

The background of the entire slide is a repeating geometric pattern in blue and white. It features stylized leaves and circular motifs. A white rectangular box with a thin blue border is centered on the slide, containing the title and author information. A solid blue rectangular tab is positioned at the top center of this white box.

PENILAIAN PEMBELAJARAN MATEMATIKA

Oleh
Dewi Mardhiyana



RELIABILITAS

Reliabilitas

- Kata realibilitas diambil dari bahasa Inggris *reliability* dari kata asal *reliable* yang artinya dapat dipercaya
- Reliabilitas dalam pengukuran merujuk kepada kekonsistenan atau kestabilan dari suatu hasil pengukuran
- Reliabilitas suatu tes menunjukkan bagaimana suatu tes dapat menunjukkan keajegan hasil tes, ajeg (tetap) disini tidak diartikan sama tetapi bagaimana tes tersebut dapat menunjukkan perubahan secara ajeg.

Reliabilitas

Teori tes klasik menyatakan bahwa setiap skor tes hasil pengukuran terdiri dari ***true score*** yaitu skor yang diperoleh jika tidak ada *error* (kesalahan), artinya skornya sangat reliabel dan ***error score***

Skor perolehan (*obtain score*) yang merupakan skor yang diperoleh oleh peserta tes dalam suatu tes.

Skor perolehan dapat di representasikan sebagai $X_i = T + E$

T merupakan *true score* yang menunjukkan kemampuan, pengetahuan, sikap, dan segala sesuatu dalam tes pengukuran dari individu, dan E merupakan *error score* yang diasumsikan sebagai adanya kesalahan dalam pengukuran.

Estimasi Reliabilitas

$X_i = T + E$ dapat diperluas dalam konsep varians sebagai $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$

σ_X^2 merepresentasikan varians skor total (perolehan dari hasil tes)

σ_T^2 merepresentasikan varians dari true score

σ_E^2 merepresentasikan varians yang diakibatkan oleh kesalahan pengukuran

Simbol umum untuk estimasi reliabilitas hasil pengukuran yaitu $r_{XX'} = \rho_{XX'}$

Estimasi dari reliabilitas suatu skor tes dinyatakan sebagai rasio dari variansi *true score* dengan variansi skor total, dinyatakan secara matematis sebagai berikut:

$$r_{XX'} = \rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

**Faktor/sumber
kesalahan dalam
pengukuran**

```
graph LR; A[Faktor/sumber kesalahan dalam pengukuran] --> B[Kesalahan konten sampling]; A --> C[Kesalahan waktu sampling]; A --> D[Sumber lainnya];
```

Kesalahan konten sampling

menunjukkan kesalahan dari sampel item tes apakah sudah benar-benar menunjukkan domain tes yang akan diukur secara representatif. Kesalahan konten sampling akan relatif kecil jika item tes benar-benar representatif terhadap domain yang akan diukur

Kesalahan waktu sampling

apakah hasil tes yang diperoleh pada saat pagi hari akan sama dengan siang hari, jeda waktu antar tes memungkinkan terjadinya kesalahan karena adanya perubahan sikap, mood, dan aspek lainnya secara acak

Sumber lainnya

pemberian skor yang subjektif

Metode Estimasi Reliabilitas

Metode tes ulang (*test-retest method*)

Bentuk Paralel (*alternate form*)

Simultaneous Administration

Delayed Administration

Internal Konsistensi

Belah dua (*split-half*)

Koefisien α dan Khurd Richardson

Antar penilai (*inter-rater*)

Estimasi Reliabilitas Metode Tes Ulang (*Test-retest Method*)

Metode ini didasarkan pada tes yang diberikan pada peserta ujian yang sama dengan tes yang sama diberikan dua kali.

Koefisien korelasi merupakan korelasi dari kedua skor hasil tes tersebut.

Jika peserta tes mendapatkan skor yang sama persis pada saat tes pertama dan kedua maka reliabilitasnya sempurna dengan koefisien reliabilitas 1,0. Tetapi, jika himpunan skor yang diperoleh pada saat tes pertama dan kedua tidak berelasi maka koefisien reliabilitasnya 0.

