

Reliabilitas Instrumen Penelitian

Oleh: Heri Retnawati
FMIPA Pend. Matematika UNY
Email: heri_retnawati@uny.ac.id

A. Estimasi Reliabilitas

Pada suatu instrumen yang digunakan untuk mengumpulkan data, reliabilitas skor hasil tes merupakan informasi yang diperlukan dalam pengembangan tes. Reliabilitas merupakan derajat keajegan (*consistency*) di antara dua skor hasil pengukuran pada objek yang sama, meskipun menggunakan alat pengukur yang berbeda dan skala yang berbeda (Mehrens & Lehmann, 1973; Reynold, Livingstone, & Wilson, 2010). Dalam kaitannya dengan penilaian pendidikan, prestasi atau kemampuan seorang siswa dikatakan reliabel jika dilakukan pengukuran, hasil pengukuran akan sama informasinya, walaupun penguji berbeda, korektornya berbeda atau butir soal yang berbeda tetapi memiliki karakteristik yang sama.

Allen & Yen (1979) menyatakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Selanjutnya dinyatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Dengan demikian, pengertian yang dapat diperoleh dari pernyataan tersebut adalah suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya.

Dalam pendidikan, pengukuran tidak dapat langsung dilakukan pada ciri atau karakter yang akan diukur. Ciri atau karakter ini bersifat abstrak, yang dapat diukur melalui suatu indikator. Hal ini menyebabkan sulitnya memperoleh alat ukur yang stabil untuk mengukur karakteristik seseorang (Mehrens & Lehmann, 1973). Kestabilan ini yang dikatakan sebagai reliabilitas. Untuk melihat reliabilitas suatu alat ukur, yang berupa suatu nilai, dapat dilakukan perhitungan statistik. Nilai ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*).

=====
Makalah disajikan pada kegiatan Workshop Teknik Analisis Data Fakultas Ekonomi dan
Bisnis IAIN Batusangkar di Rocky Hotel Bukittinggi, 25 Juli 2017.

Koefisien reliabilitas dapat diartikan sebagai koefisien keajegan atau kestabilan hasil pengukuran. Alat ukur yang reliabel akan memberikan hasil pengukuran yang stabil (Lawrence, 1994) dan konsisten (Mehrens & Lehmann, 1973). Artinya suatu alat ukur dikatakan memiliki koefisien reliabilitas tinggi manakala digunakan untuk mengukur hal yang sama pada waktu berbeda hasilnya sama atau mendekati sama. Dalam hal ini, reliabilitas merupakan sifat dari sekumpulan skor (Frisbie, 2005). Dalam kaitannya dengan dunia pendidikan, dengan alat ukur yang reliabel, hasil pengukuran akan sama informasinya walaupun penguji berbeda, korektornya berbeda atau butir soal yang berbeda tetapi mengukur hal yang sama dan memiliki karakteristik butir yang sama.

Allen & Yen (1979) menyatakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Selanjutnya dinyatakan bahwa koefisien reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Dengan demikian, pengertian yang dapat diperoleh dari pernyataan tersebut adalah suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya.

Reliabilitas (ρ) suatu tes pada umumnya diekspresikan secara numerik dalam bentuk koefisien yang besarnya $-1,00 \leq \rho \leq +1,00$. Koefisien tinggi menunjukkan reliabilitas tinggi. Sebaliknya, jika koefisien suatu skor tes rendah maka reliabilitas tes rendah. Jika suatu reliabilitas sempurna, berarti koefisien reliabilitas tersebut $+1,00$. Harapannya, koefisien reliabilitas bersifat positif.

Reliabilitas terkait pula dengan kesalahan pengukuran. Reliabilitas tinggi menunjukkan kesalahan yang kecil dalam memperoleh hasil pengukuran. Semakin besar reliabilitas suatu instrumen, akan semakin kecil kesalahan pengukuran, demikian pula sebaliknya, semakin kecil reliabilitas skor, akan semakin besar hasil pengukurannya. Kesalahan pengukuran dapat disebabkan oleh beberapa faktor, diantaranya karakteristik instrumen yang digunakan sendiri, misalnya penyusunan dan pelaksanaan pengukuran yang tidak mengikuti aturan baku, kualitas butir dalam instrumen tidak baik, adanya kerjasama selama melaksanakan tes atau mengisi instrumen, butir-butir instrumen yang meragukan, keadaan peserta selama merespons instrumen, seperti peserta yang sedang

lelah baik fisik maupun psikis, mempunyai problem pribadi, peserta yang mempunyai motivasi kurang, lingkungan tempat penyelenggaraan pengukuran yang kurang mendukung atau kombinasi dari segala permasalahan tersebut.

Mehrens & Lehmann (1973) menyatakan bahwa meskipun tidak ada perjanjian secara umum, tetapi secara luas dapat diterima bahwa untuk tes yang digunakan untuk membuat keputusan pada siswa secara perorangan harus memiliki koefisien reliabilitas minimal sebesar 0,85. Dengan demikian, pada penelitian ini, tes seleksi digunakan untuk menentukan keputusan pada siswa secara perorangan, sehingga indeks koefisien reliabilitasnya diharapkan minimal sebesar 0,85.

Proses penghitungan reliabilitas disebut dengan estimasi. Estimasi reliabilitas tes yang dapat dilakukan dengan beberapa cara, konsistensi eksternal, konsistensi internal, reliabilitas komposit, reliabilitas konstruk, reliabilitas interrater, dan estimasi reliabilitas dengan teori generalisabilitas (*Generalizability theory*).

B. Estimasi Konsistensi Eksternal

Estimasi reliabilitas eksternal diperoleh dengan menggunakan skor hasil pengukuran yang berbeda, baik dari instrumen yang berbeda maupun yang sama. Ada dua cara untuk mengestimasi reliabilitas eksternal suatu instrumen yaitu dengan teknik pengukuran ulang (*test-retest-method*) dan teknik paralel.

1. Metode Tes Ulang (*Test-Retest-Method*)

Untuk mengetahui keterandalan atau reliabilitas skor hasil pengukuran, pengukuran perlu dilakukan dua kali, pengukuran pertama dan pengukuran kedua atau ulangnya. Kedua pengukuran ini dapat dilakukan oleh orang yang sama atau berbeda, namun pada proses pengukuran yang kedua, keadaan yang diukur itu harus benar-benar berada pada kondisi yang sama dengan pengukuran pertama. Selanjutnya hasil pengukuran yang pertama dan yang kedua dikorelasikan dan hasilnya menunjukkan reliabilitas skor perangkat pengukuran.

Teknik *Test-Retest-Method* ini akan dapat sesuai dengan tujuannya jika keadaan subjek yang diukur tetap dan tidak mengalami perubahan pada saat pengukuran yang pertama maupun pada pengukuran yang kedua. Pada dasarnya keadaan responden itu

selalu berkembang, tidak statis ataupun berubah-ubah, maka sebenarnya teknik ini kurang tepat digunakan. Di samping itu pada pengukuran yang kedua akan terjadi adanya *carry-over-effect* atau *testing effect*, responden pengukuran atau penelitian telah mendapat tambahan pengetahuan karena sudah mengalami tes yang pertama ataupun belajar setelah pengukuran yang pertama.

Ada beberapa hal yang harus dipertimbangkan dalam mengestimasi koefisien dengan teknik tes-retes ini. Jangka waktu antara kedua pengukuran dengan menggunakan instrumen tersebut perlu menjadi pertimbangan. Jika jarak pengukuran terlalu dekat, maka *carry-over-effect* masih ada. Sementara jika jarak pengukuran terlalu jauh, korelasi kedua skor akan menjadi semakin rendah. Faktor kedua yang menjadi pertimbangan adalah stabilitas yang diharapkan dari kinerja yang diukur dengan instrumen tersebut. Semakin lama interval pelaksanaan pengukuran kedua instrumen, akan semakin rendah koefisien reliabilitasnya. Untuk mengatasi hal ini, jarak kedua pengukuran sebaiknya tidak terlalu jauh, misalnya tidak sampai satu bulan.

Estimasi reliabilitas dengan teknik tes-retes akan menghasilkan koefisien stabilitas. Untuk memperoleh koefisien reliabilitas melalui pendekatan tes-retes dapat dilakukan dengan menghitung koefisien korelasi linier antara skor pada pengukuran pertama (X) dengan skor hasil pengukuran kedua (Y).

$$r_i = \frac{N \sum XY - \sum X \sum Y}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}} \quad (1)$$

2. Metode Bentuk Paralel (*Equivalent*)

Teknik kedua untuk mengestimasi reliabilitas secara eksternal adalah dengan metode bentuk paralel. Pada teknik ini, diperlukan dua instrumen yang dikatakan paralel untuk mengestimasi koefisien reliabilitas. Dua buah tes dikatakan paralel atau *equivalent* adalah dua buah instrumen yang mempunyai kesamaan tujuan dalam pengukuran, tingkat kesukaran dan susunan juga sama, namun butir-butir soalnya berbeda, atau dikenal dengan istilah *alternate-forms method* atau *parallel forms*.

