Pedagogy p-ISSN: 2502-3802
Volume 7 Nomor 2 e-ISSN: 2502-3799

PENYETARAAN SKOR TES MATEMATIKA MENGGUNAKAN BUTIR BERSAMA (*ANCHOR)* DENGAN BERBAGAI METODE

Melly Elvira[1]
Jurusan Pendidikan Islam Anak Usia Dini, Fakultas Ilmu Tarbiah dan Ilmu Keguruan[1], Universitas Islam Negeri Maulana Malik Ibrahim Malang[1]
melly@uin-malang.ac.id[1]

Abstrak

Penelitian ini bertujuan menyelidiki kesetaraan skor try out UN mata pelajaran matematika SMP tahun 2014 Kota Makassar yang terdiri dari 2 paket soal yakni paket soal 1 yang direspons 2099 siswa dan paket soal 2 yang direspons 2068 siswa. Rancangan penyetaraan yang digunakan adalah Rancangan dengan butir bersama. Analisis yang digunakan adalah analisis dengan menggunakan pendekatan IRT dengan membandingkan seluruh metode mean-mean, mean-sigma, Haebara, dan Stocking-Lord menggunakan pendekatan IRT model 1PL, 2PL, dan 3PL. Langkah estimasi parameter penyetaraan menggunakan paket equateirt pada program R. Hasil penelitian menunjukkan bahwa hubungan soal UN paket 1 dan paket 2, tidak dapat dikatakan setara, hal ini diketahui setelah melakukan analisis dan menghitung persamaan regresi, diketahui bahwa paket 2 soal UN memilik tingkat kesukaran yang lebih tinggi dibandingkan paket 1. Penyetaraan paket 1 ke paket 2*, untuk semua model IRT, pada model 1 PL metode penyetaraan yang paling presisi adalah metode Stocking-Lord, untuk 2 PL yang paling dekat adalah metode Haebara, sedangkan untuk model 3 PL adalah model Stocking-Lord. Penyetaraan paket 2 ke paket 1*, untuk semua model IRT, pada model 1 PL, 2PL, dan 3PL metode penyetaraan yang paling presisi adalah metode Haebara.

Kata Kunci: penyetaraan, mean-mean, mean-sigma, haebara, stocking-lord

A. Pendahuluan

Tes matematika memiliki karakteristik yang berbeda dengan mata pelajaran lain. Materi matematika bersifat hierarkis dan berkaitan erat satu sama lain (Nisa & Retnawati, 2018). Artinya penguasaan siswa terhadap materi sebelumnya menjadi dasar untuk melanjutkan dan memahami materi pada tingkat berikutnya. Guru diharapkan mampu menulis tes yang baik dan juga dapat menggunakan tes tersebut untuk menghubungkan prestasi belajar siswa dalam pembelajaran yang berbeda nilai sehingga dapat diketahui informasi tentang perkembangan kemampuan siswa.

Selain mengetahui karakteristik butir-butir tes yang digunakan, guru diharapkan memastikan bahwa, agar informasi tentang pengembangan kemampuan siswa akurat, butir-butir tes tersebut harus menggambarkan kemampuan diri siswa. Penggunaan butir-butir tes yang berada di luar kemampuan siswa membuat siswa tidak mampu menjawab soal sehingga guru tidak dapat mengetahui informasi tentang perkembangan siswa (Nisa & Retnawati, 2018). Siswa dengan usia dan kelas yang sama mungkin tidak memiliki perkembangan yang sama.

Skor dari dua tes yang berbeda dari dua atau lebih kelompok yang berbeda dapat dibandingkan jika item-itemnya sama dan didasarkan pada skala yang sama (Kolen & Brennan, 2014). Penyetara antar skor dapat dilakukan secara statistik. Analisis statistik dilakukan terhadap skor dari dua tes yang berbeda untuk disesuaikan pada skala yang sama. Proses statistik yang digunakan untuk menghasilkan satu skala dari skor dua tes yang berbeda dengan skala yang sama disebut *equating* (Kolen & Brennan, 1995, hlm. 5). Hambleton, Swaminathan, dan Rogers (1991, hlm. 123) menyatakan bahwa menyamakan adalah suatu proses untuk mengubah skor X menjadi matriks skor tes Y atau sebaliknya, sehingga hasil dari proses persamaan tersebut dapat dibandingkan.

