

---

## ESTIMASI KESALAHAN PENGUKURAN DALAM BIDANG PENDIDIKAN BERDASARKAN TEORI TES KLASIK

I Wayan Gunartha

FKIP Universitas PGRI Mahadewa Indonesia  
Jln. Seroja No. 57 Tonja Denpasar Utara, Denpasar, Indonesia  
e-mail: [w.gunartha@yahoo.com](mailto:w.gunartha@yahoo.com)

### ABSTRACT

*Measurement results play an important role in providing information about the achievement of student learning outcomes. Unfortunately, accurate and objective information in the measurement of psychological aspects is not easily obtained because of measurement errors. These measurement errors need to be known by the teacher so that the measurement results can be interpreted appropriately. To determine the magnitude of measurement errors, the indicator used by classical test theory is the reliability index. The higher the reliability index, the lower the measurement error; and the smaller the reliability index, the higher the measurement error. This paper aims to describe the approach in calculating reliability index in order to find out the measurement errors of the measuring instrument used. The method used in this study is the study of literature, which is to examine various literature related to measurement errors and reliability of measuring instruments. In estimating measurement errors, there are three approaches that can be applied to estimate the amount of reliability index of a measuring instrument. That approach is the test-retest, the parallel-forms approach, and the internal consistency approach. Each approach is developed in accordance with the nature and function of the measuring instrument concerned with practical aspects in mind.*

**Keywords:** *Measurement errors, reliability, estimation, measuring tools*

### ABSTRAK

Hasil pengukuran berperan penting dalam memberikan informasi tentang pencapaian hasil belajar siswa. Sayangnya, informasi yang akurat dan objektif dalam pengukuran aspek psikologis tidak mudah diperoleh karena adanya kesalahan pengukuran. Kesalahan pengukuran ini perlu diketahui oleh guru agar hasil pengukuran dapat ditafsirkan secara tepat. Untuk mengetahui besarnya kesalahan pengukuran, indikator yang digunakan oleh teori tes klasik adalah indeks kehandalan atau reliabilitas. Semakin tinggi indeks reliabilitas, akan semakin rendah kesalahan pengukurannya; dan semakin kecil indeks reliabilitas, akan semakin tinggi kesalahan pengukurannya. Tulisan ini bertujuan untuk mendeskripsikan pendekatan dalam menghitung indeks reliabilitas dalam rangka mengetahui kesalahan pengukuran dari alat ukur yang digunakan. Metode yang digunakan dalam kajian ini adalah studi literatur, yaitu dengan mengkaji berbagai literatur yang terkait dengan kesalahan pengukuran dan reliabilitas alat ukur. Dalam mengestimasi kesalahan pengukuran, ada tiga pendekatan yang dapat diterapkan untuk menaksir besarnya indeks reliabilitas suatu alat ukur. Pendekatan itu adalah pendekatan tes-ulang (*test-retest*), pendekatan bentuk paralel (*parallel-forms*), dan pendekatan konsistensi internal (*internal consistency*). Masing-masing pendekatan dikembangkan sesuai dengan sifat dan fungsi alat ukur yang bersangkutan dengan mempertimbangkan segi praktisnya.

**Kata Kunci:** Kesalahan pengukuran, reliabilitas, estimasi, alat ukur

## PENDAHULUAN

Dalam usaha meningkatkan taraf hidupnya, manusia melakukan berbagai usaha. Semua usaha umat manusia ini tidak pernah lepas dari kegiatan pengukuran, baik dilakukan secara sistematis maupun secara insidental. Pengukuran dilakukan misalnya untuk mengetahui kualitas layanan, kualitas pendidikan, tingkat kesejahteraan, prestasi belajar, tingkat kepuasan, kejujuran, kedisiplinan, dan lain-lain. Pengembangan ilmu dalam semua bidang dan keberhasilan suatu program selalu melalui proses pengukuran, baik yang bersifat kualitatif maupun kuantitatif. Oleh karena itu, jelas bahwa peran pengukuran terhadap kehidupan umat manusia sangat besar.

Menurut Nunnally (1981), pengukuran terdiri atas aturan-aturan dalam menentukan angka atribut suatu objek. Aturan ini menjelaskan prosedur dalam menentukan angka dan harus dinyatakan secara eksplisit. Semua yang ada di dunia ini dapat menjadi objek pengukuran baik fisik maupun nonfisik. Atribut manusia yang bersifat fisik seperti tinggi badan, berat badan, warna rambut, dan lain-lain. Atribut nonfisik manusia seperti kemampuan, motivasi,

sikap, persepsi, kejujuran, kedisiplinan, kepuasan kerja.

