

Catatan Pada Butir dengan Korelasi Butir-Total yang Tinggi

Oleh : Wahyu Widhiarso
Fakultas Psikologi UGM 2010

Baru saja saya membaca pemikiran Pak Boyle (Prof. Gregory Boyle), kolega Fakultas Psikologi UGM dari University of Bond Australia. Meski tidak pernah berkomunikasi dengan intens pada waktu beliau berkunjung ke Fakultas, saya membaca beberapa tulisannya. Salah satunya berjudul "*Does item homogeneity indicate internal consistency or item redundancy in psychometric scales?*" yang dimuat dalam jurnal cukup ternama *Personality and Individual Differences*. Tulisan ini membahas tentang homogenitas butir dalam pengembangan pengukuran psikologi. Sebelum saya menceritakan apa yang beliau tulis, saya memulai menjelaskan pengertian homogenitas butir, tradisi dalam analisis butir yang biasa dilakukan oleh peneliti psikologi, baru merefleksikan pemikiran Pak Boyle. Saya menulisnya dengan bahasa santai, karena ini sebenarnya rangkuman pribadi saya saja untuk memudahkan memahami apa yang dipaparkan olehnya.

Homogenitas Aitem

Homogenitas butir menunjukkan kesamaan butir dalam mengukur atribut ukur. Homogenitas butir biasanya ditunjukkan dengan korelasi atau kovarian antar butir. Korelasi antar butir yang tinggi menunjukkan bahwa butir tersebut mengukur atribut yang sama. Sebaliknya butir yang memiliki korelasi antar butir yang rendah butir-butir tersebut adalah heterogen, alias memiliki target atribut ukur yang berbeda-beda. Istilah homogen dan heterogen ini terkadang disetarakan dengan istilah unidimensi dan multidimensi. Tujuannya sama saja, yaitu menggambarkan target atribut ukur.

Aitem-butir yang homogen mengukur satu dimensi atribut (unidimensi), sedangkan butir-butir yang heterogen mengukur banyak dimensi (multidimensi). Meski memiliki kesetaraan ada hal-hal yang berbeda yang ditekankan oleh masing-masing istilah, misalnya homogenitas lebih menekankan pada konsistensi internal. Konsistensi internal konsistensi butir dengan butir lainnya dalam satu alat ukur. Jika butir-butir dalam alat ukur dibelah menjadi beberapa bagian, maka jika bagian-bagian itu memiliki hubungan yang erat, maka alat ukur tersebut memiliki konsistensi internal yang baik. Unidimensi dan multidimensi lebih menunjukkan arah atau target ukur, kalau targetnya tunggal maka dikatakan unidimensi. Internal konsistensi atau homogenitas butir sering digunakan untuk mengestimasi reliabilitas intra skala melalui nilai varian dan kovarian butir berasal alat ukur yang mengukur atribut tunggal. Dalam hal ini Pak Boyle lebih menekankan catatannya pada homogenitas butir.

Catatan Pak Boyle dan Pendapat Pribadi

Pak Boyle mengutip pendapat Hattie (1985) bahwa nilai koefisien alpha bisa tinggi meskipun dikenakan pada pengukuran yang memiliki tidak memiliki dimensi tunggal, karena beberapa hal (1) nilai koefisien alpha meningkat seiring dengan jumlah dan pengulangan paralel butir, (2) nilai koefisien alpha meningkat seiring meningkatnya jumlah faktor yang berkaitan dengan masing-masing meningkat butir, dan (3) nilai koefisien alpha menurun ketika nilai peningkatan komunalitas butir mengalami peningkatan. Nilai komunalitas secara tidak langsung menunjukkan sumbangan butir terhadap faktor ukur.

Saya setuju dengan pendapat ini berdasarkan pengalaman saya. Dengan menggunakan data dari penelitian mahasiswa, saya mendapatkan nilai koefisien alpha yang besar meski alat ukur yang dikembangkan adalah multidimensi. Lalu saya menyimpulkan bahwa koefisien alpha tidak sensitif terhadap pengukuran multidimensi. Hal ini terjadi jika butir yang dilibatkan cukup banyak (>15 butir), tapi jika butirnya sedikit, koefisien alpha cukup sensitif. Untuk poin yang ketiga (3) pendapat Hattie, saya juga menemukan bahwa semakin tinggi kontribusi terhadap faktor ukurnya, nilai koefisien alpha melemah. Dan memang Cronbach telah menyiapkan koefisien tersendiri untuk data yang bersifat multidimensi seperti ini, yaitu Koefisien Alpha Berstrata (*alpha stratified coefficient*).

