

Program Magister Teknik dan Manajemen Industri ITB

@ 2013

KULIAH 7

PERENCANAAN DAN EVALUASI ALAT UKUR

- Pokok Bahasan:

1. Variabel: Variable Terikat, Variabel Bebas, Variabel Moderator, Variabel Antara
2. Pengukuran, Karakteristik Pengukuran yang Baik
3. Perencanaan Alat Ukur
4. Evaluasi Alat Ukur: Analisa Item, Reliabilitas, Validitas
5. Norma

1. Variabel

- Konsep:
 - Definisi yang digunakan untuk menggambarkan secara abstrak sifat suatu objek/fenomena atau generalisasi sifat sekelompok objek/fenomena, sehingga bisa digunakan untuk menggambarkan sifat objek/fenomena sejenis
- Variabel: Konsep yang mempunyai variasi nilai
- Contoh:
 - Badan: bukan variabel, tidak mempunyai variasi nilai
 - Tinggi badan: variabel, bisa bervariasi

Klasifikasi Variabel (Proses Kuantifikasi):

- **Variabel Nominal** (kategori/dikotomik):
 - variabel diukur berdasarkan golongan, tanpa memberikan nilai pada masing-masing golongan (sifat: diskrit, satu sama lain mutually exclusive), Contoh: jenis kelamin

1. Variabel

- **Variabel Ordinal:**

- variabel diukur menurut **jenjang**, mengikuti suatu atribut tertentu, sehingga diperoleh rangking/tingkatan objek jika diukur dengan variabel tersebut
- Contoh: dua kegiatan dijadikan objek yang akan diukur, yaitu: belajar dan berolahraga

- **Variabel Interval:**

- variabel yang dinyatakan dengan **interval/range**, bukan dengan nilai mutlak
- Contoh: berat badan diukur dengan kelas: 0 ~ 20 kg, 21 ~ 40 kg, 41 ~ 60 kg, dst.

- **Variabel Rasio:**

- sama dengan variabel interval dan diperoleh juga informasi tentang nilai mutlak atribut yang ada pada suatu objek. Jarak interval tidak dinyatakan dengan perbedaan kelas, tetapi terhadap **titik nol**.

1. Variabel

- **Variabel Rasio:**

- Contoh: berat badan 60 kg \Rightarrow ada jarak 60 x satuan kg dari titik nol.

Jenis Variabel (Proses Fungsi/Peran):

- **Variabel terikat** (dependent variable):

- variabel yang kondisinya dipengaruhi oleh variabel yang lain (variabel bebas)

- **Variabel bebas** (independent variable):

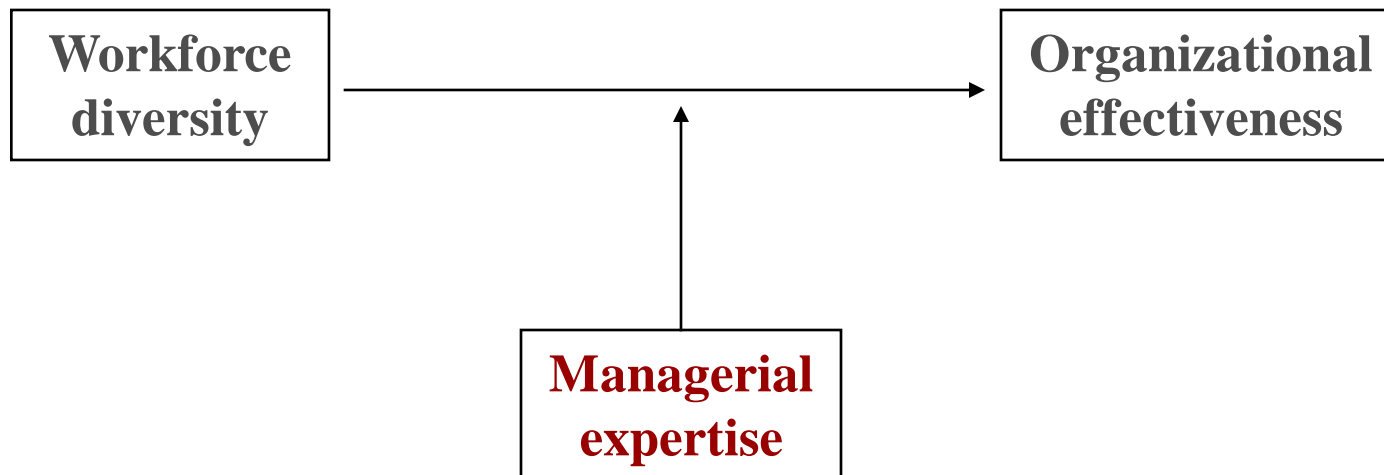
- variabel yang dipelajari pengaruhnya terhadap variabel yang lain (variabel dependen)

- **Variabel moderator** (moderating variable)

- **Variabel antara** (intervening variable)

1. Variabel

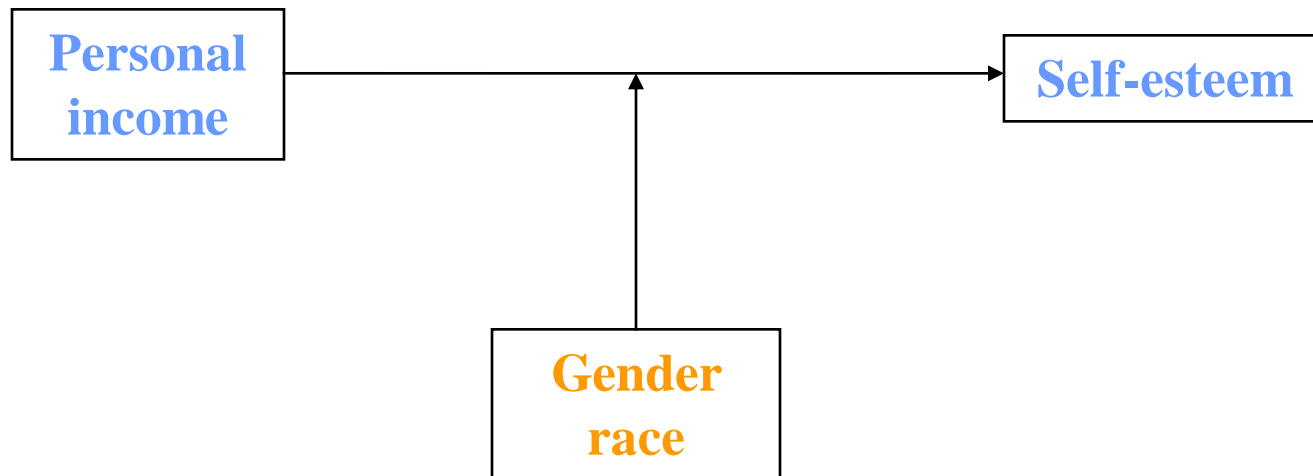
- **Variabel moderator** (moderating variable):
 - The moderating variable is one that has a **strong contingent effect** on the independent variable-dependent variable relationship.
 - The presence of a **third variable** (the moderating variable) **modifies the originally expected relationship** between the independent and dependent variable.



