

Interpretasi Koefisien Korelasi Skor-Butir dengan Skor Total Uji Kebermaknaan Koefisien Reliabilitas Kr-20 dalam Penelitian Pendidikan dan Psikologi

Kumaidi

Abstract: This article is intended as a supplement to *Ketidaktepatan Penggunaan Validitas Butir dan Koefisien Reliabilitas dalam Penelitian Pendidikan dan Psikologi* article. The author suggests following the interpretation given by earlier measurement specialists to item-total correlation indices as part of item reliability index, instead of item validity index. The index for item validity should be related to some other criteria out of the instrument being developed. Meanwhile, related to the issue of testing some hypothesized reliability coefficient (KR-20 or Cronbach Alpha), the author suggests that F distribution or its related confident interval be used, as developed and suggested by Feldt (1965).

Kata kunci: indeks validitas butir, indeks reliabilitas butir, uji keberartian koefisien reliabilitas KR-20.

Artikel ini ditulis sebagai upaya untuk memperkaya artikel Dali S. Naga berjudul *Ketidaktepatan Penggunaan Validitas Butir dan Koefisien Reliabilitas dalam Penelitian Pendidikan dan Psikologi* (Naga, 2004: 99-106). Untuk itu, dalam artikel ini dikemukakan beberapa aspek yang belum dituliskan guna memberi tambahan informasi kepada sejawat yang membutuhkannya. Penulis berharap tidak timbul kesalahpahaman.

Kumaidi adalah dosen Universitas Muhammadiyah Surakarta.

Penulis sering menemukan kekurangtepatan hal di atas. Oleh sebab itu, penulis mendukung upaya pelurusannya. Rujukan yang disajikan dalam artikel ini dapat menjadi sandaran interpretasi dan penamaan terhadap koefisien korelasi skor-butir yang dihasilkan dari suatu proses analisis butir. Dalam kesempatan terbatas, secara *oral* memang sudah sering disampaikan namun kelangkaan literatur berbahasa Indonesia sering menyusahkan banyak sejawat.

PENGGUNAAN ISTILAH *VALIDITAS BUTIR* UNTUK r_{iA}

Penulis ingin lebih menekankan bahwa r_{iA} yang sering disebut sebagai validitas butir asumsinya adalah daya beda butir (*item discrimination*). Hal ini dapat dijelaskan bahwa apabila r_{iA} positif tinggi, responden yang berskor tinggi pada butir ke-*i*, juga memperoleh skor tinggi pada total skor tesnya (atau dalam matrik Naga disimbolkan dengan *A*) dan, sebaliknya. Dalam bahasa Naga, “jika koefisien korelasi butir-total bernilai tinggi, maka skor tinggi pada butir ke-*i* berpasangan dengan skor tinggi pada responden *A* (yakni banyak responden yang menjawab betul).” Dengan kata lain, r_{iA} tinggi menunjukkan bahwa butir tersebut dapat membedakan responden yang ber-skor tinggi dan yang berskor rendah.

Penamaan r_{iA} bukan sebagai validitas butir tersebut juga dapat dikaitkan dengan interpretasi lain, yaitu lebih dekat dengan (koefisien) reliabilitas. Hal ini dapat dicari rujukannya pada uraian Gulliksen (1950: 375-377), Allen & Yen (1979: 124), Thorndike (1982: 77), Crocker & Algina (1986: 320), dan juga Azwar (1997: 172). Mereka ini mengaitkan r_{iA} sebagai bagian dari indeks reliabilitas butir. Alasannya, sebagaimana dikemukakan lima (kelompok) penulis ini (Gulliksen; Allen & Yen; Thorndike; Crocker & Algina; dan Azwar), r_{iA} tidak terkait (langsung) dengan reliabilitas, tetapi terkait langsung dengan koefisien reliabilitas (KR-20 atau Cronbach Alpha). Sebagaimana diungkapkan Azwar (1999: 172), indeks reliabilitas butir (Azwar menyebutnya sebagai indeks reliabilitas aitem, dipendekkan menjadi *ira*), yang dirumuskan sebagai (notasi disesuaikan dengan Naga) berikut:

$$ira = S_i r_{iA}$$

Komponen yang baru dari rumusan *ira* adalah s_i , yaitu deviasi standar skor butir (aitem, dalam istilah Azwar). Selanjutnya, *ira* memang hanya terkait dengan koefisien reliabilitas (KR-20 atau Cronbach Alpha) dan tidak terkait dengan (koefisien) validitas (berdasar kriteria) atau lainnya.

