

Penerapan Teori Respons Butir Dalam Penyetaraan Tes

Kana Hidayati
Jurusan Pendidikan Matematika FMIPA UNY

ABSTRAK

Penyetaraan tes perlu dilakukan khususnya bagi kegiatan pengujian dalam skala besar yang mempersiapkan lebih dari satu perangkat tes mengingat bahwa menyusun tes yang benar-benar paralel tidaklah mudah. Kegiatan penyetaraan tes dilakukan dengan mengembangkan konversi suatu system unit tes ke system unit tes yang lain sehingga setelah dikonversi skor yang berasal dari dua atau lebih perangkat tes menjadi setara dan dapat dipertukarkan. Kegiatan ini dapat dilakukan dengan menggunakan teori respons butir (*Item Response Theory/IRT*).

Teori respons butir merupakan teori tes dengan berbagai keunggulan yang mampu mengatasi kelemahan pada teori tes yang berkembang sebelumnya yakni teori tes klasik. Sebagaimana teori tes klasik, teori ini juga sangat berguna terutama dalam hal penentuan estimasi parameter butir maupun parameter kemampuan peserta tes. Selain itu, teori ini ternyata juga dapat diterapkan untuk kegiatan penyetaraan tes.

Penerapan teori respons butir dalam penyetaraan tes harus memenuhi dua asumsi dasar yakni unidimensi dan independensi lokal (*local independence*). Langkah atau proses melakukan kegiatan penyetaraan tes berdasarkan teori respons butir meliputi: (1) pengestimasi parameter, (2) pengestimasi skala *IRT* dengan menggunakan transformasi linier, dan (3) penyamaan skor. Ada 3 rancangan penyetaraan tes menurut teori respons butir yaitu: (1) rancangan kelompok tunggal (*single-group design*), rancangan kelompok ekuivalen (*equivalent-group design*), dan (3) rancangan tes jangkar (*anchor test design*). Metode-metode yang dikembangkan untuk melakukan penyetaraan tes menurut teori respons butir adalah: (1) metode regresi, (2) metode rerata dan sigma, (3) metode rerata dan sigma tegar, dan (4) metode kurva karakteristik.

Kata kunci: Teori respons butir, Penyetaraan tes

PENDAHULUAN

Pada program pengujian, khususnya dalam skala besar, penyusunan beberapa perangkat tes yang setara adalah salah satu kegiatan penting karena salah satu tugasnya adalah menjaga keamanan perangkat tes–perangkat tes tersebut. Pada taraf tertentu kesetaraan beberapa perangkat tes dapat dilaksanakan pada saat mengembangkan tes itu sendiri, tetapi biasanya bervariasi antara perangkat tes yang satu dengan perangkat tes lainnya terutama dalam hal tingkat kesukaran. Hal

ini dapat diatasi dengan melakukan kegiatan penyetaraan skor antar perangkat tes dengan cara yang tepat dan benar. Selain itu, penyetaraan tes perlu dilakukan mengingat bahwa menyusun tes yang benar-benar paralel tidaklah mudah.

Kegiatan penyetaraan tes dapat dilakukan dengan mengembangkan konversi suatu system unit tes ke system unit tes yang lain sehingga setelah dikonversi skor yang berasal dari dua perangkat tes menjadi setara dan dapat dipertukarkan. Kegiatan ini dapat dilakukan dengan menggunakan teori respons butir (*Item Response Theory/IRT*). Penerapan teori respons butir dalam penyetaraan tes sangat berguna terutama bagi pengembangan bank soal. Oleh karena itu, dalam tulisan ini akan dikaji tentang penerapan teori respons butir dalam penyetaraan tes yang meliputi asumsi dasar apa yang harus dipenuhi?, langkah-langkah dan rancangan apa dapat digunakan?, dan metode-metode apa saja yang dikembangkan?.

PEMBAHASAN

Teori Respons Butir (*Item Response Theory/IRT*)

Teori pengukuran yang berkembang saat ini ada dua yaitu teori tes klasik dan teori respons butir (teori modern). Teori tes klasik telah banyak berjasa dalam bidang pengukuran dan bahkan masih digunakan sampai sekarang. Hal ini mengingat keunggulan teori tes klasik yang terletak pada kemudahan dalam pemahaman konsepnya. Namun demikian, oleh karena teori tes klasik memiliki berbagai keterbatasan seperti adanya sifat *group dependent* dan *item dependent* maka munculnya teori respons butir menjadi sangat berguna dan terus dikembangkan karena mampu mengatasi keterbatasan tersebut.