1. Variabel

■ Mediating Effects:

- It is important to test whether **third variables** are mediators of the effect we are examining in our primary research hypothesis. Average effect sizes across studies could be calculated and evaluated for these potential mediating variable categories



1. Variabel

■ Intervening Variable:

- Setiap kali menentukan variabel penyebab munculnya suatu gejala/fenomena, ada kemungkinan bahwa variabel yang dianggap sebagai penyebab, sesungguhnya **bukan penyebab langsung**, sehingga seringkali perlu diteliti kemungkinan **penyebab lain**
- Adanya **variabel antara** merupakan **konsekuensi** dari tidak diketahuinya secara pasti cara mengidentifikasi variabel-variabel suatu fenomena dan tidak diketahui secara pasti rangkaian hubungan antar variabel-variabel suatu fenomena

■ Time: T_1 \longrightarrow T_2 \longrightarrow T_3



1. Variabel

Definisi Operasional Variabel:

- Definisi yang didasarkan pada sifat atribut yang diamati pada objek penelitian, bisa berbentuk kuantitatif maupun kualitatif
- Contoh: orang kerdil adalah orang dewasa yang tinggi badannya kurang dari 120 cm

3 pola penentuan definisi operasional variabel:

- **Kegiatan:** kegiatan yang harus dilakukan/terjadi agar hal yang didefinisikan dianggap telah terjadi
 - Contoh: orang pemalas adalah orang yang tidur minimal selama 60 jam/minggu
- **Operasi yang dilakukan Objek:** menunjukkan bagaimana hal yang didefinisikan beroperasi
 - Contoh: tukang tik yang mahir adalah yang sanggup mengetik tanpa melihat
- **Tampaknya Objek:** menunjukkan bagaimana objek tampak
 - Contoh: mangga yang mentah dagingnya berwarna putih dan kulitnya berwarna hijau

2. Pengukuran

Definisi:

- pencantuman bilangan terhadap karakteristik berdasarkan peraturan tertentu (Tyler, 1971)
- membandingkan sesuatu yang sedang diukur dengan suatu alat pengukur secara deskriptif dan tidak evaluatif (Azwar, 1987)

Psychological Versus Physical Measurement:

- Psychological measurement mempunyai presisi yang lebih rendah (less precise) & pengukuran yang kurang langsung (less direct) dibandingkan physical measurement:
 - Psychological measurement hanya mengukur sampel perilaku (karakteristik yang akan dipelajari)
 - Psychological measurement mempunyai skala yang lebih terbatas

2. Pengukuran

- Psychological measurement lebih mudah dipengaruhi variabel extraneous
 - Psychological measurement dirancang untuk menggambarkan inferensi mengenai atribut atau karakteristik
 - Psychological measurement dirancang untuk mengukur construct
-
- **Karakteristik Pengukuran yang Baik:**
 - Alat ukur yang baik harus direncanakan sebaik mungkin dan di evaluasi secara empiris untuk meyakinkan bahwa alat ukur tersebut menghasilkan informasi yang akurat dan berguna

3. Perencanaan Alat Ukur

Perencanaan Alat Ukur:

- Suatu spesifikasi alat ukur biasanya meliputi:
 1. Identifikasi tujuan pengukuran
 2. Pembatasan cakupan isi (content) alat ukur
 3. Penentuan format alat ukur yang digunakan
 4. Penentuan banyaknya item
 5. Penentuan tabel spesifikasi

- 1. Alat ukur yang baik harus mempunyai tujuan yang jelas:
 - Untuk mengukur apa alat ukur tersebut ? (domain)
 - Siapa yang akan diukur dengan alat ukur tersebut ? (audience)
 - Bagaimana hasil alat ukur digunakan ? (tipe skoring)

3. Perencanaan Alat Ukur

2. Pembatasan cakupan isi (content) alat ukur:

- Alat ukur harus **komprehensif** dan berisi **item-item yang relevan**
- **Komprehensif:** alat ukur itu meliputi **keseluruhan bagian** dan **representatif** dalam jumlah item yang **proporsional** sesuai dengan urgensi dan bobot masing-masing bagian
- **Relevan:** item yang ditulis benar-benar menanyakan mengenai bagian yang **bersangkutan** dan segala sesuatu yang **berhubungan** dengan bagian tersebut
- Suatu cara yang dapat ditempuh dalam usaha menyusun alat ukur yang berisi item yang komprehensif dan relevan adalah dengan melakukan **pembagian kelompok**, dan kemudian kelompok itu diberi bobot sesuai dengan pentingnya bagian tersebut

3. Perencanaan Alat Ukur

2. Pembatasan cakupan isi (content) alat ukur:

- Cara ini berarti bahwa isi alat ukur harus diidentifikasi dengan jelas, **dibatasi construct** atau konsepsi teoritiknya, kemudian **didefinisikan secara operasional** dalam bentuk **indikator-indikator** agar dapat diukur
- **Operasionalisasi konsep** atau definisi operasional dari suatu konsep adalah menurunkan konsep-konsep yang abstrak sehingga dapat diukur, yang dilakukan dengan melihat dimensi-dimensi atau ciri-ciri perilaku dari konsep tersebut, dan kemudian mengkategorikannya menjadi elemen-elemen yang dapat diukur atau diamati
- Pada definisi operasional suatu konsep tidak boleh menggambarkan alasan, penyebab, akibat ataupun keterkaitan konsep dengan yang lain