Secara khusus r_{iA} , terkait dengan S_A atau deviasi standar skor total (notasi A , berasal dari Gambar 1 dalam artikel Naga (2004), yaitu:

$$S_A = \sum ira$$

$$S_A = \sum S_i r_{iA}$$

Mengacu kepada hal ini, rumus koefisien reliabilitas untuk KR-20 atau Cronbach Alpha dapat ditulis ulang menjadi:

$$r_{AA} = \frac{K}{K-1} \left[1 - \frac{\sum S_i^2}{S_A^2} \right]$$

Karena, $S_A^2 = \left[\sum S_i \cdot r_{iA} \right]$ maka $r_{AA} = \frac{K}{K-1} \left[1 - \frac{\sum S_i^2}{\left(\sum S_i \cdot r_{iA} \right)^2} \right]$

Penjelasan ini menunjukkan bahwa tidak ada keterkaitan antara r_{iA} dan koefisien validitas berdasar kriteria, tetapi hanya terkait dengan koefisien reliabilitas untuk KR-20 atau Cronbach Alpha, sehingga disebut bagian dari **ira** (indeks reliabilitas butir). Bagi yang ingin mengkaji lebih mendalam, silahkan membaca beberapa literatur yang dipakai sebagai dasar penjelasan ini. Rumusan terakhir koefisien reliabilitas di atas menunjukkan bahwa dengan asumsi s_i konstan, r_{AA} akan semakin tinggi apabila r_{iA} semakin tinggi. Alasan ini memperjelas kenapa dalam seleksi butir dipakai rumusan bahwa r_{iA} harus di atas ukuran tertentu (biasanya dalam ungkapan koefisien biserial, untuk butir dengan penyekoran 0 dan 1 atau dikotomi, di atas 0,20 atau apabila dipakai koefisien poin biserial di atas 0,30). Pengembang instrumen atau tes dalam hal ini ingin memperoleh instrumen yang memberi skor dengan koefisien reliabilitas KR-20 atau Cronbach Alpha yang tinggi.

Dalam pemahaman lain, r_{iA} pada dasarnya hanyalah terkait dengan konsistensi butir dengan butir lain dalam instrumen atau tes yang ikut dibentuknya. Representasinya adalah berupa homogenitas butir, dan bukan validitas butir, mengingat r_{iA} tidak terkait dengan "... the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores." (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1985: 9).

BAGAIMANA DENGAN VALIDITAS BUTIR?

Uraian selanjutnya adalah upaya menjawab pertanyaan di mana atau bagaimana rumusan validitas butir atau yang terkait dengan itu? Jawabannya telah pula diberikan oleh Azwar (1997), Gulliksen (1950), dan Allen & Yen (1979). Di halaman 174 buku Azwar (1997) berjudul *Reliabilitas dan Validitas* ungkapan validitas butir tersebut diberikan atau di halaman 124 buku Allen & Yen ungkapan senada dapat pula ditemukan. Untuk buku Gulliksen, dapat ditemukan dalam bab yang sama dengan penjelasan r_{iA} di atas.

