



**SEMINAR NASIONAL KIMIA DAN PENDIDIKAN KIMIA VIII**  
“Peningkatan Profesionalisme Pendidik dan Periset Sains Kimia di  
Era Masyarakat Ekonomi Asean (MEA)”  
Program Studi Pendidikan FKIP UNS  
Surakarta, 14 Mei 2016



MAKALAH  
PENDAMPING

PARALEL A

ISBN : 978-602-73159-1-4

## PENINGKATAN KOMPETENSI TENAGA PENDIDIK DALAM PENGEMBANGAN TES MENGGUNAKAN METODE *EQUATING*

**Rizaldi\*, Silvia Lutasari**

Magister Penelitian dan Evaluasi Pendidikan, Program Pascasarjana, Universitas Negeri  
Yogyakarta, Yogyakarta, Indonesia

\*Keperluan korespondensi, telp:+6285375785284, email: rizal.thechosen1@gmail.com

### ABSTRAK

Artikel ini bertujuan untuk memberikan gambaran tentang urgensi pengembangan tes hasil belajar dengan menggunakan metode *equating* yang merupakan bagian *Item Response Theory*. *Equating* bertujuan menempatkan skor dari dua tes pada skala yang sama. Metode *equating* tes terbagi atas dua cara yaitu *equating vertical* dan *equating horizontal*. Desain penelitian ini merupakan studi literatur dengan menganalisis kelebihan dan keterbatasan dari metode *equating* sehingga dapat dimanfaatkan secara optimal oleh tenaga pendidik yaitu guru dalam pengembangan tes. Tes yang dikembangkan terdiri dari dua paket. Dengan demikian, dapat diketahui kesetaraan kedua paket tes tersebut. Penelitian ini menyimpulkan bahwa *equating* merupakan salah satu cara yang tepat dalam membandingkan dua paket tes yang diujikan. Dua paket tes yang *diequating* dapat diketahui kelebihan antar satu dengan yang lainnya. *Equating* perlu dikembangkan secara luas dimasa mendatang dikarenakan hal ini sangat diperlukan dalam pengembangan tes yang baik.

**Kata kunci:** *Equating, Pengembangan Tes*

### PENDAHULUAN

Penggunaan format tes dalam penilaian amat populer dan diselenggarakan dalam skala besar, bertarap lokal dan nasional. Tes-tes yang diselenggarakan dalam skala besar untuk kepentingan tertentu biasanya dibuat lebih dari satu paket. Hal ini menunjukkan adanya beberapa paket tes yang digunakan untuk mengukur variabel yang sama, namun skor

hasil tes tidak dapat diperbandingkan langsung, karena tes tersebut dibuat pada skala yang berbeda.

Untuk itu paket tes yang beragam untuk mengukur variabel yang sama harus dilakukan penyesuaian terhadap skor-skor tes dalam suatu skala yang sama, sehingga skor pada paket tes yang satu dapat diperbandingkan dengan skor pada paket tes yang lain. Proses statistik yang

digunakan untuk menyesuaikan skor-skor tersebut disebut penyetaraan (Kolen,&Brennan,1995:2). Dengan penyetaraan tes, tidak hanya skor peserta yang dapat disetarakan, tetapi parameter butir tes pun dapat disetarakan.

Suatu penyetaraan tes secara ideal memerlukan syarat-syarat teoretis yang sangat ketat, namun dalam praktik tidak pernah terjadi suatu penyetaraan yang ideal (Kolen, & Brennan, 1995: 246). Syarat-syarat teoretis antara lain menyangkut desain dan metode penyetaraan. Hal ini memiliki pengaruh yang sangat besar pada hasil penyetaraan, disamping faktor lainnya. Oleh karena itu, untuk meminimalkan ketidakstabilan hasil penyetaraan tes, perlu pemilihan desain dan metode penyetaraan yang tepat.

Kegiatan penyetaraan tes dapat dilakukan dengan mengembangkan konversi suatu sistem unit tes ke sistem 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 soal tes.

Oleh karena itu, dalam tulisan ini akan menganalisis kelebihan dan keterbatasan dari metode equating vertikal dan horizontal sehingga dapat dimanfaatkan secara optimal oleh tenaga pendidik yaitu guru dalam pengembangan tes. *Equating* perlu dikembangkan secara luas dimasa mendatang dikarenakan hal ini sangat diperlukan dalam pengembangan tes yang baik.

