

## **Reliabilitas**

**Oleh: Heri Retnawati**

**Pendidikan Matematika**

**FMIPA Universitas Negeri Yogyakarta**

**E-mail: [retnawati\\_heriuny@yahoo.co.id](mailto:retnawati_heriuny@yahoo.co.id)**

**Web: [evaluation-edu.com](http://evaluation-edu.com)**

Dalam kaitannya dengan tes, reliabilitas skor hasil tes merupakan informasi yang diperlukan dalam pengembangan tes. Mehrens & Lehmann (1973: 102) menyatakan bahwa reliabilitas merupakan derajat keajegan (*consistency*) di antara dua buah hasil pengukuran pada objek yang sama. Definisi ini dapat diilustrasikan dengan seseorang yang diukur tinggi badannya akan diperoleh hasil yang tidak berubah walaupun menggunakan alat pengukur yang berbeda dan skala yang berbeda. Dalam kaitannya dengan dunia 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: 62) 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. Hal ini menyebabkan sulitnya memperoleh alat ukur yang stabil untuk mengukur karakteristik seseorang (Mehrens & Lehmann, 1973: 103).

Berdasarkan uraian di atas, maka dalam pembuatan alat ukur dalam dunia pendidikan harus dilakukan secermat mungkin dan disesuaikan dengan kaidah-kaidah yang telah ditentukan oleh ahli-ahli pengukuran di bidang pendidikan. Untuk melihat reliabilitas suatu alat ukur, yang berupa suatu indeks reliabilitas, dapat dilakukan

penelaahan secara statistik. Nilai ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*).

Koefisien reliabilitas dapat diartikan sebagai koefisien keajegan atau kestabilan hasil pengukuran. Alat ukur yang reliabel adalah alat ukur yang mampu membuahkan hasil pengukuran yang stabil (Lawrence, 1994) dan konsisten (Mehrens & Lehmann, 1973: 102). 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 dari 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: 62) 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 suatu tes pada umumnya diekspresikan secara numerik dalam bentuk koefisien yang besarnya  $-1 < r < +1$ . Koefisien tinggi menunjukkan reliabilitas tinggi. Sebaliknya, jika koefisien suatu tes rendah maka reliabilitas tes rendah. Jika suatu reliabilitas sempurna, berarti tes tersebut mempunyai koefisien  $+1$  atau  $-1$ .

Reliabilitas tinggi menunjukkan kesalahan yang kecil. Jika suatu tes mempunyai reliabilitas tinggi maka pengaruh kesalahan pengukuran telah berkurang. Kesalahan pengukuran dapat disebabkan oleh beberapa faktor, diantaranya karakteristik tes evaluasi itu sendiri, semisal kondisi pelaksana tes yang tidak mengikuti aturan baku, butir-butir tes yang meragukan, status peserta yang mengikuti tes, seperti peserta yang sedang lelah, mempunyai problem pribadi, peserta yang mempunyai motivasi kurang, lingkungan tempat diselenggarakannya tes kurang mendukung atau kombinasi dari segala permasalahan tersebut.

Mehrens & Lehmann (1973: 104) 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.

Penghitungan reliabilitas disebut dengan estimasi. Estimasi reliabilitas tes yang dapat dilakukan dengan dua cara, baik konsistensi eksternal dan maupun konsistensi internalnya.

### **1. Pengukuran Konsistensi Eksternal**

Reliabilitas eksternal diperoleh dengan cara mengolah hasil pengesanan yang berbeda, baik dari instrument yang berbeda maupun yang sama. Ada dua cara untuk estimasi reliabilitas eksternal suatu instrument yaitu dengan teknik ulang dan teknik paralel.

#### **a. Metode Test Ulang (*Test-Retest-Method*)**

Untuk mengetahui sampai dimana suatu pengukuran dapat diandalkan, pengukuran ini dapat dilakukan dua kali, pengukuran pertama dan ulangnya. Kedua pengukuran ini dapat dilakukan oleh orang yang sama atau berbeda. Dalam hal ini perlu diatur bahwa proses pengukuran kedua, keadaan yang diukur itu harus benar-benar sama. Selanjutnya hasil pengukuran yang pertama dan yang kedua dikorelasikan dan hasilnya menunjukkan reliabilitas dari tes ini. Memang teknik ulangan ini akan dapat memenuhi sasaran bila keadaan subjek yang diukur (dites) tetap bertahan dan tidak mengalami perubahan pada saat pengukuran yang pertama maupun pada pengukuran yang kedua. Karena keadaan pribadi anak itu selalu dalam keadaan berkembang, tidak statis, maka sebenarnya teknik ini kurang tepat digunakan. Disamping itu pada pengukuran yang kedua akan dijumpai adanya "*testing effect*" anak telah mendapat tambahan pengetahuan karena sudah mengalami tes yang pertama.

