

## STUDI LITERATUR: KONSEP DASAR PENGUKURAN

Wenda Asmita<sup>1</sup>, Wahidah Fitriani<sup>2</sup>

<sup>1,2</sup>Jurusan Bimbingan dan Konseling Pendidikan Islam, Program Pascasarjana UIN Mahmud Yunus  
Batusangkar  
E-mail:wendaasmita@gmail.com/082284730828

### ABSTRAK

Dalam evaluasi dunia pendidikan, tentu tidak akan terlepas dari dari proses pengukuran dan penilaian. Lebih khusus bagi guru Bimbingan dan Konseling pelaksanaan pengukuran yang tepat dan benar perlu diimplementasikan. Namun permasalahan yang timbul ternyata masih ada lulusan bimbingan dan konseling yang belum begitu memahami mengenai konsep pengukuran. Sehingga artikel ini diharapkan mampu menjadi solusi dalam membantu pemahaman terhadap konsep pengukuran dalam bimbingan dan konseling. Metode yang digunakan adalah studi pustaka yang membahas mengenai konsep pengukuran. Baik dari sumber buku maupun jurnal terakreditasi lainnya. Sehingga diketahui bahwa kegiatan mengumpulkan data yang berkenaan dengan pelayanan bimbingan dan konseling kepada sasaran-sasaran tertentu di mana data tersebut berbentuk kuantitatif/angka disebut sebagai Pengukuran dalam Bimbingan dan Konseling.

**Kata kunci:** Pengukuran, Guru BK.

---

### ABSTRACT

*In evaluating the world of education, of course, it will not be separated from the measurement and assessment process. More specifically, for Guidance and Counseling teachers, proper and correct measurement needs to be implemented. However, the problems that arise are that there are still guidance and counseling graduates who do not really understand the concept of measurement. So that this article is expected to be a solution in helping understanding the concept of measurement in guidance and counseling. The method used is a literature study which discusses the concept of measurement. Both from book sources and other accredited journals. Which resulted in several points, namely the concept of measurement, measurement with a scale, reliability, validity and assessment development.*

**Keywords:** *Measurement, Guidance and Counseling*

## PENDAHULUAN

Dalam kehidupan sehari-hari tanpa disadari kita sering melakukan evaluasi, mengukur, dan melakukan penilaian. Namun, masih banyak orang yang belum memahami arti sebenarnya dari kata pengukuran, penilaian dan evaluasi, dan kebanyakan orang cenderung mengartikan ketiga kata tersebut dengan pengertian yang sama.

Biasanya kebanyakan orang mengidentikkan kegiatan penilaian dengan mengevaluasi karena biasanya kegiatan mengukur termasuk di dalamnya. Pengukuran, penilaian, dan evaluasi merupakan kegiatan yang berjenjang. Artinya ketiga kegiatan tersebut tidak dapat dipisahkan satu sama lain dan harus dilakukan secara berurutan.

Pengukuran merupakan suatu proses pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal atau objek tertentu menurut aturan atau formulasi yang jelas (Lupyanto and Dwikurnaningsih 2014). Salah satu tujuan pengukuran adalah untuk melihat sikap kreatif mahasiswa program studi bimbingan konseling terhadap penguasaan materi perkuliahan yang dipelajarinya. Mahasiswa pada dasarnya dapat menggunakan instrumen pengukuran untuk melakukan evaluasi diri, yaitu mengukur sejauhmana sikap kreatif mahasiswa program studi bimbingan konseling telah dikembangkan, termasuk sikap kreatif menggunakan kreativitas (Abdurrahman and Suarti, Alit, S 2016).

Pengukuran yang pernah dilakukan sebelumnya oleh konselor diantaranya adalah analisis kondisi *burn out* siswa (Supriyanto and Prasetiawan 2021), pengukuran sejauh mana kompetensi bimbingan karir kolaboratif guru BK/konselor (Afdal 2019) serta alat pengukuran data adalah kuisioner berbentuk skala yang dikembangkan berdasarkan indikator *survival and safety skills untuk SMP dari Comprehensive School Counseling Program Guide 2009* (Putri 2020). Dengan menggunakan alat ukur berupa Skala Pengukuran Karakter Adil (Setyaputri, Y, Kriphianti, Dwi, and Nawantara, Dewi 2020).

## METODE

Penelitian ini menggunakan jenis penelitian yang berupa studi kepustakaan (library research). Studi pustaka berkaitan dengan kajian teoritis dan beberapa referensi yang tidak bisa dipisahkan dari literatur-literatur ilmiah (Putri 2019). Sumber data dalam penelitian ini diperoleh dari literatur yang relevan seperti buku, majalah atau artikel ilmiah yang

berkaitan dengan topik yang dipilih: Teknik pengumpulan data yang digunakan dalam penelitian buku ini mencari data tentang item atau variabel dari catatan, buku, artikel atau artikel, jurnal dll (Arikunto, S., & Jabar, 2010; Made et al., 2022; Asmita and Fitriani 2022; Yarman and Handayani 2021).