Menurut Hambleton, Swaminathan, & Rogers (1991: 5) secara umum ciri-ciri teori respons butir adalah sebagai berikut: (1) karakteristik butir tidak tergantung pada peserta ujian, (2) skor yang digambarkan peserta ujian tidak tergantung pada tes, (3) merupakan model yang lebih menekankan pada tingkat butir daripada tingkat tes, (4) merupakan model yang tidak mensyaratkan secara ketat tes paralel untuk menaksir reliabilitas, dan (5) merupakan model yang menguraikan sebuah ukuran keputusan untuk tiap skor kemampuan yakni ada

hubungan fungsional antara peserta tes dengan tingkat kemampuan yang dimiliki. Teori respons butir juga dikembangkan atas dasar dua postulat yaitu: 1) performansi subyek pada suatu butir dapat diprediksikan oleh seperangkat faktor yang disebut *latent trait* atau kemampuan dan 2) hubungan performansi subyek pada suatu butir dan perangkat kemampuan laten yang mendasarinya digambarkan oleh fungsi naik monoton yang disebut *Item Characteristic Curve (ICC)*. Selain itu, menurut Hambleton, Swaminathan, & Rogers (1991: 9) mengemukakan bahwa asumsi-asumsi yang melandasi teori respons butir adalah unidimensi, independensi lokal, dan fungsi karakteristik butir atau kurva karakteristik butir.

Pada awalnya teori respons butir menggunakan distribusi normal, namun dalam perkembangan selanjutnya digunakan model distribusi logistik. Hal ini dikarenakan model distribusi logistik lebih sederhana analisis matematikanya (Mardapi, 1991: 7). Ada tiga macam model logistik dalam teori respons butir yaitu model logistik satu parameter, model logistik dua parameter, dan model logistik tiga parameter. Perbedaan ketiga model tersebut terletak pada banyaknya parameter yang digunakan untuk menggambarkan karakteristik butir dalam model yang bersangkutan. (Hambleton, Swaminathan, & Rogers, 1991: 7)

Kelemahan teori respons butir terletak pada tiga hal yaitu: 1) pemahaman, 2) penghitungan, dan 3) asumsi yang harus dipenuhi. Model yang digunakan adalah model statistik sehingga dibutuhkan pengetahuan tentang matematik dan statistik. Analisis butir tidak bisa dilakukan dengan manual tetapi harus menggunakan paket program komputer karena kompleksnya perhitungan. Setiap model menggunakan asumsi yang berbeda walaupun di dalamnya ada asumsi yang sama. Semakin banyak parameter dalam suatu model semakin kecil asumsi yang harus dipenuhi namun penghitungannya semakin kompleks. Selain itu ukuran sampel yang dibutuhkan pada teori ini harus banyak paling tidak 100 orang untuk model Rasch (Mardapi, 1994: 25).

Penyetaraan Tes

Penyetaraan tes merupakan pengembangan konversi suatu system satuan unit tes ke system satuan unit tes yang lain, sehingga setelah dikonversi skor yang

berasal dari dua perangkat tes menjadi setara dan dapat dipertukarkan. (Holland & Rubin, 1982: 56). Hal-hal yang perlu diperhatikan dalam melakukan kegiatan penyetaraan tes adalah sebagai berikut: a) perangkat tes yang mengukur sifat dan kemampuan yang berbeda tidak dapat disetarakan, b) skor mentah dari perangkat tes yang tidak sama reliabilitansya hendaknya jangan disetarakan, c) skor mentah dari perangkat tes yang tingkat kesukaran butirnya bervariasi tidak dapat disetarakan, c) skor pada perangkat tes X dan Y tidak dapat disetarakan tanpa adanya bukti bahwa kedua perangkat tes tersebut parallel, dan e) skor-skor yang berasal dari dua perangkat tes yang berbeda materi tidak dapat disetarakan. (Hambleton & Swaminathan, 1985: 78)