Reliabilitas tes retes ini penting ketika kita menafsirkan koefisien tes-retes untuk mengetahui : a) jangka waktu antara kedua pengambilan penilaian, b) stabilitas

yang diharapkan dari kinerja yang diukur. Secara umum, semakin lama antara interval pelaksanaan tes yang berulang, semakin rendah tingkat reliabilitasnya.

Estimasi reliabilitas dengan pendekatan tes-retes akan menghasilkan koefisien stabilitas. Untuk memperoleh koefisien reliabilitas melalui pendekatan tes-retes dapat dilakukan dengan menghitung koefisien korelasi linier antara distribusi skor subjek pada pemberian tes pertama dengan skor subjek pada pemberian skor kedua.

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

#### **b. Metode Bentuk Paralel (*Equivalent*)**

Tes paralel atau tes *equivalent* adalah dua buah tes yang mempunyai kesamaan tujuan, tingkat kesukaran dan susunan tetapi butir-butir soalnya berbeda, dalam istilah bahasa Inggris disebut *alternate-forms method* (parallel forms).

Dengan metode bentuk paralel ini, dua buah tes yang paralel, misalnya tes Matematika seri A yang akan dicari reliabilitasnya dan tes Seri B diteskan kepada sekelompok siswa yang sama, kemudian dikorelasikan. Koefisien korelasi dari kedua hasil tes inilah yang menunjukkan koefisien reliabilitas tes seri A. Jika koefisiennya tinggi maka tes tersebut sudah reliable dan dapat digunakan sebagai alat dan pengetes yang terandalkan.

Dalam menggunakan metode tes paralel pengetes harus menyiapkan dua buah tes, dan masing-masing dicobakan kepada sekelompok siswa yang sama. Penggunaan metode ini baik karena siswa dihadapkan kepada dua macam tes sehingga tidak ada factor “masih ingat-ingat soalnya” yang dalam evaluasi disebut adanya *practice-effect*- dan *carry-over-effect*. Artinya ada factor yang dibawa oleh pengikut tes karena sudah mengerjakan soal tersebut.

Kelemahan dari metode ini adalah bahwa pengetes pekerjaannya berat karena harus menyusun dua seri tes. Lagipula harus tersedia waktu yang lama untuk mencobakan dua kali tes. Mengenai pertanyaan bagaimana proses melaksanakan tes reliabilitas secara ekivalen? Berikut ini akan ditunjukkan beberapa langkah-langkah tersebut diantaranya adalah sebagai berikut:

- 1) Menentukan subjek sasaran yang hendak di tes
- 2) Melakukan tes yang dimaksud kepada sasaran subjek yang dimaksud
- 3) Diadministrasi dengan baik
- 4) Dalam waktu yang tidak begitu lama melakukan tes yang kedua pada kelompok tersebut.
- 5) Mengkorelasikan antara kedua skor tes tersebut.

Jika hasil koefisien ekivalen tinggi, berarti tes memiliki reliabilitas ekivalen baik. Sebaliknya, jika ternyata koefisien rendah maka reliabilitas ekivalen tes adalah rendah. Reliabilitas ekivalen merupakan salah satu bentuk yang diterima dan umum dipakai penelitian terutama penelitian pendidikan.

## **2. Pengukuran Konsistensi Internal**

Reliabilitas internal diperoleh dengan cara menganalisis data dari satu kali pengesanan. Pemilihan suatu teknik didasarkan atas bentuk instrument atau selera peneliti. Kadang-kadang penggunaan teknik yang berbeda menghasilkan indeks reliabilitas yang berbeda pula. Hal ini wajar saja karena kadang-kadang dipengaruhi oleh sifat atau karakteristik datanya sehingga dalam perhitungan diperoleh angka berbeda sebagai akibat pembulatan angka. Namun demikian, untuk beberapa teknik diperlukan persyaratan-persyaratan tertentu sehingga peneliti tidak begitu saja memilih teknik-teknik tersebut.