3. Perencanaan Alat Ukur

Cara menentukan isi (content) suatu alat ukur:

- Domain pengetahuan atau keterampilan (perencanaan test/test plan)
- Domain construct (construct explication)
- Domain perilaku (analisis tugas/job analysis)

3. Format alat ukur:

- Objective test: Personality & Interest:
 - Independent items, Forced-choice items
- Projective test: Personality
 - Verbal techniques, Drawing tasks, Pictorial techniques
- Attitude scales: Likert scales, Thurstone scales
- Evaluation test: Rating scale

4. Penentuan banyaknya item:

- Alat ukur haruslah berisi sebanyak-banyaknya item yang independen satu sama lain

3. Perencanaan Alat Ukur

4. Penentuan banyaknya item:

- Independen adalah bahwa masing-masing item mengungkapkan bagian terkecil bahan alat ukur yang berbeda satu sama lain
- Alat ukur yang berisi item yang lebih banyak akan mempunyai reliabilitas yang lebih tinggi
- Alat ukur yang berisi item yang banyak, memerlukan waktu yang lebih banyak dan mempengaruhi tingkat kelelahan sehingga hasil ukurnya menjadi bias

5. Penentuan tabel spesifikasi:

- Tabel yang menjadi pegangan yang akan tetap mengarahkan penulis item pada tujuan pengukuran dan menjaganya agar tidak keluar dari batasan isi pengukuran
- Tabel ini juga berguna untuk menyusun lebih dari satu buah alat ukur yang paralel

4. Evaluasi Alat Ukur

- 3 (tiga) sifat psikometrik (psychometric properties) dari alat ukur yang baik adalah:
 - Alat ukur yang baik harus reliabel
 - Alat ukur yang baik harus valid
 - Alat ukur yang baik terdiri dari item-item yang baik
- Prosedur kerja dalam melakukan pengujian seluruh item dalam alat ukur berdasarkan data empiris
 - Metode-metode analisis item:
 - Item difficulty (tingkat kesukaran)
 - Item discriminability (daya pembeda item)
 - Distractor power
 - Item characteristics

4. Evaluasi Alat Ukur: Analisis Item

- **Item difficulty** (tingkat kesukaran):
 - digunakan pada alat ukur achievement atau aptitude
 - biasanya dilakukan dengan menghitung indeks tingkat kesukaran
 - Indeks tingkat kesukaran (p):

$$\frac{\text{jumlah orang yang menjawab benar pada suatu item}}{n}$$

- nilai p optimal: 0.5 - 0.7
- **Item discriminatory** (daya pembeda item):
 - dapat digunakan untuk menganalisis item untuk setiap jenis alat ukur
 - biasanya dilakukan dengan cara menghitung item discrimination index atau **korelasi item-total**

4. Evaluasi Alat Ukur: Analisis Item

- **Item discriminatory** (daya pembeda item):
 - Item discrimination index: $D = p_T - p_B$
 - $p_T = (\text{jumlah orang yang menjawab benar pada kelompok atas}) / (\text{jumlah orang pada kelompok atas})$
 - $p_B = (\text{jumlah orang yang menjawab benar pada kelompok bawah}) / (\text{jumlah orang pada kelompok bawah})$
- **Korelasi item-total**: menggunakan rumus korelasi Pearson

4. Evaluasi Alat Ukur: Analisis Item

- Prinsip utama pemilihan item adalah dengan mencari **harga koefisien yang setinggi mungkin** dan menyingkirkan setiap item yang mempunyai **korelasi negatif (-) atau yang mendekati nol (0)**
- Friedenberg dan Kaplan menyarankan koefisien korelasi **minimal 0.3**
- Guilford menyarankan koefisien korelasi **minimal 0.4**
- Kriteria Guilford (1956):
 - < 0.20 : tidak ada korelasi
 - $0.20 - < 0.40$: korelasi rendah
 - $0.40 - < 0.70$: korelasi sedang
 - $0.70 - < 0.90$: korelasi tinggi
 - $0.90 - < 1.00$: korelasi tinggi sekali
 - 1.00 : korelasi sempurna

4. Evaluasi Alat Ukur: Analisis Item

■ Distractor power:

- dapat digunakan untuk test achievement atau aptitude, yang menggunakan format pilihan berganda
- biasanya dilakukan dengan menghitung nilai expected distractor power:

$$\frac{\text{jumlah individu yang menjawab salah pada suatu item}}{\text{jumlah distractor}}$$

■ Item characteristics:

- Cara paling sederhana untuk menyimpulkan item, dengan cara menampilkan dalam **grafik kesukaran** maupun **kurva karakteristik item**
- Grafik kesukaran item: grafik yang menggambarkan tingkat kesukaran dan daya pembeda suatu alat ukur
- Kurva karakteristik item: grafik yang menggambarkan kemungkinan jawaban yang benar sebagai fungsi dari skor total

4. Evaluasi Alat Ukur: Reliabilitas

- Tingkat keterpercayaan hasil suatu pengukuran
- Disebut juga sebagai keterpercayaan, keterandalan, keajegan, **konsistensi**, kestabilan
- Ide: **sejauh mana** skor hasil pengukuran terbebas dari galat pengukuran (measurement error)
- Secara teoritis besarnya koefisien reliabilitas adalah $-1 \leq r \leq +1$, akan tetapi koefisien yang besarnya kurang dari nol (0) tidak ada artinya karena interpretasi reliabilitas selalu mengacu pada koefisien yang positif (+)
- Reliabilitas: indeks yang menunjukkan **sejauh mana suatu alat pengukur dapat dipercaya atau dapat diandalkan** (Ancok, 1989)
- Jika suatu alat pengukur digunakan dua kali untuk mengukur sesuatu yang sama dan hasil kedua pengukuran yang diperoleh relatif **konsisten**, maka alat pengukur tersebut reliabel (Ancok, 1989)