Dengan metode bentuk paralel ini, dua buah instrumen yang paralel, misalnya instrumen paket A yang akan diestimasi reliabilitasnya dan instrumen paket B merupakan instrumen yang paralel dengan paket A, keduanya diberikan kepada sekelompok

responden yang sama, kemudian kedua skor tersebut dikorelasikan. Koefisien korelasi dari kedua skor respon responden terhadap instrumen inilah yang menunjukkan koefisien reliabilitas skor instrumen paket A. Jika koefisiennya reliabilitas skor instrumen tinggi, maka perangkat tersebut sudah dapat dikatakan reliabel dan dapat digunakan sebagai instrumen pengukur suatu konstruk yang terandalkan.

Dari sisi responden, estimasi reliabilitas dengan teknik ini ada kelemahannya. Dalam menggunakan teknik ini, diperlukan dua buah instrumen, dan masing-masing diberikan kepada sekelompok responden yang sama. Teknik ini responden tidak mengalami *practice-effect* dan *carry-over-effect* atau responden tidak mengingat pengerjaan instrumen sebelumnya.

Estimasi reliabilitas dengan cara ini merupakan pekerjaan yang cukup berat. Pada cara ini, diperlukan dua instrumen untuk digunakan, sehingga harus mengembangkan 2 instrumen dan juga mengujicobakan 2 instrumen. Membuktikan kedua instrumen tersebut merupakan tes yang paralel atau ekivalen memerlukan ilmu yang tersendiri (konsep penyetaraan tes atau *equating*).

Langkah-langkah yang ditempuh pada pembuktian reliabilitas dengan cara ini adalah: (1) menyiapkan dua instrumen yang paralel, (2) menentukan subjek untuk mengujicobakan instrumen, (3) melaksanakan pengukuran dengan mengadministrasikan instrumen tersebut, (4) melakukan penyekoran pada setiap jawaban responden terhadap kedua perangkat tersebut, (5) menghitung koefisien korelasi dari skor kedua perangkat tersebut.

Hasil koefisien korelasi yang tinggi dari skor jawaban responden pada kedua instrumen yang digunakan menunjukkan bahwa reliabilitas paralel dari perangkat tersebut berada pada kategori yang baik. Namun sebaliknya, jika ternyata koefesien korelasinya rendah, maka reliabilitas skor perangkat ekivalen adalah rendah.

C. Konsistensi Internal

Dengan teknik konsistensi internal ini, hanya dengan melakukan satu kali pengumpulan data, reliabilitas skor perangkat pengukuran dapat diestimasi. Pada pembuktian instrumen dengan cara ini ada beberapa cara, yang masing-masing dapat

memerlukan persyaratan-persyaratan atau asumsi tertentu yang harus dipenuhi oleh peneliti. Beberapa cara yang dapat digunakan untuk mengestimasi reliabilitas dengan konsistensi internal diantaranya sebagai berikut.

1. Metode Belah Dua (*Split Half Method*)

Dalam teknik belah dua ini, dalam satu instrumen dikerjakan satu kali oleh sejumlah subjek (*sample*) suatu penelitian. Butir-butir pada perangkat dibagi menjadi dua. Pembagian dapat menggunakan nomor ganjil-genap pada instrumen, atau separuh pertama maupun separuh kedua, maupun membelah dengan menggunakan nomor acak atau tanpa pola tertentu. Skor responden merespons setengah perangkat bagian yang pertama dikorelasikan dengan skor setengah perangkat pada bagian yang kedua. Teknik ini berpegang pada asumsi, belahan pertama dan belahan kedua mengukur konstruk yang sama, banyaknya butir dalam instrumen belahan pertama dan kedua harus dapat dibandingkan dari sisi banyaknya butir, atau paling tidak jumlahnya hampir sama.

Ada beberapa formula untuk mengestimasi reliabilitas dengan metode belah dua, antara lain rumus Spearman-Brown, rumus Flanagan, dan rumus Rulon. Masing formula disajikan berikut ini.

a. Reliabilitas dengan Rumus Spearman-Brown

Adapun rumus Spearman-Brown yang digunakan adalah :

$$r_i = \frac{2r_b}{1+r_b} \quad (2)$$

$$\text{Dengan } r_b = \frac{N \sum XY - \sum X \sum Y}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Dengan r_i = koefisien reliabilitas skor instrumen; r_b = koefisien korelasi antara dua belahan instrumen, N = banyaknya responden, X = belahan pertama, Y = belahan kedua.

b. Reliabilitas dengan Rumus Flanagan

Untuk mengestimasi reliabilitas dengan rumus Rulon, peneliti perlu menghitung kovarians dari skor belahan pertama dan skor belahan kedua (σ_{12}) dan varians totalnya. Koefisien reliabilitas disajikan dengan formula sebagai berikut.

$$r_i = \frac{2\sigma_{12}}{\sigma_X^2} = \frac{4r_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + 2r_{12}\sigma_{12}} \quad (\text{Walker, 2006}) \quad (3)$$

Dengan r_i = koefisien reliabilitas skor instrumen; r_{12} = koefisien korelasi antara dua belahan instrumen, σ_1^2 = varians belahan pertama, σ_2^2 = varians belahan kedua, σ_X^2 = varians skor total.

c. Reliabilitas dengan Rumus Rulon

Rulon merumuskan suatu formula untuk mengestimasi reliabilitas belah dua tanpa perlu berasumsi bahwa kedua belahan mempunyai varians yang sama. Menurut Rulon, perbedaan skor subjek pada kedua belahan instrumen akan membentuk distribusi perbedaan skor dengan varians yang besarnya ditentukan oleh varians *error* masing-masing belahan menentukan varians *error* keseluruhan instrumen, maka varians eror istrumen ini dapat diestimasi lewat besarnya varians perbedaan skor diantara kedua belahan. Dalam melakukan estimasi reliabilitas skor instrumen, varians perbedaan skor diperhitungkan sebagai sumber *error*. Untuk melakukan estimasi reliabilitas instrumen dengan rumus Rulon, peneliti juga harus menghitung dahulu varians selisih belahan pertama dan kedua dan juga varians total. Formula Rulon untuk mengestimasi reliabilitas sebagai berikut.

$$r_i = 1 - \frac{\sigma_d^2}{\sigma_t^2} \quad (4)$$

Dengan r_i = reliabilitas instrumen; σ_t^2 = varians total atau varians skor total; σ_d^2 = varians dari perbedaan skor kedua belahan (*varians difference*); d = skor pada belahan awal dikurangi skor pada belahan akhir.

C. Reliabilitas komposit

Pada suatu instrumen, sering peneliti menggunakan instrumen yang terdiri dari banyak butir. Jika butir-butir ini merupakan butir yang berbeda-beda namun membangun suatu konstruk yang sama, maka analisis untuk mengestimasi reliabilitas dapat digunakan rumus reliabilitas komposit. Komposit yang dimaksudkan yakni skor akhir merupakan

gabungan dari skor butir-butir penyusun instrumen. Ada 3 formula yang dapat digunakan untuk mengestimasi reliabilitas dengan cara ini, yaitu dengan menghitung koefisien α dari Cronbach, koefisien KR-20, dan koefisien KR-21.

1. Rumus Alpha dari Cronbach

Rumus Alpha digunakan untuk mengestimasi reliabilitas instrumen yang skornya bukan hanya 1 dan 0, namun juga skala politomus, misal misalnya angket (skala Likert 1-2-3-4-5) atau soal bentuk uraian (skor maksimum dapat tergantung peneliti). Rumus Alpha sebagai berikut.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right) \tag{5}$$

Dengan α = koefisien reliabilitas instrumen; k = banyaknya butir pertanyaan dalam instrumen; $\sum \sigma_i^2$ = jumlah varians butir instrumen; σ_t^2 = varians skor total.

2. Rumus Kuder-Richardson

Cara lain mengestimasi reliabilitas dengan reliabilitas komposit adalah dengan menggunakan formula Kuder dan Richardson yang disingkat dengan formula KR. Ada dua jenis formula KR, yaitu Kuder Richardson formula 20 (KR-20) dan Kuder Richardson formula 21 (KR-21).