Kegiatan yang melibatkan menemukan, membaca, dan meninjau hasil penelitian atas literatur yang mengandung prinsip-prinsip yang terkait disebut studi literatur dan bahkan dikenal sebagai bagian penting dari keseluruhan pendekatan penelitian ilmiah (Amalianingsih 2021). Data yang terkumpul diseleksi dan diurutkan sesuai dengan topik kajian. Kemudian dilakukan penyusunan karya tulis berdasarkan data yang telah dipersiapkan secara logis dan sistematis. Teknik analisis data bersifat deskriptif argumentative. Penulisan dilakukan dengan melihat relevansi dan sinkronisasi antar satu data/informasi satu dengan data/informasi lain sesuai dengan topik yang dikaji. Selanjutnya, kemudian dilakukan penarikan kesimpulan karya tulis. Simpulan didapatkan setelah merujuk kembali pada rumusan masalah, tujuan penulisan, serta pembahasan. Simpulan yang ditarik mempresentasikan pokok bahasan karya tulis, serta didukung dengan saran praktis sebagai rekomendasi selanjutnya. Metode ini akan memudahkan pembaca (Sari and Yustiana 2022).

## HASIL DAN PEMBAHASAN

### A. Konsep Dasar Pengukuran

Istilah pengukuran sangat sering kita dengar dalam berbagai aspek kehidupan. Terkadang kita tidak menyadari bahwa dalam hidup ini kita sering mengukur. Misalnya, ketika Anda ingin membuat pakaian, penjahit akan mengukur pinggang, lebar bahu, dll. Contoh lain adalah ketika seseorang ingin menulis surat keterangan sehat, mereka perlu mengetahui tinggi dan berat badannya.

Menurut (Ota 2015) pengukuran diartikan sebagai proses membedakan sesuatu (*A process by which things differ*), sedangkan secara operasional pengukuran adalah penerapan aturan-aturan bilangan pada objek atau fenomena tertentu, dalam penelitian kuantitatif pengukuran ditempatkan pada variabel-variabel yang kita

Dipublikasikan Oleh :

UPT Publikasi dan Pengelolaan Jurnal

Universitas Islam Kalimantan Muhammad Arsyad Al-Banjari Banjarmasin

pelajari. Pandangan lain menyatakan bahwa ketika pengukuran bersifat kontekstual dalam pendidikan, dapat disebut sebagai proses yang digunakan untuk mengumpulkan secara numerik tingkat kompetensi individu (Adom, Mensah, and Dake 2020).

Pengukuran dapat didefinisikan sebagai proses sistematis untuk mengevaluasi dan membedakan apa yang sedang diukur. Pengukuran disesuaikan dengan aturan tertentu. Aturan yang berbeda memerlukan skala dan metrik yang berbeda. Saat memproses dan menganalisis data, kami sangat memperhatikan sifat skala pengukuran yang digunakan. Pemilihan peralatan statistik yang digunakan dalam operasi matematika dan pengolahan data pada dasarnya memiliki persyaratan tertentu dalam hal skala pengukuran datanya. Ketidaksesuaian antara skala Pengukuran yang dilakukan menggunakan operasi matematika/peralatan statistik dapat menghasilkan kesimpulan yang bias dan tidak akurat/relevan (Junaidi 2015).

Dari pendapat para ahli di atas dapat disimpulkan bahwa pengukuran adalah proses atau proses "penomoran" suatu benda menurut aturan-aturan tertentu. Dalam pendampingan dan konsultasi, kegiatan pengukuran sering dilakukan oleh pihak-pihak yang berkepentingan. Dari pengertian di atas dapat disimpulkan bahwa pengukuran dalam pembinaan dan konseling mengacu pada kegiatan pengumpulan data yang berkaitan dengan layanan pembinaan dan konseling untuk suatu mata pelajaran tertentu, dimana datanya berupa angka/angka. Misalnya, ketika seorang instruktur melakukan tes psikologi pada seorang siswa, diperoleh angka tentang tes yang dia lakukan.

## B. Pengukuran dengan Skala

Menurut (Hays 2013), Bayangkan sebuah kotak peralatan yang mungkin Anda miliki untuk proyek perbaikan rumah. Dalam kotak alat ini Anda memiliki berbagai alat, beberapa lebih khusus daripada yang lain atau lebih sesuai untuk proyek tertentu. Memilih skala pengukuran seperti memilih alat dari kotak peralatan: Konselor yang mengukur variabel penilaian memilih alat atau skala yang tepat untuk fungsi tertentu. Dengan demikian, skala pengukuran adalah cara yang berbeda kita mengukur konstruksi dalam penilaian proses. Ada empat skala pengukuran, dengan skala yang lebih

kompleks (yaitu yang melakukan lebih banyak fungsi) memungkinkan pengukuran yang lebih tepat: nominal, ordinal, interval, dan rasio. Bagaimana Anda memilih untuk mengukur variabel dalam proses penilaian mengarahkan skala pengukuran apa yang Anda gunakan. Semakin tepat konstruk yang ingin dinilai oleh konselor, semakin "maju" skala yang dipilih.

## 4 Skala Pengukuran:

### 1. Skala Nominal

Skala nominal adalah skala pengukuran paling dasar yang digunakan untuk penamaan atau penggolongan saja. Bahkan juga disebut skala yang sederhana. Contoh skala nominal termasuk jenis kelamin, ras, afiliasi politik, mode dalam distribusi data, atau ada atau tidak adanya kriteria diagnostik. Skala nominal tidak memiliki besaran, internal yang sama, atau nol mutlak. Sebuah variabel dapat dikodekan dengan angka (misalnya, 0 = mahasiswa sarjana, 1 = mahasiswa pascasarjana); namun, angka-angka tersebut tidak menunjukkan besarnya (yaitu, 1 lebih besar dari 0).