Penafsiran hasil pengukuran sering disebut dengan istilah penilaian, yaitu untuk menyatakan kualitas suatu produk, prestasi belajar, mutu pendidikan, dan lain-lain. Jika penafsiran ini diikuti dengan suatu keputusan kebijakan, maka kegiatan ini sudah sampai pada evaluasi. Oleh karena itu, pengukuran, penilaian, dan evaluasi merupakan kegiatan yang berkaitan dan berurutan. Keputusan yang tepat harus didukung oleh informasi hasil pengukuran yang akurat. Untuk itu, pengukuran yang dilakukan harus dapat menyajikan informasi yang dengan kesalahan sekecil mungkin agar keputusan yang diambil juga tepat.

Sayangnya, informasi yang akurat dan objektif dalam pengukuran aspek psikologis dan sosial tidak mudah diperoleh terutama disebabkan oleh tidak mudahnya mengoperasionalkan variabel yang diukur seperti dalam penelitian mengenai aspek fisik. Walaupun operasionalisasi variabel yang hendak diteliti sudah dapat ditegakkan, prosedur pengukurannya masih meminta perhatian tersendiri karena pengukuran terhadap variabel yang diteliti menjadi penentu apakah

informasi yang dihasilkannya dapat dipercaya atau tidak.

Ada beberapa sumber kesalahan yang mempengaruhi hasil pengukuran, yaitu alat ukur, objek yang diukur, lingkungan pengukuran dan dan subjek yang mengukur (Mardapi, 1997). Indikator yang digunakan untuk menentukan besarnya kesalahan pengukuran bergantung kepada teori yang digunakan. Teori tes klasik menggunakan indeks kehandalan atau reliabilitas untuk menentukan besarnya kesalahan pengukuran. Makin tinggi indeks ini akan semakin kecil kesalahan pengukurannya. Teori respon butir menggunakan informasi tes untuk menentukan besarnya kesalahan pengukuran. Makin tinggi informasi tes, akan semakin kecil kesalahan pengukurannya.

Menurut Ebel (1979), kesalahan pengukuran adalah simpangan baku skor seseorang yang dites berulang kali, dengan asumsi tidak ada unsur kelelahan. Dijelaskan pula bahwa estimasi kesalahan pengukuran dapat dilakukan melalui informasi tentang besarnya indeks reliabilitas suatu tes.

Permasalahan yang akan digarap dalam tulisan ini adalah apakah yang dimaksud dengan kehandalan atau

reliabilitas suatu alat ukur dan bagaimanakah teknik-teknik penaksiannya dalam mengestimasi kesalahan pengukuran? Tujuan yang diharapkan adalah untuk mengetahui hakikat reliabilitas suatu alat ukur dan teknik-teknik penaksiannya dalam upaya mengestimasi kesalahan pengukuran dalam bidang pendidikan.

## **PEMBAHASAN**

### **Kesalahan Pengukuran dalam Pendidikan**

Setiap pengukuran selalu mengandung kesalahan yang disebut dengan *standar error of measurement (SEM)*. Sumber kesalahan pengukuran diantaranya adalah pada penentuan materi ujian, pihak yang diukur, pihak yang mengukur, dan lingkungan. Keadaan kesehatan fisik dan emosi. Namun, pengukuran harus memiliki kesalahan sekecil-kecilnya (Mardapi, 1997). Wright (2008, p. 130) menyatakan bahwa  $True\ score = observed\ score \pm measurement\ error$  (skor sebenarnya = skor hasil pengukuran  $\pm$  kesalahan pengukuran). Berdasarkan persamaan itu, ada dua kemungkinan yang akan terjadi, yaitu: skor hasil pengukuran lebih kecil daripada skor sebenarnya atau lebih

besar daripada skor sebenarnya. Bila salah satu dari kedua kemungkinan tersebut terjadi, berarti telah terjadi kesalahan dalam pengukuran.

Nunally (1978, p. 190-191) mengatakan bahwa dalam mengukur atribut tertentu, berbagai *error* bisa terjadi. Kesalahan itu ada bersifat acak, ada juga yang bersifat sistematis. Selanjutnya, dijelaskan oleh Mardapi (2012, p. 9) bahwa kesalahan acak disebabkan oleh kondisi fisik dan mental yang diukur dan yang mengukur bervariasi. Kesalahan acak juga bisa disebabkan pemilihan materi pengukuran yaitu bahan pelajaran yang diujikan. Kesalahan sistematis disebabkan oleh alat ukurnya, yang diukur, dan yang mengukur.

