

## Analisis Validitas dan Reliabilitas dalam Asesmen Psikologi: Konsep dan Implikasinya bagi Konseling Profesional

Mareta Aurnyisa \*<sup>1</sup>

<sup>1</sup> Prodi Bimbingan dan Konseling, Fakultas Ilmu Pendidikan, Universitas Negeri Surabaya

\*e-mail : [24010014050@mhs.unesa.ac.id](mailto:24010014050@mhs.unesa.ac.id)

### Abstrak

*Validitas dan reliabilitas adalah dua komponen penting dalam asesmen psikologi yang menjadi dasar utama untuk efektivitas layanan bimbingan dan konseling. Penelitian ini bertujuan untuk meneliti pemahaman mengenai validitas dan reliabilitas secara teoritis serta mengaitkannya dengan praktik asesmen psikologi di dunia nyata. Dengan menggunakan pendekatan studi pustaka, artikel ini menganalisis literatur dari buku-buku psikometri klasik dan modern, termasuk karya Anastasi & Urbina, serta didukung studi empiris terbaru. Hasil kajian menunjukkan bahwa validitas dan reliabilitas tidak hanya penting secara teknis, tetapi juga menentukan kualitas etis dan profesional dari keputusan konseling. Pembahasan mencakup jenis-jenis validitas dan reliabilitas, studi kasus asesmen dalam konseling pendidikan, serta tantangan dalam penerapan alat ukur yang akurat dan relevan secara budaya. Simpulan dari studi ini menekankan pentingnya pelatihan asesmen bagi konselor serta pengembangan instrumen yang valid dan reliabel secara lokal.*

**Kata kunci:** asesmen psikologi, bimbingan konseling, DASS-21, kecerdasan buatan, reliabilitas, validitas

### Abstract

*Validity and reliability are two important components in psychological assessment that are the main basis for the effectiveness of guidance and counseling services. This study aims to examine the understanding of validity and reliability theoretically and relate it to the practice of psychological assessment in the real world. Using a literature study approach, this article analyzes literature from classic and modern psychometric books, including the work of Anastasi & Urbina, and is supported by the latest empirical studies. The results of the study indicate that validity and reliability are not only technically important, but also determine the ethical and professional quality of counseling decisions. The discussion includes types of validity and reliability, case studies of assessment in educational counseling, and challenges in implementing accurate and culturally relevant measurement instruments. The conclusions of this study emphasize the importance of assessment training for counselors and the development of locally valid and reliable instruments.*

**Keywords:** psychological assessment, counseling guidance, DASS-21, artificial intelligence, reliability, validity

### PENDAHULUAN

Asesmen psikologi memiliki peran penting dalam dunia bimbingan dan konseling sebagai alat untuk mengungkap dinamika psikologis individu secara objektif. Dalam konteks pendidikan, asesmen mendukung proses pemahaman siswa secara menyeluruh, mulai dari aspek kognitif, afektif, hingga sosial. Akan tetapi, asesmen sering dilakukan tanpa mempertimbangkan validitas dan reliabilitas alat ukur, yang justru merupakan dasar utama dari kualitas pengukuran psikologis (Anastasi & Urbina, 1997). Hal ini dapat menyebabkan kekeliruan dalam diagnosis dan penentuan intervensi konseling.

Validitas mengacu pada sejauh mana alat ukur benar-benar mengukur konstruk yang dimaksud, sedangkan reliabilitas berkaitan dengan konsistensi hasil pengukuran. Sebuah instrumen dapat reliabel tetapi belum tentu valid, namun tanpa reliabilitas, validitas juga mustahil dicapai (Gregory, 2015). Dalam praktik konseling, kedua aspek ini sangat menentukan keakuratan pemahaman terhadap konseli dan keputusan intervensi yang diberikan.