### 2. Skala Ordinal

Skala ordinal mengacu pada urutan atau peringkat kategori nominal. Karena informasi dapat diperoleh tentang seorang individu yang memiliki lebih atau kurang beberapa variabel, skala ordinal dicirikan oleh besarnya. Namun, interval atau "spasi" antara peringkat tidak mungkin sama, dan dengan demikian ukuran relatif antar interval sulit untuk diketahui. Skala ordinal sepuluh digunakan dalam penilaian psikologis, seperti ketika skala Likert (misalnya, 1 = *sangat tidak setuju* sampai 6 = *sangat setuju*) merupakan penilaian. Namun jarak antara skor 1 dan 2 dibandingkan dengan skor antara 5 dan 6 tidak mungkin sama. Contoh skala ordinal termasuk tingkat kepuasan kerja, peringkat nasional universitas, dan nilai median atau peringkat persentil dari distribusi data. Konselor diperingatkan terhadap rata-rata (menghitung rata-rata) dari skor tes, seperti dalam kasus tingkat kepuasan kerja. Jika kepuasan kerja diukur pada skala Likert atau skala penilaian lain dengan interval yang tidak sama menghitung skor tipikal bermasalah. Seringkali, konselor keliru

memperlakukan skala ordinal seperti skala interval ketika menafsirkan dan melaporkan data penilaian.

### 3. Skala Interval

Skala **interval** memiliki besaran dan interval yang sama. Contoh skala interval termasuk suhu, daftar periksa perilaku, dan standar deviasi. Karena ada jarak yang sama antara titik data, konselor dapat mengasumsikan bahwa perbedaan antara dua titik sama dengan dua titik lainnya (misalnya, perbedaan antara 20° dan 25° sama dengan perbedaan antara 100° dan 105°). Meskipun Anda dapat menambah dan mengurangi nilai pada skala interval dengan percaya diri, Anda tidak boleh mengalikan atau membagi nilai karena tidak ada nol mutlak. Dengan contoh suhu, Anda dapat mengatakan bahwa 105° lebih hangat dari 100°, tetapi Anda tidak dapat mengatakan bahwa itu 5% lebih hangat, karena 0° sebenarnya tidak berarti tidak adanya suhu. Konselor harus diingatkan bahwa lebih banyak (misalnya, skor penilaian yang lebih tinggi) tidak selalu berarti lebih baik. Karena variabel tidak diukur secara sempurna dalam penilaian, sulit untuk mengetahui apa arti skor sebenarnya.

### 4. Skala Rasio

Skala **rasio** adalah skala pengukuran yang paling canggih dan tepat. Skala rasio memiliki besaran, interval yang sama, dan nol mutlak. Contoh skala rasio termasuk waktu dan tinggi. Karena 0 detik atau 0 inci dimungkinkan, misalnya, Anda dapat mengalikan dan membagi nilai-nilai ini (misalnya, satu klien menyelesaikan penilaian dua kali lebih cepat dari klien lain). Skala rasio jarang digunakan dalam penilaian konseling.

## C. Keandalan (*Reliability*)

Reliabilitas digunakan untuk menilai tingkat konsistensi dari instrument penelitian, apakah instrumen tersebut konsisten untuk digunakan mengumpulkan data. Suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya (Arifin 2017). Reliabilitas sebagai subjek teknis sangat penting bagi mereka yang menggunakan tes psikologi. Misalnya, Anda tertarik mengukur gejala

penyalahgunaan zat untuk membuat keputusan pengobatan. Anda memilih alat skrining penyalahgunaan zat (mis., Inventaris Skrining Penyalahgunaan Zat, atau SASSI-3; FG Miller & Lazowski, 1999) dan Anda menemukan individu tersebut kemungkinan hadir dengan ketergantungan zat berdasarkan kunci penilaian. Karena Anda merekomendasikan perawatan lebih lanjut untuk klien ini, Anda ingin memastikan bahwa skor akan konsisten jika SASSI-3 diberikan lagi. Dan, Anda ingin memastikan klien merespons secara konsisten selama administrasi tunggal. Reliabilitas mengacu pada seberapa konsisten suatu tes mengukur dan sejauh mana tes tersebut menghilangkan peluang dan faktor asing lainnya dalam hasilnya. Sinonim untuk keandalan meliputi ketergantungan, reproduktifitas, stabilitas, dan konsistensi.

