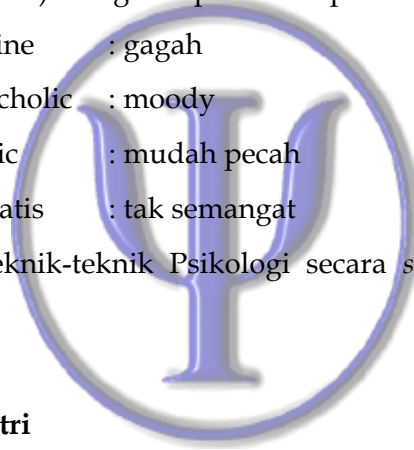

SEJARAH DAN OBJEK PSIKOMETRI

Teknik yang digunakan pertama kali dalam bidang psikologi adalah observasi dan wawancara.

1. Hipocrates (460 SM) mengelompokkan kepribadian menjadi 3 :
 - mania
 - melancholia
 - phrenelis
2. Galenus (200 SM) mengelompokkan kepribadian menjadi 4:
 - sanguine : gagah
 - melancholic : moody
 - choleric : mudah pecah
 - plegmatis : tak semangat

Pemakaian teknik-teknik Psikologi secara *scientific* baru dimulai pada tahun 1800.



A. Perintis Psikometri

1. Francis Galton (1822-1911)
 - ♦ mendirikan laboratorium antropometri dan pengukuran terhadap ketrampilan sensori-motorik
 - ♦ galton bar → alat untuk membedakan jarak visual
2. Alfred Binet (1857-1911)
 - ♦ membuat tes inteligensi yang terstandarisasi yang nantinya akan menjadi modal dan model pola pengukuran psikologi selanjutnya
 - ♦ ada 2 sub skala :
 - ♦ performance : terdiri dari 5 sub skill
 - ♦ verbal : terdiri dari 6 sub skill
3. Karl Person (Bapak Statistika Psikologi)
 - ♦ menguji hasil-hasil pengukuran
 - ♦ merumuskan product moment : menguji hubungan

- ♦ chi-square : untuk menguji normalitas sebaran data, yang digunakan adalah frekuensinya.

Populasi sebaran data jika kita ambil secara random, diasumsikan bahwa sampelnya normal. Jika non random maka akan diasumsikan tidak normal sehingga perlu diuji kenormalannya dengan menggunakan chi-square.

Yang diharapkan : $P > 0,05 \rightarrow$ tidak signifikan, karena ada aturan :

$P < 0,01 \rightarrow$ SS, H0 ditolak dan H1 diterima

$P < 0,05 \rightarrow$ S, H0 ditolak dan H1 diterima

$P > 0,05 \rightarrow$ TS, H0 diterima = tidak ada perbedaan, H1 ditolak ; artinya sebaran data pada sampel bisa digunakan pada populasi

Rumus :

$$r_{xy} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left\{ \sum X^2 - \frac{(\sum X)^2}{N} \right\} \left\{ \sum Y^2 - \frac{(\sum Y)^2}{N} \right\}}}$$

4. Edward Lee Thorndike (1874-1949)

- ♦ trial and error learning
- ♦ hand writing scale
- ♦ penulis buku tentang pengukuran yang pertama
- ♦ pernyataan Thorndike (prinsip dasar penting bagi psikometri) :
"Jika sesuatu itu ada, sesuatu tersebut pasti ada dalam suatu besaran. Jika sesuatu itu ada dalam suatu besaran, sesuatu tersebut pasti dapat diukur/diketahui besarannya"

5. Charles Spearman

- ♦ tokoh pertama yang mendefinisikan reliabilitas (Spearman-Brown)
- ♦ koefisien korelasi
- ♦ reliabilitas, dicari dengan cara : belah dua, tunggal (tes-retes) dan paralel form

B. Kajian Psikometri

Objek material : perilaku manusia

Objek formal : pengukuran secara ilmiah untuk perilaku manusia

Psikometri mempelajari model tes dan mengembangkan teori pengukuran psikologis dengan menggunakan dasar pengukuran. Dalam pengukuran, atribut pengukuran harus jelas.



PENGUKURAN

Pengukuran adalah :

- aturan pemakaian bilangan terhadap objek pengukuran yang mempresentasikan kuantitas.
- perbandingan alat ukur dengan apa yang diukur

A. Jenis Pengukuran

Pengukuran dibagi menjadi 2 bagian :

1. Pengukuran terstandar

- ♦ objektif : siapa, kapan → menunjukkan hal yang relatif sama
- ♦ kuantitatif : akan ditunjukkan dalam angka sehingga dapat dibandingkan antara yang satu dengan yang lain
- ♦ komunikasi : memudahkan untuk mengkomunikasikan hal yang didapat
- ♦ ekonomis : menghemat biaya, waktu dan tenaga

2. Pengukuran tidak terstandar

- ♦ belum baku akan tetapi akurasinya hampir tepat jika disertai dengan pengalaman

Pengukuran dilakukan pada atribut tertentu yang spesifik dan bukan secara keseluruhan, misalnya meja, yang akan diukur adalah panjang, lebar, tinggi, kepadatan kayu.

Oleh karena itu dalam pengukuran harus dipenuhi :

1. Atribut pemakaian bilangan harus eksplisit

- a. bilangan itu menunjukkan kuantitas tertentu
- b. fungsi : agar pengukuran itu terstandarisasikan

- Progressive Matrix

Misal : ada 10 soal, no. 1-5 dijawab, no. 6 tidak, no. 7-8 dijawab :
maka yang dinilai hanya no. 1-5 saja dan jawaban selanjutnya diabaikan

-
-
- Lari : seseorang yang maksimal lari 1 km, maka apabila suatu saat bisa 1,5 km itu hanya kebetulan saja
 2. Objeknya pada atribut /sifat tertentu, bahkan bisa lebih spesifik
ex : - tes Weschler terdiri atas 11 atribut
 - mengukur meja → bisa diukur dimensinya, atribut tingginya, atribut lebarnya, atribut kekuatannya.
 3. Atribut yang diukur harus dipertimbangkan masak-masak
 4. Proses pengukuran adalah proses abstrak
Untuk memahami proses kuantifikasi tersebut, maka perlu diketahui tingkat hasil pengukuran.

