

SELEKSI AITEM DALAM PENYUSUNAN SKALA PSIKOLOGI

Saifuddin Azwar

Salah satu langkah pokok dalam penyusunan dan pengembangan skala psikologi adalah penulisan dan pemilihan aitem. Hal ini menjadi sangat penting dikarenakan kualitas skala psikologi sangat tergantung pada kualitas aitem-aitemnya.

Penulisan aitem dilakukan dengan berpedoman pada *blue-print* skala dan dibimbing oleh kaidah-kaidah penulisan aitem yang berlaku bagi setiap jenis dan format instrumen yang sedang disusun. Aitem-aitem yang ditulis mengikuti *blue-print* dan bimbingan kaidah penulisan yang benar akan berfungsi sebagaimana seharusnya dan akan lebih mendukung validitas skala secara keseluruhan.

Artikel ini ditulis terutama karena banyaknya pertanyaan mengenai cara seleksi aitem yang diajukan oleh para mahasiswa S1 dan S2 yang sedang melakukan penelitian untuk skripsi dan thesis mereka. Karenanya akan dibahas salah-satu prosedur analisis aitem yang sangat populer di gunakan dalam memilih aitem-aitem untuk skala psikologi. Pendekatan yang digunakan berorientasi pada teori tes klasikal, tidak mengacu pada teori respons aitem (*Item Response Theory - IRT*) dikarenakan fasilitas *software* untuk prosedur teori respons aitem masih langka bagi praktisi penelitian, terutama bagi para mahasiswa.

Prosedur seleksi aitem menyangkut beberapa tahap kerja. Prosedur yang paling sederhana meliputi dua tahap berikut ini. Tahap pertama adalah seleksi aitem berdasarkan evaluasi kualitatif. Evaluasi ini melihat apakah aitem yang ditulis sudah sesuai dengan *blue-print* dan indikator perilaku yang hendak diungkapnya, melihat apakah aitem telah ditulis sesuai dengan kaidah penulisan yang benar, dan melihat apakah aitem-aitem yang ditulis masih mengandung *social desirability* yang tinggi. Evaluasi dan seleksi aitem dalam tahap ini dikerjakan oleh suatu panel ahli. Sedapat mungkin panel ini terdiri atas ahli pengukuran (psikometri) dan ahli dalam masalah atribut yang hendak diukur oleh skala yang sedang disusun. Apabila prosedur tahap pertama selesai dan telah diperoleh kumpulan aitem dalam jumlah yang cukup (umumnya tiga kali lebih banyak jumlahnya daripada jumlah aitem yang direncanakan akan dijadikan skala) maka kumpulan aitem ini dikompilasikan dalam bentuk semifinal dan siap diujicobakan secara empiris (*field-tested*) pada kelompok subjek yang seharusnya.

Tahap ke dua adalah seleksi aitem berdasarkan data empiris (data hasil ujicoba skala pada kelompok subjek yang setara dengan subjek yang hendak dikenai skala itu nantinya) dengan melakukan analisis kuantitatif terhadap parameter-parameter aitem. Pada tahap ini paling

tidak akan dilakukan seleksi aitem berdasarkan daya diskriminasinya. Seleksi aitem yang lebih lengkap akan meliputi analisis validitas dan reliabilitas aitem, distribusi jawaban, dan aplikasi analisis faktor. Pada tahap yang lebih lanjut, terdapat prosedur analisis *item bias*, fungsi informasi aitem, dan lain-lain.

Berapa Besar Sampel?

Dalam kaitannya dengan pengambilan data empiris, pertanyaan pertama yang seringkali diajukan oleh penyusun skala adalah mengenai berapa banyak subjek yang diperlukan sebagai sampel. Jawaban terhadap pertanyaan seperti itu tidak dapat diberikan secara tegas.

