

EFEK SELEKSI AITEM BERDASAR DAYA DISKRIMINASI TERHADAP RELIABILITAS SKOR TES

Saifuddin Azwar

Fakultas Psikologi Universitas Gadjah Mada
E-mail: sfazwar@ugm.ac.id

Abstract

Seventy three items of IPA (science) test used in the 2006 University of Gadjah Mada under-graduate admission procedure (Utul UM UGM) were analyzed to reveal the effect of eliminating items based on corrected item-total correlation coefficient on reliability of the test.

Computations of reliability estimates for four different lengths of test ($n=11.287$) showed that reducing less discriminating items did not lower reliability coefficients as predicted by Spearman-Brown prophecy. In fact, length of test was practically irrelevant to reliability coefficient when the test consisted of highly discriminating items.

Keywords: Spearman-Brown Prophecy, Reliability, Item-total correlation

Dalam pengembangan tes sebagai instrumen pengukuran untuk riset psikologi dan pendidikan, seleksi aitem berdasarkan statistik daya diskriminasi aitem merupakan salah-satu tehnik guna meningkatkan reliabilitas skor tes (Azwar, 1997). Hal tersebut adalah penting terutama dalam pendekatan teori skor-murni klasik (Allen & Yen, 1979; Crocker, & Algina, 1986). Salah-satu statistik daya diskriminasi aitem adalah koefisien korelasi antara skor aitem dan skor total tes yang dikenal dengan nama korelasi aitem-total (r_{ix}) dan koefisien r-point biserial (r_{pbis}).

Sudah diketahui bahwa estimasi terhadap perubahan reliabilitas yang disebabkan oleh perubahan panjang tes dapat dilakukan dengan komputasi formula *Spearman-Brown prophecy* (Thompson, 2003) atau formula *stepped-up reliability* (Allen & Yen, 1979), yaitu:

$$r_{xx'} = k(r_{xx}) / (1 + (k-1)r_{xx})$$

r_{xx} = reliabilitas tes sebelum perubahan jumlah aitem

$r_{xx'}$ = reliabilitas tes setelah perubahan jumlah aitem
 k = perbandingan jumlah aitem setelah dan sebelum perubahan

yang menghasilkan kenaikan reliabilitas bila harga k (yaitu rasio banyaknya aitem tes setelah perubahan dibanding sebelum perubahan) lebih besar daripada 1.

Estimat perubahan reliabilitas tersebut tidak bersifat linear dengan perubahan besarnya k dan akurasinya tergantung pada terpenuhi-tidaknya asumsi paralelisme antar aitem yang ada. Dalam kasus estimasi reliabilitas belah-dua, misalnya, koefisien alpha yang dihasilkan akan terhindar dari underestimasi bila kedua belahan bersifat τ -equivalent (Allen & Yen, 1979) sedangkan koefisien Spearman-Brown (S-B) akan akurat bila kedua belahan bersifat paralel.

Dalam hal penambahan jumlah aitem, yaitu menambah panjang tes, peningkatan reliabilitas dapat diprediksi secara akurat bila aitem-aitem yang ditambahkan berkualitas setara dengan aitem-aitem yang sudah

ada dalam tes. Bila aitem yang ditambahkan ke dalam tes berkualitas lebih rendah (memiliki daya diskriminasi lebih rendah) dibanding aitem yang sudah ada, maka peningkatan yang diharapkan oleh formula *Spearman-Brown prophecy* tidak dapat tercapai. Begitu pula dalam hal pengurangan jumlah aitem atau pemendekan tes, perubahan reliabilitas tes hanya dapat diprediksi secara akurat bila kualitas aitem yang dikeluarkan dari tes adalah setara dengan aitem-aitem yang ditinggalkan. Bila kualitas aitem yang dikeluarkan dari tes lebih baik dibanding kualitas aitem yang disisakan maka penurunan reliabilitas akan lebih besar daripada prediksi formula *Spearman-Brown prophecy* dan sebaliknya bila kualitas aitem yang dikeluarkan dari tes lebih buruk dibanding kualitas aitem yang disisakan maka penurunan reliabilitas tidak akan sebesar prediksi bahkan mungkin saja menaikkan reliabilitas. Kondisi yang terakhir inilah yang akan diverifikasi secara empiris dalam penelitian ini.

Metode

1. Sumber Data

Data skor bersumber dari 11.287 orang peserta Ujian Tulis (Utul) UM UGM Tahun 2006 yang merespon tes IPA yang soalnya berupa pilihan-ganda dengan lima pilihan. Tes IPA terdiri dari 75 aitem, namun dikarenakan aitem nomor 18 dan nomor 52 tidak diskor maka keseluruhan aitem yang disertakan dalam analisis berjumlah 73 buah. Skor aitem IPA adalah 4 untuk jawaban benar, -1 untuk jawaban salah, dan 0 untuk tidak menjawab namun untuk analisis dalam penelitian ini digunakan data skor yang telah didikotomikan menjadi 1 untuk jawaban yang benar dan 0 untuk jawaban yang salah.