B. Tingkat Hasil Pengukuran

Tingkat hasil pengukuran dibagi menjadi 4 macam, yaitu :

1. Nominal

- ♦ untuk identifikasi dan pemberian label
- ♦ besaran angka tidak mempunyai kuantitas tetapi hanya sebagai cap saja
- ♦ tidak dapat dikenai operasi hitung
- ♦ untuk untuk pencacahan, pemilahan/diskrit

misalkan:

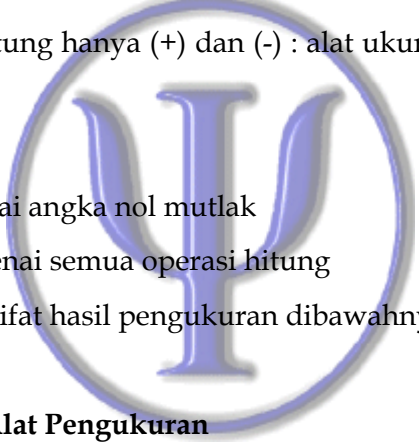
status (menikah, belum menikah)

tempat lahir (propinsi)

jenis kelamin (laki-laki = 1 dan perempuan = 0, angka tersebut bisa diganti dengan laki-laki = 0 dan perempuan = 1 → transformasi isomorfik)

2. Ordinal

- ♦ untuk perjenjangan/*ranking* dan pengurutan/*ordering*
- ♦ angka yang berbeda menunjukkan jenjang dan urutan yang berbeda
- ♦ dua angka yang berurutan tidak diketahui perbedaannya
- ♦ tidak mempunyai angka nol mutlak
- ♦ operasi hitung hanya < atau > dan =

-
- ♦ jika kita membuat urutan dari angka yang berbeda, kita harus mengurutkan dari hal yang sama
 - ♦ angka boleh berubah akan tetapi urutan harus tetap, inilah yang dinamakan transformasi monotonik
 - ♦ misal : tinggi badan, IQ, pemberian nomor pada juara tinju amatir
3. Interval
- ♦ mempunyai jarak antar jenjang yang selalu tetap (selalu sama)
 - ♦ tidak mempunyai angka nol mutlak
 - ♦ perbedaan angka mempunyai arti perbedaan kuantitatif dan kualitatif
 - ♦ bersifat invarian bila dikenai transformasi linier
 - ♦ operasi hitung hanya (+) dan (-) : alat ukur psikologis dengan analisis statistika
4. Ratio
- ♦ mempunyai angka nol mutlak
 - ♦ dapat dikenai semua operasi hitung
 - ♦ memiliki sifat hasil pengukuran dibawahnya
- 

C. Macam-Macam Alat Pengukuran

1. Skala

- ♦ Suatu seri progresif nilai atau besaran sesuai dengan gejala yang dapat dikuantifikasikan
- ♦ Seri progresif bergerak dari yang rendah ke yang tinggi atau sebaliknya, dari yang tinggi ke yang rendah
Sangat sering – tidak pernah
Sangat setuju – sangat tidak setuju
Hampir selalu – tidak pernah
Sesuai – sangat tidak sesuai
- ♦ Ordinalnya : 3, 5, 7, 11 dan yang paling sering biasanya 7

2. Inventory

- ♦ Adalah suatu instrumen untuk mengungkap ada tidaknya suatu perilaku tertentu, interest, sikap dsb

- ♦ Biasanya berbentuk daftar pertanyaan yang harus dijawab
- ♦ Pertanyaan-pertanyaan ini akan memunculkan sesuatu yang sudah ada dalam diri subjek, keinginan-keinginan subjek akan terungkap
Misal : - aku berharap aku bisa mengubah ...
- yang terbaik pilihan sekolah adalah ...
- aku betul-betul tidak suka pada ... dll

3. Rating

- ♦ Adalah pemberian skor/ranking pada individu atau gejala/data tertentu
- ♦ Rating biasanya dikaitkan dengan skala
- ♦ Yang penting pada rating adalah adanya rater yang bertugas untuk memberi skor atau ranking pada gejala tertentu/subjek (biasanya 3)
- ♦ Biasanya dibuat dalam bentuk check list

Misalkan.

aktivitas	t	s	r
.....		✓	

4. Questionnaire

- ♦ Adalah suatu set pertanyaan-pertanyaan mengenai suatu topik tertentu, yang harus dijawab oleh subjek
- ♦ Untuk mengukur interest, personality problem dan opinions, dan merekam info biografis

Misal kita menyuruh subjek untuk memilih (berilah tanda)

Jenis kelamin : laki-laki
perempuan

PENGUKURAN DALAM PSIKOLOGI

A. Pengertian Tes

Tes berasal dari bahasa latin, *testum*, alat untuk mengetahui kandungan-kandungan tanah. Dalam bahasa perancis, tes adalah alat untuk membedakan antara emas dan perak. Dalam Psikologi tes dapat berfungsi sebagai prosedur dan alat.

1. Menurut Anastasi, tes adalah *“essentially an objective and standardisasi measure of a sample behaviour”*.
2. Menurut L. S. Cronbach, tes adalah :
 - a. *Is a system procedure for comparing the behaviour of two or more person.*
 - b. *A syatematic procedure for observing a person behaviour and describing it with the aid of a numerical scale or a category system.*
3. Tes adalah sekumpulan pertanyaan dan/atau sekumpulan perintah yang harus dijawab dan/atau dilaksanakan yang akan memberi informasi aspek psikologis tertentu berdasarkan bagaimana subjek menjawab dan/atau melaksanakan perintah.

Tes disebut prosedur yang sistematis karena :

- ♦ item disusun berdasarkan cara dan aturan tertentu
- ♦ administrasi dan skoring harus jelas dan spesifik
- ♦ tiap individu yang terkena harus diperlakukan sama dan sebanding kondisinya

Tes disebut mengukur sampel perilaku karena :

- ♦ berapapun banyaknya aitem, tetap mengukur sampel perilaku. Anggapan semakin banyak semakin baik, salah jika hanya berupa duplikasi.
- ♦ kelayakan sesuatu tergantung pada apakah suatu aitem mewakili secara representatif kawasan yang hendak diukur

B. Jenis Tes

1. Tes performansi maksimal

- ♦ apa yang dapat dilakukan oleh subjek dan seberapa baik subjek melakukannya
- ♦ lebih pada kemampuan dan prestasi
- ♦ bisa ditingkatkan dengan latihan
- ♦ ada batasannya

Misalkan : UMPTN, TOELF, ulangan harian dll

2. Tes performansi tipikal

- ♦ kecenderungan yang dilakukan subjek dalam situasi tertentu atau bagaimana subjek bereaksi terhadap kondisi tertentu
- ♦ diungkap berkaitan dengan pengalaman hidupnya
- ♦ tidak ada benar salah
- ♦ keluar sesuai dengan tipikal seseorang
- ♦ stimulus kabur, ambigu, tidak jelas → merupakan proyeksi pengalaman kehidupan

Misalkan : tes kepribadian, minat, sikap dll

C. Karakteristik Pengukuran Psikologis

1. Indirect

- ♦ dilakukan secara tidak langsung
- ♦ harus diberikan stimulus agar menghasilkan respon yang merupakan representasi dari atribut yang diukur
- ♦ pada pengukuran atribut tertentu dilakukan dengan melihat keterkaitan fungsinya

2. Tidak ada satuan ukuran

- ♦ instrumen bound : hasil pengukuran harus dikaitkan dengan latnya
- ♦ person bound : hasil pengukuran berlaku hanya pada individu tertentu.
Misal EPPS, bersifat ipsaptive = komposisi need hanya dapat dibandingkan dalam diri 1 orang

-
-
3. Deskriptif
 - ♦ hasilnya disajikan dalam bentuk psikogram
 4. Selalu mengandung kemungkinan adanya “error”
 - ♦ selalu mengandung error tetapi bisa diminimalkan, karena hanya sampai bentuk data interval dan individu yang satu berbeda dengan individu yang lain.
 - ♦ harus diikutsertakan validitas dan reliabilitas
 5. Multi approach
 - ♦ tidak ada pendekatan tunggal karena atribut yang sama dapat diambil konsep yang berbeda