Perlu diperhatikan bahwa parameter-parameter aitem yang diperoleh dari data empiris ini nanti akan menentukan apakah suatu aitem dapat dianggap layak menjadi bagian skala ataukah harus mengalami revisi lebih dahulu, atau malah harus dibuang sama sekali karena kualitasnya tidak memenuhi syarat. Apabila parameter yang kita peroleh tidak akurat, tentu kita akan keliru dalam menentukan aitem-aitem yang harus diambil dan yang harus dibuang. Apalagi bila mengingat bahwa parameter aitem yang dihitung dengan rumusan teori klasik merupakan parameter yang *sample dependent*, artinya statistiknya sangat tergantung pada sampel mana yang kita gunakan. Bila sampelnya berbeda, maka parameternya akan berbeda pula walaupun aitemnya tetap sama.

Oleh karena itu, untuk memperoleh parameter yang cukup akurat dan stabil antar kelompok sampel, hendaknya sampel yang digunakan dalam ujicoba ini berjumlah besar dan skornya cukup bervariasi (heterogen).

Berapa besar? Sebagai suatu gambaran umum, menggunakan 200 orang sebagai sampel biasanya sudah cukup memadai (Crocker & Algina, 1986). Dapat juga digunakan pedoman kasar, yaitu banyaknya subjek untuk sampel adalah 5 sampai 10 kali lipat banyaknya aitem yang hendak dianalisis (Nunnally, 1967 dalam Crocker & Algina, 1986).

Jadi, banyaknya subjek yang perlu dijadikan sampel pengujian aitem dapat seratus, dua ratus, empat ratus, bahkan mungkin seribu atau beberapa ribu orang. Prinsipnya, dalam keterbatasan sumber daya yang ada dan mengingat pertimbangan teknis pelaksanaannya, harus tetap diusahakan untuk mengambil subjek dalam jumlah sebesar mungkin. Jumlah subjek yang terlalu sedikit akan mendatangkan keraguan mengenai distribusi skor subjek, mengenai akurasi hasil perhitungannya, dan mengenai kestabilan parameter yang diperoleh.

Parameter Aitem untuk Skala Afektif

Skala afektif adalah skala psikologi yang disusun untuk mengungkap aspek-aspek kepribadian berupa minat, sikap, dan berbagai variabel lain seperti agresivitas, *self-esteem*, *locus of control*, motivasi belajar, kepemimpinan, dan lain sebagainya. Ciri utama skala afektif adalah stimulusnya yang tidak berstruktur (gambar atau kalimat yang mudah difahami akan tetapi tidak dapat langsung diketahui maksud yang tersirat di dalamnya oleh subjek atau penjawab) sehingga jawaban yang diberikan oleh subjek merupakan refleksi atau proyeksi perasaan dan fikiran dalam diri mereka masing-masing. Oleh karena itu respon yang

diberikan oleh subjek tidak dapat diklasifikasikan sebagai jawaban yang benar atau jawaban yang salah.

Dalam seleksi aitem skala afektif, parameter yang paling penting adalah **daya diskriminasi aitem**. Daya diskriminasi aitem (sering diberi nama yang salah tapi kaprah sebagai validitas aitem) adalah sejauh mana aitem mampu membedakan antara kelompok yang memiliki dan tidak memiliki atribut yang diukur.

Misalnya, pada skala agresivitas, maka aitem yang berdaya diskriminasi tinggi adalah aitem yang mampu menunjukkan mana individu atau kelompok subjek yang memiliki agresivitas tinggi dan mana yang tidak. Untuk skala sikap, aitem yang berdaya diskriminasi tinggi adalah aitem yang mampu membedakan mana subjek yang bersikap positif dan mana subjek yang bersikap negatif.

Daya diskriminasi aitem ditunjukkan oleh statistik r_{iy} yang diperoleh dengan menggunakan teknik komputasi korelasi *product-moment* yang dihitung pada data skor aitem dan data skor skala (korelasi aitem-total). Formulanya adalah :

$$r_{iy} = \frac{s_{iy}}{s_i \cdot s_y}$$

- r_{iy} = indeks daya diskriminasi aitem i
- s_{iy} = kovarians skor aitem dan skor total skala y
- s_i = deviasi standar skor aitem i
- s_y = deviasi standar skor total skala y

Perhitungan ini dilakukan bagi masing-masing aitem satu-persatu sehingga diperoleh indeks daya diskriminasi bagi setiap aitem dalam skala.