Formula KR-20 dapat digunakan untuk analisis butir dikotomi. Pada butir instrumen dengan penskoran dikotomi, misal 1-0, benar-salah, ya-tidak, hidup-mati, dan lain-lain, estimasi reliabilitas dapat dilakukan dengan menggunakan rumus KR-20.

Rumus KR-20 sebagai berikut:

$$r_{ii} = \frac{k}{(k-1)} \left\{ \frac{s_t^2 - \sum p_i q_i}{s_t^2} \right\} \tag{6}$$

Dengan r_{ii} = reliabilitas skor instrumen; k = banyaknya butir pertanyaan atau banyaknya soal; s_t^2 = varians skor total; p_i = proporsi subjek yang menjawab betul pada suatu butir (proporsi subjek yang mendapat skor 1) yang dihitung dengan

$$p_i = \frac{\text{banyaknya subjek yang skornya 1}}{N}; \text{ dan } q_i = 1 - p_i$$

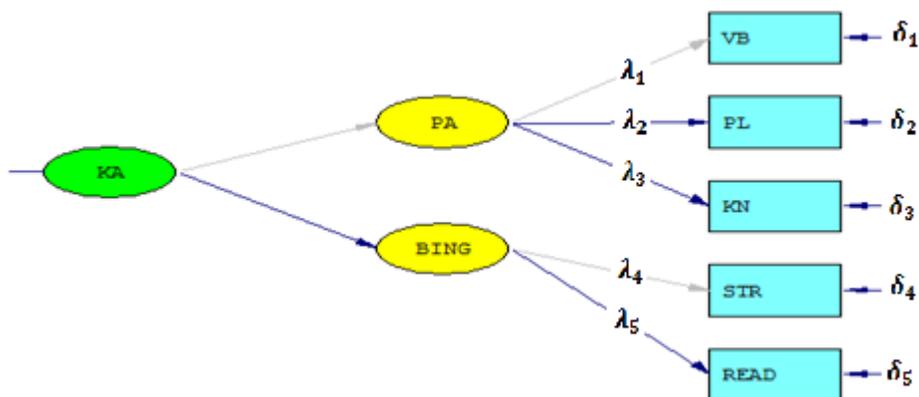
Rumus KR-21 dapat digunakan untuk instrumen dengan skornya tiap butirnya 1 dan 0, dan juga skala politomus, misal misalnya angket (skala Likert 1-2-3-4-5) atau soal bentuk uraian. Formula KR-21 sebagai berikut.

$$r_{ii} = \frac{k}{k-1} \left(1 - \frac{\bar{X}(k-\bar{X})}{k\sigma_t^2} \right) \quad (7)$$

Dengan r_{ii} = koefisien reliabilitas skor instrumen; k = banyaknya butir pertanyaan atau banyaknya soal; σ_t^2 = varians total; \bar{X} = skor rata-rata (Allen & Yen, 1979).

D. Reliabilitas Konstruk

Cara keempat untuk mengestimasi reliabilitas adalah dengan mengestimasi reliabilitas konstruk. Reliabilitas konstruk ini dapat diestimasi setelah peneliti membuktikan validitas konstruk dengan analisis faktor konfirmatori sampai memperoleh model yang cocok (model yang fit). Dengan analisis faktor ini, peneliti dapat memperoleh muatan faktor (*factor loading*) tiap indikator yang menyusun instrumen (λ) dan indeks kesalahan unik dari tiap indikator (δ). Sebagai contoh pada pembuktian validitas konstruk kemampuan akademik, diperoleh model yang fit yang disajikan pada Gambar 6.1.



Gambar 6.1. Hasil Analisis Faktor Konfirmatori

Estimasi reliabilitas dapat dilakukan dengan tiga cara, estimasi reliabilitas dengan reliabilitas konstruk (*construct reliability, CR*), reliabilitas ω , dan reliabilitas maksimal (Ω).

Estimasi CR menggunakan muatan faktor (*factor loading*) tiap indikator yang menyusun instrumen (λ) dan indeks kesalahan unik dari tiap indikator (δ). Formulasnya sebagai berikut (Geldhof, Preacher, Zyphur, 2014).

$$CR = \frac{\left(\sum_{i=1}^i \lambda_i \right)^2}{\left(\sum_{i=1}^i \lambda_i \right)^2 + \left(\sum_{i=1}^i \delta \right)} \quad (8)$$

Estimasi dengan reliabilitas ω dilakukan hanya dengan menggunakan muatan faktor (λ) saja. Formula untuk estimasi reliabilitas ω sebagai berikut (Kamata, Turhan, Darandari, 2003).

$$\omega = \frac{\left(\sum_{i=1}^i \lambda_i \right)^2}{\left(\sum_{i=1}^i \lambda_i \right)^2 + \left(\sum_{i=1}^i 1 - \lambda_i^2 \right)} \quad (9)$$

Pada estimasi reliabilitas maksimal, muatan faktor disimbulkan dengan l . Formulasnya sebagai berikut (Penev & Raykov, 2006).

$$\Omega_w = \frac{\sum_{i=1}^p \frac{l_i^2}{(1 - l_i^2)}}{1 + \sum_{i=1}^p \frac{l_i^2}{(1 - l_i^2)}} \quad (10)$$

E. Reliabilitas Inter-rater

Jika dalam suatu instrumen penskoran butir dilakukan dengan memanfaatkan dua orang rater, peneliti dapat mengestimasi reliabilitas dengan *inter-rater agreement*. Hasil estimasi reliabilitas dengan cara ini disebut dengan reliabilitas *inter-rater*. Adapun cara mengestimasi dengan menghitung terlebih dahulu banyaknya butir atau kasus yang cocok atau butir atau kasus yang diskor sama oleh kedua rater. Banyaknya butir yang cocok ini kemudian dibandingkan dengan butir total, kemudian disajikan dalam

persentase. Estimasi reliabilitas skor dengan inter-rater dapat disajikan dengan formula sebagai berikut.

$$\text{inter-rater agreement} = \frac{\text{Banyaknya kasus yang diskor sama oleh kedua rater}}{\text{Banyaknya kasus}} \times 100 \quad (11)$$

Cara tersebut mudah dilakukan untuk penyekoran dengan skala yang mudah, misalnya 1-5 saja, itupun hasil penskoran berupa bilangan bulat. Namun jika hasil penskoran berada pada rentang yang panjang, misalnya 1-100, rumus tersebut akan menghasilkan koefisien kesepakatan interater yang kecil. Cara lain yang dapat dilakukan adalah dengan membuat urutan skor setiap rater dari yang rendah ke yang tinggi. Persentase banyaknya kasus yang sama peringkatnya dari kedua rater merupakan koefisien kesepakatan rater. Formulasnya sebagai berikut.

$$\text{inter-rater agreement} = \frac{\text{Banyaknya kasus yang peringkatnya sama oleh kedua rater}}{\text{Banyaknya kasus}} \times 100 \quad (12)$$

F. Reliabilitas dengan Teori Generalizabilitas

Teori generalisabilitas (*Generalizability Theory*) terkait dengan 2 hal, generalizability (G) study dan decision (D) study. Peneliti yang melakukan *G-Study* mengutamakan generalisasi dari suatu sampel pengukuran ke keseluruhan pengukuran. Studi tentang stabilitas respons antarwaktu, equivalensi skor dari 2 atau lebih instrumen yang berbeda, hubungan antara skor sub-kemampuan dengan skor butir terkait dengan *G-study*. Pada *D-study*, data dikumpulkan untuk tujuan khusus terkait dengan membuat keputusan. Studi ini menyediakan data mendeskripsikan peserta tes, baik seleksi atau penempatan, maupun menyelidiki hubungan 2 variabel atau lebih (Crocker & Algina, 2008). Sebagai contoh, pada suatu tes seleksi, panitia akan menggunakan dua penilai atau lebih perlu diperiksa terlebih dahulu efisiensinya. Untuk hal tersebut, perlu dilakukan *D-study*.

Koefisien reliabilitas dalam teori ini disebut dengan koefisien *generalizability*. Dalam mengestimasi koefisien *generalizability*, ada beberapa desain, termasuk banyaknya bentuk tes, kesempatan melakukan tes atau administrasi tes, banyaknya rater,

yang sering disebut dengan facet. Banyaknya variabel yang digunakan menentukan banyaknya facet. Desain yang dapat dipilih misalnya desain facet tunggal (*single facet design*) dan facet ganda.