Dilanjutkan oleh Pak Boyle bahwa dengan mengutip pendapat Kline (1986) yang mengatakan bahwa butir yang memiliki korelasi butir-total yang rendah (<0.3) pasti mengukur sesuatu yang berbeda, namun jika terlalu tinggi (>0.7) maka butir tersebut terlalu mengerucut dan terlalu spesifik. Validitas yang maksimal adalah ketika butir-butir memiliki korelasi butir-butir tidak memiliki korelasi yang tinggi dengan sesama butir namun justru memiliki korelasi yang tinggi dengan kriterianya. Namun karena rendahnya korelasi antar butir, butir ini dicap memiliki konsistensi internal yang rendah.

Membaca ini saya menjadi berpikir, benar juga ya. Korelasi item 1 dengan item 2 yang sangat tinggi kemungkinan menunjukkan tumpang tindih (*redundancy*), sehingga bisa saja titik tembaknya sama dalam domain ukur. Kalau sasaran tembaknya sama, ngapain butir salah satu dipertahankan? Mbok cari butir lain yang memiliki sasaran lain dalam domain ukur yang sama. Agar alat ukur kita bisa menjangkau semua wilayah dalam domain ukur secara komprehensif. Aitem yang memiliki korelasi yang tinggi, perlu dikaji kembali kontennya, bisa jadi dua butir tersebut sama saja tapi beda wordingnya. Misalnya “saya termasuk anak pandai” dan “saya anak cerdas”, pasti butir ini memiliki korelasi yang tinggi. Kata Cattell (1978), butir seperti ini adalah pengulangan saja (*repetition*).

Diceritakan oleh Pak Boyle bahwa tingginya konsistensi internal bisa berlawanan dengan tingginya validitas. Jika skala isinya butir-butir pengulangan melulu pastilah nilai konsistensinya tinggi, tapi validitasnya rendah. Akhirnya Pak Boyle mengecam koefisien alpha yang memang merupakan koefisien yang menunjukkan konsistensi internal. Dimulai dari mengutip pendapat Kline, ahli psikometri kawakan Gulliksen, McDonald dan Lord. Kline menulis, ah internal konsistensi itu hanyalah sesuatu yang dibesar-besarkan dalam bidang psikometri (Kline, 1986). Lalu McDonald (1981) mengatakan bahwa koefisien alpha tidak dapat dipakai dalam mengestimasi reliabilitas, tetapi mengestimasi konsistensi internal dan homogenitas butir.

Gulliksen (1950) dan Lord (Lord & Novick, 1968) menjelaskan dengan sedikit halus: koefisien alpha hanya tepat dikenakan pada pengukuran dengan model nilai tau setara saja (*tau-equivalent*). Model pengukuran nilai tau setara adalah model yang mengasumsikan bahwa setiap butir memiliki kapasitas ukur yang sama terhadap atribut ukur akan tetapi memiliki eror pengukuran yang berbeda. Sebagai kritiknya terhadap koefisien alpha Pak Boyle menulis : “...it may often be more appropriate to regard estimates such as the alpha coefficient as indicators of item redundancy and narrowness of a scale” (Boyle, 1985). Wah, disimpulkan olehnya bahwa koefisien alpha bukanlah estimator reliabilitas, akan tetapi estimator redundancy dan kesempatan target ukur skala.

Trus bagaimana yang benar Pak? “*Begini Mas Weje*”, katanya, “*Items should be selected which are loaded maximally by the factor representing that scale, but which exhibit moderate to low item inter-correlations in order to maximise the breadth of measurement of the given factor*”.

“Pakai Bahasa Indonesia aja Pak”, protesku. “Okay, butir yang harus dipilih adalah butir yang secara maksimal menjelaskan faktor ukur, akan tetapi yang memperlihatkan korelasi antar item yang moderat hingga rendah untuk memaksimalkan luasnya area pengukuran dari faktor-faktornya. Untuk menganalisis ini, anda dapat melakukannya dengan menggunakan analisis faktor model konjenerik (*congeneric*). Model konjenerik adalah model yang mengasumsikan bahwa setiap butir dalam alat ukur memiliki kapasitas ukur yang berbeda dan nilai eror yang berbeda pula. Untuk menganalisisnya kita dapat menggunakan pendekatan persamaan model struktural (SEM).