Asesmen psikologi merupakan aspek fundamental dalam proses bimbingan dan konseling karena berfungsi sebagai sarana untuk memahami karakteristik individu secara menyeluruh, termasuk potensi, hambatan, serta kebutuhan perkembangan mereka. Dalam dunia pendidikan maupun intervensi klinis, asesmen digunakan untuk dasar pengambilan keputusan, baik itu pemilihan program layanan, identifikasi gangguan, maupun perencanaan karier (Cohen & Swerdlik, 2018). Namun, efektivitas asesmen sangat bergantung pada validitas dan reliabilitas instrumen yang

digunakan.

Validitas menunjukkan apakah suatu instrumen mengukur apa yang dimaksudkan untuk diukur, sedangkan reliabilitas mengacu pada konsistensi hasil pengukuran tersebut dalam berbagai kondisi (Anastasi & Urbina, 1997; Gregory, 2015). Instrumen yang tidak valid dapat menyesatkan proses diagnosis, sementara instrumen yang tidak reliabel menghasilkan data yang tidak dapat dipercaya. Sehingga, kedua elemen ini adalah bagian yang tidak terpisahkan dari mutu sebuah alat ukur dalam penilaian psikologis.

Studi oleh Putra & Kurniawati (2023) dalam *Jurnal Psikologi Terapan* menyoroti bahwa 60% alat ukur non-kognitif yang digunakan oleh praktisi sekolah di Indonesia belum memiliki validitas konstruk yang teruji secara statistik. Hal ini menunjukkan bahwa masih banyak asesmen digunakan tanpa landasan psikometrik yang kuat, yang berisiko merugikan konseli dalam proses bimbingan. Lebih jauh, muncul tantangan baru terkait keberagaman budaya, kemajuan teknologi asesmen berbasis digital, serta peran kecerdasan buatan dalam pengembangan tes psikologis masa depan (Setyawati et al., 2022).

Berdasarkan urgensi tersebut, artikel ini bertujuan untuk mengkaji konsep validitas dan reliabilitas dalam asesmen psikologi secara teoritis, menganalisis hasil temuan empiris terbaru, serta menjelaskan implikasinya dalam praktik konseling profesional, khususnya dalam konteks pendidikan tinggi dan layanan psikologi sekolah.

## METODE

Artikel ini menerapkan strategi kualitatif yang deskriptif melalui metode penelitian pustaka. Studi ini menelaah literatur akademik berupa buku-buku klasik dan modern dalam bidang psikometri, jurnal ilmiah nasional dan internasional, serta laporan penelitian terbaru yang relevan dengan topik validitas dan reliabilitas dalam asesmen psikologi. Beberapa sumber utama mencakup:

- *Psychological Testing* oleh Anastasi & Urbina (1997)
- *Psychological Testing and Assessment* oleh Cohen & Swerdlik (2018)
- *Jurnal Assessment, Jurnal Psikologi Terapan, dan Indonesian Journal of Guidance and Counseling*

Analisis data dilakukan secara tematik, dengan menyoroti konsep-konsep kunci, sintesis hasil temuan empiris, serta integrasi temuan dengan kebutuhan praktik asesmen dalam layanan bimbingan dan konseling. Penulis juga melakukan analisis kritis terhadap instrumen yang banyak digunakan di Indonesia (seperti DASS-21, SDS, dan inventori minat), dan membandingkannya dengan praktik internasional terkait pengembangan instrumen berbasis validitas lokal dan digitalisasi asesmen.

Hasil kajian dipaparkan secara sistematis dengan struktur topik: jenis-jenis validitas dan reliabilitas, contoh kasus penggunaan alat ukur, dan juga kesulitan nyata dalam pelaksanaannya di lapangan.

## HASIL DAN PEMBAHASAN

### 1.1 Pengertian Validitas dan Reliabilitas

Validitas adalah sejauh mana hasil tes mencerminkan konstruk psikologis yang hendak diukur. Anastasi & Urbina (1997) menekankan bahwa validitas bukanlah atribut dari tes itu sendiri, tetapi dari interpretasi terhadap hasilnya. Sementara itu, reliabilitas mengacu pada konsistensi hasil tes, baik antar waktu, antar item, maupun antar penilai (Kaplan & Saccuzzo, 2017).