Penerapan Teori Respons Butir dalam Penyetaraan Tes

Penerapan teori respons butir dalam kegiatan penyetaraan tes harus memenuhi dua asumsi dasar yakni unidimensi dan independensi lokal (*local independence*) (Kolen & Bremann, 1989: 48). Unidimensi artinya bahwa dimensi karakter peserta yang diukur oleh suatu tes itu tunggal. Independensi lokal adalah bahwa apabila kemampuan-kemampuan yang mempengaruhi kinerja tes dianggap konstan maka respons subjek terhadap setiap butir secara statistik tidak saling terkait. Adapun langkah-langkah melakukan kegiatan penyetaraan tes menurut teori respons butir meliputi:

- (1) Mengestimasi parameter, dapat dilakukan dengan menggunakan program BILOG 3 atau LOGIST.
- (2) Mengestimasi skala *IRT* dengan menggunakan transformasi linier.
- (3) Penyamaan skor; jika menggunakan skor jawaban yang benar maka dilakukan konversi ke skala jawaban yang benar dan kemudian ke skala skor.

Oleh karena kegiatan penyetaraan tes memiliki prosedur yang empiris, maka kegiatan ini memerlukan rancangan tertentu yang harus diperhatikan. Berbagai rancangan penyetaraan tes yang dapat digunakan menurut teori respons butir adalah:

(1) Rancangan kelompok tunggal (*single-group design*)

Menurut rancangan kelompok tunggal ini, kegiatan penyetaraan dilakukan dengan menggunakan satu kelompok peserta yang merespons dua perangkat tes misalnya X dan Y. Parameter butir dari kedua perangkat tes diestimasi secara terpisah dengan mengkalibrasi parameter kemampuan peserta atau parameter butir.

Berdasarkan rancangan ini, dengan mengkalibrasi parameter kemampuan peserta, maka parameter butir dari perangkat tes X dan Y sudah berada pada skala yang sama. Sebaliknya, jika dilakukan kalibrasi parameter butir, estimasi parameter kemampuan peserta pada kedua perangkat tes memenuhi hubungan:

$$\theta_x^* = \alpha\theta_y + \beta \dots\dots\dots(1)$$

keterangan:

θ_x^* : parameter kemampuan peserta pada perangkat tes X,

θ_y : parameter kemampuan peserta pada perangkat tes Y,

α, β : konstanta konversi penyetaraan tes.

Idealnya untuk menyetarakan skor dari beberapa perangkat tes, maka perangkat tes-perangkat tes tersebut diberikan pada responden yang sama. Dengan membandingkan kemampuan peserta tes dari dua/lebih perangkat tes maka penyetaraan dua perangkat tes dapat dilakukan. Kenyataan di lapangan, rancangan ini sulit dilakukan karena adanya faktor kelelahan, belajar, dan adanya faktor latihan untuk tes kedua atau berikutnya. Selain itu, akan terdapat kesulitan dalam hal merencanakan waktu yang cukup bagi responden untuk mengikuti tes lebih dari satu kali.

(2) Rancangan kelompok ekuivalen (*equivalent-group design*)

Desain ini merupakan kebalikan dari desain pertama, yaitu dua perangkat tes diberikan pada dua kelompok yang sama kemampuannya atau ekuivalen. Proses secara spiral digunakan dalam desain ini, dimana peserta tes dibagi dua secara acak kemudian masing-masing mendapat perangkat tes 1 dan perangkat tes 2.

Sebagai ilustrasi, misalnya terdapat dua kelompok K1 dan K2 dan dua perangkat tes misalnya X dan Y. Kelompok K1 mengerjakan perangkat tes X dan

kelompok K2 mengerjakan perangkat tes Y. Mengingat kelompok K1 dan K2 adalah ekivalen maka kedua kelompok dianggap tunggal. Dalam hal ini, jika digunakan ukuran sampel yang besar maka perbedaan *mean* dari kedua perangkat tes menunjukkan langsung perbedaan rata-rata dari tingkat kesukaran antara kedua perangkat tes tersebut. Keuntungan dari rancangan ini adalah dapat menghindari efek negatif yang disebabkan karena adanya latihan dan kelelahan peserta tes, sedangkan kekurangannya adalah adanya kemungkinan bias yang disebabkan oleh perbedaan distribusi kemampuan dari kedua kelompok peserta tes.