Pada bagian pertama dari bagian ini, informasi konseptual tentang keandalan disajikan. Informasi konseptual ini mengacu pada pengertian kesalahan pengukuran; reliabilitas adalah koefisien korelasi dari kesalahan pengukuran itu. Setelah informasi dasar ini, informasi tentang jenis reliabilitas yang ditemui konselor dalam proses penilaian disajikan

### 1. Kesalahan Pengukuran

Konselor tertarik untuk mengukur karakteristik manusia yang jauh lebih rumit daripada aspek fisik manusia, dan kualitas kompleks seperti kecemasan, kecerdasan, depresi, dan pencapaian sulit untuk didefinisikan secara tepat maupun untuk diukur. Kesalahan pengukuran, kemudian, adalah bias positif atau negatif dalam skor yang diamati. Artinya, skor yang diterima seseorang pada suatu tes terdiri dari dua elemen: skor sebenarnya orang tersebut dan kesalahan skor yang mungkin menambah atau mengurangi dari skor sebenarnya. Skor sebenarnya tidak pernah diketahui. Karena kesalahan yang lebih besar menghasilkan lebih banyak ketidakkonsistenan dalam skor, kesalahan pengukuran harus diminimalkan. Berikut adalah ilustrasi tentang bagaimana kesalahan memengaruhi skor:  $X = T + e$ , dimana skor yang diamati ( $X$ ) = skor sebenarnya ( $T$ ) + skor kesalahan ( $e$ ). Sebuah

*Dipublikasikan Oleh :*

*UPT Publikasi dan Pengelolaan Jurnal*

*Universitas Islam Kalimantan Muhammad Arsyad Al-Banjari Banjarmasin*

tes dengan reliabilitas sempurna akan sama dengan tidak ada kesalahan pengukuran. Situasi ini sebenarnya tidak mungkin, karena kesalahan berasal dari individu, pengujian, dan / atau pengujian kondisi. Mari kita pertimbangkan lagi contoh SASSI-3 dengan sekelompok klien. Individu kesalahan mungkin termasuk kecemasan ujian, motivasi, minat dalam menanggapi yang diinginkan secara sosial cara, heterogenitas kelompok yang diuji, dan uji keakraban. Kesalahan pengujian mungkin ditemukan jika, Misalnya, tidak semua gejala penyalahgunaan zat diwakili secara memadai oleh instrumen. Item atau jika ada homogenitas item (item yang bervariasi akan memiliki error yang lebih besar) atau panjang tes tidak ideal (misalnya, tes yang lebih pendek menghasilkan perkiraan keandalan yang lebih rendah untuk sampel). Kesalahan kondisi esting mungkin mengacu pada kesalahan penilaian atau gangguan dalam lingkungan pengujian, seperti waktu administrasi yang tidak memadai atau agensi yang bisung.

Mencoba untuk mempertahankan kondisi pengujian yang seragam dengan mengendalikan instruksi, batas waktu, sebuah lingkungan pengujian dilakukan untuk mengurangi varian kesalahan dan membuat nilai tes lebih dapat diandalkan. Tidak ada tes yang menghasilkan skor yang sangat dapat diandalkan, dan karena psikologika. Item pengukuran sering tidak tepat, penting untuk memeriksa keakuratan dan konsistensinya. Item instrumen secara konstan untuk memastikan bahwa ketidakandalan disimpan dalam batas wajar batas. The Standards (AERA, APA, & NCME, 1999) menekankan bahwa pengembang pengujian harus melakukannya provide menguji pengguna dengan sejumlah besar informasi tentang keandalan tes dan pengukuran-laki-laki kesalahan. Informasi ini harus mencakup perincian spesifik tentang populasi tempat kepercayaan kemampuannya data diperoleh, kesalahan standar pengukuran untuk semua jenis skor yang dilaporkan, sebuah interval antara tes

ulang dan konsistensi antar penilai jika sesuai.

## 2. Jenis Keandalan/Reliabilitas

### a. Test – Uji Ulang Keandalan

Test-tes ulang keandalan mengukur konsistensi dari waktu ke waktu. Korelasi Koefisien dalam hal ini menunjukkan hubungan antar skor yang diperoleh individu dalam kelompok yang sama pada dua administrasi pengujian. Korelasi tes-tes ulang cenderung menurun saat interval antara administrasi pengujian memanjang. Jika intervalnya singkat, ada efek praktik dan memori potensial, yang cenderung membuat estimasi keandalan tes palsu tinggi. Jika interval waktu terlalu lama, variasi dapat dipengaruhi oleh peristiwa yang terjadi pada peserta antara dua administrasi tes (misalnya, pematangan dan nya-efek tory), dan perkiraan keandalan yang semu rendah dapat diperoleh.

### b. Keandalan Bentuk Alternatif

Keandalan bentuk alternatif, atau keandalan bentuk paralel, dihitung dengan membandingkan konsistensi skor individu dalam kelompok yang sama pada dua alternatif tetapi setara bentuk tes yang sama. Karena item tes berbeda, efek memori dan efek sisa lainnya dieliminasi. Pertanyaan krusial tetap apakah sebenarnya file dua bentuk tes alternatif sebenarnya setara.

Dua tes yang mengukur konten atau variabel yang sama dan yang setara di tingkat kesulitan dapat dilakukan pada hari yang sama atau sangat dekat satu sama lain tanpa memperhatikan efek latihan. Mereka dapat berganti-ganti sehingga Tes A diberikan kepada satu kelompok terlebih dahulu dan Tes B kepada kelompok lain terlebih dahulu, dan efek latihan dengan demikian dapat dikontrol. Masalah dengan jenis reliabilitas tes ini adalah seringkali cukup sulit untuk menghasilkan satu bentuk tes yang baik, apalagi dua bentuk tes yang baik. Oleh karena itu, kecuali ada

program pengujian nasional dengan staf yang bekerja untuk mengembangkan formulir tes — seperti halnya dengan beberapa program pengujian nasional, seperti tes MCAT atau ACT — harapan untuk bentuk reliabilitas ini tidak realistis.