Apabila koefisien korelasi skor aitem-total itu dihitung pada suatu skala yang berisi hanya sedikit aitem maka sangat mungkin diperoleh koefisien korelasi yang *overestimated* dikarenakan adanya *overlap* antara skor aitem dengan skor total skala (Guilford, 1956). Demi kecermatan analisis, apabila skala terdiri atas kurang dari 30 aitem sebaiknya koefisien korelasi itu dikoreksi dengan rumusan berikut:

$$r_{i(y-i)} = \frac{r_{iy} \cdot s_y - s_i}{[s_y^2 + s_i^2 - 2 r_{iy} s_i s_y]}$$

$$r_{i(y-i)} = \text{indeks daya diskriminasi aitem } i \\ \text{(setelah dikoreksi terhadap overlap)}$$

Perhitungan-perhitungan korelasi ini walaupun banyak akan tetapi tidaklah sulit dan tidak memakan waktu lama untuk dilakukan dengan bantuan program-program komputer seperti SPSS, Microstat, Minitab, SAS, dll.

Memilih aitem

Apabila indeks-indeks daya diskriminasi (r_{iy} atau $r_{i(y-i)}$) sudah diperoleh dari semua aitem, selanjutnya akan ditentukan mana aitem yang gugur dan yang tidak. Indeks daya diskriminasi bergerak dari 0 sampai dengan 1,00 dengan tanda positif atau negatif. Semakin baik daya diskriminasi aitem maka indeksnya semakin mendekati angka 1,00. Indeks yang mendekati angka 0 atau yang memiliki tanda negatif mengindikasikan daya diskriminasi yang tidak baik.

Aitem yang memiliki harga r_{iy} atau $r_{i(y-i)}$ kurang dari 0,30 dapat diinterpretasikan sebagai aitem yang memiliki daya diskriminasi rendah. Batasan ini merupakan suatu konvensi. Penyusun skala boleh menentukan sendiri batasan indeks daya diskriminasi aitemnya dengan mempertimbangkan isi dan tujuan skala yang sedang disusun.

Apabila aitem yang memiliki indeks daya diskriminasi sama dengan atau lebih besar daripada 0,30 jumlahnya melebihi jumlah aitem yang direncanakan untuk dijadikan skala, maka kita dapat memilih aitem-aitem yang memiliki indeks daya diskriminasi tertinggi. Sebaliknya apabila jumlah aitem yang lolos ternyata masih tidak mencukupi jumlah yang diinginkan, kita dapat mempertimbangkan untuk menurunkan sedikit batas kritis 0,30 menjadi, misalnya, 0,25 sehingga jumlah aitem yang diinginkan tercapai. Apabila hal ini tidak juga menolong, maka sangat mungkin kita harus merevisi seluruh aitem-aitem yang indeksnya rendah bahkan menulis aitem-aitem baru sama sekali dan kemudian melakukan *field-testing* kembali.

Indeks daya diskriminasi hendaknya jangan dijadikan satu-satunya pertimbangan dalam seleksi aitem. Pertimbangan lain yang juga harus diperhatikan adalah komposisi aspek atau komponen yang diukur oleh skala yang bersangkutan. Kalau dalam perancangan skala semula digunakan konsep teoretis atau *construct* yang mengatakan bahwa aspek yang hendak diukur terdiri atas beberapa komponen atau aspek yang terpisah maka pemilihan aitem harus mempertimbangkan aspek-aspek tersebut. Perhatikan contoh berikut :