## METODE PENELITIAN

Desain penelitian ini merupakan studi literatur, dengan menganalisis kelebihan dan keterbatasan dari metode *equating* vertikal dan horizontal sehingga dapat dimanfaatkan secara optimal oleh tenaga pendidik yaitu guru dalam pengembangan tes. *Equating* perlu dikembangkan secara luas dimasa mendatang dikarenakan hal ini sangat diperlukan dalam pengembangan tes yang baik.

## HASIL DAN PEMBAHASAN

### 1. Karakteristik Tes

Tes yang berkualitas baik memiliki karakteristik butir soal dan perangkat tes yang baik pula. Karakteristik soal dapat dilihat dari parameter tingkat kesukaran, daya pembeda, reliabilitas, dan kesalahan pengukuran.

Tingkat kesukaran satu butir soal didefinisikan sebagai proporsi peserta tes yang menjawab benar soal tersebut (teori tes klasik). Angka yang menunjukkan mudah, sedang, atau sukar suatu butir soal disebut indeks tingkat kesukaran dengan nilai antara 0 dan 1. Daya Pembeda adalah kemampuan satu butir soal untuk membedakan antara peserta tes yang pandai dengan peserta tes yang kurang pandai. Untuk mengetahui daya pembeda butir soal biasanya menggunakan indeks korelasi antara skor butir dengan skor totalnya, seperti teknik *point biserial* dan teknik *biserial*. *Reliabilitas* adalah tingkat kepercayaan dari suatu alat ukur, artinya seberapa jauh pengukuran dilakukan berulang-ulang terhadap sekelompok subyek

yang sama memberikan hasil yang sama pula. Ukuran reliabilitas yang baik adalah antara 0,60 sampai dengan 0,85 (Grondlund, 1982). Kesalahan pengukuran (*standart error of measurement*) biasanya disebabkan oleh kesalahan dalam pengambilan sampel. Semakin kecil angka kesalahan semakin baik, sehingga skor yang diperoleh dari perangkat tersebut semakin mendekati skor yang sebenarnya.

## 2. Definisi Penyetaraan (*Equating*)

Weiss (1983) mendefinisikan penyetaraan skor sebagai suatu prosedur empiris karena data skor diperoleh dari hasil pekerjaan peserta didik yang selanjutnya diperlukan untuk mentransformasi skor. Menurut Hambleton (1991) penyetaraan skor adalah membandingkan skor yang diperoleh dari perangkat tes yang satu (X) dan perangkat tes lainnya (Y) yang dilakukan melalui proses penyetaraan skor pada kedua perangkat tes tersebut. Crocker dan Algina (1986), menyatakan bahwa dua skor hasil pengukuran yang menggunakan instrumen X dan instrumen Y dapat disetarakan skornya jika kedua instrumen mengukur kemampuan atau *trait* yang sama. Menurut Kolen (2004) penyetaraan skor dapat dilakukan jika kelompok peserta tes setara, karena kesetaraan yang ekstrim akan berpengaruh dalam perhitungan.

Berdasarkan pengertian di atas dapat disimpulkan bahwa penyetaraan *equating* merupakan prosedur secara empiris dalam rangka menyetarakan skor dari perangkat tes yang satu ke perangkat tes lainnya sehingga dapat melakukan perbandingan atau konversi secara langsung

hasilhasil individu yang mengikuti perangkat tes yang berbeda tersebut.

## 3. Tipe-tipe *Equating*/Penyetaraan.

Ada beberapa teknik dan metodologi yang dapat digunakan dalam penyetaraan tes untuk menetapkan transformasi. Secara umum teknik dan metodologi ini dapat dibagi atas dua jenis, yaitu yang metode teori tes klasik (*classical test theory*) dan metode teori modern (*item response theory*).

Menurut Hambleton, Swaminathan, & Roger (1991), pada teori tes klasik dikenal dua metode, yaitu penyetaraan linear (*linear equating*) dan penyetaraan ekuipersentil (*equipercentile equating*). Penyetaraan linear akan menghubungkan skor konversi dengan skor asalnya melalui suatu fungsi linear. Prinsip dasar metode ini adalah distribusi skor pada dua perangkat tes sama dalam hal rerata dan simpangan baku. Angoff (1971) menyatakan bahwa definisi untuk penyetaraan linear adalah skor dua perangkat tes menjadi ekuivalen jika ada hubungan yang setara dengan standar skor deviasinya. Penyetaraan skor yang menggunakan metode linear memungkinkan adanya tingkat kesulitan relatif bervariasi pada skor di antara beberapa perangkat tes tersebut. Pada penyetaraan linear diperlukan kesamaan distribusi probabilitas antara skor X dan skor Y. Jika skor X dan skor Y memiliki rerata dan simpangan baku yang berbeda, maka distribusi probabilitas yang sama dari kedua skor tersebut dapat digunakan untuk mentransformasi nilai dari satu distribusi probabilitas ke distribusi probabilitas berikutnya.