2. Cara Analisis

Formula estimasi reliabilitas alpha (Cronbach, 1951) digunakan untuk menghitung reliabilitas skor ketujuh puluh tiga aitem dalam tes IPA beserta statistik daya diskriminasi aitem masing-masing yang diperlihatkan oleh koefisien korelasi aitem-total. Sekalipun dengan adanya 73 aitem yang dianalisis tidak diperlukan koreksi terhadap *spurious overlap* (Guilford, 1953; Guilford, 1956; Wolf, 1967; Azwar, 1997) namun dalam penelitian ini tetap digunakan korelasi aitem-total yang telah dikoreksi sehingga estimasinya menjadi lebih cermat. Daya beda yang ideal diperoleh bila r_{pbis} atau r_{ix} mendekati angka 1,00 (Scorepak®, 2005). Pada umumnya daya beda dianggap memuaskan bila mencapai angka 0,30 (Azwar, 1997). Angka ini lebih tinggi dibanding rekomendasi Thorndike sebesar 0,20 (Thorndike et. al., 1991) dan rekomendasi ahli lain (Crocker & Algina, 1986) dan masih jauh lebih tinggi daripada yang disarankan oleh Kehoe yaitu 0,15 (Kehoe, 1997).

Untuk melihat efek perubahan panjang tes terhadap reliabilitas dilakukan dengan mengeliminasi secara bertahap aitem-aitem yang tidak mencapai batas daya diskriminasi tertentu. Pertama-tama korelasi aitem-total dan reliabilitas dihitung pada ketujuh puluh tiga aitem. Kemudian, aitem-aitem yang memiliki korelasi aitem-total kurang dari 0,15 dieliminasi dan reliabilitas dihitung kembali. Proses ini dilakukan untuk eliminasi korelasi aitem-total kurang dari 0,20; kurang dari 0,25; dan kurang dari 0,30 mengikuti rekomendasi masing-masing ahli di atas. Pada setiap tahap eliminasi, reliabilitas tes dihitung ulang. Perubahan reliabilitas yang diperoleh dari komputasi dibandingkan dengan perubahan yang diprediksikan oleh formula *Spearman-Brown prophecy*.

Hasil dan Kesimpulan

Skor aitem dari sampel penelitian berukuran $n=11.287$ yang merespon 73 aitem tes IPA menghasilkan koefisien korelasi aitem-total yang bervariasi mulai dari terendah $r_{ix} = -0,063$ sampai tertinggi $r_{ix}=0,462$. Ketujuh puluh tiga aitem tersebut menghasilkan koefisien reliabilitas $r_{xx}=0,833$.

Hasil estimasi reliabilitas pada sisa aitem setelah dilakukan eliminasi bertahap terhadap aitem-aitem yang tidak mencapai kriteria koefisien korelasi aitem-total $r_{ix}=0,15$; $r_{ix}=0,20$; $r_{ix}=0,25$; dan $r_{ix}=0,30$; beserta hasil prediksi Spearman-Brown (S-B) disajikan dalam Tabel 1.

Tampak jelas sebagaimana formulasi Spearman-Brown memprediksikan bahwa secara teoretis pengurangan jumlah aitem (pemendekan tes) akan berakibat penurunan drastis koefisien reliabilitas ketika jumlah aitem menjadi $2/3$ jumlah semula (dari 73 aitem menjadi 48 aitem) dan reliabilitas turun dari $r_{xx}=0,833$ menjadi hanya $r_{xx}=0,426$. Penurunan tersebut secara bertahap terjadi dengan makin sedikitnya jumlah aitem yang disisakan berdasar kriteria eliminasinya sehingga mencapai serendah $r_{xx}=0,193$ sewaktu jumlah aitem hanya tersisa 26 buah. Prediksi pengurangan ini, sekali lagi, akan terjadi secara akurat bila aitem yang dieliminasi dan yang disisakan adalah paralel, atau paling

tidak dalam hal ini memiliki daya diskriminasi yang setara.