4. Evaluasi Alat Ukur: Reliabilitas

- Reliabilitas mengukur **sampai seberapa jauh tingkat konsistensi** suatu alat ukur dalam mengukur sesuatu (berhubungan dengan stabilitas dalam pengukuran)
- **Metode penghitungan** reliabilitas:
 - Test retest reliability
 - Parallel form reliability
 - Internal consistency
 - Split half reliability (spearman brown correction)
 - Kuder Richardson 20
 - Kuder Richardson 21
 - Alpha Cronbach
- Besarnya koefisien reliabilitas:
 - Koefisien reliabilitas yang berkisar antara **0.70 - 0.80** dianggap baik untuk digunakan
 - **Clinical setting**: koefisien reliabilitas di atas 0.90

4. Evaluasi Alat Ukur: Validitas

- Validitas menentukan sampai **seberapa bagus suatu alat ukur yang didesain mampu mengukur suatu konsep tertentu yang ingin diukur** (Ancok, 1989; Sekaran, 2000)
- Dalam penelitian survei, maka kuesioner yang disusun oleh peneliti harus mengukur apa yang ingin diukurnya (Ancok, 1989)
- Ukuran yang benar-benar mengukur apa yang akan diukur
- Ketepatan & kecermatan alat ukur dalam menjalankan fungsi pengukurannya
- Hal-hal yang **mengurangi atau menurunkan validitas data** (Ancok, 1989):
 - Apakah si pewawancara mengikuti petunjuk yang telah ditetapkan di dalam kuesioner ?
 - Keadaan responden pada saat diwawancarai (malu, takut, cemas) ?
 - Alat pengukur yang digunakan (valid) ?

4. Evaluasi Alat Ukur: Validitas

- Tipe validitas:
 - Content validity: Face validity
 - Criterion related validity
 - Concurrent validity
 - Predictive validity
 - Construct validity
- Bukti adanya construct validity:
 - Alat ukur harus homogen dan mengukur konsep yang tunggal
 - Hasil pengukuran berubah karena adanya pengalaman (intervensi)
 - Hasil pengukuran menunjukkan perbedaan pada kelompok yang berbeda
 - Konvergen, Diskriminan, Analisis faktor

4. Evaluasi Alat Ukur

Daftar pertanyaan untuk mengevaluasi alat ukur:

- **Sampel:**
 - Berapa banyak sampel yang digunakan untuk menentukan reliabilitas, validitas dan norma alat ukur ?
 - Apakah alat ukur tersebut memberikan hasil yang sama untuk karakteristik pribadi dan demografik yang ada pada sampel penelitian ?
- **Reliabilitas:**
 - Metode apakah yang digunakan untuk mengestimasi koefisien reliabilitas dari alat ukur ?
 - Apakah koefisien reliabilitas sudah cukup baik untuk alat ukur tersebut ?
- **Validitas:**
 - Apakah ada cukup bukti yang menunjukkan alat ukur tersebut meaningful sesuai tujuannya ?

4. Evaluasi Alat Ukur

- **Validitas:**

- Kriteria spesifik apakah yang digunakan dalam validasi alat ukur tersebut ?