Untuk menanggulangi masalah tersebut, para ahli *educational measurement* mengembangkan metode statistik yang disebut *equating* (penyetaraan). Metode ini adalah metode ilmiah yang digunakan untuk menyetarakan nilai dari skor mentah satu perangkat ke skor mentah perangkat lainnya, yang pada akhirnya akan diperoleh sebuah tabel konversi nilai. Hambleton & Swaminathan (1985) mengatakan bahwa tidak pernah ada soal dari dua perangkat tes dengan butir soal yang berbeda walaupun berdasarkan kisi-kisi yang sama mempunyai tingkat kesukaran yang sama. Penyetaraan skor dapat dijadikan sebagai teknik penyamaan skor yang mampu membedakan peserta didik yang pandai dan peserta didik yang kurang pandai. Penyetaraan skor dimungkinkan penggunaan perangkat tes yang berbeda terhadap kelompok yang berbeda, sesuai dengan tingkat kemampuannya, sehingga skor yang diperoleh dapat dibandingkan dan peserta tes tidak merasa dirugikan atau diuntungkan karena kebetulan mendapat perangkat tes yang lebih mudah atau yang lebih sukar. Dengan demikian, menjadi suatu keharusan bagi para pengembang tes atau lembaga pengembang tes untuk melakukan penyetaraan terhadap perangkat tes yang digunakan.

Tujuan utama dilakukannya *equating* adalah menegakkan keadilan bagi peserta tes maupun pengguna hasil tes. Asumsi bahwa suatu tes paralel dari segi materi (berasal dari kisi-kisi yang sama) diyakini tidak benar, oleh karena itu perlu suatu proses yang menyetarakan skor perangkat tes paralel dengan mengeliminasi faktor perbedaan tingkat kesukaran

antarperangkat tersebut. Secara proses, *equating* adalah prosedur pemberian skor peserta tes sesuai kemampuannya dengan meniadakan efek perbedaan tingkat kesukaran antarperangkat tes. Hal ini sesuai dengan tuntutan keadilan, jangan sampai peserta didik mendapat nilai jelek karena mengerjakan perangkat tes sukar ataupun peserta didik nilainya baik karena mengerjakan tes yang mudah. Secara psikometris, *equating* merupakan suatu proses yang bertujuan untuk memperoleh skor konversi yang dapat memperbandingkan hasil beberapa perangkat tes yang paralel. Petersen (1989) mendefinisikan *equating* sebagai prosedur empirik yang dilakukan untuk menghasilkan hubungan antara skor mentah dua perangkat tes yang paralel, yaitu A dan B, sehingga skor perangkat tes A dapat diartikan dalam sudut pandang skor perangkat tes B, sedangkan Angoff (1984) mendefinisikan *equating* sebagai proses untuk mengonversi unit di satu perangkat tes menjadi unit di perangkat tes lain yang paralel.

Proses *equating* adalah hal yang mutlak harus dilakukan dalam menangani pengolahan hasil tes agar diperoleh pemetaan mutu pendidikan yang akurat dan valid, tanpa distorsi perbedaan tingkat kesulitan walaupun mendapat perangkat tes yang berbeda.

Dari perspektif yang berbeda, penyetaraan tes dapat dibedakan atas dua tipe, yaitu penyetaraan vertikal dan penyetaraan horizontal. Penyetaraan vertikal didefinisikan sebagai sebuah metode pengukuran nilai pada dua tes yang berbeda

tingkatan kelas. Kolen (1984) menyatakan bahwa *vertikal equating* memuat perbandingan perkembangan kemampuan peserta didik pada kedua level pada saat tes dan perkembangan kemampuan dari waktu ke waktu dapat pula dibandingkan. Dengan kata lain pada penyetaraan vertikal dimaksudkan untuk menentukan padanan skor-skor yang diperoleh dari dua kelompok peserta tes dalam tingkat atau jenjang pendidikan yang berbeda, tetapi dikenakan perangkat tes yang sama.

Penyetaraan horizontal didefinisikan sebagai metode skor penempatan peserta tes pada dua tes yang sama di level yang sama, mengukur hal yang sama, dan untuk populasi yang sama sehingga skor peserta tes dapat dibandingkan. Penyetaraan dua skor yang masing-masing diperoleh dari dua perangkat tes yang berbeda, tetapi mengukur hal yang sama. Penyetaraan horizontal lebih jelas karena tujuan dari penyetaraan adalah membandingkan dua atau lebih kelompok peserta tes yang memiliki kemampuan yang sama menggunakan dua atau lebih perangkat tes yang berbeda mengukur hal yang sama dan tingkat kesukaran yang sama.

