

Konsep Klasik Reliabilitas Instrumen

Ir. Ignas A. Sidik, MBA, DBA

Reliabilitas alat ukur dalam penelitian sosial adalah kondisi prasyarat untuk memperkirakan validitas hasil penelitian. Makalah ini membahas konsep klasik mengenai reliabilitas dan teori-teori dasarnya.

Dalam semua penelitian, reliabilitas memegang peranan yang sangat penting karena reliabilitas adalah syarat dasar yang harus dipenuhi sebelum validitas suatu penelitian sosial dapat diperkirakan (Budihardjo dan Sidik, 1996). Dengan perkataan lain, reliabilitas adalah *conditio sine qua non* dari penelitian (Peter, 1979). Hal ini makin penting artinya pada penelitian dalam ilmu-ilmu sosial, karena sifat pengukurannya yang tidak sematerial pada ilmu-ilmu alamiah (Ghiselli 1964, Thorndike 1967).

Pada dasarnya, reliabilitas adalah kemampuan suatu instrumen pengukur untuk menunjukkan kekonsistenan pengukuran. Misalnya, bila suatu instrumen digunakan untuk mengukur sesuatu dari ciri mental seseorang pada dua titik waktu yang berbeda dan hasil kedua pengukur tadi ternyata tidak jauh berbeda, maka instrumen tersebut adalah reliabel (*test-retest reliability*). Sebagai contoh, seorang psikolog mengukur IQ seorang mahasiswa dan mendapatkan angka 130 dengan instrumen tertentu. Instrumen tersebut dikatakan memiliki reliabilitas yang tinggi jika pada pengukuran di saat yang lain menunjukkan angka yang tidak banyak berbeda dengan pengukuran pertama (Budihardjo dan Sidik 1996).

Secara formil, umumnya reliabilitas suatu alat ukur dalam penelitian ilmu-ilmu sosial didefinisikan dalam bentuk sejauh mana adanya kesalahan acak (*random error*) yang muncul dalam pengukuran yang bersangkutan. Menurut Judd, Smith, dan Kidder (1991), reliabilitas adalah "sejauh mana (suatu alat ukur) bebas dari komponen kesalahan acak" (halaman 51). Definisi lain dari reliabilitas adalah pada penekanan "tingkat variasi non-sistematik dalam (besarnya) nilai-nilai (hasil

pengukuran pada suatu ciri mental tertentu dari seseorang jika ciri tersebut diukur beberapa kali" (Ghiselli, 1964, halaman 218).

Definisi-definisi di atas sebenarnya mengacu pada masalah yang sama, yaitu sejauh mana angka pengukuran yang diperoleh dari penggunaan suatu instrumen tertentu benar-benar mengukur angka sesungguhnya dari ciri mental yang diukur pada seseorang. Menggabungkan definisi-definisi di atas, Peter (1979) mendefinisikan reliabilitas sebagai "sejauh mana (suatu) instrumen-instrumen pengukur bebas dari kesalahan dan karenanya memberikan hasil-hasil yang konsisten" (halaman 6).

Secara konseptual, ada tiga teori utama yang menjelaskan konsep reliabilitas pengukuran pada ilmu-ilmu sosial (Ghiselli, 1964; Nunnally, 1967). Yang pertama adalah **teori angka sejati dan angka salah** (*theory of true and error score*). Di alam semesta, ada angka-angka sejati yang menunjukkan kondisi sebenarnya dari hal yang akan diukur. Angka sejati ini ada, tapi tidak bisa diukur karena munculnya berbagai kesalahan pada proses pengukurannya. Teori ini didasarkan pada dua asumsi. Pertama, bahwa individu atau entitas yang akan diukur memiliki suatu angka sejati tertentu yang tidak berubah sepanjang masa. Asumsi ini disebut juga sebagai kestabilan perilaku atau *the stability of behavior* (Epstein 1980). Kedua, kesalahan-kesalahan yang muncul selama pengukuran bersifat acak dengan ekspektasi besarnya adalah nol. Berdasar kedua teori ini, suatu angka sejati dapat selalu diperkirakan besarnya berdasarkan angka-angka yang didapat dari pengukuran dan perkiraan besarnya kesalahan-kesalahan yang muncul.