D. Langkah-Langkah Penyusunan Alat Ukur

1. Spesifikasi tujuan

- ♦ Menetapkan batasan atribut yang hendak diukur
Makin jelas batasannya, makin mudah spesifikasinya
- ♦ Menetapkan konsep yang digunakan
Atribut yang sama dapat berbeda konsep
Misal. Intelligensi → educational relationship
→ speed respon

Spesifikasi tujuan ke pembatasan sifat/traits atau penguraian isi dan kecakapan, memiliki 3 strategi :

- a. Jalur literatur
- b. Pendekatan para ahli, skala atau interviu
- c. Pendekatan awam (bottom-up) : emic approach, menyebarkan angket

2. Penguraian isi dan kecakapan

- ♦ Isi apa yang hendak diukur dan kecakapan apa yang hendak diungkap
- ♦ Kecakapan ada 3 macam, yaitu : kognitif, afektif dan psikomotor

Kognitif

- o Knowledge
- o Comprehension
- o Application

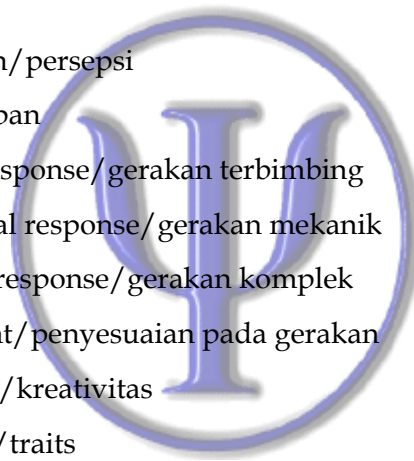
- o Analysis
- o Syntesis
- o Evaluation

Afektif

- o Receiving/penerimaan
- o Responding/partisipasi
- o Valuing/penentuan sikap
- o Organization
- o Characterization by a value/value complex/pembentukan pola hidup

Psikomotorik

- o Perception/persepsi
- o Set/kesiapan
- o Guided response/gerakan terbimbing
- o Mechanical response/gerakan mekanik
- o Complex response/gerakan kompleks
- o Adjustment/penyesuaian pada gerakan
- o Creativity/kreativitas



3. Pembatasan sifat/traits

4. Penulisan aitem

Cara penyusunan aitem :

a. Berseri

Unsur 1	Unsur 2	Unsur 3
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	11, 12, 13 ... 20	21, 22, 23 ... 30

Proses ini sesuai dengan penyusunan batu baterai □▪□▪□▪

b. Simultan

1) Biasa

❁❁❁❁❁❁❁❁❁	Aspek 1	Aspek 2
Bentuk		
Favourable	1, 5, 9	2, 6, 10
unfavourable	3, 7, 11	4, 6, 8

Konsep favorable dan unfavourable mendukung tidaknya item terhadap suatu konsep.

Misal : konsep tentang perjudian → harus diberantas, mengganggu masyarakat, memberi harapan palsu.

Aitem yang bisa kita beri :

- ♦ Tidak ada orang bahagia karena berjudi (favourable)

Sehingga jika ditanya :

- Sangat setuju
- Setuju
- Ragu-ragu
- Tidak setuju
- Sangat tidak setuju

Dilihat lagi definisi operasionalnya : semakin tinggi skor semakin menolak terhadap perjudian, maka kalau jawaban (a) maka skor tinggi.

- ♦ Daripada berdiam diri lebih baik berjudi (unfavourable)

Maka kalau jawaban (a) maka skor rendah.

2) Zigzag

	Aspek 1		Aspek 2		Aspek 3	
Favourable	1	7	5	11	3	9
Unfavourable	4	10	2	8	6	12

Nantinya akan dipasangkan secara zigzag sebagai berikut :

1-2, 7-8, 4-5, 10-11, 5-6, 11-12, 2-3, 8-9.

c. Acak atau random

Contoh :

Strategi konflik : solusi yang cenderung dipilih oleh seseorang ketika terjadi konflik. Ada 5 strategi konflik yang dimungkinkan muncul, yaitu :

1. *Withdrawing* : mundur, menunda dan mendiskusikan ketika kedua belah pihak ingin membicarakan.

Contoh aitem :

- ♦ Menjaga jarak dengan orang yang tidak sepaham
 - ♦ Menghindari pertengkaran dengan menghindarinya
 - ♦ Tidak ada yang penting dari pertengkaran
2. Smoothing : siapa yang lebih membutuhkan biarkan memiliki → mengalah.

Contoh aitem :

- ♦ Tutur kata halus menenangkan hati yang keras
 - ♦ Pembicaraan yang lembut membuka jalan yang mulus
 - ♦ Kalahkan lawanmu dengan kebaikan hati
3. Confronting : hadapi konflik, negosiasi, pikirkan, bicarakan solusinya.

Contoh aitem :

- ♦ Datanglah dan mari bicarakan bersama
 - ♦ Tidak seorompokun memiliki gambaran final, tapi setiap orang memiliki kontribusi pemikiran
4. Forcing : menekan, memakan, menguasai lawan konflik.

Contoh aitem :

- ♦ Kacah dikuasai oleh seseorang yang menang
 - ♦ Seseorang yang tidak gentar membuat orang gentar
 - ♦ Ada satu macam orang didunia, yaitu pemenang atau pecundang
5. Compromising : ambil kesempatan, bicarakan bersama, win-win solution.

Contoh aitem :

- ♦ Lebih baik mendapatkan setengah daripada tidak sama sekali
- ♦ Seseorang yang bersedekah, membina teman baik
- ♦ Kejujuran akan menghindarkan pertengkaran

Option jawaban strategi konflik yang berkaitan dengan :

- ✦ Kognitif → Sangat setuju
- Setuju
- Netral
- Tidak setuju
- Sangat tidak setuju

-
- ✦ Afektif → Sangat sesuai
Sesuai
Ragu-ragu
Tidak sesuai
Sangat tidak sesuai
 - ✦ Psikomotorik → Sangat sering
Sering
Jarang
Tidak pernah



VALIDITAS

Validitas adalah sejauh mana suatu alat ukur mampu mengukur apa yang ingin diukur. Terdapat tiga macam validitas, yaitu, validitas isi, validitas konstruk dan validitas berdasarkan kriteria.