Beberapa metode yang dapat digunakan untuk mencari besarnya reliabilitas antara lain adalah :

### **a. Metode Belah Dua (*Split Half Method*)**

Dalam teknik belah dua ini, dalam pengesanan hanya menggunakan satu tes yang dicobakan satu kali kepada sejumlah subjek (*sample*). Item-item pada tes dibagi dua. Skor dari setengah item-item tes pada bagian yang pertama dikorelasikan dengan skor setengah item-item tes pada bagian yang kedua.

Mencari reliabilitas dengan menggunakan teknik belah dua sekurang-kurangnya ada dua persyaratan yang harus dipenuhi, yaitu :

- a. Banyaknya butir pertanyaan atau butir soal dalam instrument harus genap agar dapat dibelah
- b. Antara belahan pertama dengan belahan kedua harus seimbang.

Belahan instrument dikatakan seimbang jika jumlah butir pertanyaannya sama dan pertanyaan tersebut mengungkap aspek yang sama. Untuk memperoleh belahan yang seimbang, peneliti harus membuat pertanyaan dalam jumlah genap untuk setiap aspek atau factor. Dengan demikian, letak butir dapat disebar sedemikian rupa agar kalau dalam analisis data akan melakukan pembelahan sudah diketahui dengan pasti manakah pasangan-pasangan butir pertanyaannya. Itulah sebabnya perencanaan penelitian harus terpadu dalam memperhatikan tabel, pembuatan instrument, uji coba, pengujian reliabilitas, analisis data, dsb.

Ada dua cara membelah butir soal ini yaitu :

- 1) Membelah atas item-item genap dan item-item ganjil yang selanjutnya disebut belahan ganjil-genap.
- 2) Membelah atas item-item awal dan item-item akhir yaitu separoh jumlah pada nomor-nomor awal dan separo pada nomor-nomor akhir yang selanjutnya disebut belahan awal-akhir.

Cara pembelahan ini dapat menghindari kemungkinan terjadinya pengelompokan item-item tertentu kedalam salah satu belahan saja. Ada beberapa pengujian reliabilitas dengan metode belah dua, antara lain:

#### 1. Reliabilitas dengan Rumus Spearman-Brown

Adapun rumus spearman brown yang digunakan adalah :

$$r_i = \frac{2r_b}{1 + r_b}$$

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

keterangan :  $r_i$  = reliabilitas instrument

$r_b$  = indeks korelasi antara dua belahan instrument

N = banyaknya responden

X = belahan pertama

Y = belahan kedua

(Nitko, 2007:73)

## 2. Reliabilitas dengan Rumus Flanagan

Dalam teknik ini peneliti juga harus melakukan analisis butir terlebih dahulu dan menggunakan teknik belah dua ganjil-genap. Rumus Flanagan yang dimaksud adalah :

$$r_i = 2 \left( 1 - \frac{V_1 - V_2}{V_t} \right)$$

Dengan keterangan :

$r_i$  = reliabilitas instrument

$V_1$  = varians belahan pertama (varian skor butir-butir ganjil)

$V_2$  = varians belahan kedua (varian skor butir-butir genap)

$V_t$  = varians skor total

Untuk semua varians rumusnya adalah :

$$V = \frac{\sum X^2 - \frac{(\sum x)^2}{N}}{N}$$

## 3. Reliabilitas dengan Rumus Rulon

Rulon (1939) 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 tes akan membentuk distribusi perbedaan skor dengan varians yang besarnya ditentukan oleh varians eror masing-masing belahan menentukan varians eror keseluruhan tes, maka varians eror tes ini dapat diestimasi lewat besarnya varians perbedaan skor diantara kedua belahan. Dengan demikian, dalam melakukan estimasi terhadap reliabilitas tes, varians perbedaan skor inilah yang perlu diperhitungkan sebagai sumber eror. Untuk estimasi reliabilitas instrument dengan rumus Rulon, peneliti juga harus melakukan analisis skor butir. Kemudian rumusnya adalah :

$$r_i = 1 - \frac{V_d}{V_t}$$

Dengan keterangan :