Tabel 1. Model *blue-print* untuk Skala *Coping Strategy*

| Orientasi | Komponen | Muatan | Aitem | Persen |
|-----------|---------------------|--------|-------|--------|
| Problem | cautiousness | 16,3 | 19 | 46,84 |
| | instrumental action | 13,6 | 16 | 39,00 |
| | negotiation | 4,9 | 5 | 14,16 |
| | Jumlah | | 40 | 100,00 |
| Orientasi | Komponen | Muatan | Aitem | Persen |
| Emosi | escapism | 4,1 | 7 | 18,72 |
| | minimization | 8,1 | 15 | 36,99 |
| | self-blame | 5,4 | 9 | 24,66 |
| | seeking meaning | 4,3 | 8 | 19,63 |
| | Jumlah | | 40 | 100,00 |

Dalam penyusunan skala *Coping Strategy* digunakan konsep Folkman & Lazarus (Aldwin & Revenson, 1987) yang membagi *coping strategy* dalam dua tipe orientasi dengan komponen-komponennya seperti disajikan dalam Tabel 1.

Tabel 1. memperlihatkan bahwa skala *coping strategy* direncanakan terdiri atas 80 aitem yang masing-masing 40 aitem untuk tipe orientasi problem dan 40 aitem untuk orientasi emosi. Perbandingan banyaknya aitem pada masing-masing komponen ditentukan oleh muatan (*load*) setiap komponen sebagai hasil dari suatu analisis faktor, sehingga penyusun skala tinggal mengikutinya saja. Apabila muatan ini belum ditentukan, biasanya penyusun skala dapat menyamakan saja muatan bagi setiap komponen bila tidak ada alasan untuk membuatnya berbeda-beda. Sekarang andaikan kita telah membuat aitem dalam jumlah banyak dan telah menghitung indeks daya diskriminasinya masing-masing seperti disajikan dalam Tabel 2.

Pada Tabel 2., aitem yang indeks diskriminasinya diletakkan di antara tanda () merupakan aitem yang akan disisihkan dari skala. Perhatikan bahwa dalam menentukan aitem mana yang akan disisihkan, pertimbangan pertama adalah batasan r_{iy} minimal 0,30. Apabila dalam komponen yang bersangkutan ternyata jumlah aitem yang memenuhi syarat tersebut masih kurang dari jumlah aitem yang direncanakan, maka diambil aitem dalam komponen itu juga yang r_{iy} nya sedikit lebih rendah. Hal ini tampak misalnya pada komponen *minimization* dan *instrumental action*.

Perhatikan juga bahwa meskipun dalam suatu komponen terdapat aitem yang daya diskriminasinya tinggi dalam jumlah lebih daripada yang direncanakan, maka aitem yang dipilih adalah aitem-aitem yang memiliki daya diskriminasi tertinggi sedangkan yang lain akan disisihkan walaupun indeksnya di atas 0,30. Hal ini dicontohkan pada komponen *cautiousness* dan *seeking meaning* di mana aitem yang memiliki r_{iy} sebesar 0,34 tetap tidak diikutsertakan karena jumlah yang diperlukan sudah cukup.

Tabel 2. Indeks Daya Diskriminasi untuk Skala *Coping Strategy*

| Komponen | Aitem direncanakan | Aitem dibuat | r_{iy} |
|---------------------|--------------------|--------------|--------------------------------------------------------------------------------------------------------------|
| Cautiousness | 19 | 24 | (.34) .56 (.32) .45 .44 .64 .65 .51 .55 .66 (.19) .43 .54 .50 .42 .40 .39 .67 .55 .37 .45 .65 (.33) (.21) |
| Instrumental Action | 16 | 20 | (.26) .44 .32 .32 (.11) .45 .47 .38 .29 .39 .42 .40 .33 .51 .29 (.24) .43 .48 (.25) .35 |
| Negotiation | 5 | 10 | .44 .56 (.32) (.19) (.32) (.22) .64 (.22) .38 .57 |
| Escapism | 7 | 11 | (.32) (.34) .54 (.34) .50 .44 (.27) .39 .54 .64 .66 |
| Minimization | 15 | 21 | .26 .65 .32 (.21) .29 .36 .42 (.15) .36 .43 .44 .32 .30 (.20) (.11) .50 .28 (.09) .31 .30 (.07) |
| Self-blame | 9 | 14 | (.08) .43 (.23) .45 (.32) .46 (.29) (.38) .57 .54 .44 .65 .72 (.22) |
| Seeking meaning | 8 | 15 | (.22) .45 .43 (.24) .51 .42 .71 .38 (.34) (.32) (.29) (.35) .46 (.33) .65 |