Keandalan split-half adalah bentuk populer dari tes reliabilitas tabulating karena dapat diperoleh dari satu administrasi dengan membagi tes menjadi dua bagian yang sebanding dan membandingkan dua skor yang dihasilkan untuk setiap individu (Gambar 5.4). Ini dikelola sekaligus, jadi tidak ada fluktuasi waktu ke waktu yang terjadi. Dari sudut pandang ini, ini dapat dianggap sebagai kasus khusus keandalan bentuk-alternatif. Di sebagian besar tes, paruh pertama dan paruh kedua tidak akan terjadi sebanding karena perbedaan dalam kesulitan item serta efek latihan dan kelelahan yang cenderung bervariasi dari awal hingga akhir tes. Oleh karena itu, sebagian besar tes dibagi menjadi item ganjil dan genap, kecuali jika beberapa item berhubungan dengan masalah tertentu, dalam hal ini seluruh kelompok item ditugaskan ke satu atau separuh lainnya.

Kelemahan penting dalam pendekatan split-half terletak pada prinsip umum pengambilan sampel — yaitu, biasanya semakin banyak jumlah item, semakin stabil konsep yang diukur. Semua hal dianggap sama, semakin lama tesnya, semakin dapat diandalkan nilainya. Prosedur split-half memotong panjang tes menjadi dua, sehingga menurunkan estimasi reliabilitas. Untuk mengoreksi reliabilitas yang dihitung berdasarkan panjang yang lebih pendek, rumus Spearman – Brown dapat digunakan untuk menghasilkan perkiraan seperti apa reliabilitas jika diperoleh pada tes panjang penuh. Misalnya, jika Anda mengoreksi estimasi reliabilitas split-half (dihitung sebagai koefisien korelasi Pearson) sebesar 0,80, estimasi

reliabilitas yang dikoreksi adalah 0,89 (lihat Lampiran C untuk rumus statistik). Namun, rumus Spearman-Brown tidak akan mengoreksi perkiraan keandalan secara substansial untuk pengujian dengan jumlah item yang lebih banyak, sebagian karena ada penurunan kesalahan pengukuran dengan lebih banyak item.

c. Konsistensi Internal

Konsistensi antar item adalah ukuran konsistensi internal yang menilai sejauh mana item-item pada tes terkait satu sama lain dan dengan skor total (Gambar 5.5). Pengukuran reliabilitas skor tes ini memberikan perkiraan interkorelasi rata-rata antara semua item pada tes. Bergantung pada jenis respons yang diminta pada instrumen, rumus yang dikenal sebagai Rumus Kuder – Richardson (KR) 20 untuk jawaban dua-respons (misalnya, benar atau salah, ya atau tidak) atau koefisien reliabilitas alfa Cronbach (yaitu, koefisien alfa) untuk lebih dari dua alternatif yang komputed (lihat Lampiran C). Semua respon item individu untuk setiap orang di seluruh sampel dianalisis, dan koefisien reliabilitas yang dihasilkan menunjukkan konsistensi item sampel sifat yang sedang diukur. Angka berkorelasi dengan skor total untuk klien dan kemudian dibandingkan dengan variabilitas skor keseluruhan. Pada dasarnya, ini adalah korelasi dari semua reliabilitas split-half.

Koefisien reliabilitas interitem (serta koefisien reliabilitas split-half) juga dapat dikuadratkan untuk menjelaskan varians, dengan cara yang mirip dengan bagaimana koefisien determinasi dihitung sebelumnya untuk dua set skor. Sebagai contoh, mari kita asumsikan kita menghitung estimasi reliabilitas 0,75. Jika kami mengkuadratkan koefisien, kami mencatat bahwa sekitar 56% varian dari konstruk itu (misalnya, kecemasan) dijelaskan oleh penilaian itu.

*Dipublikasikan Oleh :*

*UPT Publikasi dan Pengelolaan Jurnal*

*Universitas Islam Kalimantan Muhammad Arsyad Al-Banjari Banjarmasin*

#### D. Validitas

Menurut (Simmamora 2012) validitas adalah suatu ukuran yang menunjukkan tingkat kevalidan dan kesahihan suatu instrument, suatu instrument dianggap valid apabila mampu mengukur apa yang mau diukur, dengan kata lain mampu memperoleh data yang tepat dari variabel yang diteliti. Pertimbangkan dua set skor berikut untuk inventaris harga diri untuk 10 anak (skor yang lebih tinggi menunjukkan harga diri yang lebih tinggi).

##### 1. Jenis Validitas

Bukti salah satu validitas tes untuk tujuan tertentu dapat dinilai dengan cara yang berbeda, sebagaimana dicatat dalam Standar (AERA, APA, & NCME, 1999; Urbina, 2004). Penekanan ditempatkan pada sifat dan kekuatan bukti untuk interpretasi tertentu. Berbagai jenis bukti validitas dibahas di bawah ini: konten, terkait kriteria, konstruksi, dan perlakuan. Konten dan kriteria terkait dapat dianggap dimasukkan dalam validitas konstruk (Messick, 1995).

Validitas wajah, yang sebenarnya bukan bukti validitas, ditentukan jika penilaian "tampak seperti" mengukur apa yang seharusnya diukur.