1. Desain Facet Tunggal

Desain faet tunggal terdiri dari 4 desain, yakni 1) setiap peserta tes atau jawaban peserta tes dinilai oleh satu rater, dan rater ini menilai semua peserta tes, 2) setiap peserta dinilai oleh beberapa rater, dan semua rater menilai peserta tes, 3) setiap peserta tes dinilai oleh rater yang berbeda, hanya satu rater untuk setiap peserta tes, dan 4) setiap penilaia dinilai oleh beberapa rater, ada rater yang berbeda-beda untuk setipa peserta tes (Crocker & Algina, 2008).

$$\rho_i^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} \quad (13)$$

Pelaksanaan mengestimasi dilakukan dengan bantuan analisis varians (ANAVA).

Misalkan MS adalah mean square pada ANAVA, dengan sumber variasi peserta tes (persons, P), dan rater (R).Rangkuman tabel ANAVA disajikan pada Tabel 6.1.

Tabel 6.1. Rangkuman Tabel Anava

SV	SS	df	MS	EMS
Examinee (P)	$n_i \sum_p (X_{pi} - X_{PI})^2$	$n_p - 1$	$\frac{SS_p}{n_p - 1}$	$\sigma_e^2 + n_i \sigma_p^2$
Raters (I)	$n_p \sum_i (X_{pi} - X_{PI})^2$	$n_i - 1$	$\frac{SS_i}{n_i - 1}$	$\sigma_e^2 + n_p \sigma_i^2$
Residual (R)	$\sum_i \sum_j (X_{pi} - X_{PI})^2 - SS_p - SS_i$	$(n_p - 1)(n_i - 1)$	$\frac{SS_r}{(n_p - 1)(n_i - 1)}$	σ_e^2
	$X_{PI} = \sum_i \frac{X_{pi}}{n_i}$	$X_{pi} = \sum_p \frac{X_{pi}}{n_p}$	$X_{PI} = \sum_i \sum_p \frac{X_{pi}}{n_i n_p}$	

EMS_p dapat disubstitusikan dari

$\sigma_e^2 + n_p \sigma_i^2$ dan EMS_r yaitu

σ_e^2 untuk memperoleh

$n_i \sigma_p^2 = (EMS_p - EMS_r)$ sehingga $\sigma_p^2 = \frac{(EMS_p - EMS_r)}{n_i}$ dan juga

$$\hat{\sigma}_p^2 = \frac{(MS_p - MS_r)}{n_i}$$

Sehingga diperoleh

$$\hat{\rho}_i^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2}$$

$$\hat{\rho}_i^2 = \frac{MS_p - MS_r}{MS_p + (n_i - 1)MS_r} \quad (14)$$

yang merupakan formula koefisien generalizability untuk desain facet tunggal yang pertama.

Pada desain facet tunggal yang kedua, banyaknya rater ditingkatkan menjadi n_i' untuk mengetahui banyaknya rater yang sesuai pada D-study. Koefisien generalizability diestimasi dengan

$\hat{\rho}_{i*}^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2/n_i'}$ yang nilainya diestimasi dengan rumus

$$\hat{\rho}_{i*}^2 = \frac{MS_p - MS_r}{MS_p + (n_i - n_i')MS_r/n_i'} \quad (15)$$

Desain facet tunggal yang ketiga, setiap peserta tes dinilai oleh rater yang berbeda, hanya satu rater untuk setiap peserta tes.

$\hat{\rho}_i^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_i^2 + \hat{\sigma}_e^2}$ yang nilainya diestimasi dengan rumus

$$\hat{\rho}_i^2 = \frac{MS_p - MS_r}{MS_p + n_i MS_i/n_p + (n_i n_p - n_i - n_p)MS_r/n_p} \quad (16)$$

Desain facet tunggal yang keempat, setiap penilaian dinilai oleh beberapa rater, ada rater yang berbeda-beda untuk setiap peserta tes.

$$\hat{\rho}_i^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + (\hat{\sigma}_i^2 + \hat{\sigma}_e^2)/n'_i}$$

$$\hat{\rho}_i^2 = \frac{MS_p - MS_r}{MS_p + n_i MS_i / n_p n'_i + (n_i n_p - n_p n'_i - n_i) MS_e / n_p n'_i} \quad (17)$$

2. Generalisabilitas untuk Desain Facet Ganda

Seperti pada ANAVA, pada teori ini dikenal dengan istilah tersilang (*crosses*) dan istilah tersarang (*nested*). Tersilang jika setiap kondisi pengukuran pada *facet* pertama terjadi dalam kombinasi dengan setiap pengukuran pada faktor yang kedua. Suatu *facet* dikatakan tersarang dalam *facet* kedua jika himpunan yang berbeda dari kondisi pengukuran pada *facet* pertama terjadi dalam kombinasi dengan setiap kondisi pengukuran pada *facet* yang kedua.

Untuk mengestimasi koefisien generalisabilitas (*generalizability coefficient*), ada beberapa varians skor yang digunakan. Pada desain dua *facet*, melibatkan varians peserta tes, varians kondisi *facet* I, varians kondisi *facet* J, varians interaksi, dan varians residual. Mengestimasi koefisien ini dapat dilakukan dengan menggunakan bantuan analisis varians (ANAVA) yang disajikan pada Tabel 6.2, dengan sumber varians *person* (p), butir (*question*, q), dan *rater* (r).

Tabel 6.2 Rangkuman ANAVA 3 jalur untuk generalizability 2 facet

Sumber Varians	Derajat Kebebasan	Sum Square (SS)	Mean Square (MS)	F	Sig
p (person)	p-1	SS _p	MS _p	.	.
q	q-1	SS _q	MS _q	.	.
R	r-1	SS _r	MS _r	.	.
pq	pq-1	SS _{pq}	MS _{pq}	.	.
pr	pr-1	SS _{pr}	MS _{pr}	.	.
qr	qr-1	SS _{qr}	MS _{qr}	.	.
pqr	pqr-1	SS _{pqr}	MS _{pqr}	.	.

Dari tabel rangkuman ANAVA tersebut, kolom MS yang digunakan untuk mengestimasi reliabilitas 2 *facet*. Adapun varians-variens yang digunakan menghitung reliabilitas yaitu varians peserta tes, varians kondisi *facet* I (*question*, q), varians kondisi *facet* J (*rater*, r),

varians interaksi, dan varians residual. Masing-masing disajikan sebagai berikut (Lord, 1973).

$$\begin{aligned}
 \sigma_{pqr}^2 &= MS_{pqr} \\
 \sigma_{pq}^2 &= \frac{1}{n_r} (MS_{pq} - MS_{pqr}) \\
 \sigma_{pr}^2 &= \frac{1}{n_q} (MS_{pr} - MS_{pqr}) \\
 \sigma_{qr}^2 &= \frac{1}{n_p} (MS_{qr} - MS_{pqr}) \\
 \sigma_p^2 &= \frac{1}{n_q n_r} (MS_p - MS_{pq} - MS_{pr} + MS_{pqr}) \\
 \sigma_q^2 &= \frac{1}{n_p n_r} (MS_q - MS_{pq} - MS_{qr} + MS_{pqr}) \\
 \sigma_r^2 &= \frac{1}{n_p n_q} (MS_r - MS_{pr} - MS_{qr} + MS_{pqr})
 \end{aligned} \tag{17}$$

Selanjutnya koefisien generalizability diestimasi dengan formula

$$r_{xx'} = \frac{\sigma_{true}^2}{\sigma_{obs}^2} \tag{18}$$

Komponen varians skor murni dan varians skor amatan dimaksud dapat diuraikan sebagai berikut:

$$\frac{\sigma_{true}^2}{\sigma_{obs}^2} = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_q^2}{n_q} + \frac{\sigma_r^2}{n_r} + \frac{\sigma_{pq}^2}{n_q} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{qr}^2}{n_q n_r} + \frac{\sigma_{pqr}^2}{n_q n_r}} \tag{19}$$

B. Kesalahan Pengukuran Standar (Standard Error of Measurement, SEM)

Kesalahan Baku Pengukuran (*Standard Error of Measurement, SEM*) dapat digunakan untuk memahami kesalahan yang bersifat acak/random yang mempengaruhi skor responden dalam merespons instrumen. Kesalahan pengukuran, yang disimbolkan dengan σ_E , dapat dihitung dengan rumus pada persamaan 5, yang diturunkan dari rumus reliabilitas (Allen & Yen, 1979).

$$\sigma_E = \sigma_x \sqrt{1 - \rho_{xx'}} \dots\dots\dots(20)$$

Penafsiran SEM dilakukan karena tidak adanya prosedur penilaian yang sangat konsisten, interpretasi skor dapat ditingkatkan dengan mempertimbangkan ukuran kemungkinan kesalahan pengukuran.