Ada pendidik yang suka membuat soal yang mudah, ada juga yang cenderung membuat soal yang sulit, sehingga hasil pengukuran bisa *under* atau *over estimate* dari kemampuan yang sebenarnya. Ada juga pendidik yang murah dan ada yang mahal dalam memberikan skor untuk semua peserta didik. Apabila murah atau mahal dalam memberi skor itu terjadi pada peserta didik tertentu, akan terjadi *bias* dalam pengukuran. Rasa cemas pada peserta

didik bisa juga menjadi sumber *under estimate* dalam pengukuran.

Berdasarkan uraian di atas dapat dikatakan bahwa setiap proses pengukuran dalam pendidikan selalu ada kesalahan pengukuran. Kesalahan pengukuran adalah perbedaan yang terjadi antara skor yang diperoleh anak dengan kemampuan yang sesungguhnya. Bila skor yang diperoleh anak lebih besar daripada kemampuan sesungguhnya terjadi *over estimate*. Jika terjadi sebaliknya, akan terjadi *under estimate*. Kesalahan pengukuran bisa bersifat acak, dan bisa bersifat sistematis.

#### **Hakikat Reliabilitas Suatu Alat Ukur**

Reliabilitas merupakan terjemahan dari kata *reliability*, yang mempunyai asal kata *rely* dan *ability*. Pengukuran yang memiliki reliabilitas tinggi disebut sebagai pengukuran yang reliabel (*reliable*). Walaupun reliabilitas mempunyai berbagai nama lain seperti keterpercayaan, keterandalan, kejelasan, kestabilan, konsistensi, dan sebagainya, namun ide pokok yang terkandung dalam konsep reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya (Aswar, 2020, p. 4). Suharsimi Arikunto (1997, p. 83) juga mengatakan hal yang hampir sama,

bahwa reliabilitas berhubungan dengan masalah kepercayaan. Suatu tes dapat dikatakan memiliki tingkat kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Mehrens & Lehmann (1973, p.102) mengatakan bahwa reliabilitas adalah taraf konsistensi antara dua pengukuran terhadap sesuatu yang sama.

Jadi, berdasarkan uraian di atas dapat dikatakan bahwa pengukuran dapat dipercaya hanya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, selama aspek yang diukur dalam diri subjek memang belum berubah. Relatif sama artinya tetap ada toleransi terhadap perbedaan-perbedaan kecil di antara hasil beberapa kali pengukuran. Dengan kata lain, indeks kehandalan tes adalah konsistensi seseorang dalam memberikan respon terhadap tes atau nontes yang dikerjakan dua kali dalam selang waktu tertentu, bila kemampuannya tidak berubah. Bila perbedaan itu sangat besar dari waktu ke waktu, maka hasil pengukuran tidak dapat dipercaya dan dikatakan tidak reliabel.

Secara empirik, tinggi rendahnya reliabilitas ditunjukkan oleh suatu angka

yang disebut koefisien atau indeks reliabilitas. Pada awalnya, tinggi rendahnya reliabilitas tes dicerminkan oleh koefisien korelasi antara skor pada dua tes yang paralel, yang dikenakan pada sekelompok individu yang sama. Semakin tinggi koefisien korelasi termaksud berarti konsistensi antara hasil pengenaan dua tes tersebut semakin baik dan hasil ukur kedua tes itu dikatakan semakin reliabel. Sebaliknya, apabila dua tes yang dianggap paralel ternyata menghasilkan skor yang satu sama lain berkorelasi rendah, maka dapat dikatakan bahwa reliabilitas hasil ukur tes tersebut tidak tinggi. Koefisien korelasi antara dua variabel dilambangkan dengan huruf  $r$ . Apabila skor pada tes pertama diberi lambang  $X$ , dan skor pada tes yang kedua yang paralel diberi lambang  $X'$ , maka koefisien korelasi antara kedua tes tersebut adalah  $r_{XX'}$ . Simbol inilah yang biasa digunakan sebagai simbol koefisien reliabilitas.