(3) Rancangan tes jangkar (*anchor test design*).

Desain ini biasanya digunakan jika masalah keamanan tes menjadi salah satu pertimbangan penting dan memungkinkan untuk menyelenggarakan beberapa tes dalam satu waktu. Pada desain ini masing-masing perangkat tes mempunyai beberapa item yang sama (*common item*) dan masing-masing kelompok mengerjakan perangkat tes yang berbeda. Pada desain ini terdapat dua variasi yakni pertama, jika *common item* diperhitungkan dalam pemberian skor disebut *internal common item* dan kedua, jika *common item* tidak diperhitungkan dalam pemberian skor disebut *external common item*.

Dalam rancangan ini, apabila digunakan dua perangkat tes yakni X dan Y dan dua kelompok peserta yakni K1 dan K2, maka masing-masing perangkat tes ditambahkan item-item tes jangkar Z sehingga kedua perangkat tes menjadi X+Z item dan Y+Z item. Kelompok peserta K1 mengerjakan perangkat tes X+Z dan kelompok K2 mengerjakan Y+Z sehingga item-item tes anchor Z dikerjakan oleh dua kelompok peserta tes (*common item*).

Penyamaan skala penyetaraan dilakukan dengan kalibrasi parameter kemampuan atau parameter butir tes jangkar. Apabila pada rancangan tes jangkar dengan kalibrasi parameter butir, maka parameter kemampuan peserta kedua kelompok sudah berada pada skala yang sama. Sebaliknya jika penyamaan skala dilakukan dengan kalibrasi kemampuan peserta, maka estimasi parameter butir tes jangkar dari kelompok K1 ke kelompok K2 memenuhi persamaan:

$$b^*_{K1} = \alpha b_{K2} + \beta \dots\dots\dots(2)$$

$$a^*_{K2} = \alpha a_{K1} \dots\dots\dots(3)$$

Keterangan:

b^*_{K1} : parameter tingkat kesukaran butir tes jangkar pada kelompok 1,

a^*_{K2} : parameter daya pembeda butir tes jangkar pada kelompok 2,

b_{K2} : parameter tingkat kesukaran butir kelompok 2,

a_{K1} : parameter daya pembeda butir kelompok 1.

α, β : konstanta konversi penyetaraan tes.

Metode Penyetaraan Menurut Teori Respons Butir

Metode penyetaraan menurut teori respons butir berfungsi untuk menentukan konstanta konversi. Hal ini mengingat bahwa penyetaraan antara dua perangkat tes atau lebih dapat dilakukan jika konstanta konversi telah diketahui (Hambleton & Swaminathan, 1985: 25). Nilai konversi yang dihasilkan kemudian disubstitusi dalam persamaan skala pada rancangan penyetaraan yang digunakan. Metode penyetaraan untuk menentukan konstanta konversi menurut teori respons butir adalah sebagai berikut:

1) Metode regresi

Penentuan konstanta konversi α dan β menggunakan metode regresi dilakukan dengan memperhatikan respons peserta tes pada kedua perangkat tes X dan Y. Estimasi parameter butir dan parameter kemampuan peserta memenuhi persamaan regresi linier sbagai berikut:

$$y = \alpha x + \beta + \varepsilon \dots\dots\dots(4)$$

$$\alpha = \frac{r_{xy}S_y}{S_x} \dots\dots\dots(5)$$

$$\beta = \bar{y} - \alpha\bar{x} \dots\dots\dots(6)$$

Keterangan:

y : estimasi kemampuan atau estimasi parameter butir pada perangkat tes Y,

x : estimasi kemampuan atau estimasi parameter butir pada perangkat tes X,

r_{XY} : koefisien korelasi antara X dan Y,

\bar{y}, \bar{x} : rerata dari y dan x,

s_x, s_y : simpangan baku dari x dan y.

Penggunaan metode ini bersifat tidak timbal balik (asimetris) sehingga kurang memadai untuk penentuan konstanta konversi apalagi mengingat bahwa penyetaraan dua perangkat tes atau lebih sangat memerlukan syarat invariansi dan timbal balik dari perangkat tes yang disetarakan.