Menurut Azwar (1997: 174) *iva* (indeks validitas aitem atau butir) atau Allen & Yen (1979: 124) atau Crocker & Algina (1986: 320) disebut *item-validity index* (cetak miring asli Allen & Yen dan juga asli Crocker & Algina) dirumuskan sebagai berikut.

$$iva = S_i \cdot r_{iY}$$

di mana r_{iY} adalah korelasi antara skor butir i sebagaimana uraian sebelumnya dari instrumen atau tes yang dikembangkan dengan skor pada kriteria Y , yaitu tes selain tes yang butir i dibahas atau dianalisis. Tes kriteria ini **bukan** tes yang sedang dikembangkan. Tes ini tes kriteria, sebagaimana dikenal dalam pembahasan validitas berdasarkan kriteria, baik validitas konkuren maupun validitas prediktif. Sebagai contoh, apabila seorang peneliti mengembangkan instrumen atau tes, misalnya Tes Potensi Akademik (TPA) untuk seleksi masuk perguruan tinggi, sebagai kriterianya dapat berupa angka prestasi di sekolah atau hasil UAN yang telah dimiliki peserta ujian sebagai presentasi validitas konkuren atau Indeks Prestasi Semester di universitas sebagai representasi validitas prediktif. Jadi penentuan validitas ini tidak menggunakan diri sendiri sebagai kriteria, sebagaimana sering dipakai oleh banyak peneliti pendidikan dan psikologi selama ini. Dengan penjelasan ini mudah-mudahan peneliti dapat lebih berhati-hati dalam melakukan analisis validitas instrumen, tes, atau bahkan butir.

Selanjutnya koefisien validitas instrumen (berdasarkan kriteria Y) dirumuskan oleh Allen & Yen (1979: 125) sebagai berikut:

$$r_{aY} = \frac{\sum S_i \cdot r_{iY}}{\sum S_i \cdot r_{iA}}$$

di mana, r_{aY} adalah koefisien validitas instrumen yang menghasilkan total skor A dengan kriteria skor tes Y , sedangkan S_i , r_{iY} , dan r_{iA} sama dengan pemahaman sebelumnya. Rumusan ini selanjutnya dapat juga dituliskan sebagai berikut.

$$r_{ay} = \frac{\sum iva}{\sum ira}$$

di mana, *iva* dan *ira* masing-masing adalah indeks validitas butir dan indeks reliabilitas butir.

KEBERMAKNAAN KOEFISIEN RELIABILITAS

Bagian lain dari artikel Naga (2004: 99-106)) adalah kekurangtepatan penerapan strategi uji keberartian koefisien reliabilitas (KR-20 atau Cronbach Alpha). Dalam naskah dan banyak peneliti memakai uji keberartian koefisien re-liabilitas (KR-20 atau Cronbach Alpha) menggunakan uji t (lihat Naga, 2004: 1-8)), sebagaimana biasa diterapkan terhadap koefisien korelasi *product moment*. Pendekatan ini tidak tepat bukan sekadar karena n (jumlah responden) besar, sebagaimana diungkapkan dalam Naga, tetapi juga pertimbangan model distribusi samplingnya atau teori samplingnya (Feldt, 1965; Feldt & Brennan, 1989).

Kritik kekurangtepatan pemakaian model distribusi sampling t di atas, dalam bahasa Feldt & Brennan (1989: 126) adalah: “When test reliability is estimated via coefficient alpha, the sampling theory is not the same as that which applies to product moment coefficients”. Untuk itu ada baiknya dipelajari model distribusi sampling yang dikembangkan dan dianjurkan oleh Feldt (1965). Feldt (1965) menggunakan pendekatan Hoyt (1941) sebagai dasar model KR-20 (atau juga Cronbach Alpha) pada rumusan koefisien reliabilitas. Secara matematis, KR-20 merupakan kondisi khusus (*special case*) dari koefisien reliabilitas (Hoyt, 1941).