Walaupun secara teoretik besarnya koefisien reliabilitas berkisar mulai dari 0,0 sampai dengan 1,0, akan tetapi pada kenyataannya koefisien sebesar 1,0 dan sekecil 0,0 tidak pernah dijumpai. Di samping itu, walaupun koefisien korelasi dapat saja bertanda

negatif (-), koefisien reliabilitas selalu mengacu pada angka positif (+) karena angka negatif tidak ada artinya bagi interpretasi reliabilitas alat ukur. Koefisien reliabilitas  $r_{xx}' = 1,0$  berarti adanya konsistensi yang sempurna pada hasil ukur yang bersangkutan. Hal ini tidak pernah terjadi dalam pengukuran aspek psikologis manusia.

Pengertian reliabilitas alat ukur dan reliabilitas hasil ukur biasanya dianggap sama. Namun penggunaannya masing-masing perlu diperhatikan. Konsep reliabilitas dalam arti *reliabilitas alat ukur* erat berkaitan dengan kesalahan pengukuran (*error of measurement*). Kesalahan pengukuran menunjuk pada sejauh mana inkonsistensi hasil pengukuran terjadi apabila pengukuran dilakukan ulang pada kelompok subjek yang sama. Konsep reliabilitas dalam arti *reliabilitas hasil ukur* erat berkaitan dengan kesalahan dalam pengambilan sampel (*sampling error*), yang mengacu kepada inkonsistensi hasil ukur apabila pengukuran dilakukan pada kelompok individu yang berbeda. Oleh karena itu, dalam penelitian yang menggunakan alat ukur yang sebelumnya telah teruji reliabilitasnya, komputasi koefisien reliabilitas hasil ukur bagi subjek

penelitian tersebut masih tetap perlu dilakukan.

### **Estimasi Kesalahan Pengukuran**

Estimasi kesalahan pengukuran menurut teori tes klasik menggunakan indeks kehandalan atau reliabilitas tes. Makin besar indeks kehandalan tes semakin kecil kesalahan pengukuran. Kesalahan pengukuran ini berlaku untuk melompok tidak untuk setiap individu. Namun perkembangan selanjutnya, dengan menggunakan model kesalahan binomial, besarnya kesalahan pengukuran untuk individu dapat dihitung.

Penaksiran terhadap tingginya indeks reliabilitas dapat dilakukan dengan berbagai pendekatan. Tiap-tiap pendekatan menggunakan asumsi tertentu. Masing-masing pendekatan dikembangkan sesuai dengan sifat dan fungsi alat ukur yang bersangkutan dengan mempertimbangkan segi praktisnya. Oleh karena itu pemilihan pendekatan dan teknik yang paling tepat bergantung pada karakteristik distribusi skor yang paling mendekati asumsi tersebut. Secara tradisional, terdapat tiga pendekatan reliabilitas, yaitu: pendekatan tes-ulang (*test-retest*), pendekatan bentuk paralel (*parallel-*

*forms*), dan pendekatan konsistensi internal (*internal consistency*).

### ***Pendekatan Tes Ulang (Test-retest)***

Pendekatan tes-ulang dilakukan dengan menyajikan tes dua kali pada satu kelompok subjek dengan tenggang waktu di antara penyajian tersebut. Asumsi yang menjadi dasar dalam cara ini adalah bahwa suatu tes yang reliabel tentu akan menghasilkan skor-tampak yang relatif sama apabila dikenakan dua kali pada waktu yang berbeda. Semakin besar variasi perbedaan skor subjek antara kedua pengenaan tes itu, berarti semakin sulit untuk mempercayai bahwa tes itu memberikan hasil ukur yang konsisten.

Untuk memperoleh koefisien reliabilitas dengan cara ini, kita menghitung koefisien korelasi linier antara distribusi skor subjek pada pemberian tes pertama dan kedua. Jika semua subjek memperoleh skor sama antara tes pertama dan kedua, maka reliabilitas tes akan 1,0 ini mengindikasikan reliabilitas yang sempurna (Allen & Yen (1979, p. 76). Koefisien korelasi kedua skor itu merupakan koefisien reliabilitas tes yang bersangkutan. Asumsi yang menjadi dasar dalam cara ini adalah bahwa suatu tes yang reliabel akan

menghasilkan skor-tampak yang relatif sama apabila dikenakan dua kali pada waktu yang berbeda.