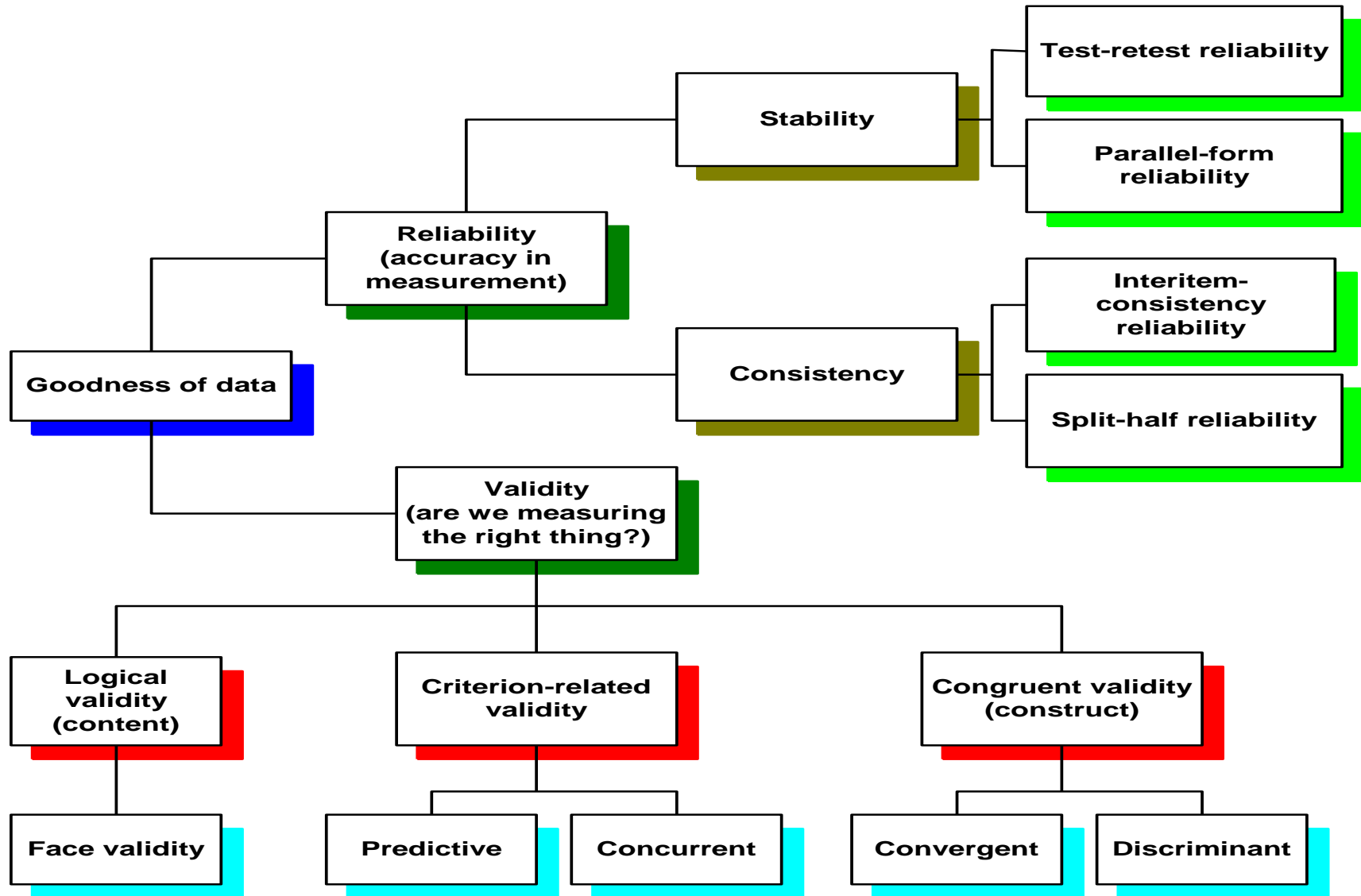
- **Skoring:**

- Apakah tersedia kunci untuk skoring ?
- Jika ukur tersebut diskor menggunakan suatu mesin khusus, berapa biaya yang diperlukannya ?

- **Pertimbangan pemakaian:**

- Berapa lama administrasi pengukuran tersebut ?
- Apakah alat ukur tersebut harus dibacakan ?
- Berapa banyak latihan yang diperlukan untuk administrasi pengukuran tersebut ? Apa yang diperoleh dari latihan tersebut ?

4. Evaluasi Alat Ukur



Sumber: Sekaran (2000)

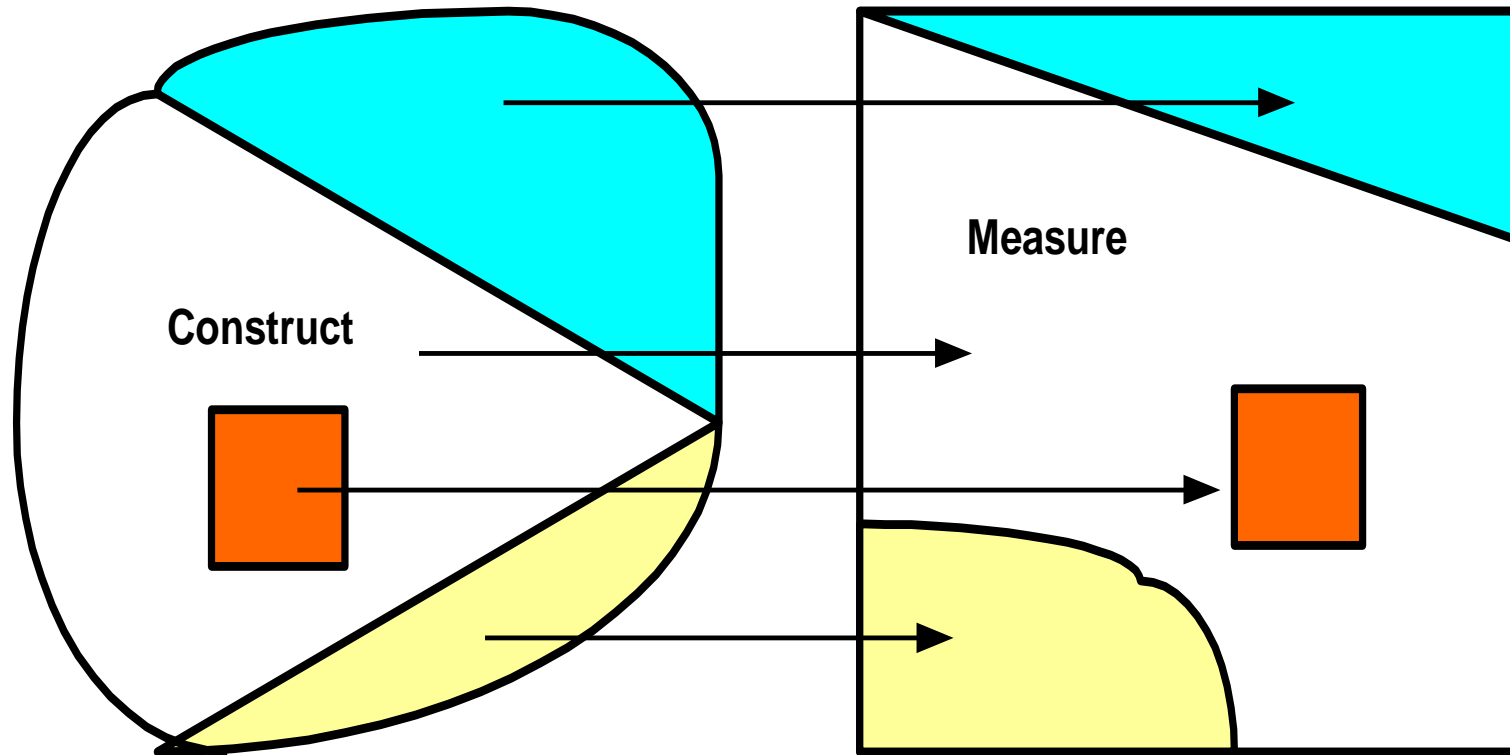
4. Evaluasi Alat Ukur

Content Validity (captures the entire meaning)	Does the measure adequately measure the concept?
Face validity (in the judgment of others)	Do “experts” (scientific community) validate that the instrument measures what its name suggests it measures?
Criterion-related validity (agrees with an external source)	Concurrent: agrees with a pre-existing measure Predictive: agrees with future behavior
Construct validity (multiple indicators are consistent)	Convergent - alike ones are alike Discriminant - different ones differ

4. Evaluasi Alat Ukur: Content Validity

- Se jauh mana isi alat pengukur mewakili semua aspek yang dianggap sebagai aspek-aspek kerangka konsep (Ancok, 1989)
- Hanya bisa diuji secara logika (subjektif).
- Misal:
 - Seorang peneliti di dalam menyusun kuesioner hanya memasukkan 3 dari 5 aspek yang merupakan kerangka suatu konsep, maka alat ukur (kuesioner) yang disusun tidak memiliki validitas isi yang tinggi