A. Validitas isi

- ♦ diestimasi melalui pengujian terhadap isi tes dengan analisis rasional
- ♦ untuk menjawab : sejauh mana aitem dalam tes mencakup keseluruhan kawasan isi objek yang tidak diukur oleh tes yang bersangkutan
- ♦ lebih bersifat subjektif, tergantung penyusun dan penganalisanya
- ♦ validitas isi mendasari validitas selanjutnya karena sebelum menguji validitas konstruk dan validitas berdasarkan kriteria kita sudah menguji validitas isi terlebih dahulu

Dibagi menjadi 2 yaitu :

- a. Face validity : melihat secara sepintas mengenai isi tes
- b. Logical validity : apakah isi tes merupakan representasi dari aspek yang hendak diukur

Contoh :

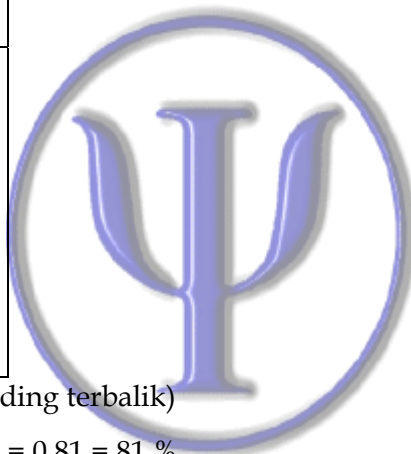
X	Y
4	4
3	3
4	4
2	2
3	3

$r_{xy} = 1,00$ maka korelasi $xy =$ semua

X	Y
4	3
3	3
4	3
2	3
3	3

$r_{xy} = 0$ (karena tidak ada hubungan diantara keduanya/y-nya selalu tetap tidak ada perubahan)

X	Y
4	2
5	1
3	3
2	4
4	2
1	5



$r_{xy} = -1,00$ (berbanding terbalik)

$r_{xy} = 0,9$ maka $r^2_{xy} = 0,81 = 81\%$

Besarnya varians X yang didapat dijelaskan dari variabel Y

81 % disebabkan oleh Y

19 % disebabkan oleh yang lain

Sehingga 19 % milik X dan 19 % milik Y isinya berbeda.

“Jadi semakin kecil korelasinya maka semakin kecil overlappingnya. Dan semakin besar prosentase serta makin besar korelasinya maka pasti akan semakin overlapping”.

Contoh lain :

Pola kepemimpinan tidak bisa diukur secara statistik, tapi hanya bisa secara validitas isi.

Pola kepemimpinan ada 3 : demokratis, laissez fair dan otoriter

Ketiganya hanya bisa dianalisis dengan validitas logic (content validity)

-
- a. jawaban mencerminkan demokratis
 - b. jawaban mencerminkan laissez faire
 - c. jawaban mencerminkan otoriter

Contoh aitem :

Ketika saya melihat buku saya yang hilang di bawa teman, maka :

- a. saya akan menanyakan dimana buku tersebut diperoleh
- b. saya biarkan saja, siapa tahu kesadaran untuk mengembalikannya muncul
- c. langsung saya minta untuk dikembalikan

B. Validitas konstruk

- ♦ menunjukkan sejauh mana suatu tes mengungkap suatu traits/sifat atau konstruk teoritis yang hendak diukur
- ♦ sangat berguna untuk mengukur trait yang tidak mempunyai kriteria
- ♦ dilihat konstruknya (aitem-aitemnya bagaimana, dll) → trus diuji

Multitrait method : mendasarkan pada teknik korelasi. Kita yakin apabila ada korelasi, berarti ada varians bersama antar 2 variabel. Bila tidak ada korelasi/korelasi rendah berarti antar 2 variabel adalah independent/tidak saling terkait

Bila kita uji dengan :

- a. Monotrait Multimethod → korelasi tinggi
- b. Monotrait Monomethod → korelasi lebih tinggi
- c. Multitrait Monomethod → korelasi rendah
- d. Multitrait Multimethod → korelasi lebih rendah/ada korelasi

Sehingga (a) > (b) dan (c) > (d)

	A ₁	B ₁	A ₂	B ₂
A ₁	1,00	0,34	0,87	0,26
B ₁		1,00	0,26	0,82
A ₂			1,00	0,31
B ₂				1,00

Dimana :

A_1 dan A_1 : monotrait monomethod

A_1 dan B_1 : multitrait monomethod

A_1 dan B_2 : multitrait multimethod

B_2 dan B_2 : monotrait monomethod

$A_1 A_2$: korelasi tinggi

: convergent validity

: korelasi terhadap trait yang sama meskipun metode berbeda

$A_1 B_1$: korelasi rendah

: discriminant validity

: membuktikan bahwa tidak ada korelasi dengan alat ukur yang mengukur trait berbeda

Misal. Jika kita mengkorelasikan antara kestabilan emosi dan aritmatik, ternyata korelasinya rendah jadi tidak ada overlap antara kestabilan emosi dan aritmatika.

$A_1 B_1$: discriminant validity $\rightarrow A_1 B_1 > A_1 B_2$

Jika kita melakukan $A_1 - A_2$ (traitnya sama)



(tunggal/criterion related validity)

Criterion related validity : jika kita hanya mengukur A_1 dan A_2 saja, tapi kalau kita juga mengukur $A_1 B_1$ maka $A_1 A_2$ disebut convergent validity, karena hasil pengukuran $A_1 A_2$ itu akan dibandingkan dengan hasil pengukuran $A_1 B_1$.

C. Validitas berdasarkan kriteria

a. Validitas prediktif

Prediktor masa mendatang

Misal :

- ♦ Tes Finger Dextivity menjadi prediktor terhadap kecepatan magnetik. Jika korelasi keduanya tinggi maka tes tersebut mempunyai nilai prediktor yang tinggi.

-
- ♦ UMPTN merupakan prediktor bagi penguasaan materi diperguruan tinggi. Jika korelasi antara keduanya tinggi maka UMPTN mempunyai nilai prediktor yang tinggi terhadap penguasaan materi perguruan tinggi.
 - ♦ TPA (Tes Potensial Akademik). Jika validitas prediktif TPA baik maka yang lulus adalah yang memiliki kemampuan akademik yang tinggi.

b. Validitas concurrent

Mengkorelasikan suatu alat ukur dengan alat ukur lain yang sudah terstandar. Jika korelasi tinggi maka suatu alat ukur itu dapat dikatakan mempunyai validitas konkuren.

I : alat ukur yang diuji validitasnya

II : alat ukur lain yang mengukur atribut yang sama

I dan II dikenakan pada sekelompok subjek lalu dikorelasikan. Jika korelasi tinggi berarti alat ukur I memiliki validitas konkurent.

Kelemahan :

- ♦ Jika ada alat ukur yang telah baku buat apa disusun alat ukur baru. Karena biasanya : lebih singkat, lebih mudah/ sederhana dan dipakai spesifik.
- ♦ Jika dikorelasikan dengan alat ukur baku sudah mengandung error sehingga korelasi antara alat ukur baru dengan alat ukur baku menjadi double error.

RELIABILITAS

Reliabilitas bisa diartikan sebagai tingkat keajegan, dijabarkan berdasarkan *classical true score theory* dan memakai estimasi reliabilitas.

A. Classical True Score Theory

$$X = T + E \text{ (classical true score theory)}$$

X : obtained score (skor perolehan)

T : true score (skor sesungguhnya)

E : error

Misal :

Tutur kata yang halus menenangkan hati yang keras → SS, S, N, TS, STS

Dalam pilihan jawaban dari skala pengukuran tersebut diberi skor : 1, 2, 3, 4, 5 (harga yang diberikan pada jawaban responden yang merupakan representasi atribut laten "X" observed score/obtained score = skor perolehan).