Untuk menentukan koefisien korelasi dapat digunakan formula korelasi *product – moment* dari Pearson

Masalah yang Timbul dengan Metode Tes Ulang

- *Carry-over effects* antara tes, yaitu ada kemungkinan tes pertama berpengaruh pada tes kedua. Misalnya: masih mengingat jawaban, disebabkan oleh pengaruh latihan (tes ketangkasan dan tes kemampuan), perubahan sikap peserta ujian (mood, kondisi fisik, dll), mencari informasi bagaimana meningkatkan skor tes mereka.
- Lamanya waktu antara pemberian kedua tes.
Selang waktu yang sangat singkat akan menimbulkan *carry over effects* yang disebabkan oleh ingatan, latihan, atau suasana hati. Selang waktu yang lama akan memberikan pengaruh karena perubahan informasi atau suasana.



**cocok digunakan untuk mengukur tentang tes
diskriminasi sensori (tes kemampuan audio atau visual)**

Estimasi Reliabilitas Metode Bentuk Paralel (*Alternate Forms*)

Metode bentuk parallel melibatkan pengembangan dari dua tes yang ekuivalen (paralel atau setara), dalam konteks isi, tingkat kesulitan atau karakteristik lain yang relevan diantara kedua tes

Jenis bentuk paralel

Simultaneous Administration

penyajian dua bentuk tes yang setara kepada kelompok subjek yang sama **dalam satu sesi atau waktu yang simultan/berurutan**

Delayed Administration

penyajian dua bentuk tes yang setara kepada kelompok subjek yang sama tetapi **dengan jeda waktu atau tidak pada sesi yang sama**

Untuk menentukan koefisien korelasi dapat digunakan formula korelasi *product – moment* dari Pearson

Kelebihan dan Kelemahan Bentuk Paralel (*Alternate Forms*)

Kelebihan

tidak ada faktor “masih mengingat soal”

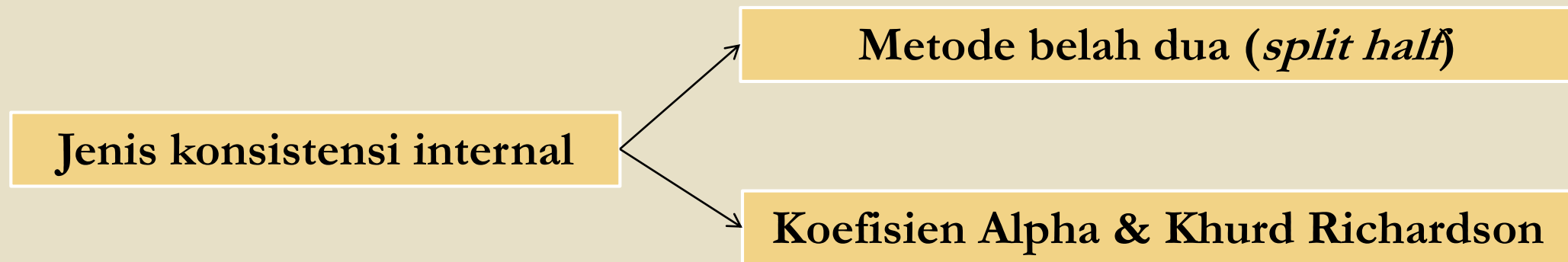
Kelemahan

- pekerjaan mengetes lebih berat karena harus menyusun dua seri tes
- waktu yang dibutuhkan lebih lama untuk mencobakan dua kali tes

Estimasi Reliabilitas Konsistensi Internal (*Internal Consistency*)

Reliabilitas konsistensi internal hanya memerlukan satu kali penyajian tes saja.

Estimasi reliabilitas dengan konsistensi internal berdasarkan pada hubungan antara item dalam tes dan berasal dari satu kali penyajian tes.