Penyetaraan tes adalah proses statistik yang menyediakan interkonversi skor yang diperoleh dari berbagai bentuk tes yang mengukur struktur yang sama. Skor yang disamakan memiliki arti yang sama terlepas dari kapan dan kepada siapa tes itu diterapkan. Oleh karena itu, skor yang diperoleh dari suatu bentuk tes dapat dibandingkan dengan skor yang diperoleh dari bentuk tes yang lain (Kolen & Brennan, 2014). Dalam hal siswa secara bersamaan diberikan bentuk tes yang berbeda dan tingkat kesulitan antar bentuk tidak sama, kemungkinan individu yang diberikan tes sulit mendapatkan nilai yang lebih rendah daripada individu yang diberikan tes mudah. Situasi ini membuat sulit untuk membandingkan poin yang diperoleh dari bentuk tes yang berbeda. Penyetaraan tes mencegah kemungkinan ketidakadilan terhadap individu yang diberikan tes yang sulit serta menghilangkan masalah bias yang disebabkan oleh perbedaan tingkat kesulitan bentuk tes (Kilmen & Demirtasli, 2012)

Ada dua macam proses penyetaraan yang dapat dilakukan untuk menguji skor tes: penyetaraan horizontal dan penyetaraan vertikal. Penyetaraan horizontal adalah

penyetaraan yang dilakukan terhadap nilai tes yang memiliki indeks kesukaran yang sama pada kelas yang sama (Hambleton & Swaminathan, 1985), sedangkan penyetaraan vertikal adalah proses penyetaraan yang dilakukan untuk mengungkapkan kemampuan siswa yang diukur dengan instrumen tes yang indeks kesukarannya berbeda dan pada kelas yang berbeda. , tetapi mereka mengukur sifat yang sama (Crocker et al., 2008).

Implementasi penggunaan Metode *Item* Respons*e Theory* (IRT) untuk penyetaraan tes bentuk dikotomi dengan paket soal yang berbeda. Pada umumnya model yang digunakan dalam IRT meliputi model 3PL, 2,PL, 1PL, dan model Rasch. Perangkat tes dapat di setarakan ketika memiliki sejumlah item yang sama (*common item*) untuk penyetaraan langsung atau ketika kedua perangkat tes dapat di *link*-an dalam rantai perangkat tes yang memuat item bersama pada pasangan perangkat *tes* (*equiting* tak langsung/*cahain*).

Jika dua perangkat tes dengan menggunakan pola yang berbeda, maka hasil konversi memuat hasil yang merupakan rerata koefisien penyetaraan. Paket R yang digunakan pada kesempatan ini menghitung koefisien penyetaraan langsung ataupun tak langsung. Paket yang digunakan akan menghasilkan estimasi koefisien penyetaraan dan *standard error* dari penyetaraan langsung.

Metode penyetaraan yang digunakan adalah metode rerata -rerata, metode rerata - sigma, Haebara, dan Stocking-Lord. Ujian nasional yang dilaksanakan terdiri atas beberapa paket soal yang menjadi masalah apakah paket-paket soal tersebut dapat mengukur secara obyektif terhadap peserta didik. Agar peserta didik yang satu dengan peserta didik yang lainnya tidak ada yang dirugikan atau diuntungkan dari paket soal yang berbeda maka dalam evaluasi pembelajaran aspek keadilan merupakan salah satu prinsip yang penting, artinya peserta didik yang menghadapi ujian nasional di berbagai wilayah Indonesia dan waktu yang berbeda harusnya mendapatkan perlakuan yang adil.

1. Metode Penyetaran dengan Teori Respons Butir

Terdapat berbagai metode yang dapat diterapkan untuk menghubungkan skor antara dua tes atau lebih. Menurut Heri Retnawati (2014) Terdapat beberapa metode yang dapat digunakan dalam penyetaraan tes ini. Metode tersebut antara lain adalah

metode Rarata & Sigma, metode Rerata & Rerata, metode Rerata & Sigma Tegar dan metode Kurva Karakteristik.

a. Metode Rerata dan Rerata

Metode ini adalah metode yang dikembangkan oleh Loyd dan Hoover (2016) dan digunakan untuk mengubah skala skor parameter kesulitan dan diskriminasi. Metode ini menggunakan parameter diskriminasi dalam penentuan kurva; dan rata-rata parameter kesulitan digunakan dalam penentuan konstanta penyetaraan.. Kolen & Brennan (2014) menjelaskan hubungan antara estimasi parameter butir dari kedua perangkat tes yang disetarakan berdasarkan pada persamaan berikut:

$b_2 = \alpha b_1 + \beta$ *(1)*

$a_2 = \frac{a_1}{\alpha}$ *(2)*

maka diperoleh nilai :

$\bar{b}_2 = \alpha \bar{b}_1 + \beta$ *(3)*

jadi

$\alpha = \frac{\bar{a}_1}{\bar{a}_2}$ *(4)*

$\beta = \bar{b}_2 - \overline{\alpha b_1}$ *(5)*

Keterangan:

a_1, $\overline{a_2}$: rata-rata daya pembeda pada perangkat tes 1 dan 2

b1, $\overline{b2}$: rata-rata tingkat kesulitan butir pada perangkat tes 1 dan 2

b. Metode Rerata dan Sigma

Metode ini adalah proses konversi skor yang didefinisikan oleh Marco (1977). metode ini menggunakan standar deviasi dalam penentuan kurva persamaan penyetaraan; dan rata-rata kesulitan tes digunakan dalam penentuan konstanta persamaan seperti dalam kasus metode *mean-mean*. Pada metode penyetaraan Rerata & Sigma ini, penentuan konstanta α dan konstanta β dilakukan dengan memperhatikan nilai rerata (μ) dan simpangan baku (S). Kolen & Brennan (Kolen & Brennan, 2014) menjelaskan hubungan antara estimasi parameter butir dari kedua perangkat tes yang disetarakan berdasarkan pada persamaan berikut ini.

$b_2 = \alpha b_j + \beta$ *(6)*

maka diperoleh nilai

$\bar{b}_2 = \alpha \bar{b}_1 + \beta$, dan $S_2 = \alpha S_1$ *(7)*

jadi

$$\alpha = \frac{S_2}{S_1} \qquad (8)$$

$$\beta = \bar{b}_2 - \alpha \bar{b}_1 \qquad (9)$$

Keterangan:

b1, b2 : estimasi parameter tingkat kesukaran butir pada perangkat tes 1 dan tes 2

$\bar{b}1, \bar{b}2$: rata-rata tingkat kesukaran butir pada perangkat tes 1 dan 2

S1 dan S2 : simpangan baku tingkat kesukaran perangkat 1 dan 2

α dan β : konstanta yang digunakan untuk menyetarakan perangkat tes

c. Metode Haebara

Pendekatan ini dikembangkan oleh Haebara (1980) yang dinamai menurut metode tersebut. Metode Haebara adalah metode kurva karakteristik yang penyetaraan parameter butirnya berdasarkan pada fungsi karakteristik butir. Jumlah kuadrat dari selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteristik butir dari dua skala yang sudah disetarakan dinyatakan dengan $H(\theta_i)$ yaitu:

$$H(\theta_i) = \sum_{j=1}^{n} \left(T_{ij} - T_{ij}^*\right)^2 \qquad (10)$$

$$T_{ij} = P_j(\theta_i) \qquad (11)$$

$$T_{ij}^* = P_j^*(\theta_i) \qquad (12)$$

Dengan

n : banyaknya butir

$P_j(\theta_i)$: probabilitas menjawab benar butir ke j oleh peserta berkemampuan θ_i

$P_j^*(\theta_i)$: probabilitas hasil estimasinya

d. Metode Stocking-Lord

Stocking-Lord (Stocking & Lord, 1983) memodifikasi metode Haebara. Koefisien penyetaraan (α dan β) pada metode Stocking-Lord diperoleh dengan cara meminimalkan rerata kuadrat jumlah dan selisih antara estimasi skor sejati fungsi-fungsi respons butir tes bersama dua tes, menurut metode Srocking-Lord, fungsi kriteria F adalah

$$F = \frac{1}{N}\sum_{i=1}^{N}(T_i - T_i^*)^2 \quad (13)$$

Dengan N adalah Jumlah peserta i (I = 1, 2, …, N); T_i adalah skor sejati, yakni jumlah dari probabilitas respons benar peserta i terhadap butir j tes bersama, dan T^*_i adalah hasil transformasi skor sejati T_i pada butir tes bersama ke-2 menjadi skala tes bersama ke-1.