Jawaban yang diberikan oleh subjek itu akan berbeda intensitasnya sehingga ada errornya (error tersebut dapat saja terjadi walaupun skornya sama, karena beberapa true score dan errornya).

Misalkan A dan B → skornya sama-sama 4 (tidak setuju)

$$X = T + E$$

$$A : 4 = 3,8 + 0,2$$

$$B : 4 = 3,1 + 0,9$$

Disini skor keduanya sama akan tetapi intensitasnya berbeda. Yang A lebih ke Tidak Setuju sedangkan B lebih ke Netral, tapi tetap saja Tidak Setuju.

B. Estimasi Reliabilitas Mengendalikan Error

Error positif : lebih tinggi dari true score (X)

Error negatif : lebih rendah dari true score (X)

$$9 = 7 + E \rightarrow E = 9 - 7 = 2 (+)$$

$$6 = 7 + E \rightarrow E = 6 - 7 = -1 (-)$$

Error bisa dibagi menjadi 2 macam, yaitu :

1. Error acak, karena :
 - ♦ Kondisi individu testee
 - ♦ Kondisi individu tester
 - ♦ Kesalahan sampling
 - ♦ *Hallo efek* (efek penyerta, error disebabkan karena penampilan, pribadi testee dll)
2. Error systematik, karena :
 - ♦ Tidak dapat diestimasi, dimana alat ukur menjadi under/over estimate
 - ♦ Terkena pada semua subjek

C. Asumsi-Asumsi Hubungan Antara Error Pengukuran Dan Skor Murni

1. $\epsilon(X) = T$, dimana T adalah nilai expected value X

tes 1	$\rightarrow I$	$\rightarrow X_1$		$X_1 - T = E_1$
2	$\rightarrow n$	$\rightarrow X_2$		$X_2 - T = E_2$
3	$\rightarrow d$	$\rightarrow X_3$	rata-rata $X = T$	$X_3 - T = E_3$
4	$\rightarrow i$	$\rightarrow X_4$		$X_4 - T = E_4$
5	$\rightarrow v$	$\rightarrow X_5$		$X_5 - T = E_5$
n	$\rightarrow idu$	$\rightarrow X_n$		$X_n - T = E_n$
	\downarrow			jadi rata-rata $E = 0$

masing-masing independent

$$\text{rata-rata } X = T$$

$$X_1 - T = E_1$$

Kurang lebihnya error itu (rata-rata error itu = 0). Error itu ada yang kurang tapi ada juga yang lebih. Bila error itu dijumlahkan trus dirata-rata = 0.

Karena kira-kira rata-rata error itu = 0

maka data itu = true score

2. $\rho_{et} = 0$

E dan T tidak saling berkorelasi

E bisa (+) dan juga (-)

Korelasi antara true score dan error = 0

3. $\rho_{e1e2} = 0$

Error pada pengukuran pertama tidak berhubungan dengan error pada pengukuran kedua sehingga bila kedua error dikorelasikan = 0.

(error pada suatu tes tidak bergantung pada error yang lainnya)

4. $\rho_{e1t2} = 0$

Error kedua tidak terkait dengan pengukuran yang pertama (saling independent). Error suatu tes tidak tergantung pada true score tes yang lain.

Semakin besar varians semakin besar pula kaki kurva itu.

✧ Error Acak ✧

Kesalahan sampling

Contoh : berat badan

Populasi	Sampel
A = 46	B & F → rata2 : 2,58 → E1 = 6,6
B = 40	A & F → rata2 : 48 → E2 = 1,16
C = 64	D & F → rata2 : 50 → E3 = -0,84
D = 50	A & B → rata2 : 43 → E4 = 6,16
E = 45	A & C → rata2 : 55 → E5 = -5,84
F = 50	A & D → rata2 : 48 → E6 = 1,16
	A & E → rata2 : 45,5 → E7 = 3,66
	B & C → rata2 : 52 → E8 = -2,84
	B & D → rata2 : 45 → E9 = 4,16
	B & F → rata2 : 45 → E10 = 4,16
	C & D → rata2 : 57 → E11 = -7,84
	C & E → rata2 : 54,5 → E12 = -5,34
	C & F → rata2 : 57 → E13 = -7,84
	D & E → rata2 : 47,5 → E14 = 1,66
	E & F → rata2 : 47,5 → E15 = 1,66
$\mu = 49,16$	Rata-rata E = 0 (seharusnya!)

Salah satu penyebab error adalah faking.

Faking adalah jawaban yang tidak jujur untuk menutupi keadaan yang sesungguhnya (baik faking good atau faking bad).

Faking dilakukan karena :

1. untuk menghindari ancaman
 - berkaitan dengan pekerjaan
misal. ketika diminta pengkritikan tentang atasan
 - berkaitan dengan nama baik
misal. soal suap menyuap
2. ingin tampil lebih baik
misal. para pelamar pekerjaan
3. ingin menutupi kekurangannya
4. ingin sama dengan norma masyarakat/ social desirability

Untuk menghindari faking/mengatasi faking :

1. teknik meloncat
misal. Apakah anda melakukan masturbasi ? (ya/tidak)
biasanya akan menjawab tidak, maka pertanyaan akan diubah menjadi :
Berapa kali dalam seminggu anda melakukan masturbasi?
2. mengetahui nama/tujuan alat ukur
 - tujuan : demi perkembangan ilmu pengetahuan
 - "tes ketaatan" beribadah diberi nama "tes kejujuran" dengan harapan agar menjawab dengan jujur
3. tidak diberi waktu mengerjakan yang lama
karena jawaban yang spontan biasanya adalah jawaban yang lebih jujur dari pada yang dipikir-pikir lebih dahulu
4. menggunakan forced choice
memberikan jawaban yang sudah ada jawabannya, dan disusun memilih diantara jawaban itu

D. INTEPRETASI TERHADAP $X = T + E$

Se^2 semakin besar maka tes semakin reliable

(Se^2 : varian : yang menentukan besar tidaknya kaki sebuah kurva)

1. $\rho_{xx'}$

Reliabilitas adalah korelasi antara 2 obtained score dari 2 tes yang paralel.

Tes paralel = 2 true score akan sama = $\sigma_{e1} = \sigma_{e2}$

2. $\rho_{xx'^2}$

Besarnya proporsi varians X yang dijelaskan oleh hubungan liniernya dengan X' .

3. $\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2}$

Reliabilitas adalah perbandingan antara varians skor murni dengan skor perolehan.

Jika $T = X$ maka $\sigma_X^2 = \sigma_T^2 \rightarrow r_{xx'} = 1$

Semakin X mendekati T maka semakin tinggi koefisien reliabilitasnya.

Semakin jauh X dari T maka semakin rendah koefisien reliabilitas, semakin besar error.

$\sigma^2 = S^2 = V = \text{Varian} = SD^2$

4. $\rho_{xx'} = \rho_{xT^2}$

Reliabilitas merupakan koefisien determinasi korelasi antara X dan T.

5. $\rho_{xx'} = 1 - \rho_{xE^2}$

Semakin besar errornya maka semakin kecil reliabilitasnya. Jika tidak ada error, maka reliabilitasnya sempurna.

