Penentuan validitas isi umumnya melibatkan ahli yang mengevaluasi kejelasan, relevansi, dan representatif item (Suprpto dan Suryani, 2021).

2. Validitas Konstruk (Construct Validity)

Validitas ini berfokus pada seberapa baik tes menggambarkan struktur teoritis atau konstruk psikologis yang dimaksud. Validitas konstruk sering diuji melalui analisis statistik seperti analisis faktor eksploratori (EFA) atau konfirmatori (CFA). Analisis ini membantu menentukan apakah item-item dalam tes benar-benar berkumpul dalam satu faktor atau lebih sesuai dengan hipotesis teoritik yang ada. Jika sebuah tes dimaksudkan untuk mengukur satu konstruk tunggal, maka item-item tersebut seharusnya memiliki korelasi yang tinggi satu sama lain dan membentuk satu dimensi utama (Trianingih, 2023).

3. Validitas Kriteria (Criterion-related Validity)

Validitas ini mengacu pada hubungan antara skor tes dengan indikator eksternal atau kriteria yang terkait. Validitas kriteria terdiri dari dua bentuk, yaitu validitas prediktif (hubungan dengan kriteria di masa depan, seperti prestasi akademik) dan validitas konkuren (hubungan dengan kriteria yang diukur pada waktu bersamaan). Misalnya, tes minat karier seharusnya berkorelasi dengan jurusan yang dipilih oleh siswa.

4. Validitas Wajah (Face Validity)

Walaupun bukan validitas secara teknis, validitas wajah mengacu pada sejauh mana sebuah tes tampak "valid" bagi pengguna atau partisipan. Ini penting untuk menjaga kepercayaan dan kenyamanan responden, meskipun validitas ini bersifat subjektif dan tidak menjamin kualitas psikometrik.

B. Konsep Reliabilitas dalam Teori Pengukuran

Reliabilitas adalah ukuran sejauh mana instrumen dapat menunjukkan konsistensi. Sebuah tes dianggap reliabel jika hasilnya tetap konsisten di bawah kondisi yang sama, baik dalam periode waktu yang berbeda (stabilitas temporal), antar item (konsistensi internal), maupun antara versi yang berbeda (paralelisme versi). Menurut teori tes klasik (Classical Test Theory - CTT), setiap skor individu terdiri dari dua elemen, yaitu skor nyata (true score) dan kesalahan pengukuran (error) (Ramadhan et al., 2024). Berbagai pendekatan untuk mengukur reliabilitas antara lain:

1. Reliabilitas Uji-Ulang

Pengukuran ini dilakukan dengan memberikan tes yang sama kepada responden pada dua waktu yang berbeda, kemudian menghitung koefisien korelasi antara kedua hasil tersebut. Semakin tinggi tingkat korelasinya, semakin dapat diandalkan tes tersebut dari sisi waktu.

2. Reliabilitas Konsistensi Internal

Ini adalah metode yang paling sering digunakan untuk mengukur reliabilitas biasanya dengan koefisien alpha Cronbach. Semakin tinggi nilai alpha (idealnya di atas 0,70), semakin konsisten antar item dalam tes yang mengukur konstruk yang sama. Namun, alpha yang sangat tinggi (> 0,90) bisa menunjukkan bahwa beberapa item terlalu serupa atau berlebihan (Tavakol dan Dennick, 2011).

3. Reliabilitas Split-Half

Dalam teknik ini, item dibagi menjadi dua bagian, dan skor dari masing-masing bagian dikorelasikan. Korelasi ini mencerminkan ketangguhan internal tes, meskipun cara pembagian item dapat memengaruhi hasilnya.

#### 4. Reliabilitas Bentuk Paralel

Dua versi tes yang setara diberikan kepada kelompok yang sama, dan hasilnya dibandingkan. Metode ini jarang diterapkan karena sulit untuk menciptakan dua versi tes yang benar-benar setara.

### METODE

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan metode studi pustaka dan studi kasus. Subjek kajian adalah instrumen IMK. Data diperoleh dari literatur dan dokumentasi hasil pengembangan. Validitas isi diuji menggunakan CVI oleh para ahli, validitas konstruk diuji dengan CFA, dan reliabilitas diukur menggunakan Cronbach's Alpha. Analisis dilakukan secara deskriptif untuk menilai kualitas psikometrik dan aplikasinya dalam layanan konseling.

### HASIL DAN PEMBAHASAN

#### A. Analisis Teori Keterkaitan Validitas dan Reliabilitas

Reliabilitas dan validitas saling berkaitan satu sama lain, tetapi keduanya tidaklah sama. Sebuah instrumen dapat menunjukkan reliabilitas meskipun tidak valid, yang berarti hasilnya konsisten namun tidak akurat atau relevan dalam mengukur konstruk yang dimaksud. Di sisi lain, meskipun tes memiliki validitas secara teoritis, jika tidak ada reliabilitas, hasil yang diperoleh akan tidak stabil dan sulit untuk diinterpretasikan dengan cara yang konsisten (Taherdoost, 2016; Tavakol dan Dennick, 2011).