Bila seseorang mengukur suatu ciri tertentu, yang diukur sebenarnya bukan ciri tersebut, tetapi hanya jabaran dari sebagian jelajah makna yang mewakili ciri yang diukur tersebut. Jadi, instrumen apa pun yang dirancang dan digunakan untuk mengukur suatu ciri tidak akan pernah mampu mengukur seluruh jelajah ciri tadi, melainkan hanya akan mampu "menangkap" sebagian saja dari jelajah tersebut. Dengan demikian, reliabilitas suatu instrumen menyatakan sejauh mana berbagai instrumen akan menangkap angka-angka di dalam jelajah suatu ciri yang diukur.

Teori kedua adalah **konsep eklektik mengenai angka sejati dan pengukuran paralel**. Menurut teori ini, seorang peneliti sebenarnya menyadari sepenuhnya bahwa angka sejati tidak pernah dapat diukur, tetapi dengan menggunakan berbagai pengukuran paralel akan diperoleh lebih dari satu angka hasil pengukuran yang didapat dari berbagai alat ukur. Maka, angka sejati dapat diperkirakan dengan menghitung angka rata-rata dari berbagai pengukuran tersebut.

Teori ketiga adalah **konsep jelajah pengumpulan sampel** (*the concept of domain sampling*). Bila seseorang mengukur suatu ciri tertentu, yang diukur sebenarnya bukan ciri tersebut, tetapi hanya jabaran dari sebagian jelajah makna yang mewakili ciri yang diukur tersebut. Jadi, instrumen apa pun yang dirancang dan digunakan untuk mengukur suatu ciri tidak akan pernah mampu mengukur seluruh jelajah ciri tadi, melainkan hanya akan mampu "menangkap" sebagian saja dari jelajah tersebut. Dengan demikian, reliabilitas suatu instrumen menyatakan sejauh mana berbagai instrumen akan menangkap angka-angka di dalam jelajah suatu ciri yang diukur.

Sebenarnya, ketiga teori tadi saling melengkapi. Sebagai contoh, Churchill dan Peter (1984) melakukan analisis-meta pada hasil pengukuran reliabilitas dari berbagai penelitian dalam bidang ilmu pemasaran. Walaupun tidak menyatakan secara eksplisit, mereka menggunakan ketiga teori di atas secara bersama-sama. Yang jelas, teori pertama dari ketiga teori di atas masih tetap berlaku hingga saat ini, yakni pendekatan klasik pada pengukuran reliabilitas instrumen.

Pustaka klasik mengenai reliabilitas dimulai sejak dipublikasikannya karya-karya klasik Spearman (1904, 1910). Pembahasan berikut akan menggunakan pendekatan klasik ini (Peter, 1979). Istilah instrumen dalam pembahasan ini mewakili alat-alat ukur verbal yang terdiri lebih dari satu *item* (*multi-item scales*). Dengan menggunakan lebih dari satu *item*, kesalahan-kesalahan pengukuran yang muncul dapat diharapkan untuk saling menghilangkan (*cancelling out*).

Pengukuran dengan lebih dari satu *item* akan menghasilkan beberapa angka pengukuran yang dapat dihitung rata-ratanya (*mean*) maupun variansinya. Pendekatan klasik mengenai reliabilitas ini didasarkan pada pandangan bahwa baik rata-rata (*mean*) maupun variansi (*variance*) angka-angka hasil pengukuran dapat dipecah menjadi dua bagian terpisah. Angka pengamatan tersebut dapat dipecah menjadi angka sejati dan angka kesalahan (*error*), atau

$$(1) \quad X_{\text{pengamatan}} = X_{\text{sejati}} + X_{\text{error}}$$

Menurut teori pertama yang sudah dibahas di muka, angka sejati adalah hasil pengukuran sempurna atas hal yang diukur, namun angka sejati ini tidak pernah bisa diukur. Angka sejati ini diasumsikan sebagai angka rata-rata dari berbagai pengukuran dengan *item* yang berbeda-beda. Dengan demikian, angka kesalahan (*error*) merupakan kenaikan atau penurunan dibandingkan angka sejati tadi yang disebabkan oleh kesalahan pengukuran. Kesalahan pengukuran inilah yang merupakan sumber ketidakreliabelan. Ini dikarenakan berbagai *item* dalam satu instrumen sebenarnya bisa saja mengukur hal-hal yang sebenarnya memang berbeda.