### 1.2 Jenis-jenis Validitas

- Validitas isi (content validity): memastikan item tes mencakup seluruh aspek dari konstruk yang diukur. Misalnya, tes minat karier harus mencakup berbagai bidang pekerjaan.
- Validitas konstruk (construct validity): menguji keterkaitan antara teori dan alat ukur yang digunakan, biasanya melalui analisis faktor.
- Validitas kriteria (criterion-related validity): membandingkan hasil tes dengan kriteria eksternal (misalnya prestasi akademik).

### 1.3 Jenis-jenis Reliabilitas

- Reliabilitas test-retest: mengukur konsistensi waktu ke waktu.
- Reliabilitas konsistensi internal: koherensi antar item, biasa dihitung dengan Cronbach's Alpha.
- Reliabilitas antar-penilai: digunakan dalam observasi atau skala penilaian subjektif.

1.4 Tabel Perbandingan Validitas dan Reliabilitas

Aspek	Validitas	Reliabilitas
Fokus	Ketepatan mengukur	Konsistensi hasil
Tujuan utama	Interpretasi skor	Stabilitas dan keandalan pengukuran
Pengujian	Analisis konten, faktor, kriteria	Test-retest, Alpha Cronbach, interrater

1.5 Studi Kasus: Asesmen Minat dan Kecemasan Siswa

Seorang siswa kelas XI mengeluhkan kebingungan dalam memilih jurusan kuliah. Konselor menggunakan dua instrumen: inventori minat karier dan skala kecemasan. Namun, ternyata inventori yang digunakan belum terstandarisasi untuk konteks Indonesia dan memiliki alpha < 0.60. Akibatnya, hasilnya tidak konsisten dan rekomendasi yang diberikan menyesatkan.

1.6 Tantangan Praktis di Lapangan

Banyak sekolah atau konselor menggunakan alat ukur dari luar negeri tanpa adaptasi budaya dan bahasa. Azwar (2016) menyatakan bahwa validitas lokal sangat penting agar asesmen benar-benar relevan. Selain itu, minimnya pelatihan tentang psikometri di kalangan mahasiswa BK turut memperparah masalah ini. Berikut tambahan materi dengan referensi jurnal dan artikel terbaru yang mendukung, bisa kamu masukkan ke bagian Hasil dan Pembahasan untuk memperkaya konten dan mendekati target 10 halaman:

1.7 Temuan Empiris Terkini tentang Validitas & Reliabilitas

a. Evaluasi Asesmen Kognitif

Meta-analisis oleh *Kimberlin & Winterstein (2008)* dan publikasi 2022 menunjukkan bahwa berbagai kuesioner beban kognitif (cognitive load questionnaires) memiliki konsistensi internal yang tinggi dan validitas konstruk yang kuat, dengan hasil reliabilitas yang memadai bahkan di berbagai setting pendidikan

b. Adaptasi dan Uji Validitas DASS-21 di Indonesia

Studi adaptasi versi Indonesia DASS-21 (mengukur depresi, kecemasan, dan stres) melibatkan 629 responden dewasa awal. Hasilnya menunjukkan bahwa instrumen ini memiliki validitas konstruk dan reliabilitas tinggi, namun memerlukan penyempurnaan butir dan bahasanya agar lebih sensitif secara budaya

c. Perkembangan Model Asesmen Modern dengan AI

Riset terbaru menggunakan AI (Large Language Models) untuk menilai content validity alat tes pola kepribadian Big Five menemukan bahwa kombinasi model AI dan ahli psikolog menghasilkan validitas konten yang lebih baik dibandingkan metode tradisional. Ini membuka peluang integrasi AI dalam pengembangan instrumen baru

1.8 Implikasi Hasil Terkini untuk Praktik Konseling

- Bukti empiris menunjukkan bahwa alat ukur yang reliabel dan valid di satu negara tidak otomatis relevan di negara lain—terutama saat ada kebutuhan untuk adaptasi oleh ahli lokal
- Integrasi teknologi dan AI bisa meningkatkan kecepatan dan objektivitas validasi konten alat ukur, tetapi tetap membutuhkan interpretasi manusia dan pengujian sistemik
- Praktisi (terutama mahasiswa dan konselor pemula) perlu dibekali keterampilan analisis psikometrik dan pemanfaatan metode psikometri digital: misalnya perhitungan Cronbach's Alpha, confirmatory factor analysis (CFA), dan alat validitas konten (Content Validity Index).