Interpretasi dari SEM digunakan untuk memprediksikan rentang skor sebenarnya (*true score*) yang diperoleh responden. Skor sebenarnya (*true score*, τ) dari hasil pengukuran $X \pm SEM$, atau dengan simbol matematika sebagai berikut:

$$X - SEM < \tau < X + SEM \quad (21)$$

C. Faktor-Faktor yang mempengaruhi Reliabilitas

Ada beberapa faktor yang mempengaruhi reliabilitas. Faktor tersebut dapat dikategorikan faktor-faktor yang mempengaruhi secara langsung dan secara tidak langsung. Faktor-faktor tersebut yaitu:

- 1) Panjang tes dan kualitas butir-butir instrumen. Instrumen yang terdiri dari banyak butir, tentu lebih reliabel dibandingkan dengan instrumen yang hanya terdiri dari beberapa butir. Jika panjang instrumen ditambah dengan menambah butir-butir yang baik maka semakin panjang suatu instrumen maka reliabilitas skornya semakin tinggi. Namun jika instrumen terlalu panjang, maka responden akan terlalu lelah mengerjakannya. Faktor kelelahan responden ini yang akan menurunkan reliabilitas.
 - 2) Kondisi penyelenggaraan pengumpulan data atau administrasi.
 - a. Sebagai contoh pada pelaksanaan tes, petunjuk yang diberikan sebelum tes dimulai dan petunjuk ini disajikan dengan jelas, penyelenggaraan tes akan berjalan lancar dan tidak akan banyak terdapat pertanyaan atau komentar dari responden. Hal ini akan menjamin pelaksanaan tes yang tertib dan tenang sehingga skor yang diperoleh lebih reliabel.
 - b. Pengawas yang tertib akan mempengaruhi skor hasil perolehan responden. Pengawasan yang terlalu ketat ketika pengumpulan data menyebabkan responden merasa kurang nyaman atau merasa takut dan tidak dapat dengan leluasa dalam
-

merespon instrumen, namun jika pengawasan kurang, maka peserta akan bekerjasama sehingga hasil pengumpulan data kurang dapat dipercaya.

- c. Suasana lingkungan dan tempat pengumpulan data (tempat duduk yang tidak teratur, suasana disekelilingnya gaduh atau tidak tenang, dan sebagainya) akan mempengaruhi reliabilitas. Sebagai contoh pada pelaksanaan tes, suasana yang panas dan dekat sumber kegaduhan akan mempengaruhi hasil tes.

Adapun faktor-faktor yang mempengaruhi secara langsung hasil estimasi reliabilitas adalah

- a. waktu penyelenggaraan pengumpulan data pertama dan kedua. Faktor ini terutama pada estimasi reliabilitas dengan menggunakan teknik tes-retes. Interval penyelenggaraan yang terlalu dekat atau terlalu jauh, akan mempengaruhi koefisien reliabilitas.
- b. Panjang instrumen, semakin panjang suatu instrumen pengumpul data, semakin banyak butir yang termuat di dalamnya. Hal ini akan memberikan dampak hasil pengumpulan data akan semakin mendekati keadaan yang sebenarnya, yang akan mempertinggi koefisien reliabilitas.

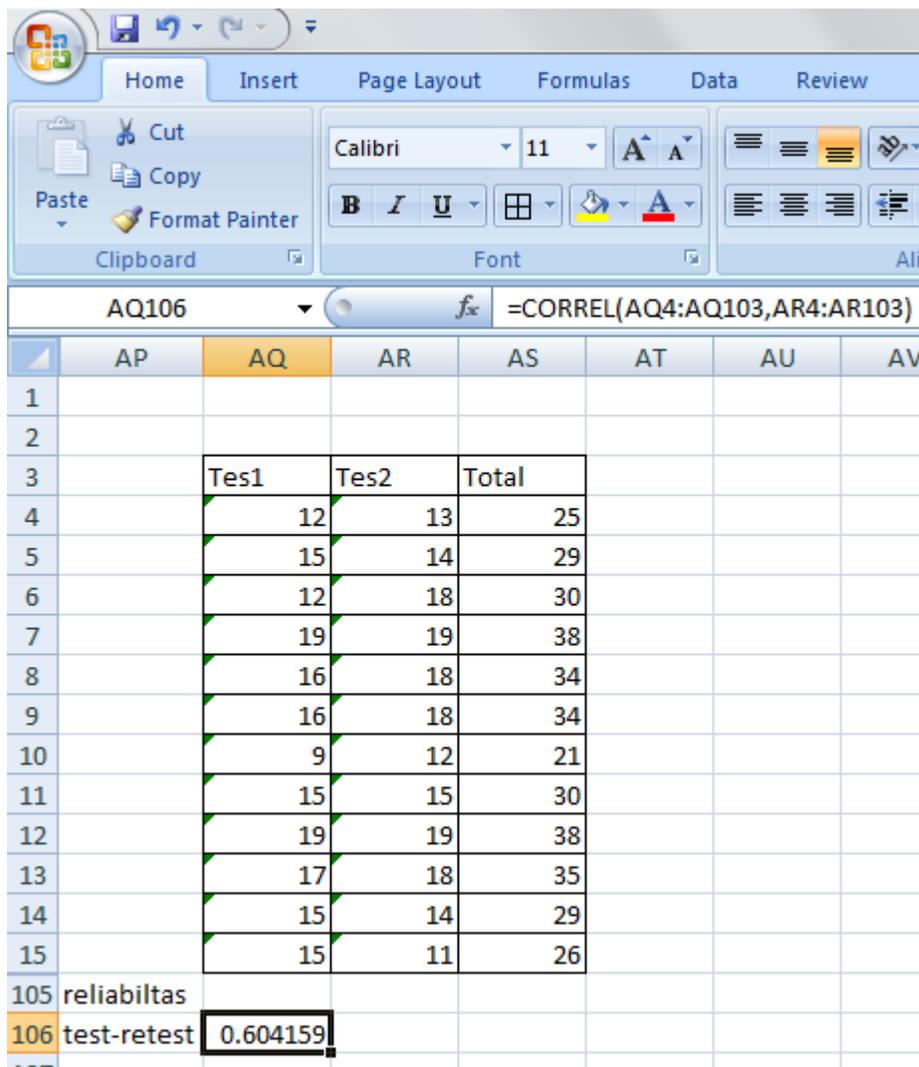
$$r_n = \frac{nr}{(n-1)r+1} \quad (22)$$

- c. Penyebaran skor perolehan responden. Koefisien reliabilitas secara langsung dipengaruhi oleh bentuk sebaran skor (variansi) dalam kelompok responden yang diukur. Semakin tinggi varians skor hasil pengukuran, semakin tinggi estimasi koefisien reliabilitas.
 - d. Tingkat kesulitan butir instrumen. Butir yang terlalu mudah dan butir terlalu sulit tidak memberikan tambahan variansi sebaran skor hasil pengukuran, sehingga akan mempengaruhi reliabilitas.
 - e. Objektivitas penskoran. Objektivitas penskoran terhadap respons responden terhadap instrumen akan mempengaruhi reliabilitas. Semakin objektif penskoran suatu instrumen, maka skor perolehannya akan menjadi semakin reliabel.
-

E. Mengestimasi Reliabilitas

1. Tes-retes dan paralel

Untuk mengestimasi reliabilitas dengan formula ini, diperlukan 2 skor tes, baik untuk teknik tes-retes (tes ulang) maupun tes yang paralel. Skor kedua tes tersebut dikorelasikan dan hasil perhitungan korelasi tersebut merupakan koefisien reliabilitas tes ulang atau koefisien reliabilitas bentuk paralel. Teknik ini akan sangat mudah dilakukan dengan menggunakan program excel dengan menggunakan rumus korelasi **=CORREL**.



The screenshot shows the Microsoft Excel interface. The formula bar displays the formula `=CORREL(AQ4:AQ103,AR4:AR103)`. The spreadsheet contains the following data:

	AP	AQ	AR	AS	AT	AU	AV
1							
2							
3		Tes1	Tes2	Total			
4		12	13	25			
5		15	14	29			
6		12	18	30			
7		19	19	38			
8		16	18	34			
9		16	18	34			
10		9	12	21			
11		15	15	30			
12		19	19	38			
13		17	18	35			
14		15	14	29			
15		15	11	26			
105	reliabilitas						
106	test-retest	0.604159					

2. Belah dua

Pada estimasi dengan teknik belah dua ini, dipilih dulu skor belahan pertama dan skor belahan kedua, misalnya pada contoh dalam buku ini, skor belahan ganjil dan skor belahan genap. Setelah dikorelasikan keduanya dengan fungsi (=CORREL(ARRAY1,ARRAI2)), reliabilitas dengan rumus belah dua atau dikenal juga dengan rumus Spearman-Brown dapat dihitung menggunakan rumus (6.2).