Di sisi lain, variansi dari angka teramati ($X_{\text{pengamatan}}$) juga diasumsikan memiliki komponen sejati dan komponen yang muncul karena kesalahan, atau

$$(2) \quad V_{\text{pengamatan}} = V_{\text{sejati}} + V_{\text{error}}$$

Komponen variansi sejati mencakup semua variansi sistematis. Di sisi lain, komponen variansi karena kesalahan mencakup semua variansi non-sistematis yang muncul secara acak. Variansi sejati tidak mengubah urutan jarak

antarsubjek. Sebagai contoh, subjek A memiliki angka sejati lebih besar daripada subjek B. Karena adanya variasi sistematis, maka baik angka pengamatan A maupun B bisa berbeda dari angka sejatinya masing-masing. Karena sifatnya sistematis, maka variasi sistematis tersebut akan mengubah angka pengamatan tersebut secara proporsional, sehingga urutannya tidak berbeda. Dengan kata lain, walaupun muncul variasi, tapi angka A akan tetap lebih besar daripada angka B.

Sebaliknya, variasi acak dapat mempengaruhi bukan saja besarnya angka pengamatan, tetapi juga urutan subjek. Jadi, makin rendah variasi kesalahan, maka makin tinggi variasi sejati yang sistematis. Dengan kata lain, semakin tinggi variasi sejati, maka semakin tinggi reliabilitas suatu instrumen. Memang, reliabilitas adalah tidak lebih daripada rasio dari variasi sistematis terhadap variasi total, Secara lebih formalnya:

$$(3) \quad r_{tt} = \frac{V_{\text{sejati}}}{V_{\text{pengamatan}}}$$

Persamaan (3) ini tidak dapat dipakai untuk mengukur reliabilitas karena V_{sejati} , per definisi, tidak pernah dapat diukur, tetapi menurut persamaan (2), $V_{\text{sejati}} = 1 - V_{\text{kesalahan}}$, dan $V_{\text{kesalahan}}$ dapat diukur. Dengan demikian,

$$(4) \quad r_{tt} = 1 - \frac{V_{\text{kesalahan}}}{V_{\text{pengamatan}}}$$

bila dikalikan dengan $V_{\text{pengamatan}}$, maka

$$(5) \quad r_{tt} = \frac{V_{\text{pengamatan}} - V_{\text{kesalahan}}}{V_{\text{pengamatan}}}$$

Kedua persamaan terakhir ini memungkinkan diukurnya reliabilitas secara praktis, misalnya dengan menggunakan Anova. Anova dapat memunculkan perkiraan akan besarnya $V_{\text{kesalahan}}$ dengan menghitung angka kuadrat rata-rata dari residual, dan memperkirakan $V_{\text{pengamatan}}$ melalui angka kuadrat rata-rata dari tiap subjek.

Daftar Pustaka

1. Budihardjo, Andreas, dan Sidik, Ignas G. (1995). "Validitas Instrumen," *Forum Manajemen Prasetiya Mulya*, Tahun ke-IX, No. 61.
2. Churchill, Gilbert A., Jr., dan Peter, J. Paul (1984). "Research Design Effect on the Reliability of Rating Scales: A Meta Analysis," *Journal of Marketing Research*, 21 November, h. 360-375.
3. Ghiselli, Edwin E. (1964). *Theory of Psychological Measurement*. New York, NY: McGraw Hill.
4. Judd, Charles M., Smith, Elliot R., dan Kidder, Louis H. (1991). *Research Methods in Social Sciences*, edisi ke-6. Forth Worth, TX: Holt, Reinhart, and Winston.
5. Nunnally, Jum C. (1967). *Psychometric Theory*. New York, NY: McGraw-Hill.
6. Peter, J. Paul (1979), "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research*, 16 Februari, h. 6-17.

Ir. Ignas G. Sidik, MBA, DBA, adalah Faculty Member Sekolah Tinggi Manajemen Prasetiya Mulya.