Dalam kerangka kerja teori pengukuran klasik, reliabilitas diartikan sebagai proporsi variasi dalam skor yang mencerminkan skor sejati. Sementara itu, validitas memiliki dimensi yang lebih rumit karena membutuhkan bukti empiris dan argumen yang menunjukkan bahwa skor tes dapat digunakan dengan tepat sesuai dengan tujuan yang dimaksudkan (AERA, APA, dan NCME, 2014). Oleh karena itu, dalam proses pengembangan instrumen, langkah pertama yang sangat penting adalah memastikan konsistensi internal melalui pengujian reliabilitas, kemudian dilanjutkan dengan evaluasi validitas yang meliputi validitas isi, konstruk, dan kriteria.

Dalam proses pengembangan instrumen psikologis, aspek budaya sangat berpengaruh terhadap validitas. Tes yang dirancang dalam konteks budaya Barat tidak selalu memiliki tingkat validitas yang sama ketika diterapkan di Indonesia. Oleh karena itu, pengujian validitas lintas budaya sangat diperlukan, terutama pada tes kepribadian atau tes yang terpengaruh oleh norma sosial dan bahasa. Validitas konstruk harus diuji ulang melalui penelitian eksploratori, dan kadang-kadang perlu dilakukan modifikasi item untuk menyesuaikan dengan konteks lokal (Suprpto dan Suryani, 2021).

#### B. Studi Kasus Implementasi Validitas dan Reliabilitas dalam Bimbingan dan Konseling

Salah satu contoh kasus yang penting dan mencerminkan dalam bidang Bimbingan dan Konseling (BK) adalah pembuatan Inventori Minat Karier (IMK) oleh Sari dan Kurniawan (2022). Penelitian ini bertujuan untuk menciptakan alat ukur yang sah dan akurat mengenai minat karier berdasarkan teori Holland, yang mencakup enam tipe kepribadian karier, yaitu: Realistic, Investigative, Artistic, Social, Enterprising, dan Conventional (RIASEC). Pengembangan IMK ditujukan untuk membantu siswa SMA dalam menemukan minat karier mereka dengan menggunakan pendekatan yang berfokus pada psikologi kepribadian. Banyak siswa yang merasa bingung dalam memilih jurusan kuliah atau jalan profesi karena kurangnya pemahaman tentang minat dan karakteristik diri mereka.

1. Langkah awal dalam pengembangan instrumen ini adalah mengidentifikasi elemen dari teori Holland, lalu dilanjutkan dengan membuat indikator untuk masing-masing tipe RIASEC. Untuk memastikan bahwa isi alat ukur ini valid, item-item yang telah dihasilkan diajukan kepada tiga orang ahli dalam bidang Bimbingan dan Konseling serta seorang spesialis psikometri. Penilaian dilakukan dengan menggunakan metode Content Validity Index (CVI), yang merupakan indeks untuk mengukur persentase kesepakatan para ahli mengenai relevansi setiap item. Hasil CVI pada tahap ini menunjukkan bahwa lebih dari 90% item dinilai "sangat relevan" atau "relevan," dengan nilai I-CVI (Item-Level CVI) berkisar antara 0,80 hingga 1,00. Ini

mengindikasikan bahwa item-item tersebut memiliki validitas isi yang tinggi dan sesuai dengan konstruk teoritis yang ingin diukur (Sari dan Kurniawan, 2022).

2. Tahap berikutnya adalah pengujian validitas konstruk dengan menggunakan teknik Confirmatory Factor Analysis (CFA), dengan melibatkan 200 siswa SMA sebagai peserta uji coba. Temuan CFA menunjukkan bahwa model yang diterapkan memiliki kesesuaian (goodness of fit) yang memadai. Nilai-nilai indeks kesesuaian seperti Comparative Fit Index ( $CFI > 0,90$ ), Root Mean Square Error of Approximation ( $RMSEA < 0,5$ ), menunjukkan bahwa kontribusi masing-masing item terhadap konstruk yang diukur adalah signifikan dan dapat diterima. Ini menunjukkan bahwa alat ukur ini tidak hanya memiliki validitas isi yang baik, tetapi juga validitas konstruk yang memadai.