#### 4. Desain Penyetaraan/Equating Tes

Menurut Peterson (1989), penyetaraan tidak sesederhana seperti regresi, karena metode penyetaraan adalah sebuah prosedur empiris yang melibatkan sebuah desain untuk pengumpulan data dan sebuah aturan untuk menetapkan transformasi. Beberapa desain dapat digunakan untuk memperoleh data pada proses penyetaraan *equating*. Ada tiga desain yang sering

digunakan oleh lembaga testing, yaitu *single group design*, *common item non equivalent*, dan *random group design*.

Pada *single group design*, satu sampel yang sama diuji dua kali dengan paket tes yang berbeda. Misal paket tes tersebut paket X dan paket Y. Paket X diadministrasikan pertama pada semua peserta tes, kemudian disusul dengan paket Y. Cara ini mengakibatkan paket Y lebih sukar soalnya dari paket X karena diujikan belakangan (*fatigue effect*).

Untuk mengurangi *fatigue effect* dapat digunakan *Counterbalancing*, yaitu dengan cara sampel dibagi atas subgrup 1 dan subgrup 2. Paket Tipe yang kedua metode penyetaraan ekuipersentil (*equipercentile equating*). Penyetaraan ekuipersentil adalah metode penyetaraan dua paket tes, misal X dan Y, dengan mengasumsikan bahwa kedua paket tersebut mengukur variabel laten yang sama dan nilai persentil rank distribusi skor kedua paket tes hampir sama. Apabila distribusi skor kedua paket tes persis sama, maka hasil *equating* skor di paket X akan persis sama dengan skala skor paket tes Y.

Livingstone (1984) menjelaskan prosedur penyetaraan ekuipersentil sebagai berikut: pertama membuat tabulasi ranking persentil untuk distribusi skor pada masing-masing paket yang disetarakan. Kedua, transformasikan skor di paket tes yang baru terhadap paket tes acuan sedemikian rupa sehingga skor yang berpasangan memiliki nilai ranking persentil yang sama.

Prosedur estimasi frekuensi penyetaraan ekuipersentil dikerjakan dengan langsung menyetarakan dua

perangkat tes yaitu paket X dan paket Y. Namun demikian, paket X dan paket Y dapat juga disetarakan langsung dengan tes yang ketiga (tes V). Prosedur ini lebih baik digunakan dalam penyetaraan ekuipersentil.

Satu masalah yang sering dihadapi dalam proses penyetaraan *equating* dengan metode ekuipersentil adalah ketidakteraturan distribusi skor, terutama pada kasus sampel yang kecil. Ketidakteraturan ini menimbulkan masalah karena nilai ranking persentil menjadi tidak stabil saat digeneralisasikan ke populasi. Untuk mengatasi masalah ini perlu dilakukan proses *smoothing* distribusi skor, terutama untuk jumlah sampel yang kecil. *Smoothing* adalah proses pemulusan atau penghalusan ketidakteraturan distribusi skor dengan cara mengganti distribusi tersebut dengan distribusi lain yang memiliki bentuk, lokasi, penyebaran, *skewness*, dan *kurtosis* yang sama, tetapi meminimalisir ketidakteraturan.

Penerapan teknik pemulusan pada distribusi skor sampel yang mengandung kekeliruan acak dan memiliki bentuk ketidakberaturan (melonjak-lonjak), akan menghasilkan bentuk distribusi skor yang halus. Dengan demikian hasil penyetaraan yang lebih akurat dapat diperoleh dengan cara *smoothing* (pemulusan).

Metode penyetaraan dengan *item response theory* (IRT) atau teori respon butir, didasarkan asumsi bahwa ada sebuah fungsi matematika yang menggambarkan hubungan antara kemampuan peserta tes dan kemungkinan peserta tes menjawab soal dengan benar. Ada tiga model penyetaraan dengan *item response theory*, yaitu *Rasch Model* (satu parameter logistik),

model dua parameter logistik, dan model tiga parameter logistik (Hambleton, 1991).