Dalam penggunaan pendekatan tes-ulang, harus diperhatikan kemungkinan adanya perubahan kondisi subjek sejalan dengan berbedanya waktu di antara kedua penyajian tes. Perubahan kondisi subjek yang terjadi tidak pada keseluruhan subjek dan tidak searah sedikit-banyak akan ada pengaruhnya terhadap koefisien reliabilitas yang diperoleh. Sebagai contoh, apabila subjek tidak bersungguh-sungguh, kelelahan, atau memang tidak siap ketika dikenai tes yang pertama kali, kemudian ia belajar sehingga betul-betul siap, sungguh-sungguh dalam mengerjakan tes yang kedua, maka skor subjek pada kedua pemberian tes akan berbeda. Kalau hal itu terjadi pada sebagian subjek, dan perubahan skor itu tidak searah bagi semua subjek, maka reliabilitas yang diperoleh akan rendah. Tentu saja koefisien tersebut tidak merupakan estimasi yang benar terhadap reliabilitas tes, akan tetapi merupakan estimasi yang lebih rendah daripada semestinya (*underestimasi*). Itulah salah satu kasus terjadinya *carry-over effects* (efek bawaan) yang seringkali menjadi

masalah serius dalam pendekatan reliabilitas tes-ulang. Di samping itu, masih banyak lagi efek bawaan yang lain.

Memang, biasanya tidak mudah untuk menentukan berapa lama tenggang waktu yang perlu disediakan di antara kedua kali pemberian tes. Bila tenggang waktu terlalu singkat, sangat mungkin terjadi efek bawaan, sedangkan bila tenggang waktunya terlalu panjang, sangat mungkin terjadi perubahan pada aspek psikologis yang diukur dalam diri subjek. Mungkin juga lamanya tenggang waktu akan menyebabkan perubahan suasana hati, motivasi, dan sikap terhadap pengetesan. Oleh karena itu, pendekatan tes-ulang hanya cocok digunakan untuk tes yang mengukur aspek psikologis yang relatif stabil dan tidak mudah berubah.

### ***Pendekatan Bentuk Paralel***

Estimasi reliabilitas bentuk paralel adalah korelasi skor-tampak pada dua tes paralel (Allen & Yen (1979, p. 77)). Jadi, dalam pendekatan bentuk paralel, tes yang akan diestimasi reliabilitasnya harus ada paralelnya, yaitu tes yang sama tujuan ukurnya dan setara isi aitemnya, baik secara kualitas maupun kuantitasnya. Sebenarnya, tes yang

paralel hanya ada secara teoretik, tidak benar-benar paralel secara empirik. Dari segi praktis, dua tes yang mempunyai syarat-syarat dan asumsi tertentu dapat dianggap paralel.

Untuk membuat dua tes yang paralel, penyusunannya haruslah didasarkan pada satu spesifikasi yang sama. Spesifikasi ini meliputi antara lain tujuan ukur, batasan objek ukur, indikator prilakunya, banyaknya aitem, taraf kesukaran. Secara empirik, dua tes yang paralel haruslah menghasilkan mean skor dan varians yang setara dan korelasi yang juga tidak berbeda dengan suatu variabel ketiga. Hanya itulah bukti terpenuhinya sifat paralel antara dua tes yang dapat diperoleh dalam penyusunan tes.

Bila telah diperoleh dua tes yang paralel, maka estimasi reliabilitasnya dilakukan dengan mengenakan dua tes tersebut bersamaan pada sekelompok subjek. Dalam pelaksanaannya, kedua tes paralel itu dapat digabungkan terlebih dahulu sehingga seakan-akan merupakan satu bentuk tes. Setelah selesai dijawab oleh subjek, barulah butir pada masing-masing tes semula dipisahkan kembali untuk diberi skor masing-masing. Dengan demikian, masing-masing akan menghasilkan



distribusi skor seluruh subjek. Penggabungan cara penggabungan ini adalah tidak terkesannya beban berat mengerjakan dua tes bagi subjek dan bila urutan nomor butir gabungan itu disusun sedemikian rupa akan dapat mengurangi efek *carry-over* dari satu bentuk ke bentuk tes yang lain. Kalau penggabungan butir dari masing-masing tes tidak dimungkinkan, maka kedua tes tersebut harus diberikan secara berturut-turut dengan tenggang waktu yang hanya sebentar. Dengan menghitung korelasi antara kedua distribusi skor dengan teknik korelasi *Product Moment*, akan diperoleh koefisien reliabilitas tes tersebut. Walaupun pendekatan ini dapat menghilangkan masalah penentuan tenggang waktu yang tepat, kelemahannya adalah sulitnya menyusun dua tes yang paralel.