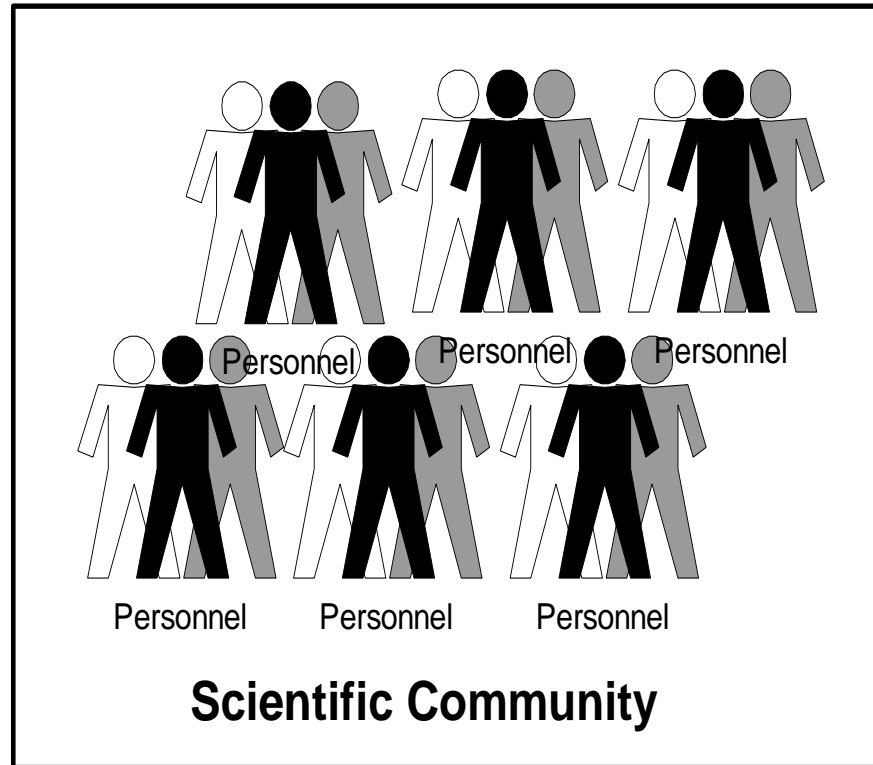
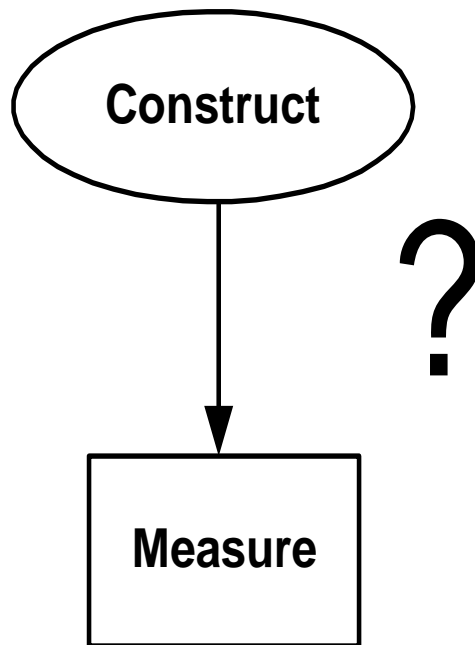
4. Evaluasi Alat Ukur: Content Validity (Neuman, 1994)



4. Evaluasi Alat Ukur: Face Validity

- Face Validity tidak menunjukkan apa yang ingin diukur, tetapi hanya menunjukkan bahwa dari segi rupanya suatu alat pengukur tampaknya mengukur apa yang ingin diukur (Ancok, 1989)
- Face validity is ultimately a matter of judgment (Selltiz et al., 1976)
- Face validity is simply assessed by the evaluator's studying the concept to be measured and determining, in his or her judgment, whether the instrument arrives at the concept adequately (Bailey, 1987)
- Major problems with face validity arise when (Bailey, 1987):
 1. there is no consensus about the definition of the concept to be measured;
 2. the concept is a multidimensional one consisting of several sub-concepts;
 3. the measure is lengthy and complex.

4. Evaluasi Alat Ukur: Face Validity (Neuman, 1994)



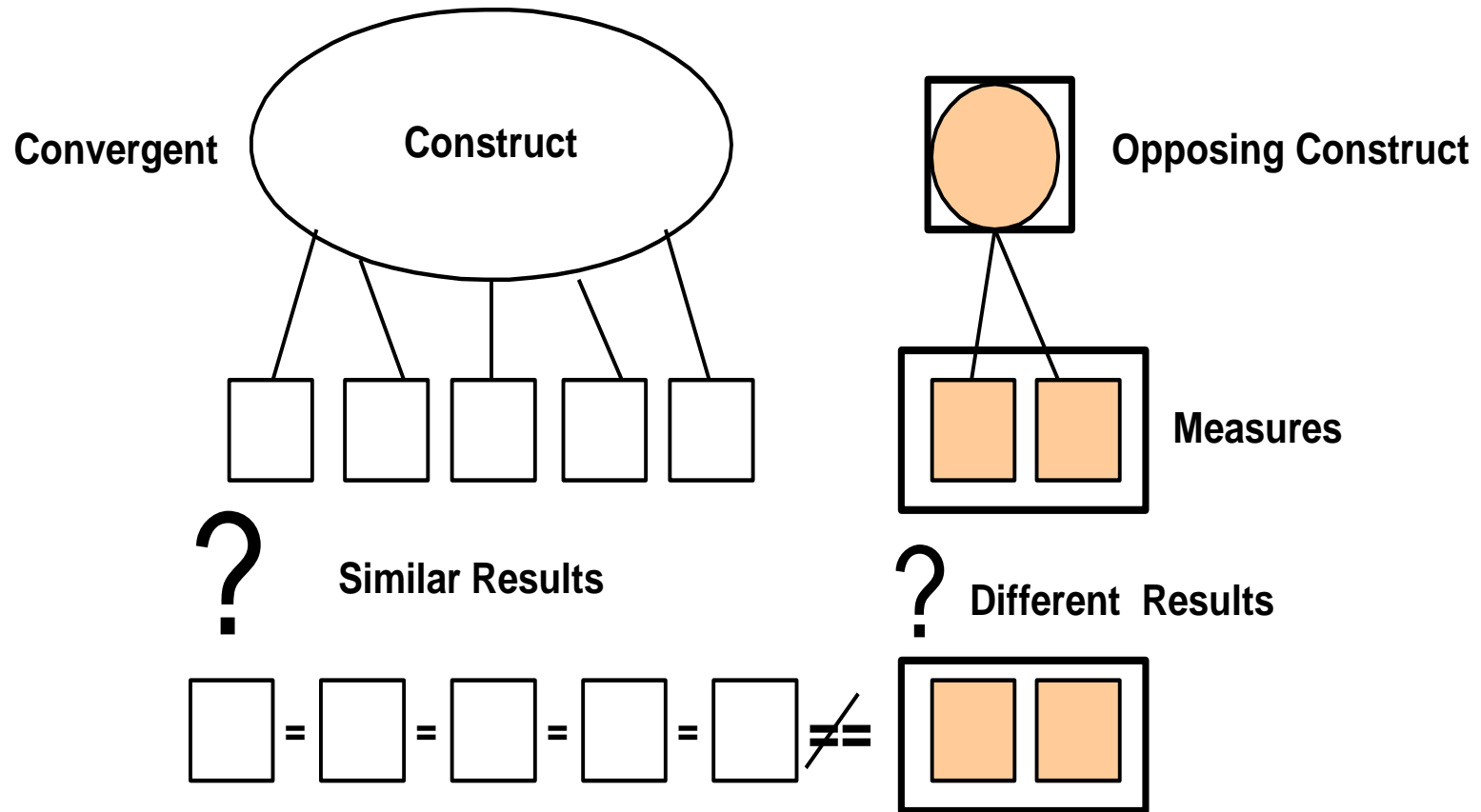
4. Evaluasi Alat Ukur: Construct Validity