AZ107		f _x = (2*AZ106)/(1+AZ106)			
	AY	AZ	BA	BB	BC
1					
2					
3		ganjil (X)	genap (Y)		
4		10	15		
5		15	14		
6		14	16		
7		19	19		
8		16	18		
9		16	18		
10		10	11		
11		14	16		
12		18	20		
13		18	17		
14		15	14		
15		9	17		
105					
106	korelasi XY	0.697159			
107	reliabilitas SB	0.82156			
108					

3. Rulon

Pada estimasi reliabilitas dengan rumus Rulon, skor butir ganjil dan genap tetap dihitung, berikut nilai d_i = skor butir ganjil – skor butir genap (boleh sebaliknya asal konsisten). Variansi d_i keudian dihitung demikian pula variansi total (skor butir ganjil + skor butir genap). Hasil-hasil ini digunakan untuk mengestimasi reliabilitas dengan menggunakan rumus Rulon.

latihan reliabilitas pmat intel - Microsoft Excel non-c...

Home Insert Page Layout Formulas Data Review View

Clipboard Font Alignment Number Styles Cells

B104 fx =1-D102/E102

	A	B	C	D	E	F
1	Responden	Ganjil (X)	Genap (Y)	d	Total	
2	1	10	15	-5	25	
3	2	15	14	1	29	
4	3	14	16	-2	30	
5	4	19	19	0	38	
6	5	16	18	-2	34	
7	6	16	18	-2	34	
8	7	10	11	-1	21	
98	97	15	13	2	28	
99	98	13	15	-2	28	
100	99	12	11	1	23	
101	100	20	18	2	38	
102	Var			5.601111	27.74495	
103						
104	reliabilitas (Rulon)	0.7981				
105						

4. Alpha

Pada estimasi reliabilitas dengan rumus allfa dari Cronbach, diestimasi dulu varians tiap butir dan varians total. Koefisien reliabilitas selanjutnya dapat dihitung.

B110		fx					=(40/39)*(1-B107/B108)	
	A	B	C	D	AO	AP	AQ	
1								
2								
3	Responden	Butir 1	Butir 2	Butir 3	Butir 40	Total		
4	1	0	1	1	0	25		
5	2	1	1	1	0	29		
6	3	1	1	1	1	30		
7	4	1	1	1	1	38		
8	5	1	1	1	1	34		
9	6	1	1	1	1	34		
102	99	1	1	1	1	23		
103	100	1	1	1	1	38		
104								
105	vari	0.090909	0.173333	0.074343	0.065758	27.74495		
106								
107	Sigma Si^2	6.14798 (jumlah varians butir)						
108	Sx^2	27.74495 (variens total)						
109								
110	alfa	0.79837						
111								
112								

5. KR-20

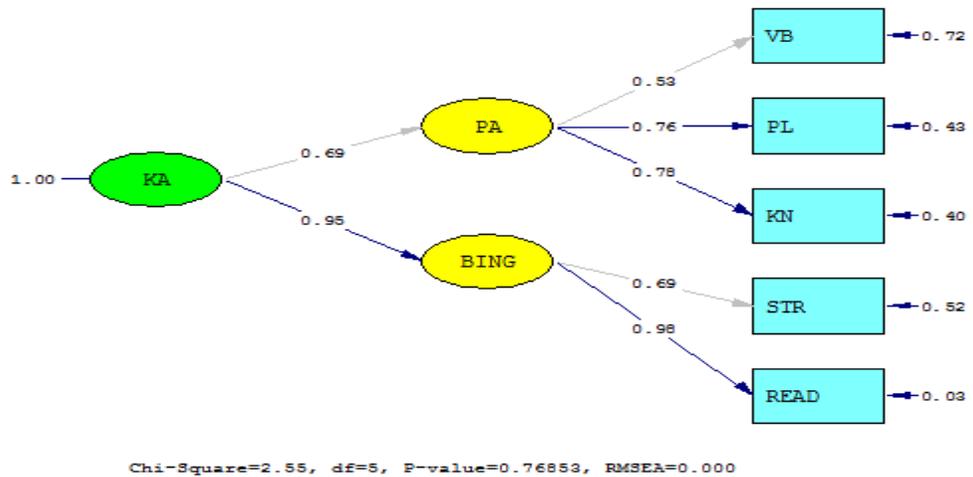
Pada estimasi reliabilitas dengan rumus KR-20, perlu dihitung terlebih dahulu proporsi menjawab benar tiap butir (p = banyaknya responden yang benar dibagi banyaknya siswa yang mengikuti tes), proporsi menjawab salah ($q = 1-p$), dan varians skor butir, pq (p dikali q). Selanjutnya dapat dihitung reliabilitas dengan KR20.

B112		fx =(40/39)*(1-(B109/27))						
	A	B	C	D	E	AN	AO	AP
1								
2								
3	Responden	Butir 1	Butir 2	Butir 3	Butir 4	Butir 39	Butir 40	Total
4	1	0	1	1	0	1	0	25
5	2	1	1	1	0	1	0	29
6	3	1	1	1	0	1	1	30
7	4	1	1	1	1	1	1	38
8	5	1	1	1	0	1	1	34
9	6	1	1	1	0	1	1	34
10	7	1	1	0	0	1	1	21
102	99	1	1	1	0	1	1	23
103	100	1	1	1	0	1	1	38
104								
105	pi	0.9	0.78	0.92	0.42	0.72	0.93	27.74495
106	qi	0.1	0.22	0.08	0.58	0.28	0.07	
107	pi.qi	0.09	0.1716	0.0736	0.2436	0.2016	0.0651	
108								
109	Sigma pi.qi	6.0865 (jumlah pi.qi)						
110	Sx^2	27.74495 (varians total)						
111								
112	KR20	0.7944						
113								

6. Reliabilitas Konstruk

Reliabilitas konstruk dapat dihitung setelah muatan faktor pada analisis faktor konfirmatori (λ) dan juga *error* dari tiap variable *observable* (δ). Dengan bantuan program Excel, menggunakan output dari analisis faktor konfirmatori dapat dihitung reliabilitas konstruk, reliabilitas ω , dan reliabilitas maksimal.

Hasil output analisis faktor konfirmatori:



Hasil analisis faktor konfirmatori tersebut kemudian dimasukkan dalam daftar di Excel untuk menghitung reliabilitas konstruk menggunakan rumus 6.8, 6.9, dan 6.10.

No.	λ	δ	λ^2	$1-\lambda^2$	$\lambda^2/(1-\lambda^2)$
1	0,53	0,72	0,2809	0,7191	0,390627
2	0,76	0,43	0,5776	0,4224	1,367424
3	0,78	0,4	0,6084	0,3916	1,553626
4	0,69	0,52	0,4761	0,5239	0,908761
5	0,98	0,03	0,9604	0,0396	24,25253
Jumlah	3,74	2,1	2,9034		28,47296
Reliabilitas Konstruk	0,869465				
		Reliabilitas	0,82811		
		Omega		Reliabilitas	
				Maksimal	0,966071

7. Reliabilitas Inter-rater

Reliabilitas interrater dapat diestimasi dengan menghitung persentase kecocokan skor hasil penilaian oleh rater 1 dan rater 2. Teknik ini hanya dapat digunakan untuk melihat kecocokan 2 rater saja.

Responden	Rater 1	Rater 2
1	3	3
2	2	3
3	1	1
4	3	2
5	4	4
6	2	2
7	5	5
8	5	4
9	4	4
10	5	4

Responden	Rater 1	Rater 2
11	5	5
12	4	3
13	4	4
14	5	5
15	4	4
16	3	3
17	2	2
18	1	1
19	1	2
20	5	5

Dari skor penilaian rater 1 dan rater 2 dibuat daftar, kemudian dihitung skor-skor yang sama hasil penilaian 2 rater yang berbeda. Pada kasus tersebut, yang cocok ada 14 dari 20 butir, sehingga reliabilitas skornya 70% atau 0,7.

$$\text{Inter-rater agreement} = (14/20) \times 100\% = 70\%.$$

8. Koefisien Generalisabilitas

Estimasi reliabilitas dengan menggunakan koefisien generalisabilitas (*generalizability coefficient*) melibatkan analisis varians. Analisis varians ini dalam rangka mempermudah menghitung mean square (MS). Pada contoh ini, rangkuman analisis varians untuk memperoleh MS dilakukan dengan menggunakan SPSS. Sebelumnya, diinputkan datanya dengan variabel *person* (peserta tes), *score* (perolehan skor), dan *rater* (penilai).