2. Penyetaraan Tes Berdasarkan karakteristik Butir

Penyetaraan tes dapat dilakukan setelah koefisien α dan β diketahui, hasil estimasi parameter butir dan parameter kemampuan dari perangkat tes 1 disetarakan pada skala yang sama dengan perangkat tes 2 dengan menggunakan persamaan berikut:

$$b_2^* = \alpha b_1 + \beta \quad (14)$$
$$a_2^* = \frac{a_1}{\alpha} \quad (15)$$
$$\theta_2^* = \alpha\theta_1 + \beta \quad (16)$$

Keterangan:

b_2^* : tingkat kesukaran butir pada perangkat tes 1 setelah disetarakan pada skala tes 2

a_2^* : daya pembeda butir pada perangkat tes 1 setelah disetarakan pada skala tes 2

θ_2^* : kemampuan siswa pada perangkat tes 1 setelah disetarakan pada skala tes 2

3. Akurasi Metode Penyetaraan

Adanya penerapan lebih dari satu metode dalam proses penyetaraan maka perlu diketahui bagaimanakah akurasi hasil penyetaraan dari masing-masing metode penyetaraan. Akurasi hasil dari penyetaraan dapat dilihat dengan cara membandingkan rata-rata nilai *Root Mean Square Different* (RMSD) dari karakteristik butir sebelum dan setelah disetarakan. Kilmen & Demirtasli (2012) melakukan penyetaraan dengan empat metode pada pendekatan IRT. Keempat metode tersebut dihitung keakuratannya dengan melihat nilai RMSD terkecil dengan menggunakan rumus sebagai berikut.

$$\text{RMSD } (a) = \sqrt{\frac{\sum_{i=1}^{N}(a_2^* - a_1)^2}{N}} \quad (17)$$

$$\text{RMSD } (b) = \sqrt{\frac{\sum_{i=1}^{N}(b_2^* - b_1)^2}{N}} \quad (18)$$

Keterangan:

RMSD = *Root Mean Square Different*

a_2^* = daya pembeda tes 1 setelah disetarakan ke tes 2

a_1 = daya pembeda tes 1

b_2^* = tingkat kesukaran tes 1 setelah disetarakan ke tes 2

b_1 = tingkat kesukaran tes 1

B. Metode Penelitian

1. Jenis Penelitian

Penelitian ini adalah penelitian deskriptif eksploratif dengan tujuan menyelidiki kesetaraan skor *try out* UN mata pelajaran matematika SMP tahun 2014 Kota Makassar.

2. Subjek dan Objek Penelitian

Subjek terdiri atas kelompok siswa dari SMP Kota Makassar yang mengikuti *try out* UN Matematika Tahun Pelajaran 2013/2014. Objek yang dipilih adalah butir soal matematika objektif pilihan ganda yang terdiri atas masing-masing terdiri atas 40 butir soal yang diskor dikotomi.

3. Teknik pengumpulan data

Dalam penelitian ini, Teknik yang dipergunakan dalam mengumpulkan data adalah teknik dokumentasi, dengan mengumpulkan respons siswa pada *try out* UN matematika se-Kota Makassar yang terdiri dari 2 paket soal yakni paket soal 1 yang direspons 2099 siswa dan paket soal 2 yang direspons 2068 siswa.

4. Teknik Analisis Data

Rancangan penyetaraan yang digunakan adalah Rancangan dengan butir bersama dengan rancangan sebagai berikut:

Tabel 1. Rancangan Penyetaraan dengan menggunakan butir bersama

Populasi	Sampel	Paket 1	*Anchor*	Paket 2
P	1	37	3	
Q	2		3	37

Analisis yang digunakan adalah analisis dengan menggunakan pendekatan IRT dengan membandingkan seluruh metode *mean-mena, mean-sigma, Haebara, dan Stocking-Lord*. Langkah estimasi parameter penyetaraan menggunakan paket

equateIRT pada program R. Berikut langkah-langkah estimasi parameter menggunakan 2 paket soal dan estimasi parameter butir menggunakan pendekatan IRT menggunakan model 1PL, 2PL, atau 3 PL.

a. Mempersiapkan Data

   Sebelum melakukan estimasi terlebih dahulu mempersiapkan data dalam format csv. Paket program R yang digunakan adalah paket equateIRT dengan sintak sebagai berikut:

   ```
   > library (ltm)
   > library(equateIRT)
   > paket1<-read.csv(file="Paket 1.csv",header=T)
   > paket2<-read.csv(file="Paket 2.csv",header=T)
   ```

b. Melakukan estimasi parameter butir

   1) Estimasi dengan model Rasch/1PL

   ```
   > m1<-rasch(paket1)
   > m2<-rasch(paket2)
   ```

   2) Estimasi dengan model 2PL

   ```
   > n1<-ltm(paket1~z1)
   > n2<-ltm(paket2~z1)
   ```

   3) Estimasi dengan model 3PL

   ```
   > o1<-tpm(paket1)
   > o2<-tpm(paket2)
   ```

c. Melakukan Varian dan Kovarian untuk semua Model IRT (1PL, 2PL, dan 3PL)