Dalam pendekatan bentuk paralel, tidak sempurnanya reliabilitas tes juga disebabkan oleh adanya *varians error*. Namun, *error* tersebut dapat dikatakan tidak berkaitan dengan masalah tenggang waktu di antara pemberian kedua tes yang paralel tersebut. *Error* yang terjadi lebih disebabkan oleh faktor-faktor yang ada dalam tes itu

sendiri atau berasal dari subjek yang diukur, dan dari pihak pemberi tes.

### ***Pendekatan Konsistensi Internal (Internal Consistency)***

Pendekatan konsistensi internal dalam estimasi reliabilitas dimaksudkan antara lain untuk menghindari masalah-masalah yang biasanya timbul pada pendekatan tes ulang dan pendekatan paralel. Pendekatan konsistensi internal dilakukan dengan menggunakan satu bentuk tes yang dikenakan hanya sekali saja pada sekelompok subjek (*single-trial administration*). Oleh karena itu, pendekatan ini mempunyai nilai praktis dan efisiensi yang tinggi. Dengan menyajikan satu tes hanya satu kali, maka masalah yang mungkin timbul pada dua pendekatan terdahulu dapat dihindari.

Pendekatan konsistensi internal berujuan melihat konsistensi antaritem atau antar bagian dalam tes itu sendiri. Untuk itu, setelah skor setiap item diperoleh, dari sekelompok subjek, tes dibagi menjadi beberapa belahan. Tes yang akan diestimasi reliabilitasnya, dapat dibelah menjadi dua bagian, tiga, empat, bahkan dapat dibagi menjadi belahan-belahan sebanyak jumlah itemnya sehingga setiap belahan berisi satu item saja. Bila suatu tes dibelah

menjadi belahan-belahan yang masing-masing berisi lebih dari satu aitem, maka dianjurkan agar jumlah aitem tiap belahan seimbang. Ada beberapa formula untuk estimasi reliabilitas dengan pendekatan konsistensi internal, seperti formula *Spearman-Brown* untuk belah dua, Formula *Rulon*, Formula *Alpha*, Formula *Kuder-Richardson*, Formula *Kristof* untuk belah tiga, dan masih banyak lagi. Dari formula-formula itu hanya akan dibahas beberapa, yaitu yang sering digunakan dalam estimasi reliabilitas tes prestasi belajar.

#### **Formula Spearman-Brown untuk belah Dua**

Formula *Spearman-Brown* merupakan sebuah formula komputasi yang sangat populer untuk estimasi reliabilitas tes yang dibelah menjadi dua bagian yang relatif paralel. Formula ini dapat digunakan pada tes yang aitem-aitemnya diberi skor dikotomi maupun bukan. Formulanya adalah sebagai berikut:

$$S-B = r_{xx'} = \frac{2(r_{12})}{1 + r_{12}}$$

#### Keterangan :

$r_{xx'}$  = Koefisien reliabilitas

$r_{12}$  = koefisien korelasi antara kedua belahan, dimana  $r_{12} =$

$$\frac{n \cdot \sum XY - (\sum X)(\sum Y)}{\sqrt{\{n \cdot \sum X^2 - (\sum X)^2\} \{n \cdot \sum Y^2 - (\sum Y)^2\}}}$$

#### Keterangan rumus:

$r_{12}$  = koefisien korelasi antarbelahan

n = jumlah sampel

X = skor belahan 1

Y = skor belahan 2

Misalnya, korelasi antarkedua belahan diperoleh 0,957 maka koefisien reliabilitasnya adalah :

$$S-B = r_{xx'} = \frac{2 \cdot (0,957)}{1 + 0,957} = 0,978$$

Formula ini hanya dapat digunakan bila apabila kita percaya bahwa asumsi paralelisme di antara kedua belahan terpenuhi. Ciri terpenuhinya asumsi tersebut antara lain adalah apabila kedua belahan tes menghasilkan rata-rata skor (*mean*) yang setara dan varians skor yang sebanding. Di samping itu, formula ini akan menghasilkan estimasi reliabilitas yang cermat bila koefisien korelasi di antara kedua belahan tersebut cukup tinggi. Jika tidak, sebaiknya formula ini jangan digunakan.