Pada penelitian ini, teknik *equating* yang digunakan adalah *equipercentile equating*. Paket tes X yang akan disetarakan dengan Y, diasumsikan mengukur hal yang sama karena kedua paket tes tersebut berasal dari kisi-kisi yang sama, serta nilai persentil rank distribusi skor kedua paket tes hampir sama. Jika distribusi skor tidak beraturan sehingga nilai ranking persentil menjadi tidak stabil saat digeneralisasikan ke populasi, maka dalam analisis ini dilakukan *smoothing* (penghalusan) distribusi skor. Selanjutnya diperoleh hasil transformasi skor paket tes baru terhadap paket tes acuan.

## 5. 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:

- Mengestimasi parameter, dapat dilakukan dengan menggunakan program BILOG 3 atau LOGIST.
- Mengestimasi skala IRT dengan menggunakan transformasi linier.



c. Penyesuaian skor

Oleh karena kegiatan penyetaraan tes memerlukan rancangan tertentu yang harus diperhatikan. Berbagai rancangan penyetaraan tes yang dapat digunakan menurut teori respons butir adalah:

a. 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:

$$\theta^*x = \alpha\theta y + \beta \dots \dots \dots (1)$$

Idealnya untuk menyetarakan skor dari beberapa perangkat tes, maka perangkat tes tersebut diberikan pada responden yang sama. 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.

b. 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.

c. 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:

$$b \cdot K1 = abK2 + \beta \dots (2)$$

$$a \cdot K2 = \alpha aK1 \dots (3)$$

## 6. 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:

### a. Metode regresi

Penentuan konstanta konversi  $\alpha$  dan  $\beta$  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:

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

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

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

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.

### b. Metode rerata dan sigma.

Penentuan konstanta konversi  $\alpha$  dan  $\beta$  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:

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

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

$$s_y \alpha = \dots (9)$$

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

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.



c. 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:

- 1) Menentukan bobot parameter butir
- 2) Menentukan bobot terskala  $w_i$
- 3) Menghitung estimasi berbobot tes X dan Y
- 4) Menentukan rerata dan simpangan baku dari estimasi berbobot tes X dan Y.
- 5) Menentukan konstanta konversi  $\alpha$  dan  $\beta$  dengan menggunakan rerata dan simpangan baku estimasi berbobot dengan mensubstitusikan rerata dan simpangan baku estimasi berbobot pada persamaan penyamaan skala.

d. Metode kurva karakteristik.

Penentuan konstanta konversi  $\alpha$  dan  $\beta$  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 perangkat tes maka dengan metode kurva karakteristik, hubungan antara parameter-parameter daya pembeda kedua perangkat 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.

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 mem-

perhitungkan 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.

## UCAPAN TERIMAKASIH

Prof. Djemari Mardapi, Ph.D; Prof. Badrun Kartowagiran; (Dosen Program Studi Magister Penelitian dan Evaluasi Pendidikan Universitas Negeri Yogyakarta) atas motivasinya untuk menulis artikel.

## DAFTAR RUJUKAN

- [1] 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.

- [2] Hambleton, R.K. & Swaminathan H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Inc.
- [3] Hambleton, R.K., Swaminathan H. & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- [4] Holland, P. W. & Rubin, D. B. (1982). *Test equating*. New York: Academic Press, Inc.
- [5] Angoff, W.H., (1971). Scale, norms, and equivalent scores, In RL Thorndike (ed), *Educational Measurement*, Washington DC: American Council on Education.
- [6] Crocker, L., Algina, J., (1986). *Introduction to Classical and Modern Test Theory*, NY: Holt, Rinehart and Winston, Inc.
- [7] Grondlund, E.N., (1982). *Constructing Achievement Test*, EC: Prentice Hall, Inc.
- [8] Kolen. (2004). *Software Common Item Program for Equating (CIPE) versi 2.0*
- [9] Kolen, (1984), *Effectiveness of Analysis in Equipercentile Equating*, Journal of Educational Statistic, 9, pp.25-44
- [10] Kolen, M. J., & Brennan, R. L., (2004). *Test equating, scaling, and linking: Methods and practices*. New York: Springer
- [11] Livingstone, (2004). *Equating Test Scores (Without IRT)*, Princeton, Nj: ETS

- [12] Petersen, N.S.. (1989). *Educational Measurement, Scaling, Norming, and Equating, in R.L Linn (ed), Educational Measurement*, NY: Macmillan
- [13] Weiss, D.J.,(1983). *New Harizons in testing : A Test of adequacy of curvilinear score equating models*, NY: Academic Press.
- [14] Kolen M. J. & Bremann, R. I. (1995). *Test Equiting: Methods and Practices*. New York: Springer.
- [15] Djemari Mardapi. (2012). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Nuha Litera
- [16] Djemari Mardapi. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*.