Variabel untuk menginput data

The screenshot shows the SPSS Data Editor window with the following variable definitions:

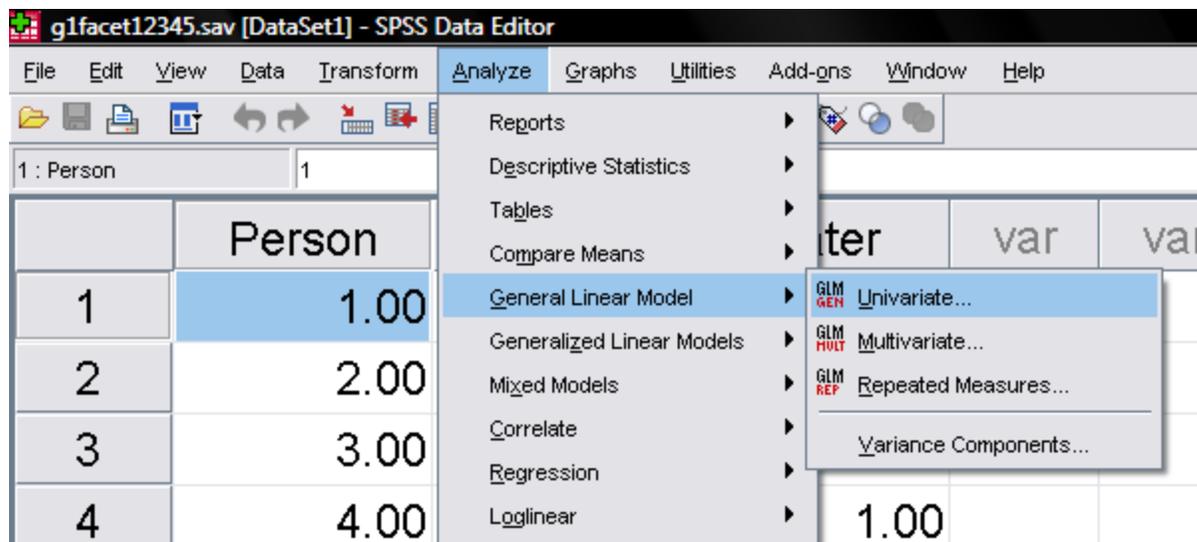
	Name	Type	Width	Decimals
1	Person	Numeric	8	2
2	Score	Numeric	8	2
3	Rater	Numeric	8	2

Hasil input data sebagai berikut.

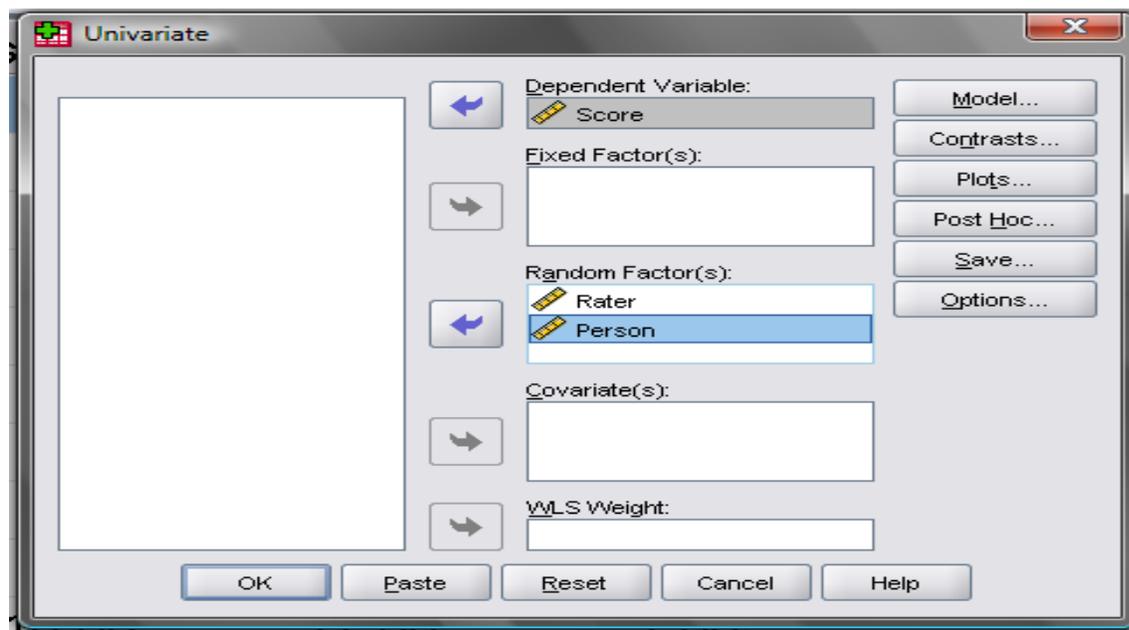
The screenshot shows the SPSS Data Editor window in Data View, displaying the following data:

	Person	Score	Rater	var	var	var	var
1	1.00	20.00	1.00				
2	2.00	20.00	1.00				
3	3.00	23.00	1.00				
4	4.00	24.00	1.00				
5	5.00	17.00	1.00				
6	6.00	19.00	1.00				
7	7.00	13.00	1.00				
8	8.00	22.00	1.00				
9	9.00	25.00	1.00				
10	10.00	17.00	1.00				

Selanjutnya klik **Analyze**→**General Linear Model**→**Univariat**.



Masukkan **Score** pada variabel terikat (**Dependent Variable**) dan **Rater** dan **Person** sebagai variabel bebas (**Random Factor(s)**), kemudian klik **OK**.



Setelah itu akan muncul output rangkuman tabel analisis varians.

Tests of Between-Subjects Effects

Dependent Variable: Score

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	71923.205	1	71923.205	879.287	.000
	Error	3152.882	38.545	81.797 ^a		
Rater	Hypothesis	4.513	4	1.128	2.386	.054
	Error	71.887	152	.473 ^b		
Person	Hypothesis	3083.395	38	81.142	171.569	.000
	Error	71.887	152	.473 ^b		
Rater * Person	Hypothesis	71.887	152	.473		
	Error	.000	0	. ^c		

a. $MS(\text{Rater}) + MS(\text{Person}) - MS(\text{Rater} * \text{Person})$

b. $MS(\text{Rater} * \text{Person})$

c. $MS(\text{Error})$

Output tersebut dapat dicopy, kemudian dibuka program Excel.

Copy tersebut kemudian di-Paste di Excel, sehingga kita dapat melakukan perhitungan untuk mengistemasi reliabilitas dengan 1 facet.

Source	Type III Sum of Squares	df	Mean Square						
Intercept	71923,21	1	71923,21						
Rater	4,512821	4	1,128205	nr	5	var(r)			
Person	3083,395	38	81,14197	np	39	var(p)			
Rater * Person	71,88718	152	0,472942			var(pr)			
Error	3152,882	38,5451	81,79723						
								Komponen Penyebut	
								0,016801619	
								16,13380567	
								0,47294197	
								0	
								0	
								Pembilang	16,13380567
								Penyebut	16,62354926
								reliabilitas	0,970539168

9. Estimasi Koefisien Geralisabilitas 2 Facet

Seperti halnya pada 2 facet, data diinputkan pada SPSS terlebih dahulu. Pada 2 facet ini, yang perlu diinputkan adalah *person*, *score* (dapat berupa skor butir), *rater* dan *question*.

Person	Score	Rater	Question	var	var
1	1.00	2.00	1.00	1.00	
2	2.00	4.00	1.00	1.00	
3	3.00	4.00	1.00	1.00	
4	4.00	5.00	1.00	1.00	
5	5.00	2.00	1.00	1.00	
6	6.00	0.00	1.00	1.00	
7	7.00	2.00	1.00	1.00	
8	8.00	2.00	1.00	1.00	
9	9.00	5.00	1.00	1.00	
10	10.00	2.00	1.00	1.00	

	Name	Type	Width	Decimals
1	Person	Numeric	8	2
2	Score	Numeric	8	2
3	Rater	Numeric	8	2
4	Question	Numeric	8	2

Output yang diperoleh dapat diperoleh kemudian dapat dicopy, kemudian di-paste di Excel.