Diceritakan oleh Pak Boyle bahwa validitas yang tinggi akan diikuti dengan reliabilitas yang tinggi pula. Tetapi reliabilitas tinggi belum tentu validitasnya tinggi. Hal ini dibuktikan dengan reliabilitas yang akan menaik hingga level tinggi hanya dengan menambahkan butir-butir yang tidak seberapa valid, maupun pengulangan butir-butir yang mirip dengan butir-butir yang lain. Reliabilitasnya meningkat, akan tetapi validitasnya tetap.

Keuntungan dari butir dengan jangkauan atau diversitas yang luas terlihat dari analisis regresi berganda ketika sumbangan efektif terhadap kriteria banyak dikontribusikan oleh butir-butir yang bervariasi.

Membaca ini saya membayangkan jika butir yang sempit (*narrow*) adalah air 10 ml dan butir yang luas adalah 100 ml, makamemasukkan air sebanyak 100 ml ke dalam gelas akan mampu memenuhi gelas dibanding dengan memasukkan 10 ml saja. Sumbangan efektif air 100ml lebih besar dibanding dengan 10ml. Cattell menambahkan lagi bahwa butir yang memiliki jangkauan lebih luas mampu menggambarkan berbagai sampel perilaku dari atribut ukur. Namun demikian apabila kita jangkauan butir terlalu luas bisa jadi akan masuk faktor-faktor baru yang tidak relevan dengan atribut ukur kita.

Sebagai penutup, mengutip saran dari Cattell, Pak Boyle menyarankan agar mendapatkan butir yang memiliki jangkauan yang luas akan tetapi tetapi valid, maka diharapkan melakukan beberapa hal 1) memahami kompleksitas dan nature dari domain ukur dengan baik, 2) memahami kompleksitas teori-teori psikometri dan 3) analisis faktor pada banyak butir untuk mendapatkan butir yang kita inginkan. Saya cukup tercengang setelah membaca ini. Biar tulisan ini tidak terlalu dramatis, saya mempertegas bahwa ketercengangan saya terlalu karena beberapa pendapat ahli yang dikutip oleh Pak Boyle juga saya pakai sebagai referensi penelitian saya mengenai koefisien alpha sebelumnya.

REFLEKSI PRIBADI

Pelajaran yang saya dapatkan setelah membaca artikel ini adalah bahwa korelasi butir yang terlalu tinggi tidak selalu sempurna, ia bisa juga memiliki kelemahan yaitu bisa saja overlap dengan butir lainnya. Tapi mahasiswa saya beda Pak! Saya mengajarkan pada mahasiswa saya bahwa meskipun kita melipatgandakan jumlah butir yang diujicobakan untuk mengantisipasi butir-butir yang gugur, saya menyarankan kepada mereka untuk tidak melakukan repetisi. Jadi semua butir yang ditulis oleh mahasiswa saya adalah butir yang benar-benar memiliki keunikan sendiri-sendiri akan tetapi masih dalam domain ukur yang ditargetkan.

Yang belum saya ajarkan kepada mahasiswa saya adalah mengenai model pengukuran konjenerik dengan mendalam. Saya melihat bahwa dari sebagian besar alat ukur yang mereka buat adalah model konjenerik (sebagai konsekuensi dari butir yang tidak redundant), namun

estimasi reliabilitas yang mereka gunakan langsung mengarah kepada koefisien alpha yang sebenarnya cocok untuk model nilai tau setara.

Dari tulisan Pak Boyle dapat saya tarik beberapa kesimpulan, yaitu :

1. Butir yang memiliki korelasi butir-total yang tinggi (>0.7) agar diperiksa tata tulis dan konten semantiknya, apakah ia tumpang tindih dengan aitem lainnya. Apabila ditemukan bahwa ia memiliki konten yang overlap dengan aitem lainnya, maka kita pertimbangkan untuk menggugurkannya.

2. Lebih baik menggunakan istilah bahwa koefisien alpha sebagai koefisien konsistensi internal dibanding koefisien reliabilitas. Hal ini dikarenakan masih ada beberapa ahli yang belum setuju.

3. Sebelum melakukan analisis aitem, lebih baik mengawalinya dengan analisis faktor untuk melihat struktur data. Jika terbukti data kita adalah heterogen, maka estimasi reliabilitas hendaknya dilakukan secara terpisah berdasarkan faktornya.

REFERENSI

Karena artikel ini hanyalah rangkuman, semua kutipan pada artikel diambil dari tulisan Pak Boyle.

Gregory J. Boyle. "Does item homogeneity indicate internal consistency or item redundancy in psychometric scales?" *Personality and Individual Differences* (1991). Available at: http://works.bepress.com/greg_boyle/12