##### a. Validitas Isi

Content validitas lihat untuk keterwakilan item dari "populasi" item. Barang Misalnya, pertanyaan tentang tes atau pertanyaan dalam wawancara klinis. Apakah item representative dari domain yang diminati? Seberapa baik Anda mengambil sampel? Untuk pendidikan penilaian di sekolah, item tes dipilih dengan memeriksa kurikulum, buku teks, dan lainnya. Jika tes dirancang untuk mengukur prestasi fisika sekolah menengah atas, sejumlah guru fisika sekolah menengah atas, dan mungkin beberapa perguruan tinggi fisika. Untuk guru, periksalah item-item pada tes untuk menentukan apakah item-item ini benar-benar adamengukurg pengetahuan tentang apa yang biasanya diajarkan dalam fisika sekolah menengah. Kemudian, item ditimbang dengan cara

tertentu agar sesuai dengan kepentingan relatifnya terhadap pencapaian fisika. Validitas konten, kemudian, melibatkan pendefinisian domain minat (yaitu, prestasi fisika), mengidentifikasi item yang merujuk ke domain tersebut (yaitu, kurikulum dan / atau tujuan kinerja), melibatkan juri untuk memastikan bahwa item tersebut sesuai dengan domain dan item tersebut diambil sampelnya secara memadai, dan kemudian menentukan berapa banyak item dari komponen domain harus disertakan untuk mewakili domain itu.

##### b. Validitas Terkait Kriteria

Validitas terkait kriteria berkaitan dengan bukti validitas yang diperoleh dengan membandingkan tes skor dengan kinerja pada ukuran kriteria. Kriteria dapat mencakup skor pada as- penilaian atau mungkin beberapa kriteria eksternal, seperti nilai, diagnosis, atau kepuasan kerja. Ada dua jenis validitas yang terkait dengan kriteria, yang dibedakan menurut waktu saat kriteria diukur.

Validitas serentak mengacu pada kapan skor tes dan kinerja kriteria skor dikumpulkan pada waktu yang sama. Koefisien korelasi dihitung antara skor tes dan skor pada variabel kriteria. Misalnya, uji ap-titude mungkin diberikan kepada sekelompok ahli mesin yang bekerja, dan kemudian peringkat yang mereka berikan receive oleh supervisor mereka mungkin diperiksa untuk menentukan apakah mekanik ap-skor titude terkait dengan pekerjaan mereka saat ini. Seringkali ukuran validitas konkuren adalah memperoleh karena tes akan digunakan di masa depan untuk memprediksi beberapa jenis perilaku seperti kemampuan melakukan pekerjaan masinis.

Jenis kedua dari validitas terkait kriteria adalah validitas prediktif. Dalam kasus ini, klien kinerja atau ukuran kriteria diperoleh beberapa saat setelah skor tes. Untuk seorang skolastik tes

bakat yang dirancang untuk memprediksi nilai perguruan tinggi, nilai yang diperoleh siswa di perguruan tinggi diperiksa untuk menentukan apakah tes bakat skolastik yang diberikan di sekolah menengah memiliki validitas prediksi.

c. Bangun Validitas

Jenis lain dari bukti validitas menanyakan pertanyaan, Apakah hasil tes berhubungan dengan variabel bahwa mereka seharusnya terkait dan tidak terkait dengan variabel yang seharusnya tidak mereka miliki? Untuk Misalnya, apakah hasil tes berubah sesuai dengan apa yang diketahui tentang perkembangan berubah? Apakah siswa yang lebih tua mengerjakan tes lebih baik daripada siswa yang lebih muda; misalnya, lakukan siswa kelas enam melakukan tes aritmatika lebih baik daripada siswa kelas tiga? Bukti yang berkaitan dengan landasan teoritis dari suatu tes kadang-kadang disebut sebagai konstruk

keabsahan-jenis validitas yang telah digunakan sebagai sarana untuk menjelaskan makna psikologis-dig variabel ("konstruksi") yang diukur dengan tes. Intinya, validitas konstruk adalah sinonim dengan istilah validitas itu sendiri, yang berfokus pada luas dan sifat bukti menggunakan untuk mendukung semua interpretasi tes (AERA, APA, & NCME, 1999). Seperti yang disebutkan sebelumnya, konten validitas dan bukti validitas terkait kriteria pada akhirnya menginformasikan validitas konstruk. Pola hubungan dengan variabel lain menghasilkan bukti validitas yang dikenal sebagai konver-validitas gent dan validitas diskriminan.

Di satu sisi, nilai tes harus diharapkan untuk menunjukkan korelasi substansial dengan tes dan penilaian lain yang mengukur karakteristik serupa (validitas konvergen). Ukuran bakat

matematika harus dikaitkan dengan nilai dalam studi matematika. Di sisi lain, nilai tes tidak boleh secara substansial berkorelasi dengan tes lain yang seharusnya membedakannya; artinya, mereka harus menunjukkan validitas diskriminan. Tes kemampuan matematika mungkin seharusnya tidak menunjukkan korelasi yang kuat dengan tes kecepatan dan akurasi klerikal. Ukuran sosialitas harus berhubungan negatif dengan skor pada skala skizofrenia dan berhubungan positif dengan skor pada skala ekstraversi. Kebanyakan studi validasi melaporkan validitas konvergen.