6. $\rho_{xx'} = 1 - (\frac{\sigma_e^2}{\sigma_x^2})$

Reliabilitas sangat tergantung pada varians error maupun varians obtained score. Sebenarnya yang kita ketahui adalah σ_x^2 (obtained score). Semakin besar varians obtained score maka semakin besar reliabilitasnya. Jika kita mempunyai alat ukur yang heterogen maka semakin reliabilitas (semakin banyak subjek semakin reliabel).

$$\frac{\sigma_e^2}{\sigma_x^2} = 1 - \rho_{xx'}$$

$$\sigma_e^2 = \sigma_x^2 (1 - \rho_{xx'})$$

$$\sigma_e = \sigma_X \sqrt{1 - \rho_{XX'}}$$

7. $r_{XX'} = \text{antara } 0 - 1,00 \quad 0 \leq r_{XX'} < 1$

maka : $X = T + E$

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

$$\rho_{XT} = \sqrt{\rho_{XX'}^2}$$

Reliabilitas semuanya menggunakan teknik korelasi.

E. Metode Estimasi Reliabilitas

1. Metode tes ulang (tes-retes)

Percobaan I : Percobaan II

Subjek	Skor	Subjek	Skor
A		a	
b		b	
....		
K		k	

Antara percobaan I dan percobaan II ada tenggang waktu (tenggang waktu pada apa yang ingin diukur). Lalu akan dikorelasikan dan diperoleh hasil koefisien reliabilitas.

Contoh :

I		II		N = 5				
S	X	S	X	X	X'	X ²	X' ²	XX'
A	9	a	8	9	8	81	64	72
b	8	b	9	8	9	64	81	72
c	8	c	8	8	8	64	64	64
d	9	d	8	9	8	81	64	72
e	6	e	6	6	6	36	36	36

$$\Sigma = 40 \quad 36$$

Dengan kalkulator :

```

INV      AC
9        XdYd 8      RUN
8        » 9      data
  
```

8	»	8	»
9	»	8	»
6	»	6	»

hasilnya :

kout 3 → N	kout 5 → ΣY
kout 2 → ΣX	kout 4 → ΣY ²
kout 3 → ΣX ²	kout 6 → ΣXY
INV 9	→ r = 0,745

Ada kemungkinan pengaruh :

- a. Carry over effect
 - efek-efek yang terbawa dari pengukuran pertama menuju pengukuran kedua
 - dimana subjek akan ingat pengukuran pertama dan sekedar mengulang jawaban kedua tanpa memikir lagi
 - didapatkan koefisien reliabilitas tinggi, tapi koefisien reliabilitas itu semu
- b. Practise effect
 - terutama pada pengukuran kognisi, performansi maksimal, yang semakin lama semakin tinggi
 - terdapat suatu titik optimal
 - ex : tes TOEFL ; kelemahan bangsa Indonesia = listening, PPA
 - menggunakan pengalaman pertama untuk tes 2
- c. Gangguan suasana hati
 - subjek akan merasa bahwa pengukuran yang pertama sudah pernah dilakukan maka pada pengukuran yang kedua akan menjawab yang seenaknya sehingga akan terjadi under estimated

Kelemahan metode tes-retes :

- ♦ boros, tidak efisien, tidak hemat biaya
- ♦ butuh tenaga yang banyak
- ♦ butuh waktu yang lama

- ♦ mengandung kemungkinan error
2. Metode paralel (paralel form)
- ♦ Bisa untuk menghemat waktu karena menggunakan 2 alat tes paralel digunakan secara bersama-sama lalu dikorelasikan tapi belum efektif (boros kertas, tenaga, biaya)
 - ♦ Bisa menghemat waktu

Kelemahan :

- ♦ harus membuat 2 alat ukur paralel, padahal 1 saja sudah sulit
- ♦ belum bebas dari : carry over effect, practise effect dan gangguan suasana hati

alat ukur yg diuji alat ukur alternatif/ paralel

Subjek	Skor	Subjek	Skor
A		a	
b		b	
....		
K		k	

Dikorelasikan

Paralel/alternatif adalah untuk mengukur atribut yang sama, mempunyai varians yang sama tetapi item beda.

(-) : susah dilakukan/ digunakan (mulai ditinggalkan)

(+) : > singkat, > mudah, > praktis

3. Metode internal consistency

Hanya memerlukan 1 kali pengujian (single trial administration) sehingga pengaruh-pengaruh dapat dihindarkan. Prinsipnya seperti tes paralel, dimana 1 alat ukur dibelah menjadi 2 dan dianggap sebagai alat ukur paralel. Ada beberapa cara, yaitu split half, odd even, belah tiga dan belah banyak sesuai dengan jumlah aitem.

a. Split half (ditengah, penyusunan aitem yang berseri)

Split half (belah tengah), tidak bisa di kenakan pada penyusunan aitem yang berseri, simultan karena prinsip paralel tidak di peroleh.

S	1 2 3 ... 20	Y1	21 22 23 ... 40	Y2 dst
a		a		a	
b		b		b	
....		
k		k		k	

Hasil antara Y1 dan Y2 dikorelasikan .

($r_{y_1y_2}$) → hasil korelasi adalah hasil dari ½ alat ukur oleh karena itu perlu di"utuh"kan kembali dengan rumus Spearman-Brown (Rumus dasar untuk efek perubahan panjang)

Sperman Brown

$$r_{xx'} = \frac{2r_{y_1y_2}}{1 + r_{y_1y_2}}$$

b. Odd-even (belahan genap-ganjil)

B I : aitem-aitem ganjil

B II : aitem-aitem genap

S	1 2 3 4 5 6 7 ...	Y1	Y2
a			
b			
....			
k			

Y 1 = aitem-aitem no 1 + 3 + 5 + 7

Y 2 = aitem-aitem no 2 + 4 + 6

Pembelahan baik split half maupun odd-even harus dapat dipertanggungjawabkan tingkat keparalelannya.

Beberapa macam teknik split half dan odd-even :

Split half Rulon

Kedua belahan tidak perlu mempunyai nilai varians yang sama.

Rulon :

$$r_{xx'} = 1 - \frac{Sd^2}{Sx^2}$$

Dimana :

Sd^2 : varians perbedaan kedua belahan

$Sd^2 \rightarrow$ sebagai estimasi

Sx^2 : varians skor total

Keistimewaannya :

- bisa untuk skor dikotomi (misal ya-tidak, benar-salah)
- bisa untuk skala

Cronbach Alpha

Dapat digunakan jika varians kedua belahan tidak sama atau tidak paralel pada kedua belahan.

$$\alpha = 2 \left[1 - \frac{S_1^2 - S_2^2}{S_x^2} \right]$$

Dimana :

S_1^2 = varians belahan I

S_2^2 = varians belahan II

S_k^2 = varians total

Varians = SD^2

Kuder Richardson (KR)

KR_{20} untuk split half dengan skor dikotomi

KR_{21} sama dengan KR_{20} namun mempunyai taraf kesukaran yang setara (TK)

$$TK = \frac{b}{N}$$

Dimana b = jawaban betul

- ini digunakan untuk mengukur kemampuan (true-false)
- soal semakin mudah TK semakin besar

$$KR_{20} = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum P(1-P)}{S_x^2} \right)$$

$$KR_{21} = \left(\frac{K}{K-1} \right) \left(1 - \frac{K.P(1-P)}{S_x^2} \right)$$

$$P = \frac{\sum P}{K}$$

Dimana :

K : banyaknya aitem

S_x^2 : varians total

p : porposi subjek yang mendapat skor 1 dibagi jumlah subjek

c. Belah tiga (Kuder-Richardson) oleh Kristoff, jarang di gunakan.