   ```
   > estm1 <- import.ltm(m1, display = FALSE)
   > estm2 <- import.ltm(m2, display = FALSE)
   ```

d. Membuat daftar matriks Koefisien dan Kovarian untuk semua Model IRT

   ```
   estc <- list(estm1$coef, estm2$coef)
   estv <- list(estm1$var, estm2$var)
   paket <- paste("paket", 1:2, sep = "")
   ```

e. Membuat objek kelas dengan sintak modIRT untuk semua Model IRT

   1) Model 1PL

   ```
   > modRasch<-modIRT(coef = estc, var = estv, names = paket, display = FALSE)
   ```

2) Model 2PL

```
> mod2PL<-modIRT(coef = estc, var = estv, names = paket, display = FALSE
```

3) Model 3PL

```
> mod3PL<-modIRT(coef = estc, var = estv, names = paket, display = FALSE
```

f. Melakukan Estimasi semua koefisien Penyetaraan langsung untuk Paket 1 dan Paket 2 dengan sintaks sebagai berikut:

1) Metode mean-mean

```
> MM1PL<-alldirec(mods = modRasch, method = "mean-mean")#1PL
> summary(MM1PL)
> MM2PL<-alldirec(mods = mod2PL, method = "mean-mean")#2PL
> summary(MM2PL)
> MM3PL<-alldirec(mods = mod3PL, method = "mean-mean")#3PL
> summary(MM3PL)
```

2) Metode mean-sigma

```
> MS1PL<-alldirec(mods = modRasch, method = "mean-sigma")#1PL
> summary(MS1PL)
> MM2PL<-alldirec(mods = mod2PL, method = "mean-sigma")#2PL
> summary(MM2PL)
> MS3PL<-alldirec(mods = mod3PL, method = "mean-sigma")#3PL
> summary(MS3PL)
```

3) Metode Haebara

```
> HA1PL<-alldirec(mods = modRasch, method = "Haebara")#1PL
> summary(HA1PL)
> HE2PL<-alldirec(mods = mod2PL, method = "Haebara")#2PL
> summary(HE2PL)
> HE3PL<-alldirec(mods = mod3PL, method = "Haebara")#3PL
> summary(HE3PL)
```

4) Stocking-Lord

```
> SL1PL<-alldirec(mods = modRasch, method = "Stocking-Lord")
> summary(SL1PL)
> SL2PL<-alldirec(mods = mod2PL, method = "Stocking-Lord")
> summary(SL2PL)
```

```
> SL3PL<-alldirec(mods = mod3PL, method = "Stocking-Lord")
> summary(SL3PL)
```

C. Hasil Dan Pembahasan

1. Hasil Penelitian

a. Karakteristik Item Tes dan Estimasi Koefisien Penyetaraan

Analisis butir dengan pendekatan *item response theory* (IRT) menggunakan Program R. Komponen psikometri yang menjadi fokus pada bagian ini adalah fungsi informasi dan jumlah item fit untuk masing-masing paket dan parameter logistik yang digunakan. Secara umum untuk melihat perbandingan karakteristik butir dari masing-masing paket berdasarkan model yang digunakan serta total informasi dari paket adalah disajikan pada tabel 2.

Tabel 2. Jumlah Item Fit dan Fungsi Informasi Tes

Model	Paket 1		Paket 2	
	Jumlah Item Fit	Informasi Tes	Jumlah Item Fit	Informasi Tes
1PL	32	63,43%	36	51,99%
2PL	28	77,97%	22	73,31%
3PL	25	97,05%	28	95,04%

Sumber: Data Primer, **Tahun** : 2021

Ringkasan Hasil estimasi Koefisien penyetaraan dengan program R dengan menggunakan Langkah estimasi pada bagian analisis data disajikan pada tabel 3.