Jika suatu instrumen terkait dengan teori psikologis tertentu, maka hasilnya harus sesuai dengan teori itu. Analisis faktor dapat menentukan apakah item-item tes termasuk dalam faktor-faktor yang berbeda seperti yang disarankan oleh teori. Jika tes dibangun di sepanjang garis teori Jung, seperti Indikator Tipe Myers-Briggs, faktor-faktor yang dihasilkan dari analisis faktor harus terkait dengan konsep Jung seperti introversi versus ekstraversi, penginderaan versus intuisi, dan pemikiran versus perasaan.

d. Validitas Pengobatan

Jenis validitas lain yang penting bagi konselor dan dokter telah disebut pengobatan-validitas ment: Apakah hasil yang diperoleh dari tes membuat perbedaan dalam pengobatan? (Holland, 1997). Jika hasil tes bermanfaat, jika mereka membuat perbedaan dalam konseling proses, maka tes tersebut dapat dikatakan memiliki validitas pengobatan. Misalnya, Finn dan Ton-sager (1992) menemukan bahwa klien yang ditafsirkan skor MMPI-2-nya peningkatan yang signifikan pada beberapa variabel pengobatan. Dengan cara yang sama, Randahl, Hansen, dan Haverkamp (1993) menemukan bahwa klien yang memiliki Inventaris Minat Kuat profil yang ditafsirkan kepada

mereka membuat kemajuan yang signifikan dalam perencanaan karir mereka.

e. Catatan tambahan: Skala Validitas

Sementara konselor berusaha untuk memastikan bahwa penilaian dirancang dengan baik untuk merepresentasikan konten penting, kriteria, dan konsekuensi yang sesuai untuk klien, juga penting untuk dilakukan menentukan keakuratan tanggapan klien. Artinya, pola dalam respons yang tidak relevan yang tidak sesuai dengan maksud sebenarnya dari penilaian (rangkaian tanggapan) mungkin muncul.

f. Skala validitas

Merupakan alat yang digunakan untuk menentukan tiga jenis distorsi respons: klien yang berpura-pura memilikinya beberapa masalah atau gangguan (berpura-pura buruk), klien merespons dengan cara yang diinginkan secara sosial tampak lebih menguntungkan atau kurang bergejala (berpura-pura baik), dan klien menanggapi secara sengaja atau tidak sengaja (Van Brunt, 2009).

### E. Pengembangan Asesmen

Untuk mengetahui pemahaman dan kebutuhan perkembangan siswa maka dilakukan dengan menyusun asesmen. Bagi pelayanan bimbingan dan konseling, asesmen merupakan hal yang penting. Asesmen dalam bimbingan dan konseling terbagi menjadi asesmen tes dan non tes. Asesmen non-tes lebih sederhana karena tidak memerlukan penyusunan norma yang ketat sebagaimana asesmen tes contohnya observasi, wawancara, angket dan skala. Asesmen tes dalam bimbingan dan konseling umumnya menggunakan alat-alat tes psikologi yang beragam, seperti tes inteligensi, tes, bakat, tes minat, dan tes kepribadian yang penggunaannya tentu disesuaikan dengan pertimbangan konselor dan kebutuhan klien.

Sebagaimana situasi pandemi seperti saat sekarang ini, Menurut (Aji et al. 2020) perlu

adanya “*Pengembangan Asesmen Berbasis Teknologi untuk Keberlangsungan BK ditengah Pandemi Covid-19*”. Karena seorang konselor perlu berinovasi dalam memberikan asesmen kepada peserta didik, dimana salah satunya dengan memanfaatkan online form application di dalamnya memuat asesmen dengan memanfaatkan online form application. Beliau memfokuskan pada google form dan jotform sebagai fasilitas online form yang mumpuni dalam mendukung pelaksanaan asesmen bimbingan dan konseling. Aplikasi ini sangat sesuai digunakan untuk dosen, mahasiswa, guru, professional dan pegawai kantor yang sering membuat kuisioner dan survey online. platform tersebut bisa digunakan sebagai referensi media dalam memberikan asesmen BK kepada peserta didik dengan pemanfaatan teknologi. Untuk itu formulir daring jotform dan google form bisa menjadi solusi bagi guru BK dalam melakukan asesmen kepada peserta didik. Karena selain mudah digunakan banyak kelebihan-kelebihan yang ditawarkan salah satunya adalah data asesmen cepat diidentifikasi.

### PENUTUP

Pengukuran dalam bimbingan dan konseling adalah kegiatan mengumpulkan data yang berkenaan dengan pelayanan bimbingan dan konseling kepada sasaran-sasaran tertentu di mana data tersebut berbentuk kuantitatif/angka. Sebagai contoh ketika Guru Pembimbing memberikan tes psikologis kepada siswa kemudian diperoleh angka-angka berkenaan dengan tes yang telah dilakukannya. Ada empat skala pengukuran, dengan skala yang lebih kompleks (yaitu yang melakukan lebih banyak fungsi) memungkinkan pengukuran yang lebih tepat: nominal, ordinal, interval, dan rasio. Validitas merupakan sebuah bentuk dari pengacuan kepada sejauh apa sebuah tes dari melakukan pengukuran dan juga apa yang dimana akan dilakukan diterima untuk melakukan pengukuran. Reliabilitas merupakan sebuah pengacuan kepada sebuah ke konsistensian dari sebuah hasil tes.