Dengan cara ini proporsi aitem yang direncanakan akan tercapai, kontrak yang mendasari pengukuran tetap terpelihara, dan kualitas aitem juga terjaga.

Contoh lain adalah dalam perencanaan suatu skala sikap. Dalam skala sikap (dan banyak skala-skala psikologis lain) konsep teoretis yang mendasari objek ukur merupakan suatu konsep tunggal. Merupakan konsep yang menunjuk satu kesatuan atribut, walaupun memiliki beberapa dimensi. Misalnya didapat data seperti pada Tabel 3.

Tabel 3. Indeks daya diskriminasi (koefisien yang dikoreksi $r_{i(x-i)}$) pada skala sikap yang terdiri dari tiga aspek.

Item-Total Statistics

| Aitem | Corrected Item-Total Correlation | Aitem | Corrected Item-Total Correlation | Aitem | Corrected Item-Total Correlation |
|-------|----------------------------------|-------|----------------------------------|-------|----------------------------------|
| A1 | (.1810) | B1 | .3743 | C1 | .5334 |
| A2 | .3970 | B2 | (.2324) | C2 | .6646 |
| A3 | (.1554) | B3 | (.2727) | C3 | .6196 |
| A4 | (.2612) | B4 | .3367 | C4 | .5713 |
| A5 | (.2716) | B5 | .3263 | C5 | .5312 |
| A6 | .3301 | B6 | (.2539) | C6 | .6686 |
| A7 | .3335 | B7 | (.2729) | | |
| A8 | .3154 | B8 | .6201 | | |
| A9 | (.1899) | B9 | .4190 | | |

Pada Tabel 3. tampak bahwa aitem yang semula dibuat adalah sebanyak 9 untuk aspek A, 9 untuk aspek B, dan 6 untuk aspek C. Hasil komputasi indeks daya diskriminasi memperlihatkan bahwa untuk aspek A hanya 4 aitem yang terpilih, untuk aspek B ada 5 aitem, dan untuk aspek C adalah seluruhnya yang berjumlah 6 aitem. Dengan demikian diperoleh 15 aitem.

Dalam kasus seperti ini, dikarenakan masing-masing aspek tidak mempunyai tujuan ukur yang berbeda secara spesifik satu sama lain melainkan merupakan dimensi saja dari satu tujuan ukur yang lebih luas, maka penyusun skala dapat menetapkan saja 15 aitem yang berdaya diskriminasi tinggi itu sebagai aitem final tanpa harus risau mengenai komposisi jumlah aitem dalam setiap aspeknya. Hal ini cukup dapat dipertanggungjawabkan sejauh 15 aitem itu nantinya terbukti memiliki reliabilitas yang memuaskan.

Atribut Tunggal vs Atribut Komposit

Skala psikologi dapat mengukur suatu variabel yang dikonsepsikan sebagai atribut tunggal walaupun memiliki beberapa aspek atau dimensi. Aspek atau dimensi ini merupakan bagian dari suatu kawasan (*domain*) ukur sesuai dengan kontrak yang mendasarinya.

Dari segi prosedur penyusunan atau konstruksinya, pada skala yang dibuat untuk mengukur atribut tunggal seperti ini interkorelasi antar aspek atau dimensi-dimensi itu diharapkan untuk tinggi karena memang mereka dirancang untuk mengukur hal yang sama. Dalam seleksi aitemnya pun kita memilih daya beda aitem tertinggi dengan membandingkan indeksinya secara keseluruhan. Lebih lanjut, untuk pengujian reliabilitasnya cukup dilakukan satu pengujian saja bagi seluruh aitem.