$r_i$  = reliabilitas instrument

$V_t$  = varians total atau varians skor total

$V_d$  = varians (variens difference)

$d$  = skor pada belahan awal dikurangi skor pada belahan akhir

**b. Kuder-Richardson-Reliability**

Metode kedua untuk estimasi reliabilitas tes adalah penggunaan salah satu metode Kuder dan Richardson (1937). Kedua metodenya akan dibahas disini, Kuder Richardson formula 20 (KR-20) dan Kuder Richardson formula 21 (KR-21) yang digunakan ketika skorbutir tes 0 dan 1.

**1. KR-20**

Apabila peneliti memiliki instrument dengan jumlah butir pertanyaan ganjil, maka peneliti tersebut tidak mungkin menggunakan teknik belah dua untuk pengujian reliabilitasnya. Untuk itu maka peneliti dapat menggunakan rumus KR-20.

Rumus:

$$r_i = \frac{k}{(k - 1)} \left\{ \frac{s_t^2 - \sum p_i q_i}{s_t^2} \right\}$$

$r_i$  = reliabilitas instrument

$k$  = banyaknya butir pertanyaan atau banyaknya soal

$s_t^2$  = varians total

$p_i$  = proporsi subjek yang menjawab betul pada suatu butir (proporsi subjek yang mendapat skor 1)

$p_i = \frac{\text{banyaknya subjek yang skornya 1}}{N}$

$q_i = 1 - p_i$

**2. KR 21**

Rumus KR-21



$$r_i = \frac{k}{k-1} \left( 1 - \frac{\bar{p}(k-\bar{p})}{k\sigma_t^2} \right)$$

$r_i$  = reliabilitas instrument

$k$  = banyaknya butir pertanyaan atau banyaknya soal

$\sigma_t^2$  = varians total

$\bar{p}$  = skor rata-rata (Allen, 1979:84)

### c. Cronbach Alpha

Rumus Alpha digunakan untuk mencari reliabilitas instrument yang skornya bukan 1 dan 0, misalnya angket atau soal bentuk uraian.

*Rumus Alpha*

$$r_i = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right)$$

$r_i$  = reliabilitas instrumen

$k$  = banyaknya butir pertanyaan atau banyaknya soal

$\sum \sigma_b^2$  = jumlah varians butir

$\sigma_t^2$  = varians total

### Faktor-Faktor yang mempengaruhi Reliabilitas

Beberapa hal yang sedikit banyak mempengaruhi hasil tes banyak sekali. Namun secara garis besar dapat dikelompokkan menjadi 3 hal, yaitu :

- 1) Hal yang berhubungan dengan tes itu sendiri, yaitu panjang tes dan kualitas butir-butir soalnya. Tes yang terdiri dari banyak butir, tentu saja lebih valid dibandingkan dengan tes yang hanya terdiri dari beberapa butir soal. Tinggi rendahnya validitas merupakan tinggi rendahnya reliabilitas tes. Dengan demikian maka semakin panjang tes maka reliabilitasnya semakin tinggi.
- 2) Hal yang berhubungan dengan terdoba. Suatu tes dicobakan kepada kelompok yang terdiri dari banyak siswa akan mencerminkan keragaman hasil yang menggambarkan

besar kecilnya reliabilitas tes-tes yang dicobakan bukan kepada kelompok tidak terpilih, akan menunjukkan reliabilitas yang lebih besar daripada yang dicobakan pada kelompok tertentu yang diambil secara dipilih.

- 3) Hal yang berhubungan dengan penyelenggaraan tes sudah di sebutkan bahwa factor penyelenggaraan tes yang bersifat administratif, sangat menentukan hasil tes. Contoh :
  - a. Petunjuk yang diberikan sebelum tes dimulai, maka memberikan keterangan kepada para tes-tes dalam mengerjakan tes, dan dalam penyelenggaraan tidak akan banyak terdapat pertanyaan. Ketenangan ini tentu saja akan berpengaruh terhadap hasil tes.
  - b. Pengawas yang tertib akan mempengaruhi hasil yang diberikan oleh siswa terhadap tes bagi siswa-siswa tertentu adalah pengawasan yang terlalu ketat menyebabkan rasa jengkel dan tidak dapat dengan leluasa mengerjakan tes.
  - c. Suasana lingkungan dan tempat tes (duduk tidak diatur, suasana disekelilingnya ramai, dan sebagainya) akan mempengaruhi hasil tes.