### REFERENSI

- Abdurrahman, and Ni Suarti, Alit, S. (2016). “Pengembangan Instrumen Sikap Kreatif Mahasiswa Program Studi Studi Bimbingan Dan Konseling.” 1(April):1–14.
- Adom, Dickson, Jephtar Adu Mensah, and Dennis

- Atsu Dake. (2020). "Test, Measurement, and Evaluation: Understanding and Use of the Concepts in Education." *International Journal of Evaluation and Research in Education*. doi: 10.11591/ijere.v9i1.20457.
- Afdal, Afdal. (2019). "Kompetensi Bimbingan Karir Kolaboratif Konselor SMA Negeri Di Kota Payakumbuh Sumatera Barat." *JPGI (Jurnal Penelitian Guru Indonesia)*. doi: 10.29210/0288jpgi0005.
- Aji, Bayu Selo, Emilia Nurpitasari, Nuri Cholidah Hanum, Ahmat Ario Akbar, and Caraka Putra Bhakti. (2020). "Pengembangan Asesmen Berbasis Teknologi Untuk Keberlangsungan BK Ditengah Pandemi Covid-19." *Seminar Nasional Daring IIBKIN 2020 "Penggunaan Asesmen Dan Tes Psikologi Dalam Bimbingan Dan Konseling Di Era Adaptasi Kebiasaan Baru"* 98–103.
- Amalianingsih, Restu & Herdi. (2021). "Studi Literatur: Faktor Pendukung Dan Penghambat Dalam Penyelenggaraan Program Bimbingan Dan Konseling Di Sekolah Menengah Kejuruan." *Jurnal Bimbingan Dan Konseling Terapan*.
- Arifin, Zaenal. (2017). "Kriteria Instrumen Dalam Suatu Penelitian." *Jurnal Theorems (the Original Research of Mathematics)*.
- Arikunto, S., & Jabar, C. S. A. (2010). *Evaluasi Program Pendidikan*. Jakarta: Bumi Aksara.
- Asmita, Wenda, and Wahidah Fitriani. (2022). "Analisis Konsep Dasar Assesmen Bimbingan Dan Konseling Dalam Konteks Pendidikan." 8:129–34.
- Hays, Danica G. (2013). *Assesment in Counseling; Procedures and Proctices*. Virginia, United States: American Counseling Association.
- Junaidi. 2015. "Memahami Skala-Skala Pengukuran." *Research Gate*.
- Lupyanto, Limpid Sestu, and Yari Dwikurnaningsih. (2014). "Pengembangan Pengukuran Kompetensi Kepribadian Berbantuan Komputer Untuk Mahasiswa Bimbingan Dan Konseling." *Satya Widya* 30(2):71. doi: 10.24246/j.sw.2014.v30.i2.p71-81.
- Made, Ni, Dwi Narita, Ni Komang, Sri Yuliastini, Dwi Sri Rahayu, Ni Ketut, Kusuma Umbarini, and Leslie J. Briggs. (2022). "Pemanfaatan Jenis - Jenis Media Bk Di Sekolah Pada Pembelajaran Daring." 23(1). doi: 10.5281/zenodo.6390878.
- Otaya, Lian G. (2015). "Skala Pengukuran Dalam Penelitian." *Jurnal TADBIR*.
- Putri, Arum Ekasari. 2019. *Evaluasi Program Bimbingan dan Konseling: Sebuah Studi Pustaka*. Vol. 4. Bandung: Alfabeta.
- Putri, Vany Dwi. (2020). "Jurnal Konseling Dan Pendidikan Coution Journal Layanan Dasar Bimbingan Dan Konseling Untuk Peningkatan Survival And Safety Skills Siswa." *Jurnal Coution : Konseling Dan Pendidikan*.
- Sari, Syska Purnama, and Yusi Riksa Yustiana. (2022). "Cognitive Behavioral Untuk Mengembangkan Resiliensi Mahasiswa." *Jurnal Mahasiswa BK An-Nur : Berbeda, Bermakna, Mulia* 8:113–20.
- Setyaputri, Y, Nora, Y; Kriphianti, Dwi, and Rosalia Nawantara, Dewi. (2020). "Skala Pengukuran Karakter Adil:Salah Satu Instrumen Sebagai Piranti BADRANAYA (Board-Game Karakter Konselor Multibudaya)." 10.36728/ganesha.v1i02.1483.
- Simmamora. (2012). "Uji Validitas Dan Reabilitas Metode Penlitan." *Mercubuana*.
- Supriyanto, Agus, and Hardi Prasetiawan. (2021). "Pendampingan Konselor Untuk Pencegahan Burn Out Belajar Melalui Konseling Seni Kreatif Pada Kondisi Pandemi Covid-19." *Ganesha: Jurnal Pengabdian Masyarakat*. doi: 10.36728/ganesha.v1i02.1483.
- Wahidah, Nurul, Cucu Cuntini, and Siti Fatimah. (2019). "Peran Dan Aplikasi Assessment Dalam Bimbingan Dan Konseling." *Fokus: Kajian Bimbingan & Konseling Dalam Pendidikan* 2(2):25–36.

Dipublikasikan Oleh :

UPT Publikasi dan Pengelolaan Jurnal

Universitas Islam Kalimantan Muhammad Arsyad Al-Banjari Banjarmasin