2) Metode rerata dan sigma.

Penentuan konstanta konversi α dan β menurut metode rerata dan sigma dilakukan dengan memperhatikan nilai estimasi parameter tingkat kesukaran butir tes pada kedua perangkat tes yaitu b_x dan b_y . Menurut Hambleton & Swaminathan (1985: 26), hubungan antara estimasi parameter butir tes atau parameter kemampuan peserta pada kedua perangkat tes yang akan disetarakan dan penentuan konstanta konversinya memenuhi persamaan sebagai berikut:

$$y = \alpha x + \beta \dots\dots\dots(7)$$

$$\bar{y} = \alpha \bar{x} + \beta \dots\dots\dots(8)$$

$$\alpha = \frac{s_y}{s_x} \dots\dots\dots(9)$$

$$\beta = \bar{y} - \alpha \bar{x} \dots\dots\dots(10)$$

Keterangan:

y : estimasi kemampuan atau estimasi parameter butir pada perangkat tes Y,

x : estimasi kemampuan atau estimasi parameter butir pada perangkat tes X,

\bar{y}, \bar{x} : rerata dari y dan x,

s_x, s_y : simpangan baku dari x dan y.

Metode rerata dan sigma ini bersifat timbal balik sehingga dengan cara yang sama hubungan dari y ke x dapat ditentukan. Namun demikian, menurut Hambleton & Swaminathan (1991: 26) mengemukakan bahwa metode penyetaraan rerata dan sigma ini tidak mempertimbangkan variasi standar error estimasi parameter butir.

3) Metode rerata dan sigma tegar.

Berbeda dengan metode rerata dan sigma, menurut Linn, et al (Hambleton & Swaminathan, 1991: 26) menyatakan bahwa metode rerata dan sigma tegar

mempertimbangkan adanya variasi standar error estimasi parameter butir. Adapun dalam prosedur penyetaraan dengan metode rerata dan sigma tegar yang dikembangkan oleh Linn, Levin, Hastings, & Wardrop (Hambleton & Swaminathan, 1991: 27), langkah-langkah penentuan konstanta konversi dalam penyetaraan tes adalah sebagai berikut:

- a. Menentukan bobot parameter butir i (w_i) pada setiap pasangan (b_{xi} , b_{yi}), dengan persamaan sebagai berikut:

$$w_i = [\text{maks}\{v(x_i), v(y_i)\}]^{-1} \dots\dots\dots(11)$$

dengan, $v(x_i)$ dan $v(y_i)$ adalah varians estimasi parameter tingkat kesukaran butir perangkat tes X dan Y.

- b. Menentukan bobot terskala w_i' dengan persamaan:

$$w_i' = \frac{w_i}{\sum_{j=1}^k w_j} \dots\dots\dots(12)$$

dengan k adalah jumlah butir pada perangkat tes.

- c. Menghitung estimasi berbobot tes X dan Y dengan menggunakan rumus:

$$x_i' = w_i' x_i \dots\dots\dots(13)$$

$$y_i' = w_i' y_i \dots\dots\dots(14)$$

- d. Menentukan rerata dan simpangan baku dari estimasi berbobot tes X dan Y yaitu $\bar{y}, \bar{x}, s_x, s_y$

- e. Menentukan konstanta konversi α dan β dengan menggunakan rerata dan simpangan baku estimasi berbobot dengan mensubstitusikan rerata dan simpangan baku estimasi berbobot pada persamaan penyamaan skala.

4) Metode kurva karakteristik.

Penentuan konstanta konversi α dan β pada metode kurva karakteristik ini dilakukan dengan memperhatikan nilai estimasi parameter butir tes kedua perangkat soal yang akan disetarakan misalnya X dan Y. Apabila pada metode rerata dan sigma serta metode rerata dan sigma tegar dalam menghitung konstanta konversi hanya memperhitungkan hubungan antara parameter-parameter tingkat kesukaran butir perangkat tes yang satu dengan yang lainnya tanpa

mempertimbangkan hubungan antara parameter-parameter daya pembeda kedua prangkat tes maka dengan metode kurva karakteristik, hubungan antara parameter-parameter daya pembeda kedua prangkat tes dipertimbangkan.