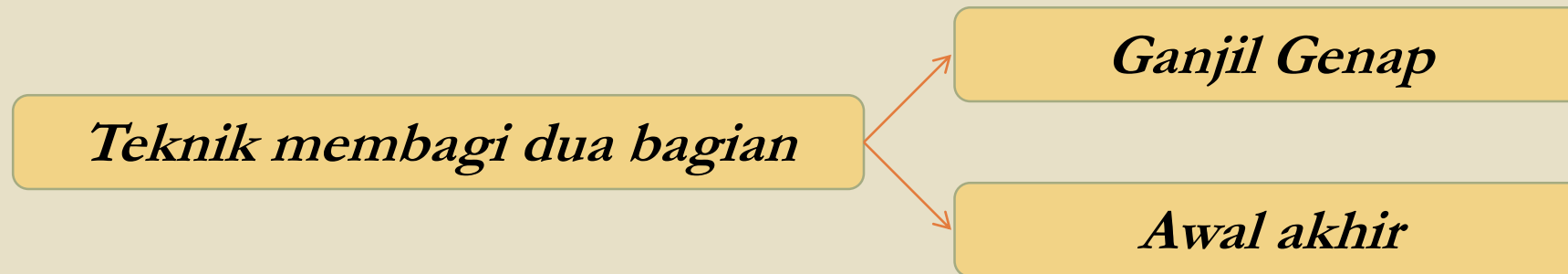
Metode Belah Dua (*Split Half*)

Komputasi koefisien reliabilitasnya dilakukan setelah keseluruhan instrumen yang telah dikenakan pada subjek itu dibelah menjadi beberapa bagian.

Metode belah dua dilakukan dengan membagi tes menjadi dua bagian yang ekuivalen. Hasil dari bagian pertama dari tes kemudian dikorelasikan dengan hasil pada bagian lainnya dari tes tersebut.

Korelasi yang digunakan adalah korelasi *product moment* dari Pearson.

Metode Belah Dua (*Split Half*)



Persyaratan yang harus dipenuhi pada metode ini:

- kedua belahan tes menghasilkan rata-rata skor (*mean*) yang setara
- varians skor yang tidak jauh berbeda satu sama lain
- materi yang diukur sama

Metode Belah Dua (*Split Half*)

Komputasi metode belah dua dilakukan dengan **formula Spearman Brown**.

Formula Spearman Brown merupakan sebuah formulasi komputasi untuk melakukan estimasi terhadap reliabilitas tes yang telah dibelah menjadi dua bagian yang relatif paralel satu dengan yang lain.

Formula ini dapat digunakan pada tes yang diberi skor dikotomi maupun bukan dikotomi.

Formula *Spearman Brown* (Allen dan Yen, 1979:79) adalah:

$$\rho_{xx'} = \frac{2\rho_{yy'}}{1 + \rho_{yy'}}$$

$\rho_{xx'}$ = koefisien reliabilitas Spearman Brown

$\rho_{yy'}$ = koefisien korelasi antara skor belahan Y dan skor belahan Y'

Koefisien Alpha dan Khurd Richardson (KR)

Pada pendekatan ini kekonsistenan dari semua item individu dilihat secara keseluruhan dalam tes, tidak membandingkan dengan cara membagi dua tes tersebut seperti metode belah dua.

Rumus koefisien alpha dikemukakan oleh Cronbach (1951) dan KR dikemukakan oleh Khurd Richardson (1937)

Koefisien Alpha

Koefisien Alpha (Cronbach, 1951) adalah bentuk yang berhubungan dengan item tes yang menghasilkan skor dengan beberapa nilai (misalnya, 0, 1, atau 2).

Rumus untuk menghitung koefisien Alpha

$$\alpha = \left(\frac{N}{N-1} \right) \left(\frac{\sigma_X^2 - \sum_{i=1}^N \sigma_{Yi}^2}{\sigma_X^2} \right)$$

Keterangan:

α = koefisien Alpha

N = jumlah item soal

σ_X^2 = varians skor tes total

$\sum_{i=1}^N \sigma_{Yi}^2$ = varians skor setiap item

Khurd Richardson-20 (KR-20)

KR-20 berlaku ketika item tes diskor secara dikotomi, yaitu, hanya benar atau salah sebagai 0 atau 1.