3. Untuk menguji reliabilitas, digunakan koefisien Cronbach's Alpha sebagai indikator konsistensi internal. Hasil analisis menunjukkan nilai alpha sebesar:

- Realistic: 0,84
- Investigative: 0,86
- Artistic: 0,83
- Social: 0,88
- Enterprising: 0,85
- Conventional: 0,82

Dengan nilai-nilai di atas ambang batas 0,70, maka instrumen ini dikategorikan sangat reliabel. Ini berarti bahwa item-item dalam masing-masing dimensi memiliki korelasi yang tinggi dan konsisten satu sama lain, sehingga instrumen dapat menghasilkan skor yang stabil bila digunakan pada waktu berbeda dalam kondisi yang sama.

Implementasi Inventori Minat Karier ini dalam setting sekolah memungkinkan konselor untuk memberikan bimbingan karier berbasis data objektif. Misalnya, ketika seorang siswa menunjukkan minat yang tinggi pada tipe Investigative dan Artistic, konselor dapat membantu siswa tersebut mengeksplorasi jurusan seperti biologi, psikologi, seni rupa, atau desain komunikasi visual.

### C. Kelebihan dan Kekurangan Pengembangan dalam Praktik

#### 1. Kelebihan

Penggunaan alat ukur yang sah dan dapat diandalkan dalam layanan Bimbingan dan Konseling (BK) memiliki beberapa manfaat. Pertama, hasil penilaian dapat dianggap bisa diandalkan untuk pengambilan keputusan, baik untuk konselor, guru, maupun orang tua. Kedua, instrumen yang bermutu meningkatkan ketepatan diagnosis dan efektivitas tindakan. Ketiga, dalam konteks institusi, data yang diperoleh bisa dimanfaatkan untuk evaluasi program dan pengembangan kebijakan pendidikan (Ramadhan et al., 2024).

#### 2. Kekurangan

Ada keterbatasan dalam sumber daya saat melakukan validasi alat ukur. Banyak konselor sekolah yang memakai instrumen tanpa memahami proses pengembangan serta kualitas psikometriknya. Situasi ini berpotensi menimbulkan bias dan kesalahan dalam interpretasi yang bisa merugikan konseli. Di samping itu, pengaruh budaya setempat sering kali tidak diperhatikan dalam pengembangan alat yang diadaptasi dari negara lain, padahal konteks nilai, bahasa, dan norma sosial sangat memengaruhi cara siswa menjawab pertanyaan dalam tes (Suprpto dan Suryani, 2021).

Masalah lainnya adalah kurangnya pelatihan dan pemahaman konselor tentang prinsip-prinsip psikometri. Dalam banyak situasi, penilaian dilakukan secara otomatis tanpa mempertimbangkan kevalidan atau keandalannya. Untuk itu, diperlukan peningkatan kemampuan profesional melalui pelatihan, lokakarya, dan kolaborasi dengan pakar pengukuran atau psikolog pendidikan.

## KESIMPULAN

Validitas dan reliabilitas adalah dua elemen penting dalam penciptaan tes psikologi. Validitas memastikan bahwa alat ukur benar-benar menilai aspek yang dimaksud, sedangkan reliabilitas memastikan hasil pengukuran tetap konsisten. Tanpa adanya kedua hal ini, hasil evaluasi menjadi tidak dapat diandalkan. Tinjauan teori memperlihatkan bahwa berbagai jenis

validitas dan metode untuk menguji reliabilitas perlu dipahami dan diterapkan dengan tepat saat mengembangkan alat ukur, terutama dalam bidang pendidikan dan bimbingan konseling.

Contoh pengembangan Inventori Minat Karier (IMK) menunjukkan bahwa pembuatan instrumen yang terencana dan berlandaskan teori dapat menghasilkan alat ukur yang valid dan dapat dipercaya. Meski demikian, masih ada kendala seperti representasi sampel dan pengujian antarbudaya yang perlu diperhatikan. Oleh karena itu, perlu adanya peningkatan kemampuan konselor dan mereka yang mengembangkan instrumen agar penilaian psikologis yang dilakukan tidak hanya tepat, tetapi juga etis serta bermanfaat bagi perkembangan siswa.

#### DAFTAR PUSTAKA

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas dan reliabilitas instrumen: Kajian teoretis dan aplikatif. *Journal on Education*, 6(2), 10967–10975.
- Sari, A. P., & Kurniawan, D. A. (2022). Pengembangan inventori minat karier berbasis teori Holland untuk siswa SMA. *Jurnal Konseling Indonesia*, 11(1),
- Suprpto, T., & Suryani, L. (2021). Validitas budaya dan adaptasi alat ukur psikologis dalam konteks Indonesia. *Jurnal Psikologi Integratif*, 5(1), 34–42.
- Taherdoost, H. (2016). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, 5(3), 28–36.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
- Trianingsih, L. (2023). Analisis kualitas tes dan butir soal integral pada evaluasi formatif matematika teknik. *Indonesian Journal of Curriculum and Educational Evaluation*, 9(2), 44–57.