- Konstruk (construct) merupakan **kerangka dari suatu konsep**
- Construct validation is often said to be the **strongest** kind of validation procedure (Bailey, 1987)

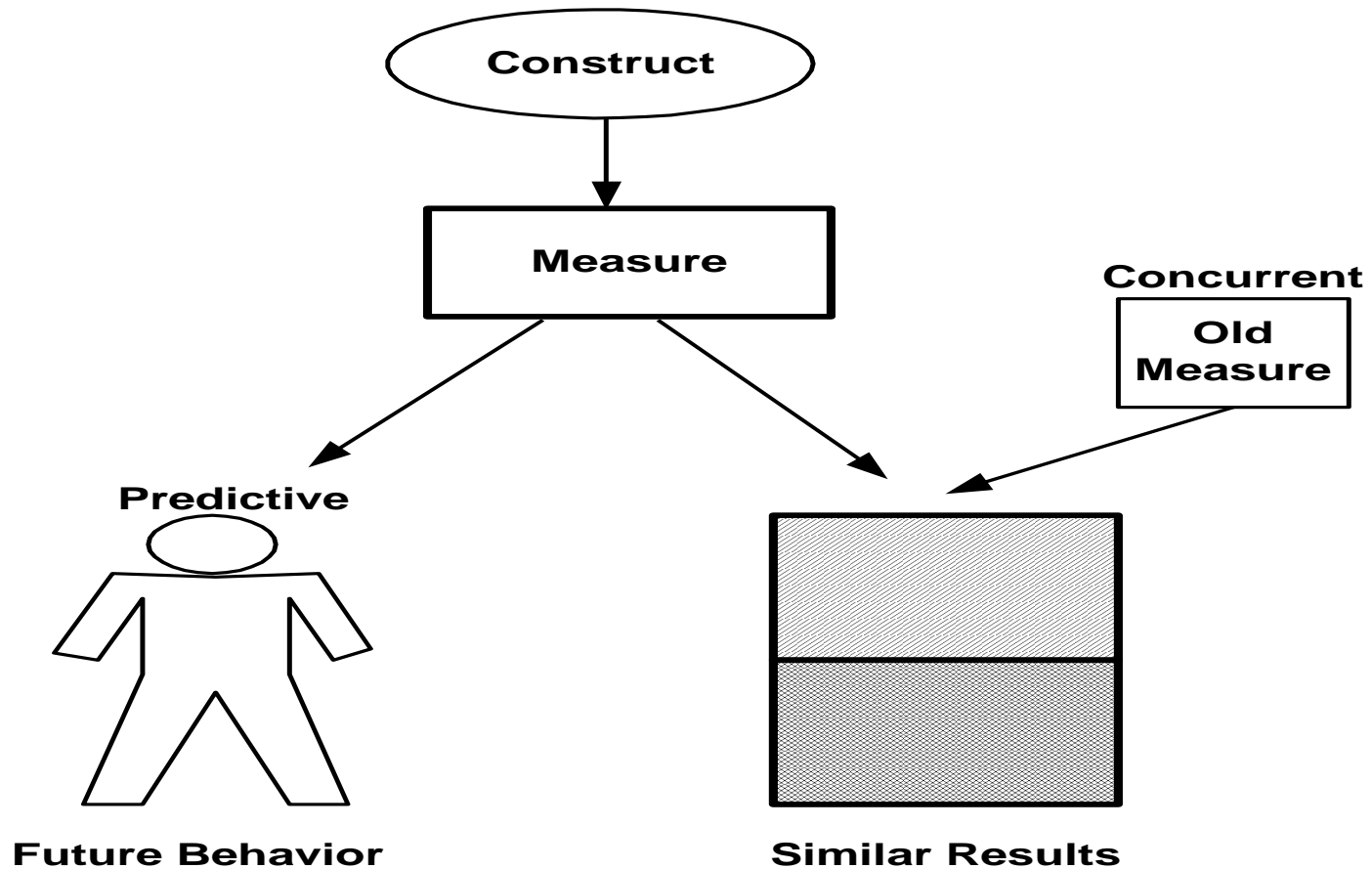
Cara mencari Kerangka suatu Konsep (Ancok, 1989):