Rumusan koefisien reliabilitas Hoyt (KR-20 atau Cronbach Alpha) menggunakan pendekatan analisis varian, yang menemukan bahwa KR-20 “... is algebraically equivalent to a function of the mean square derived from an items-by-examinees analysis of variance” (Feldt, 1965: 358). Dengan menggunakan notasi k sebagai jumlah butir dan n sebagai jumlah responden (*examinees* atau *subjects*) rumus KR-20 dituliskan sebagai berikut.

$$r_{AA} = r_{20} = \frac{MS_s - MS_{i.s}}{MS_s} = 1 - \frac{1}{\frac{MS_s}{MS_{i.s}}}$$

Feldt (1965) menjelaskan bahwa distribusi sampling dari koefisien reliabilitas (KR-20 atau Cronbach Alpha) ini dapat dipandang dalam tiga sudut.

Pertama, seseorang dapat mengkonsepsikannya sebagai distribusi koefisien (reliabilitas) yang dihasilkan dari pengujian (atau pengadministrasian) satu set butir (sebuah tes atau instrumen) kepada suatu sampel responden yang dipilih secara rambang (random) dari suatu populasi responden yang jumlahnya tak terbatas. Konsepsi pertama ini juga dikenal sebagai tipe pengujian 1. *Kedua*, seseorang dapat memandangnya sebagai distribusi dari koefisien (reliabilitas) yang dihasilkan dari pengujian berulang-ulang terhadap sekelompok responden, dan setiap kali pengujian menggunakan set butir (instrumen atau tes) baru yang dipilih secara rambang (random) dari populasi butir sejenis (*population of items*). Konsepsi kedua ini sering disebut sebagai tipe pengujian 2. *Ketiga*, seseorang dapat memandangnya sebagai distribusi sampling dari koefisien (reliabilitas) yang dihasilkan dari sampel rambang butir dan responden sekaligus. Tipe terakhir ini dikenal sebagai tipe 3, dan merupakan tipe yang lebih cocok dengan praktik pengujian atau pengembangan instrumen yang sering dijumpai di lapangan.

Dengan konsepsi seperti itu, Feldt (1965) mengembangkan model matematis dari distribusi sampling yang dipakai untuk menguji keberartian suatu koefisien reliabilitas (KR-20 atau Cronbach Alpha). Dia, menggunakan rumusan KR-20 dalam bentuk analisis varian di atas, mendefinisikan bahwa distribusi sampling KR-20 adalah distribusi sampling dari $[(MS_s)/(MS_{i,s})]$. Selanjutnya, untuk mempersingkat penjelasan, beberapa asumsi yang melandasi distribusi sampling KR-20 ini tidak dijelaskan di sini. Pembaca yang memerlukan dipersilahkan membaca artikel Feldt (1965), yang salah satunya dapat ditemukan di Perpustakaan Fakultas Psikologi UGM.

Berikut ini disajikan uraian ringkas pengembangan model distribusi sampling yang penulis kutip dari Feldt (1965: 361-362). Dalam penjabarannya ditemukan tes statistik sebagai berikut:

$$F_{(n-1),(n-1)(k-1)} = \frac{MS_s}{MS_{i,s}} \cdot \frac{\sigma_e^2}{\sigma_e^2 + k \cdot \sigma_t^2}$$

$$F_{(n-1),(n-1)(k-1)} = F_{ob} \cdot \frac{1}{F_{pop}}$$

Dalam rumusan di atas, Feldt (1965) menjelaskan, ketika σ_t^2 sama dengan nol atau lebih besar dari nol, maka F_{ob}/F_{pop} berdistribusi F dengan derajat kebebasan $(n-1)$ dan $(n-1)(k-1)$. Akibat lanjutannya, probabilitas $P[F_a < F < F_b] = 1 - \alpha$ adalah:

$$P \left[F_a < \frac{F_{ob}}{F_{pop}} < F_b \right] = 1 - \alpha$$

dan

$$P \left[F_a \cdot F_{pop} < F_{ob} < F_b \cdot F_{pop} \right] = 1 - \alpha$$

Karena $r_{20} = 1 - \frac{1}{F_{ob}}$, maka $F_{ob} = \frac{1}{1 - r_{20}}$ selanjutnya

$$P \left[F_a \cdot F_{pop} < \frac{1}{1 - r_{20}} < F_b \cdot F_{pop} \right] = 1 - \alpha$$