- hanya cocok untuk belah tiga
- tidak populer

d. Belah-banyak (Hoyt)

- dibelah sebanyak jumlah aitem
- treatment dengan subject design
- analisis varians

$$r_{xx}^1 = 1 - \frac{MK_{AS}}{MK_S}$$

Dimana :

MK_{AS} = variabel residu (S_r^2) adalah mean kuadrat/rerata kuadrat interaksi antara subjek dengan aitem

MK_S = variabel subjek (S_s^2) adalah mean kuadrat antar subjek

Formula yang cocok untuk belah banyak :

$$\alpha = \left[\frac{K}{K-1} \right] \left[1 - \frac{\sum S_j^2}{S_x^2} \right]$$

Dimana :

K : banyaknya belahan

S_j^2 : variabel belahan ke J $\rightarrow J = 1, 2, 3, \dots k$

S_x^2 : variabel total skor

PENSKALAAN

Prinsipnya mempergunakan subyek supaya bisa membuat skala yang benar.

A. Paired Comparison

Adalah proses penskalaan dengan membandingkan obyek yang diskala.

Langkah-langkah :

- a. susun objek yang hendak diskala
- b. susun pasangan-pasangan secara random
- c. ujikan pada sejumlah subjek untuk di "judge"
- d. hitung frekuensi pilihan terhadap masing-masing stimulus

- e. konversikan "f" ke "p" dimana $P = \frac{\sum f}{N * K}$

N : jumlah subjek, dan K : jumlah stimulus

- f. konfersikan "p" kenilai "z"
- g. konversi "z" terendah menjadi 0

Contoh :

4 objek : biji merah (M), biji biru (B), biji hijau (H), biji kuning (K)

$k = 4$ dan $\frac{1}{2}k(k-1) \rightarrow$ rumus perbandingan

Jumlah subjek = 10

Pasangan perbandingan (secara random) : MK, HB, BM, KH, MH, BK.

	M	B	K	H
M	5	2	1	4
B	8	5	4	4
K	9	6	5	6
H	6	6	4	5
f	28	19	14	19
p	0,7	0,475	0,35	0,455
z	0,52	-	-	-
		0,063	0,39	0,063

Contoh :

Perbandingan merah biru, dimana : 2 orang pilih biru dan 2 orang pilih merah

$$Z_0 = \frac{X - \bar{X}}{S}$$

Cara memilih nilai standar :

$p = 0,7 \rightarrow 70\%$ dimana $50\% + 20\%$

Jika dibawah rata-rata = (-) (dibawah 50%)

B. Method of Equal Appearing Interval

Tujuan untuk mencari statement yang diajukan kepada penilai apakah statement yang dibaca favourable atau tidak.

Digunakan untuk EPPS \rightarrow untuk memberikan skor pada setiap statementnya.

- Dikemukakan oleh Thurstone
- Statement favourable-unfavourable dianggap sebagai suatu kontinum
- Judgement : membaca state dan meletakkan dalam kontinum (garis berkesinambungan) dan dimana letak statement tersebut (favourable/unfavourable)

Kontinum ada 11

A B C D E F G H I J K

\leftarrow semakin unfav semakin fav \rightarrow

AB = BC = CD = ... = JK (jarak kontinum)

Asumsi : jarak setiap gradasi dianggap sama

Contoh :

Gus Dur memimpin secara demokratis

	A	B	C	D	E	F	G	H	I	J	K
F	1	1	2	14	16	23	18	8	10	6	1
p	0,01	0,01	0,02	0,14	0,16	0,23	0,18	0,08	0,1	0,06	0,1
PK	0,01	0,2	0,04	0,18	0,34	0,57	0,75	0,83	0,93	0,99	1

Proporsi = p = perbandingan antara jumlah pemilih dengan jumlah keseluruhan

Proporsi kumulatif = PK = meningkat dari yang bawah ke yang atas

Catatan : tiga tendensi sentral (Mean, Md, Mo)

Contoh :

Kelas interval	f	fk
91 - 100	8	100
81 - 90	19	92
71 - 80	47	73
61 - 70	21	26
51 - 60	5	5

$\Sigma = 100$

$$Bb = 70,5$$

$$Pkb = 26$$

$$fd = 47$$

$$i = 100$$

$$\text{Median} = \frac{1}{2} N = 50$$

$$= Bb + \left(\frac{\frac{1}{2} N + fkb}{fd} \right)$$

$$= 70,5 + \left(\frac{50 - 26}{47} \right) \cdot 10$$

$$= 75,6$$

$$\text{Med} = Bb + \left(\frac{50 - fcb}{fd} \right) \cdot i$$

$$= 5,5 + \left(\frac{50 - 16}{23} \right) \cdot 1$$

$$= 6,4$$

Jadi tingkat demokratis Gus Dur berada dalam skala 6,4

C. Method Interval Berurutan (Successive Intervals)

Huruf-huruf pada kontinum mewakili letak yang semakin kekanan berarti semakin tinggi kualitas obyek menurut dimensi yang diperhatikan

D. Summated Rating

- d. Model skala likert
- e. Nilai skala berdasarkan respon subjek
- f. Tiap skala dianggap mempunyai sifat favourabel atau unfavourabel, hanya belum ada skalanya
- g. Jawaban dari orang yang positif terhadap statement favourabel memiliki bobot lebih tinggi daripada orang yang negative

Contoh :

Statement : kenaikan harga BBM adalah bijaksana

STS TS N S SS

Aslinya Likert menggunakan 7 skala

f	40	35	12	8	5
P	0,4	0,35	0,12	0,08	0,05
PK	0,4	0,75	0,87	0,95	1,0
PKt	0,2	0,575	0,81	0,91	0,975
z	-0,84	0,19	0,88	1,34	1,96
z + 0,84	0	1,03	1,72	2,18	2,8

PKt = proporsi kumulati tengah = $\frac{1}{2}$ P + PK bawah

z : dibawah 0,5 berarti negatif

diatas 0,5 berarti positif

EFEK PERUBAHAN PANJANG TES

Efek perubahan panjang tes adalah perubahan pada jumlah aitem tes di mana penambahan jumlah aitem dan pengurangan jumlah aitem akan merubah mean, varians skor tes, reliabilitas dan validitas.

A. Efek Terhadap Reliabilitas

Memperbanyak aitem akan meningkatkan reliabilitas asal aitem yang di tambahkan bersifat paralel dengan aitem yang sudah ada. Dan sebaliknya, pengurangan jumlah aitem akan menurunkan tingkat reliabilitas tes bila aitem yang di hilangkan bersifat paralel dengan aitem-aitem yang di sisakan. Sebetulnya merupakan modifikasi pengayaan dari rumus Spearman-Brown.