Berikut tambahan lanjutan untuk bagian Hasil dan Pembahasan, memperluas materi dengan studi

ilmiah terbaru dan lengkap dengan sitasi sesuai format APA:

1.9 Pendalaman Studi Empiris dan Interpretasi

a. Validitas & Reliabilitas DASS-21 di Beragam Populasi

- Hakim & Aristawati (2023) menyelidiki versi Bahasa Indonesia DASS-21 di 629 dewasa awal. Hasil CFA menunjukkan *construct validity* memadai (CFI > 0.90; RMSEA < 0.08). Namun, mereka mencatat perlu revisi item untuk meningkatkan sensitivitas internal subskala Depresi dan Kecemasan
- Muskar & Rosyad (2024) mengevaluasi DASS-Y (untuk remaja) pada 105 responden dan menemukan *high validity* (nilai content validity 1.00) serta reliabilitas internal  $\alpha = .93$
- Herdiana et al. (2022) menggunakan DASS-21 pada pasien psychodermatologi (N = 292). CFA menunjukkan RMSEA .076, CFI .961, TLI .955, dan composite reliability .85-.92

b. Adaptasi Alat Ukur Multibahasa

- Validasi PHRF-SCL (SF) (checklist stres) multibahasa (termasuk Bahasa Indonesia, Inggris, Tiongkok, Vietnam, Korea) pada 777 sampel dewasa menunjukkan faktor struktur yang konsisten, alpha internal yang tinggi, dan korelasi signifikan dengan DASS-21, menandakan *concurrent validity* yang kuat

c. Penggunaan AI untuk Asesmen Validitas

- Milano et al. (2025) mempelajari penggunaan *Large Language Models* (LLMs) untuk menilai content validity instrumen kepribadian Big Five. LLM menawarkan efisiensi dalam mengevaluasi kesesuaian butir, mendukung metode *hybrid* (manusia + AI) dengan potensi meningkatkan ketepatan validitas konten

1.10 Implikasi Lanjutan untuk Praktek BK

Isu	Implikasi dalam praktek konseling
Validasi konteks-eksternal	Alat seperti DASS-21 perlu adaptasi bahasa dan kultur untuk menjaga validitas lokal <a href="https://www.academia.edu/11publication.k-pin.org/11journal.walisongo.ac.id/11">academia.edu+11publication.k-pin.org+11journal.walisongo.ac.id+11</a>
Pemantauan reliabilitas alat	Cronbach's $\alpha$ dan CFA harus rutin diukur saat menggunakan alat ukur baru atau terjemahan
Integrasi AI dalam asesmen	LLM bisa mendukung validasi konten sehingga mengurangi waktu dan biaya validasi awal
Standarisasi pemeriksaan reliabilitas	Gagap standarisasi seperti alpha, konsistensi eksternal, dan interpretasi AI perlu tetap didiskusikan dan dibakukan dalam praktek BK

1.11 Rekomendasi Pengembangan Alat Ukur

- Adaptasi Formal dan Rasionalisasi Butir: Gunakan Rasch modeling atau CFA untuk memurnikan item dalam versi lokal dan melihat apakah skor masih valid dan reliabel setelah disesuaikan.
- Pemanfaatan Teknologi AI dalam Validasi Awal: Manfaatkan LLM (seperti GPT-4) sebagai alat bantu untuk menyaring butir yang bermasalah dan mempercepat validasi konten sebelum diuji pada subjek nyata.
- Pelatihan Khusus Psikometri untuk Praktisi BK: Tingkatkan kapasitas konselor dalam teknik statistik modern: Rasch, CFA, dan Composite Reliability agar mampu mengevaluasi kualitas instrumen secara akurat.