Rumus untuk menghitung koefisien KR-20

$$KR-20 = \left(\frac{N}{N-1} \right) \left(\frac{\sigma_X^2 - \sum_{i=1}^N p_i(1-p_i)}{\sigma_X^2} \right)$$

Keterangan:

KR-20 = koefisien KR-20

N = jumlah item soal

σ_X^2 = varians skor tes total

p_i = proporsi jawaban benar setiap soal

Khurd Richardson-21 (KR-21)

Kuder Richardson juga merumuskan formula estimasi reliabilitas 21 atau yang lebih dikenal dengan KR-21

Rumus untuk menghitung koefisien KR-21

$$KR-21 = \frac{N}{N-1} \left(\frac{\sigma_x^2 - N\bar{p}(1-\bar{p})}{\sigma_x^2} \right)$$

Formula KR-21 ini menghasilkan koefisien yang lebih kecil dibanding koefisien yang dihitung oleh formula KR-20. Hal ini selalu benar bila harga p , yang menyatakan taraf kesukaran item-item dalam tes yang bersangkutan sangat bervariasi. Dengan kata lain, formula KR-21 tidak sesuai untuk digunakan pada tes yang tingkat kesukaran item-itemnya tidak homogen atau varians item-itemnya tidak setara.

Estimasi Reliabilitas Antar Penilai (*inter rater*)

Tes tersebut dilaksanakan satu kali dan dua individu secara independen menilai setiap tes. Korelasi kemudian dihitung antara skor yang diperoleh dari dua penilai tersebut.

Selain pendekatan korelasional, kesepakatan antar-penilai juga dapat dievaluasi dengan menghitung persentase berapa banyak dua penilai menetapkan skor yang sama terhadap kinerja siswa

Koefisien reliabilitas menggunakan *inter-rater* pada dasarnya menggunakan teknik analisis varians.

Standard Error of Measurement

Standard error of measurement (S_E) adalah standar deviasi dari distribusi skor eror yang diperoleh dari seseorang yang melaksanakan beberapa tes (*parallel form*).

Dengan memperoleh beberapa skor dari beberapa tes tersebut, maka distribusi skor akan dihasilkan. Rata-rata dari distribusi skor ini sering kita sebut sebagai skor sebenarnya (*True Score*) dan S_E adalah standar deviasi dari distribusi skor eror ini.

S_E ini diestimasi menggunakan rumus berikut ini

$$S_E = s_x \sqrt{1 - r_{xx'}}$$

Keterangan: s_x = standar deviasi dari skor yang diperoleh
 $r_{xx'}$ = reliabilitas dari tes

Besar kecilnya S_E merupakan indikator kepercayaan pengukuran yang komparabel. Semakin kecil harga S_E berarti pengukuran tersebut semakin terpercaya dikarenakan variasi erornya semakin kecil

Standard Error of Measurement

Untuk mengestimasi skor murni dapat digunakan **interval kepercayaan** skor murni:

$$X - z_c s_e \leq T \leq X + z_c s_e$$

Keterangan:

X = skor yang diperoleh pada tes

Z_c = nilai kritis standar deviasi normal pada taraf kepercayaan yang dikehendaki

S_E = standar eror dalam pengukuran pada kelompok di mana subjek berada

Sempit luasnya interval bagi taraf kepercayaan tertentu tergantung pada besar kecilnya standar eror dalam pengukuran.

Semakin kecil standar eror dalam pengukuran, semakin sempit interval kepercayaan skor murni, yang berarti hasil tes semakin cermat.

Sebaliknya, semakin besar standar eror dalam pengukuran maka akan semakin lebar pula interval yang terjadi dan hasil pengukuran semakin tidak cermat.

Cara Meningkatkan Reliabilitas

Untuk meningkatkan besarnya reliabilitas adalah dengan meningkatkan jumlah item pada tes.

Rumus untuk meningkatkan reliabilitas :

$$r = \frac{n \times r_{xx}}{1 + (n - 1)r_{xx}}$$

Keterangan:

r = estimasi reliabilitas pada tes dengan *item* baru

n = faktor dari banyaknya tes yang ditingkatkan

r_{xx} = reliabilitas dari tes awal