- Mencari **definisi konsep** yang dikemukakan para ahli yang tertulis di dalam literatur (kalau beruntung, dapat diperoleh atau sudah ada definisi serta kerangka konsep yang jelas; kalau belum ada, maka peneliti harus membuat definisi operasional terlebih dahulu)
- Peneliti mendefinisikan sendiri **konsep yang akan diteliti**, misalnya melalui diskusi dengan para ahli yang kompeten dan selanjutnya dicari kesamaan-kesamaannya.
- Menanyakan definisi **konsep yang akan diukur** kepada calon responden atau kepada orang-orang yang memiliki karakteristik yang sama dengan responden. Cara ini bisa dipakai untuk menghindari **bias**

4. Evaluasi Alat Ukur: Construct Validity (Neuman, 1994)



4. Evaluasi Alat Ukur: Criterion Validity (Neuman, 1994)



4. Evaluasi Alat Ukur: Predictive Validity

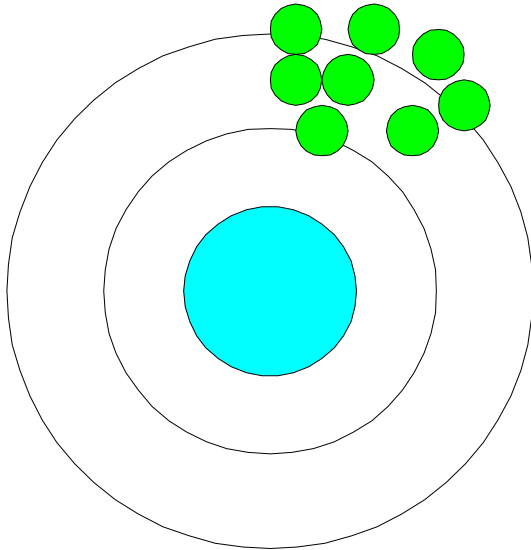
- Alat pengukur yang dibuat peneliti seringkali dimaksudkan untuk **memprediksi apa** yang akan terjadi di masa yang akan datang (to predict)
- Bertujuan untuk memeriksa apakah ada **kesesuaian** antara ramalan (prediksi) tentang perilaku obyek penelitian dengan perilakunya yang nyata.
- Validitas prediktif diberi nilai yang tinggi jika diramalkan ternyata terbukti

4. Evaluasi Alat Ukur: Validitas Budaya

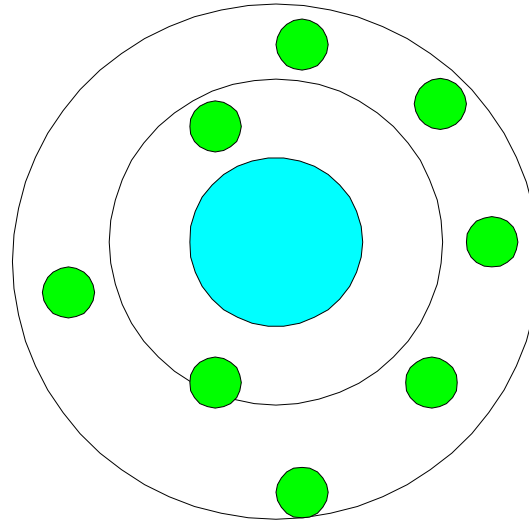
- Validitas antar budaya ini sangat penting bagi penelitian di negara yang suku bangsanya sangat bervariasi (Ancok, 1989).
- Penelitian yang dilakukan sekaligus di beberapa negara dengan alat pengukur yang sama juga akan menghadapi masalah validitas budaya (Ancok, 1989)
- Suatu alat pengukur yang sudah valid untuk penelitian di suatu negara, belum tentu akan valid jika digunakan di negara lain yang budayanya berbeda (Ancok, 1989)

4. Evaluasi Alat Ukur (Babbie, 1992)

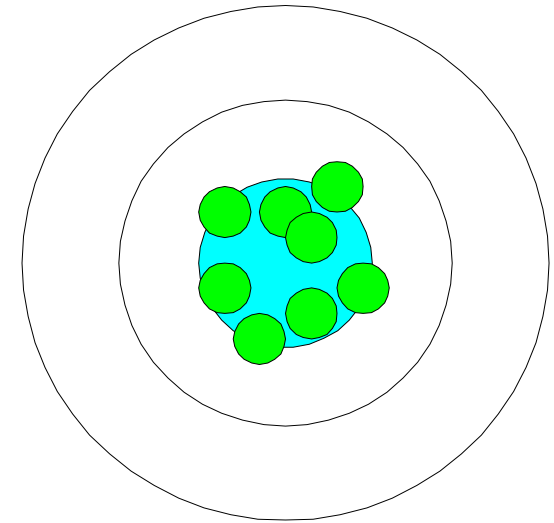
Alat ukur tidak mampu mengukur suatu konsep yang ingin diukur



Reliable but not valid



Valid but not reliable



Valid and reliable

Alat ukur tidak bisa diandalkan (tidak konsisten)

4. Evaluasi Alat Ukur: Type of Error (Bailey, 1987)

Stage of Research	Type of Error
1. Concept and hypothesis construction (including choice of operational definitions)	Lack of face validity or content validity
2. Construction of research instrument (questionnaire)	Lack of reliability (faulty or ambiguous wording on questions)
3. Sampling	Lack of external validity (sampling error)
4. Data gathering	Error due to failure to control: <ul style="list-style-type: none">▪ Environment▪ Personal characteristics of respondent▪ Relationship between respondent and researcher▪ Defects of research instrument (faulty sound recording, equipment failure, etc.)▪ Interviewer misunderstood answer
5. Coding	Incorrect information recorded due to missing data illegible data, or simple coding error
6. Data analysis	Misuse of statistics or faulty interpretation of data

5. Norma

- Norma:
 - gambaran ringkasan performansi individu-individu dalam suatu kelompok yang diukur dengan alat ukur yang baku
- Evaluasi hasil pengukuran dilakukan dengan 2 cara:
 - **norm referenced**, dimana skor pengukuran ditransformasikan menjadi skor baru yang menunjukkan dimana posisi setiap individu dibandingkan kelompoknya
 - **criterion referenced**, dimana skor suatu pengukuran ditransformasikan ke dalam standar atau kriteria yang ditentukan