Di sisi lain, ada skala psikologi yang berisi beberapa aspek yang berbeda satu sama lain akan tetapi skor-skor dari setiap aspek itu kemudian dijadikan satu skor komposit yang mengindikasikan satu konsep tujuan ukurnya.

Contoh yang paling jelas dalam hal ini adalah WAIS (skala inteligensi dewasa ciptaan Wechsler). WAIS bertujuan untuk mengukur IQ. IQ sendiri disimpulkan dari inteligensi yang konsepsinya terdiri atas 11 aspek kecakapan yang berbeda-beda. Tidak satupun di antara aspek kecakapan itu yang dinamai inteligensi. Setelah skor dari setiap aspek diperoleh dan dikompositkan sedemikian rupa, barulah skor akhirnya disebut IQ yang mencerminkan inteligensi sebagaimana tujuan ukur semula.

Dalam hal pengukuran atribut komposit seperti ini, kita mengharapkan agar interkorelasi antar aspek atau dimensi itu rendah karena hal itu berarti bahwa setiap aspek memiliki fungsi ukur yang unik dan tidak ada tumpang tindih. Dari segi pemilihan aitemnya, kita harus melakukan analisis aitem bagi setiap aspek dengan membandingkan indeks daya diskriminasinya dalam aspek masing-masing, bukan secara keseluruhan. Begitu juga dalam menguji reliabilitasnya diperlukan komputasi reliabilitas bagi masing-masing aspek, baru kemudian dihitung reliabilitas secara keseluruhan yang dikenal dengan nama reliabilitas skor komposit.

Penutup

Sebagaimana telah dijelaskan, seleksi aitem dengan menggunakan komputasi korelasi antara skor aitem dan skor total skala menghasilkan indeks daya diskriminasi aitem atau dikenal juga dengan indeks konsistensi aitem-total. Perlu diingatkan bahwa indeks daya diskriminasi ini tidak sama dengan koefisien validitas aitem. Daya diskriminasi aitem dan validitas aitem merupakan dua hal yang berbeda dan jangan dicampuradukkan (dalam artikel yang akan datang akan diuraikan penjelasan mengenai validitas dan reliabilitas aitem). Indeks diskriminasi aitem semata-mata menunjukkan sejauhmana aitem yang bersangkutan berfungsi sama seperti fungsi skala. Indeks daya diskriminasi aitem yang rendah berarti bahwa fungsi aitem yang bersangkutan tidak selaras dengan tujuan ukur skala.

Suatu skala yang seluruhnya berisi aitem dengan indeks diskriminasi tinggi berarti bahwa skala itu merupakan kumpulan aitem yang memiliki tujuan dan fungsi yang sama, tapi hal itu belum menunjukkan fungsi apa yang sebenarnya dimilikinya. Artinya, suatu skala yang daya diskriminasi aitemnya tinggi, belum tentu valid untuk tujuan ukur yang direncanakan. Daya diskriminasi aitem tidak memiliki hubungan langsung dengan validitas skala.

Untuk menjawab apakah skala itu valid, masih diperlukan pengujian dengan menggunakan kriteria validasi berupa ukuran-ukuran atau skor variabel lain yang relevan (*criterion-related validity*). Dengan demikian, diharapkan para penyusun skala dan para peneliti yang menggunakan skala akan dapat melakukan evaluasi kualitas aitem dan kualitas skalanya dengan cara yang tepat dan menempatkan prosedur seleksi aitem pada proporsi yang selayaknya.

Kepustakaan

- Aldwin, C.M. & Revenson, T.A. (1987). Does Coping Help? A Reexamination of The Relation between Coping and Mental Health, *Journal of Personality and Social Psychology*, 53, 337-348.
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*, Forth Worth: Holt, Rinehart, and Winston, INC.
- Guilford, J.P. (1956). *Fundamental Statistics in Psychology and Education (3rd edition)*, New York: McGraw-Hill.