Perubahan panjang tes :

- ♦ Perubahan jumlah aitem
- ♦ Pengurangan jumlah aitem

$$\rho_{xx'} = \frac{K \cdot \rho_{yy'}}{1 + (K - 1)\rho_{yy'}}$$

Dimana :

$\rho_{yy'}$ = reliabilitas sebelum berubah

$\rho_{xx'}$ = reliabilitas setelah berubah

k = jumlah aitem setelah berubah/jumlah aitem sebelum berubah

Misal :

$\rho_{yy'} = 0,60$

Jumlah aitem sebelum berubah = 20

Jumlah aitem setelah berubah = 30

Jadi perpanjangan $\frac{1}{2} X$ adalah ??

$$\rho_{xx'} = \frac{\frac{3}{2} * 0,60}{1 + \left(\frac{3}{2} - 1\right)0,60}$$

$$= \frac{0,90}{1,30} = 0,69$$

Jadi ketika 20 aitem, reliabilitasnya = 0,60
30 aitem, reliabilitasnya = 0,69

B. Efek Pada Validitas

Sama halnya terhadap reliabilitas, prediksi efek perubahan panjang tes terhadap validitas hanya dapat dilakukan apabila aitem yang di tambahkan atau dikurangkan bersifat paralel dengan aitem-aitem yang lainnya dalam tes yang bersangkutan.

$$\rho_{YZ} = \frac{K \cdot \rho_{YZ}}{\sqrt{[K = K(K-1) \cdot \rho_{YY'}]}}$$

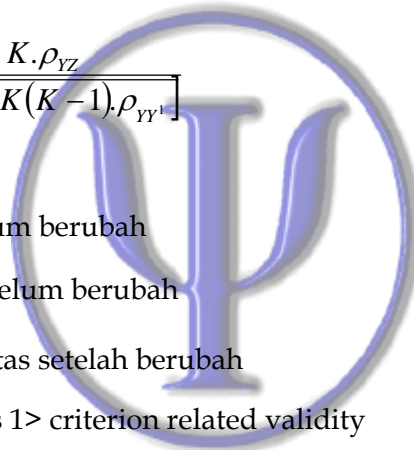
Dimana :

ρ_{YZ} = validitas sebelum berubah

$\rho_{YY'}$ = reliabilitas sebelum berubah

ρ_{XZ} = estimasi validitas setelah berubah

Hanya pada validitas 1 > criterion related validity



HETEROGENITAS KELOMPOK

Pada kelompok yang bersifat heterogen, semakin heterogen maka validitas makin kecil dan reliabilitas makin besar. Tes yang menghasilkan varians skor yang lebih besar, yaitu di kenakan pada kelompok sampel yang lebih heterogen akan menghasilkan pula koefisien reliabilitas yang lebih tinggi.

alat 1 $\sigma_e = \sigma_x \sqrt{(1 - \rho_{xx'})}$

alat 2 $\sigma_e = \sigma_y \sqrt{(1 - \rho_{yy'})}$

sehingga $\sigma_x \sqrt{(1 - \rho_{xx'})} = \sigma_y \sqrt{(1 - \rho_{yy'})} \rightarrow$ dikuadratkan

$$\rho_{yy'} = \frac{1 - \sigma_x^2 (1 - \rho_{xx'})}{\sigma_y^2}$$

di dapat dari

$$\begin{aligned} \sigma_x^2 (1 - \rho_{xx'}) &= \sigma_y^2 (1 - \rho_{yy'}) \\ \sigma_x^2 (1 - \rho_{xx'}) &= 1 - \rho_{yy'} \\ \frac{\sigma_x^2 (1 - \rho_{xx'})}{\sigma_y^2} &= 1 - \rho_{yy'} \end{aligned}$$

Menggambaran reliabilitas yang restricted dan unrestricted

Realitas yang semakin heterogen maka reliabilitasnya semakin besar.

Bisa juga digunakan pada validitas :

$$\rho_{xy} = \frac{\sigma_x \rho_{xy}}{\sqrt{(\sigma_x^2 \rho_{xy}^2 + \sigma_x^2 - \sigma_x^2 \rho_{xy})}}$$

- ♦ Heterogenitas skor kelompok yang dites mempengaruhi koefisien reliabilitas ($\rho_{xx'}$) dan koefisien validitas tes (ρ_{xy})
- ♦ Jenis kelamin :
 - heterogen (unrestricted group) : tidak dibatasi oleh karakteristik tertentu
 - homogen (restricted group) : dibatasi oleh karakteristik tertentu

Kesimpulan :

- ♦ tes yang menghasilkan varians skor yang lebih besar (dikenakan pada kelompok yang lebih heterogen), menghasilkan koefisien reliabilitas yang lebih tinggi
- ♦ varians error kelompok yang skornya homogen lebih kecil daripada varians error yang skornya heterogen atau pada kelompok sample yang heterogen dapat terjadi overestimate terhadap tingkat reliabilitas yang sesungguhnya sedangkan pada kelompok sample yang lebih homogen dapat terjadi suatu underestimate.



DAFTAR PUSTAKA

- Allen, M.J. & Yen, W.M. 1979. *Introduction to Measurement Theory*, Monterey : Brooks/Cole Publishing Company.
- Azwar, S. 1993. "Kelompok Subjek Ini Memiliki Harga Diri yang Rendah", *Kok Tahu...?. Buletin Psikologi*. Thn. I Nomor 2, Halaman 13-17
- Azwar, S. 1994. Seleksi Aitem Dalam Penyusunan Skala Psikologi, *Buletin Psikologi*. Thn. II Nomor 2 , Halaman 26-33
- Azwar, S. 1997. *Reliabilitas dan Validitas*. Edisi 3. Yogyakarta: Pustaka Pelajar.
- Azwar, S. 1997. *Tes Prestasi*. Edisi 2. Yogyakarta: Pustaka Pelajar.
- Bloom, B.S. (ed.). 1969. *Taxonomy of Educational Objectives - Hand-book I - Cognitive Domain*. New York, NY.: David McKay Company, INC.
- Brown, F.G. 1971. *Measurement and Evaluation*. Itasca, Ill.: F.E. Peacock Publisher.
- Brown, F.G. 1976. *Principles of Educational and Psychological Testing*, 2nd. edition. New York, NY: Holt, Rinehart & Winston.
- Crocker, L. & Algina, J. 1986. *Introduction to Classical and Modern Test Theory*. Forth Worth: Holt, Rinehart, and Winston, INC.
- Cronbach, L.J. 1970. *Essentials of Psychological Testing*, 3rd. edition. New York, NY: Harper and Row.
- Ebel, R.L. 1979. *Essentials of Educational Measurement*, 3rd. edition. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Sax, G. 1980. *Principles of Educational and Psychological Measurement and Evaluation*, 2nd. edition. Belmont, CA: Wadsworth Publishing Company.
- Thorndike, R.M., Cunningham, G.K., Thorndike, R.L., & Hagen, E.P. *Measurement and Evaluation in Psychology and Education*. New York, NY: Macmillan Publishing Company.