Penyetaraan tes dengan metode kurva karakteristik mempertimbangkan informasi dari parameter daya pembeda butir dan tingkat kesukaran butir dalam penentuan konstanta konversi (Haebara, 1980). Oleh karena itu, dalam metode ini diperhatikan hubungan antara parameter daya pembeda dan hubungan antara parameter tingkat kesukaran butir perangkat tes-perangkat tes yang akan disetarakan. Selain itu, dalam metode kurva karakteristik ini juga diperhatikan *true score* peserta tes pada kedua perangkat tes.

True score (τ_{xa}) dari peserta tes dengan kemampuan θ yang merespons k butir dalam perangkat tes X dan Y ditentukan dengan rumus sebagai berikut:

$$\tau_{xa} = \sum_{i=1}^k p(\theta, b_{xi}, a_{xi}, c_{xi}) \dots \dots \dots (15)$$

$$\tau_{ya} = \sum_{i=1}^k p(\theta, b_{yi}, a_{yi}, c_{yi}) \dots \dots \dots (16)$$

Adapun penentuan konstanta konversi untuk setiap butir pada perangkat tes X dan Y dilakukan dengan rumus sebagai berikut:

$$b_{yi} = \alpha b_{xi} + \beta \dots \dots \dots (17)$$

$$a_{yi} = \frac{a_{xi}}{\alpha} \dots \dots \dots (18)$$

atau,

$$\alpha = \frac{a_{xi}}{a_{yi}} \dots \dots \dots (19)$$

$$\beta = b_{yi} - \alpha b_{xi} \dots \dots \dots (20)$$

Secara keseluruhan tampak bahwa masing-masing metode memiliki kelebihan atau kekurangan. Metode regresi tidak bersifat timbal balik, metode rerata dan sigma bersifat timbal balik namun tidak mempertimbangkan variasi standar error estimasi parameter butir. Metode rerata dan sigma tegar bersifat timbal balik dan mempertimbangkan variasi standar error estimasi parameter butir

namun tidak mempertimbangkan hubungan antar daya pembeda perangkat tes yang disetarakan. Metode kurva karakteristik selain bersifat timbal balik dan mempertimbangkan variasi standar error estimasi parameter butir juga memperhitungkan hubungan parameter daya pembeda antara perangkat tes. Memperhatikan kelebihan atau kelemahan masing-masing metode tersebut, menunjukkan bahwa metode kurva karakteristik secara teoretik lebih baik dari metode lainnya.

KESIMPULAN

Berdasarkan uraian di atas, dapat disimpulkan bahwa penerapan teori respons butir dalam penyetaraan tes mengharuskan dipenuhinya asumsi unidimensi dan independensi lokal. Ada tiga rancangan penyetaraan yang dapat digunakan untuk melakukan kegiatan penyetaraan tes yakni rancangan kelompok tunggal (*single-group design*), rancangan kelompok ekuivalen (*equivalent-group design*), dan rancangan tes jangkar (*anchor test design*). Pemilihan rancangan ini akan sangat tergantung dari tujuan dan karakteristik perangkat tes yang akan disetarakan. Adapun metode penyetaraan yang dapat digunakan menurut teori respons butir ada 4 macam yakni metode regresi, rerata dan sigma, rerata dan sigma tegar, dan metode kurva karakteristik.

DAFTAR PUSTAKA

- Haebara, T. (1980). Equating logistic ability scales by weighted least square method dalam Hambleton R. K. & Swaminathan H. (1985) *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R.K. & Swaminathan H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Inc.
- Hambleton, R.K., Swaminathan H. & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Holland, P. W. & Rubin, D. B. (1982). *Test equating*. New York: Academic Press, Inc.

Kolen M. J. & Bremann, R. I. (1995). *Test Equating: Methods and Practices*. New York: Springer.

Mardapi, Djemari. (1991). Konsep dasar teori respons butir: Perkembangan dalam pengukuran pendidikan. *Cakrawala Pendidikan* 3(X). 1-16..

Pottorf, R. F. (1982). Some Issues in test equating, dalam Holland, P. W. & Rubim, D. B. (eds.) *Test equating*. New York: Academic Inc.