Adanya hal-hal yang mempengaruhi hasil tes ini semua, secara tidak langsung akan mempengaruhi reliabilitas tes. Koefisien reliabilitas dapat dipengaruhi diantaranya oleh waktu penyelenggaraan tes-tes. Interval penyelenggaraan yang terlalu dekat atau terlalu jauh, akan mempengaruhi koefisien reliabilitas. Faktor-faktor yang lain yang juga mempengaruhi reliabilitas instrument evaluasi diantaranya sebagai berikut:

1. Panjang tes, semakin panjang suatu tes evaluasi, semakin banyak jumlah item materi pembelajaran diukur. Ini menunjukkan dua kemungkinan yaitu:
  - a. Tes semakin mendekati kebenaran
  - b. Dalam mengikuti tes semakin kecil siswa menebak bearti akan semakin tinggi nilai koefisien reliabilitasnya.
2. Penyebaran skor, koefisien reliabilitas secara langsung dipengaruhi oleh bentuk sebaran skor dalam kelompok siswa yang diukur. Semakin tinggi sebaran, semakin tinggi estimasi koefisien reliabilitas. Hal ini terjadi karena posisi skor siswa secara individual mempunyai kedudukan sama pada tes-tes lain, sebagai acuan.<sup>5</sup>
3. Kesulitan tes, tes normative yang terlalu mudah atau terlalu sulit untuk siswa, cenderung menghasilkan skor reliabilitas rendah. Fenomena tersebut akan

menghasilkan sebaran skor yang cenderung terbatas pada salah satu sisi. Untuk tes yang terlalu mudah skor jawaban siswa akan mengumpul pada sisi atas, misalnya 9 atau 10. Untuk tes yang terlalu sulit, skor jawaban siswa akan cenderung mengumpul pada ujung sebaliknya, atau rendah. Dua gejala tersebut mempunyai kesamaan yaitu bahwa perbedaan diantara individu adalah kecil dan cenderung tidak relevan

4. Objektivitas, yang dimaksud dengan objektif yaitu derajat dimana siswa dengan kompetensi sama, mencapai hasil sama. Ketika prosedur tes evaluasi memiliki objektivitas tinggi, maka reliabilitas hasil tes tidak dipengaruhi oleh prosedur teknik penskoran. Item tes skor objektif yang dihasilkan tidak dipengaruhi pertimbangan atau opini dari seorang evaluator

### **Kesalahan Pengukuran Standar (Standard Error of Measurement, SEM)**

Penafsiran SEM dilakukan karena tidak adanya prosedur penilaian yang sangat konsisten, interpretasi skor dapat ditingkatkan dengan mempertimbangkan ukuran kemungkinan kesalahan pengukuran. Salah satu cara untuk menggambarkan inkonsistensi nilai dalam penilaian adalah dengan menilai siswa berulang kali dan bukan dengan banyaknya variasi nilai. Ini hanya dapat dilakukan secara hipotesis, namun jika peneliti dapat menilai seorang siswa beberapa kali (tanpa mengubah kemampuan siswa yang berhubungan dengan perlakuan yang dinilai), peneliti akan mendapatkan kumpulan nilai siswa yang diperoleh. Beberapa nilai akan lebih tinggi dari yang lain, tetapi sebagian besar akan berkelompok di sekitar rata-rata nilai. Rata-rata adalah nilai asli yang kita maksud sebelumnya. Standar deviasi adalah standar kesalahan pengukuran (SEM). Yang dipergunakan untuk mengestimasi SEM seperti perbedaan kemungkinan skor siswa yang diperoleh dari nilai asli mereka.

Dalam praktiknya tidak perlu berulang kali menilai siswa, sehingga SEM tidak dihitung dengan benar-benar menilai kembali siswa. Melainkan dengan menggunakan persamaan berikut:

$$SEM = SD_x \sqrt{1 - \text{koefisien reliabilitas}}$$

Dengan  $SD_x$  adalah standar deviasi skor yang diperoleh dari penilaian.