#### **Formula Alpha ( $\alpha$ )**

Jika kita tidak yakin asumsi paralelisme terpenuhi dalam menggunakan formula *Spearman-Brown*, maka formula *Alpha*

ini dapat digunakan. Formula koefisien *Alpha* untuk estimasi reliabilitas tes belah dua adalah sebagai berikut.

$$r_{xx}/\alpha = 2\left(1 - \frac{S_1^2 + S_2^2}{S_x^2}\right)$$

Keterangan:

$S_1^2$  dan  $S_2^2$  = Varians skor belahan 1 dan belahan 2

$S_x^2$  = Varian skor total

Misalkan dua belas aitem tes dibelah menjadi dua. Skor belahan 1 diberi simbol  $Y_1$  dan skor pada belahan kedua diberi simbol  $Y_2$ . skor total X berarti  $Y_1 + Y_2$ . Jumlah sampel adalah 10.

Jika misalnya  $\Sigma Y_1^2 = 508$ ; dan  $\Sigma Y_1 = 66$ , maka varian belaha  $Y_1$  adalah  $[508 - (66^2)/10]/9 = 8,044$ .

Jika misalnya  $\Sigma Y_2^2 = 460$ ; dan  $\Sigma Y_2 = 62$ , maka varian belaha  $Y_2$  adalah  $[460 - (62^2)/10]/9 = 8,399$ .

$\Sigma X = 1928$  dan  $\Sigma X = 128$ , maka varian skor total X =  $[1928 - (128^2)/10]/9 = 32,178$ .

Dengan demikian, koefisien *Alpha* untuk data ini adalah:

$$r_{xx}/\alpha = 2\left(1 - \frac{8,044 + 8,339}{32,178}\right) = 0,978$$

Formula ini dapat juga dipergunakan pada tes yang aitem-aitemnya diberi skor dikotomi.

Jika tes akan dibelah menjadi lebih dari dua belahan; tiga, empat, maka formula yang digunakan adalah sebagai berikut.

$$\alpha = \left(\frac{k}{k-1}\right) \left(1 - \frac{\sum S_j^2}{S_x^2}\right)$$

Keterangan:

k = banyaknya belahan tes

$S_j^2$  = Varians belahan j; j = 1, 2, ...k

$S_x^2$  = Varians skor tes.

### ***Formula Kuder-Richadson (KR)-20 dan KR-21***

Bila suatu tes berisi aitem-aitem yang diberi skor dikotomi, sedangkan jumlah aitemnya tidak banyak, membagi tes menjadi dua bagian tidak dapat menghasilkan bagian yang setara. Jika tes dibagi menjadi lebih dari dua belahan akan menyebabkan jumlah aitem dalam setiap belahan terlalu sedikit. Bila dalam belahan terdapat sedikit aitem, komputasi reliabilitasnya tidak dapat menghasilkan estimasi yang cermat.

Salah satu cara yang dapat dilakukan adalah membelah tes tersebut menjadi sebanyak jumlah aitemnya sehingga setiap belahan berisi hanya satu aitem. Kemudian estimasi reliabilitasnya dilakukan melalui formula *alpha* yang disesuaikan, yang dikenal dengan nama formula *Kuder-Richadson -20* atau *KR-20*.

Formulanya adalah sebagai berikut.

$$KR-20 = \left[ \frac{k}{k-1} \right] \left[ 1 - \frac{\sum p(1-p)}{S_x^2} \right]$$

p = proporsi subjek yang mendapat angka 1 pada suatu aitem.

Misalnya data hasil tes sebagai berikut.

Keterangan:

k = banyaknya aitem dalam tes.

$S_x^2$  = varians skor tes

**Tabel 01**  
**Contoh Perhitungan Reliabilitas dengan KR-20 dan KR-21**

Subjek	Nomor Item												X
	1	2	3	4	5	6	7	8	9	10	11	12	
A	1	1	1	1	1	1	1	1	1	1	1	1	12
B	1	1	1	1	1	1	1	1	1	1	1	0	11
C	1	1	1	1	1	1	1	1	1	1	0	0	10
D	1	1	1	1	1	1	1	1	1	0	0	0	9
E	1	1	1	1	1	1	1	1	0	0	0	0	8
F	1	1	1	1	1	1	0	1	0	0	0	0	7
G	1	1	1	1	1	0	0	1	1	1	0	0	8
H	1	1	0	0	1	1	0	0	0	0	0	0	4
I	1	1	0	0	0	0	0	0	1	0	0	0	3
J	0	0	1	1	0	0	0	0	0	0	0	0	2
p	.9	.9	.8	.8	.8	.7	.5	.7	.6	.4	.2	.1	$\Sigma p=7,4$
(1-p)	.1	.1	.2	.2	.2	.3	.5	.3	.4	.6	.8	.9	$\Sigma P(1-p)=$
p(1-p)	.09	.09	.16	.16	.16	.21	.25	.21	.24	.24	.16	.09	2,06

Berdasarkan data yang disajikan di atas, diketahui banyaknya aitem adalah atau k = 12, varians skor total = 11,599.