Atau setelah disusun ulang menjadi

$$P \left[\frac{1}{F_a \cdot F_{pop}} > 1 - r_{20} < \frac{1}{F_b \cdot F_{pop}} \right] = 1 - \alpha$$

Rumusan terakhir ini dapat disusun ulang menjadi

$$P \left[1 - \frac{1}{F_a \cdot F_{pop}} < r_{20} < 1 - \frac{1}{F_b \cdot F_{pop}} \right] = 1 - \alpha$$

F_{pop} terkait dengan populasi KR-20, yang dapat dinotasikan ρ_{20} , dan analog dengan r_{20} dengan F_{ob} , maka rumusan terakhir berubah menjadi:

$$P \left[1 - \frac{1 - \rho_{20}}{F_a} < r_{20} < 1 - \frac{1 - \rho_{20}}{F_b} \right] = 1 - \alpha$$

Rumus terakhir inilah yang dimaksudkan sebagai model distribusi sampling dalam pengujian kebermaknaan koefisien reliabilitas (KR-20 atau Cronbach Alpha), yang menguji r_{20} terhadap ρ_{20} yang dihipotesiskan. Distribusi F yang dimaksud di sini merupakan F dengan derajat kebebasan penyebut $(n-1)$ dan pembilang $(n-1)(k-1)$, di mana n adalah banyak responden, sedangkan k banyak butir pada instrumen atau tes yang dikaji koefisien reliabilitasnya.

KESIMPULAN

Mengikuti tradisi dan pendapat ahli pengukuran (*measurement specialists*) terdahulu, r_{iA} sebaiknya jangan diinterpretasikan sebagai validitas butir, tetapi sebagai daya beda butir (*item discrimination*). Apabila yang per-

tama ini belum memuaskan, paling jauh r_{iA} dapat diinterpretasikan sebagai bagian dari reliabilitas butir (*ira*). Apabila peneliti pendidikan dan psikologi menginginkan validitas butir, dapat dipakai *iva* atau dengan jalan mengungkapkan bagian dari *iva* yaitu r_{iY} . Dalam pengujian kebermaknaan koefisien reliabilitas (KR-20 atau Cronbach Alpha) sebaiknya dihindari penggunaan analogi terhadap korelasi *product moment* (diuji dengan t-test), tetapi menggunakan distribusi F yang memiliki derajat kebebasan $(n-1)$ dan $(n-1)(k-1)$ atau memakai bentuk rentang kepercayaannya sebagaimana dianjurkan Feldt (1965).

DAFTAR RUJUKAN

- Allen, M.J. & Yen, W.M. 1979. *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole Publishing Company.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. 1985. *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Azwar, S. 1997. *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Crocker, L. & Algina, J. 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart, and Winston.
- Feldt, L.S. 1965. The Approximate Sampling Distribution of Kuder-Richardson Reliability Coefficient Twenty. *Psychometrika*, 30: 357-370.
- Feldt, L.S. & Brennan, R.L. 1989. Reliability. Dalam R.L. Linn (Ed.), *Educational Measurement* (3rd Ed., hlm. 105-146). New York: American Council on Education.
- Gulliksen, H. 1950. *Theory of Mental Tests*. New York: Wiley.
- Hoyt, C.J. 1941. Test Reliability Estimated by Analysis of Variance. *Psychometrika*, 6: 153-160. Juga tersedia dalam W.A. Mehrens & R.L. Ebel (Eds.), *Principles of Educational and Psychological Measurement: A Book of Selected Readings* (hlm. 108-115). Chicago, Ill: Rand McNally & Company.
- Naga, D.S. 2004. Ketidaktepatan Penggunaan Validitas Butir dan Koefisien Reliabilitas dalam Penelitian Pendidikan dan Psikologi. *Jurnal Ilmu Pendidikan*, 11 (2): 99-106.
- Thorndike, R.L. 1982. *Applied Psychometrics*. Boston, MA: Houghton Mifflin Company.