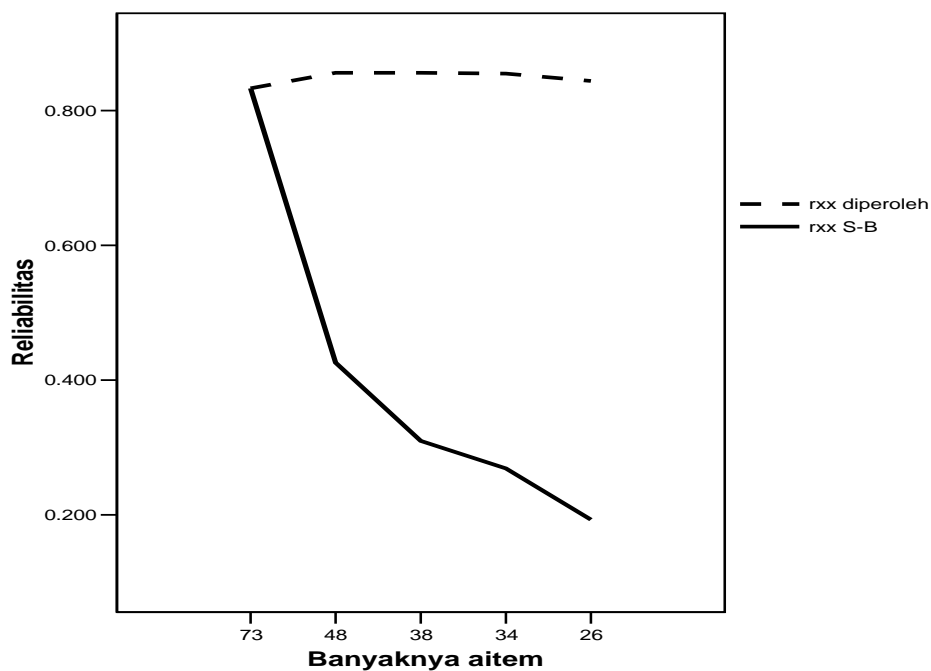
Dalam data skor penelitian ini, pengurangan aitem tidak dilakukan untuk alasan pemendekan tes (efisiensi) dari suatu tes yang berisi aitem homogen (*content homogeneous*) melainkan untuk tujuan seleksi aitem berdasar daya diskriminasi. Dengan demikian, aitem yang dieliminasi adalah aitem-aitem yang dikategorikan sebagai tidak memuaskan. Sebagaimana diduga, dikarenakan aitem-aitem yang diambil memiliki daya beda yang lebih tinggi dibanding aitem-aitem yang dieliminasi maka prediksi penurunan reliabilitas tidak terjadi dan bahkan tampak jelas adanya kenaikan sebesar 0,023 angka seperti tampak pada waktu sisa aitem sebanyak 48, 38, dan 34 buah. Bahkan ketika tes dipendekkan menjadi hanya 26 aitem, peningkatan reliabilitas tetap terjadi sekalipun hanya 0,011 angka, yaitu dari semula $r_{xx}=0,833$ menjadi $r_{xx}=0,844$.

Perbandingan perubahan reliabilitas berdasar seleksi aitem dan berdasar prediksi Spearman-Brown ini secara visual tampak lebih jelas dalam Gambar 1.

Dengan demikian analisis ini menyimpulkan bahwa seleksi aitem berdasar koefisien korelasi aitem-total dapat mempertahankan bahkan meningkatkan reliabilitas skor tes. Tes yang pendek namun berisi aitem-aitem berdaya diskriminasi

Tabel 1
Komputasi $r_{xx'}$ empirik dan $r_{xx'}$ prediksi Spearman-Brown
($n = 11.287$)

Kriteria eliminasi	Banyaknya aitem yang dieliminasi	Banyaknya aitem	k	$r_{xx'}$ diperoleh	$r_{xx'}$ prediksi S-B
-	0	73	1	0,833	0,833
$r_{ix} < 0,15$	25	48	0,657	0,856	0,426
$r_{ix} < 0,20$	35	38	0,521	0,856	0,310
$r_{ix} < 0,25$	39	34	0,466	0,855	0,269
$r_{ix} < 0,30$	47	26	0,356	0,844	0,193



Gambar 1. Perbandingan perubahan reliabilitas berdasar seleksi aitem dan berdasar prediksi Spearman-Brown ($n = 11.287$)

tinggi adalah lebih baik reliabilitasnya dibanding tes yang panjang namun berisi aitem-aitem yang berkualitas rendah. Implikasi lebih lanjut adalah bahwa efisiensi tes dengan mengurangi aitem dapat dipertimbangkan tanpa mengorbankan reliabilitas, bila didasarkan pada pemilihan aitem yang berdaya diskriminasi tinggi.

Daftar Pustaka

- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Azwar, S. (1997). *Reliabilitas dan validitas*. Edisi ke 3. Yogyakarta: Pustaka Pelajar.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Forth Worth: Holt, Rinehart, and Winston, INC.,
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Guilford, J.P. (1953). The correlation of an item with a composite of the remaining items in a test. *Educational and Psychological Measurement*, 13, 1, 87-93.
- Guilford, J.P. (1956). *Fundamental statistics in psychology and education* (3rd ed.). New York: McGraw-Hill.
- Kehoe, J. (1997). Basic item analysis for multiple-choice tests. *ERIC Digest*. <http://www.ericdigests.org/1997-1/basic.html>
- Scorepak® (2005). *Item analysis*. Office of Educational Assessment, University of Washington. http://www.washington.edu/oea/pdfs/resources/item_analysis.pdf
- Thompson, B. (2003). Understanding reliability and coefficient alpha, really. Dalam Bruce Thompson (Ed.), *Score reliability – Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage Publications.
- Thorndike, R.M., Cunningham, G.K., Thorndike, R.L., & Hagen, E.P. (1991).

Measurement and evaluation in psychology and education. New York, NY: Macmillan Publishing Company.

Wolf, R. (1967). Evaluation of several formulae for correction of item-total correlations in item analysis. *Journal of Educational Measurement*, 4, 1, 21-26.