Untuk menghitung harga p bagi suatu aitem, dihitung dari banyaknya angka 1 pada aitem yang bersangkutan dibagi dengan banyaknya subjek (dalam contoh n = 10). Sebagai contoh, aitem no. 3, banyaknya angka 1 adalah 8, lalu dibagi 10. Dengan demikian, harga p-nya adalah  $8/10 = 0,8$ .

Dengan demikian, koefisien KR-20 untuk data di atas adalah:

$$KR-20 = \left[ \frac{12}{12-1} \right] \left[ 1 - \frac{2,06}{11,599} \right] = 0,897$$

Kuder dan Richardson juga merumuskan formula yang ke-21, yang disebut dengan formula KR-21. dalam perhitungannya, formula ini menggunakan raa-rataa harga p dari keseluruhan aitem. Rumusan formula KR-21 adalah sebagai berikut.

$$KR-21 = \left[ \frac{k}{k-1} \right] \left[ 1 - \frac{k \bar{p}(1-\bar{p})}{S_x^2} \right]$$

Keterangan :

k = banyaknya aitem

$\bar{p}$  = rata-rata p

$S_x^2$  = varians skor tes

Perhitungan reliabilitas KR-21 dengan menggunakan data di atas adalah sebagai berikut.

$$k = 12; \quad \sum p = 7,4; \quad \bar{p} = 7,4/12 =$$

$$0,617; \quad 1 - \bar{p} = 0,383; \quad S_x^2 = 11,599.$$

$$KR-21 = \left[ \frac{12}{12-1} \right] \left[ 1 - \frac{12(0,617) \cdot (0,383)}{11,599} \right]$$

$$= 0,824$$

Hasilnya akan lebih kecil dibandingkan dengan menggunakan formula KR-20, bila harga p, yang menyatakan taraf kesukaran aitem sangat bervariasi.

## SIMPULAN

Berdasarkan seluruh uraian yang disampaikan dalam pembahasan di atas, maka dapat disimpulkan bahwa reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya. Suatu tes dapat dikatakan memiliki tingkat kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Reliabilitas juga diartikan sebagai tingkat stabilitas alat ukur. Menurut

teori tes klasik, reliabilitas suatu alat ukur merupakan alat estimasi kesalahan pengukuran (*error of measurement*). Artinya, semakin besar indeks reliabilitas semakin kecil kesalahan pengukuran yang terjadi, dan sebaliknya semakin kecil indeks reliabilitas semakin besar kesalahan pengukurannya.

Ada tiga pendekatan yang dapat digunakan dalam menghitung atau menaksir indeks reliabilitas, yaitu: pendekatan tes-ulang (*test-retest*), pendekatan tes parallel, dan pendekatan konsistensi internal (*internal consistency*). Dalam pendekatan konsistensi internal ini ada beberapa teknik yang dapat diterapkan, yaitu: formula *Spearman-Brown* untuk belah dua, Formula *Rulon*, Formula *Alpha*, Formula *Kuder-Richardson*, Formula *Kristof* untuk belah tiga. Semua pendekatan tersebut memiliki asumsi yang berbeda-beda.

## DAFTAR PUSTAKA

- Allen, M. J & Yen, W. J. (1979). *Introduction to measurement theory*. California: Brooks/Cole Publishing Company.
- Aswar, S. (2020). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar Offset.

- Ebel, R. L. (1979). *Essential of educational measurement*. New Jersey: Prentice Hall.
- Mardapi, D. (1997). *Ragam bentuk evaluasi*. Makalah Semiloka Evaluasi Penilaian Hasil Belajar Mahasiswa. LP3. UGM.
- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Mehrens, W. A. & Lehmann, I. J. (1973). *Measurement and evaluation in education and psychology*. Michigan: Holt, Rinehart, and Winston, Inc.
- Nunnally, J.C. (1981). *Psychometric theory*. New Delhi : Tata McGraw-Hill Publishing Company Limited.
- Wright, R. J. (2008) *Educational assessment*. Thousand Oaks: Sage Publications