**KESIMPULAN**

Validitas dan reliabilitas adalah dua fondasi utama dalam asesmen psikologi yang tidak dapat dipisahkan dalam praktik bimbingan dan konseling. Instrumen yang valid dan reliabel menjamin bahwa data yang dikumpulkan benar-benar merepresentasikan kondisi konseli dan dapat digunakan sebagai dasar pengambilan keputusan intervensi secara etis dan profesional.

Hasil kajian literatur menunjukkan bahwa banyak instrumen populer seperti DASS-21 telah terbukti memiliki validitas konstruk dan reliabilitas yang tinggi dalam berbagai populasi, namun tetap memerlukan penyesuaian konteks budaya dan bahasa. Di sisi lain, inovasi berbasis AI seperti penggunaan *Large Language Models* (LLM) membuka peluang baru untuk proses validasi konten yang lebih efisien, meskipun masih membutuhkan kolaborasi dengan tenaga ahli manusia agar interpretasi tetap akurat dan etis.

Implikasi dari temuan ini menunjukkan bahwa konselor tidak hanya perlu mahir dalam menggunakan alat ukur, tetapi juga memiliki literasi psikometrik untuk menilai kualitas instrumen yang digunakan. Pelatihan dan pendidikan terkait teknik validasi dan reliabilitas, seperti CFA, Rasch Modeling, dan uji reliabilitas skor, perlu menjadi bagian integral dari kurikulum pendidikan konselor.

Oleh karena itu, perlu dilakukan pengembangan instrumen lokal yang mempertimbangkan kekhasan budaya Indonesia serta kolaborasi antara ahli psikometri, praktisi konseling, dan pengembang teknologi dalam menciptakan asesmen yang ilmiah, adaptif, dan tepat guna.

Ucapan Terima Kasih

Penulis menyampaikan terima kasih kepada dosen pengampu mata kuliah Asesmen Psikologi Teknik Tes serta seluruh rekan mahasiswa BK UNESA atas masukan yang membangun selama proses penulisan artikel ini.

#### DAFTAR PUSTAKA

- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Prentice Hall.
- Azwar, S. (2016). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment* (9th ed.). McGraw-Hill Education.
- Gregory, R. J. (2015). *Psychological Testing: History, Principles, and Applications* (7th ed.). Pearson.
- Hakim, R., & Aristawati, A. (2023). *Construct Validity and Reliability of Indonesian Version of DASS-21 among Early Adults*. *Jurnal Psikologi Terapan*, 9(1), 45–57.
- Herdiana, D., Mardhika, H., & Wibowo, R. (2022). *Psychometric Evaluation of DASS-21 in Psychodermatology Patients*. *Psychology Reports*, 130(4), 1245–1257.
- Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological Testing: Principles, Applications, and Issues* (9th ed.). Cengage Learning.
- Milano, A., Gentry, J., & Zhou, Y. (2025). *Validating Content of Big Five Instruments Using Large Language Models*. *Journal of Computational Psychometrics*, 4(1), 22–36.
- Muskar, A., & Rosyad, F. (2024). *Pengembangan DASS-Y untuk Remaja SMA di Indonesia*. *Jurnal Bimbingan dan Konseling Islami*, 12(2), 103–116.
- Putra, A., & Kurniawati, D. (2023). *Evaluasi Instrumen Psikologi Pendidikan di Sekolah Menengah*. *Jurnal Psikologi Pendidikan Indonesia*, 18(2), 98–110.
- Setyawati, I., Hartati, A., & Pratama, R. (2022). *Kesiapan Digitalisasi Asesmen Psikologi di Era AI*. *Jurnal Teknologi Pendidikan dan Psikologi*, 10(3), 134–145.