Source	df	Mean Square	F	Sig.
Corrected Total	974	2.609	1.150	.302
Intercept	1	14737.055	6.494E3	.000
Person	38	16.682	7.352	.000
Rater	4	.082	.036	.997
Question	4	90.063	39.689	.000
Person * Question	152	.147	.065	1.000
Person * Rater	1435.068	9.441	4.161	.000
Rater * Question	6.857	.429	.189	1.000
Person * Rater * Question	77.996	.128	.057	1.000
Error	88.500	2.269		
Total	17699.000			
Corrected Total	2629.689			

a. R Squared = ,966 (Adjusted R Squared = ,126)

Selanjutnya dengan menggunakan Excel dapat dihitung pada lembar kerja Excel untuk menghitung koefisien generalisabilitas.

Referensi

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehard and Winston Inc.
- Frisbie, D. A. 2005. Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 25 (3), 21-28.
- Geldhof, G.J., Preacher, K., Zyphur, M.J. (2014). Reliability Estimation in a Multilevel Confirmatory Factor Analysis Framework. *Psychological Methods*, 19(1), 72-91.
- Kamata, A., Turhan, A., Darandari, E. (2003). *Estimating Reliability for Multidimensional Composite Scale Scores*. Paper. Presented at the annual meeting of American Educational Research Association, Chicago, April 2003.
- Lawrence, M.R. (1994). Question to ask when evaluating test. *Eric Digest. Artikel*. Diambil dari: <http://www.ericfacility.net/ericdigest/ed.385607.html> tanggal 6 Januari 2007.
- Mehrens, W.A. & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology*. New York: Hold, Rinehart and Wiston, Inc.
- Penev, S. & Raykov, T. (2006). On the Relationship between Maximal Reliability and Maximal Validity of Linear Composites. *MULTIVARIATE BEHAVIORAL RESEARCH*, 41(2), 105–126.
- Reynold, C. R., Livingstone, R. B. & Wilson, V. (2010). *Measuremet and Assesment in Education*. New York, NY: Pearson.
- Walker, David A. (2006) "A Comparison of the Spearman-Brown and Flanagan-Rulon Formulas for Split Half Reliability under Various Variance Parameter Conditions," *Journal of Modern Applied Statistical Methods*: Vol. 5: Iss. 2, Article 18. Available at: <http://digitalcommons.wayne.edu/jmasm/vol5/iss2/18>
-



KEMENTERIAN AGAMA REPUBLIK INDONESIA
INSTITUT AGAMA ISLAM NEGERI BATUSANGKAR
FAKULTAS EKONOMI DAN BISNIS ISLAM

Jl. Sudirman No. 137 Lima Kaum Batusangkar Telp. (0752) 71160, 674221, 71880 Fax. (0752) 71879
Website : www.iainbatusangkar.ac.id e-mail : info@iainbatusangkar.ac.id

Nomor : B- ~~479~~ /In.27/F.IV.1/PP.00.9/07/2017
Sifat : Biasa
Lamp. : 1 lembar
Perihal : Mohon Mengutus Narasumber

17 Juli 2017

Yth. Bapak Dekan Fakultas Matematika
dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta

di
Yogyakarta

Assalamu'alaikum wr, wb.
Dengan hormat,

Pertama sekali kami mendoakan agar Bapak selalu dalam keadaan sehat dan sukses dalam menjalankan aktivitas sehari-hari, Amin ya robbal'amin.

Sehubungan akan diladakannya Kegiatan Workshop Pembelajaran Berbasis Riset dan Teknik Analisis Data pada Fakultas Ekonomi dan Bisnis Islam IAIN Batusangkar, bersama ini kami mohon kepada Bapak/Ibu agar berkenan kiranya Mengutus Ibu **Dr. Heri Retnowati, M.Pd.** untuk Menjadi Narasumber pada kegiatan dimaksud yang akan diadakan pada:

Hari/ Tanggal : Rabu/ 26 Juli 2017
Waktu : 08.00 WIB s.d. selesai
Tempat : Rocky Hotel Bukittinggi

Demikian kami sampaikan, atas berkenannya Bapak, disampaikan terima kasih.

Wassalam,

Dekan,
Kecamatan Bidang Akademik
dan Kemahasiswaan

Heri Retnowati, M.Ag





KEMENTERIAN RISET, TEKNOLOGI, DAN PENDIDIKAN TINGGI
UNIVERSITAS NEGERI YOGYAKARTA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

Jalan Colombo Nomor 1 Yogyakarta 55281
Telepon (0274) 565411 Ponsel 217, (0274) 565411 (TU), fax: (0274) 548205
Laman : mpa.uny.ac.id, E-mail : humas_fmipa@uny.ac.id

SURAT IZIN

NO. : 2097/UN.34.13/KP/2017

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam UNY memberikan izin kepada :

No	Nama	NIP	Pangkat/Gol.	Jabatan
1	Dr. Heri Retnawati	19730103 200003 2 001	Penata Tk. I / III/d	Lektor

Keperluan : Sebagai Narasumber dalam kegiatan Workshop Pembelajaran Berbasis Riset dan Teknik Analisis Data
Hari, Tanggal : Selasa s.d. Kamis, 25 s/d 27 Juli 2017
Tempat : Rocky Hotel Bukittinggi
Keterangan : Berdasarkan surat dari 1. Dekan Fakultas Ekonomi dan Bisnis Islam IAIN Batusangkar No. B-479/In.27/F.IV.1/PP.00.9/07/2017, tanggal 17 Juli 2017 2. Ketua Jurdik Matematika No. 318/UN.34.13.M/TU/2017, tanggal 18 Juli 2017

Surat izin ini diberikan untuk dilaksanakan sebaik-baiknya dan mohon melaporkan hasilnya kepada Dekan.

Yogyakarta, 18 Juli 2017
Dekan



Dr. Hartono
NIP. 19620329 198702 1 002

Tembusan :

1. Wakil Dekan I dan II FMIPA
2. Kajurdik Matematika
3. Kasubag UKP FMIPA
4. Yang Bersangkutan



**KEMENTERIAN AGAMA REPUBLIK INDONESIA
INSTITUT AGAMA ISLAM NEGERI BATUSANGKAR
FAKULTAS EKONOMI DAN BISNIS ISLAM**

Jl. Sudirman No. 137 Lima Kaum Batusangkar Telp. (0752) 71150, 574221, 71890 Fax. (0752) 71879
Website : www.iainbatusangkar.ac.id e-mail : info@iainbatusangkar.ac.id

SURAT PERNYATAAN MELAKSANAKAN TUGAS
Nomor: B- 506 /In.27/F.IV/PP.00.9/07/2017

Yang bertanda tangan di bawah ini:

Nama : Nasfizar Guspendri, SE., M.Si.
NIP : 19750823 200312 1 004
Jabatan : Dekan Fakultas Ekonomi dan Bisnis Islam

menyatakan dengan sesungguhnya bahwa Dr. Heri Retnowati, M.Pd. berdasarkan Surat Keputusan Rektor IAIN Batusangkar Nomor. B- 474 /In.27/R/PP.00.9/07/2017 tanggal 18 Juli 2017 secara nyata telah melaksanakan tugas sebagai Narasumber Workshop Pembelajaran Berbasis Riset dan Teknik Analisis Data pada Fakultas Ekonomi dan Bisnis Islam Institut Agama Islam Negeri Batusangkar tanggal 25 s.d 27 Juli 2017 sebanyak 14 (empat belas) JPL.

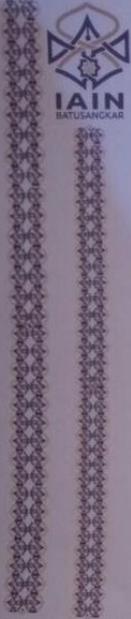
Demikian Surat Pernyataan ini dibuat untuk dapat dipergunakan seperlunya.

Batusangkar, 27 Juli 2017

Dekan



Nasfizar Guspendri, SE., M.Si. *f*



KEMENTERIAN AGAMA REPUBLIK INDONESIA
INSTITUT AGAMA ISLAM NEGERI BATUSANGKAR
FAKULTAS EKONOMI DAN BISNIS ISLAM

SERTIFIKAT

Nomor: B-305/In.27/F.IV/PP.00.9/07/2017

diberikan kepada:

Dr. Heri Retnowati, M.Pd

Sebagai Narasumber dalam kegiatan
Workshop Pembelajaran Berbasis Riset dan Teknik Analisis Data
Pada Tanggal 25 s/d 27 Juli 2017 di Rocky Hotel Bukittinggi

Bukittinggi, 27 Juli 2017

Dekan

Nasfizar Guspendri, SE., M.Si