Tabel 3. Hasil Estimasi Koefisien penyetaraan dengan berbagai model menggunakan program R

Model	Mean-mean	Mean-Sigma	Haebara	Stocking-Lord
1 PL	Link: paket1.paket2 Method: mean-mean Equating coefficients: Estimate StdErr A 0.2724 1.6662 B 6.9945 6.7895	Link: paket1.paket2 Method: mean-sigma Equating coefficients: Estimate StdErr A 24.277 17.883 B -30.075 19.425	Link: paket1.paket2 Method: Haebara Equating coefficients: Estimate StdErr A 1.1813 0.057951 B 1.6482 0.146459	Link: paket1.paket2 Method: Stocking-Lord Equating coefficients: Estimate StdErr A 1.2804 0.059595 B 1.4475 0.143125
	Link: paket2.paket1 Method: mean-mean Equating coefficients: Estimate StdErr A 3.6711 22.455	Link: paket2.paket1 Method: mean-sigma Equating coefficients: Estimate StdErr A 0.041192 0.030344	Link: paket2.paket1 Method: Haebara Equating coefficients: Estimate StdErr A 0.78753 0.037296	Link: paket2.paket1 Method: Stocking-Lord Equating coefficients: Estimate StdErr A 0.78684 0.036685

Model	Mean-mean	Mean-Sigma	Haebara	Stocking-Lord
	B -25.6775 167.515	B 1.238835 NaN	B -1.14917 0.112368	B -1.13482 0.110185
2 PL	Link: paket1.paket2 Method: mean-mean Equating coefficients: Estimate StdErr A 1.5693 0.22003 B 6.9936 8.99166	Link: paket1.paket2 Method: mean-sigma Equating coefficients: Estimate StdErr A 9.0837 8.1975 B -3.0125 1.9095	Link: paket1.paket2 Method: Haebara Equating coefficients: Estimate StdErr A 1.2614 0.18168 B 1.3890 0.20738	Link: paket1.paket2 Method: Stocking-Lord Equating coefficients: Estimate StdErr A 1.4484 0.19367 B 1.2547 0.17379
	Link: paket2.paket1 Method: mean-mean Equating coefficients: Estimate StdErr A 0.63722 0.089343 B-4.45645 5.425454	Link: paket2.paket1 Method: mean-sigma Equating coefficients: Estimate StdErr A 0.11009 0.099347 B 0.33164 0.139722	Link: paket2.paket1 Method: Haebara Equating coefficients: Estimate StdErr A 0.92365 0.14256 B -0.84544 0.15378	Link: paket2.paket1 Method: Stocking-Lord Equating coefficients: Estimate StdErr A 0.66132 0.095996 B -0.83647 0.104414
3 PL	Link: paket1.paket2 Method: mean-mean Equating coefficients: Estimate StdErr A 0.2724 1.6662 B 6.9945 6.7895	Link: paket1.paket2 Method: mean-sigma Equating coefficients: Estimate StdErr A 24.277 17.883 B -30.075 19.425	Link: paket1.paket2 Method: Haebara Equating coefficients: Estimate StdErr A 1.2918 0.17499 B 1.1211 0.31654	Link: paket1.paket2 Method: Stocking-Lord Equating coefficients: Estimate StdErr A 0.63122 0.025189 B 1.27515 0.141607
	Link: paket1.paket2 Method: mean-mean Equating coefficients: Estimate StdErr A 3.6711 22.455 B -25.6775 167.515	Link: paket2.paket1 Method: mean-sigma Equating coefficients: Estimate StdErr A 0.041192 0.030344 B 1.238835 NaN	Link: paket2.paket1 Method: Haebara Equating coefficients: Estimate StdErr A 1.32911 0.66384 B -0.87177 0.60980	Link: paket2.paket1 Method: Stocking-Lord Equating coefficients: Estimate StdErr A 2.1257 0.33865 B -2.9442 0.67034

Sumber: Data Primer, **Tahun** : 2021

Secara ringkas persamaan regresi penyetaraan untuk masing-masing metode dengan menggunakan formula (14) dan (15) disajikan pada tabel 4.

Tabel 4. Persamaan Regresi Penyetaraan Tes Paket 1 ke 2 dan Paket 2 ke Paket 1

Model	Arah	MM	MS	HA	SL
1PL	1 ke 2	bi*= 0,27bi + 6,99	bi*= 24,28bi - 0,08	bi*= 1,29bi + 1,12	bi*= 0,63bi + 1,28
		ai* = 1; ci* = 0	ai* = 1; ci* = 0	ai* = 1; ci* = 0	ai* = 1; ci* = 0
	2 ke 1	bi*= 3,67bi -25,68	bi*= 0,04bi 1,24	bi*= 1,33bi -0,87	bi*= 2,13bi -2,94
		ai* = 1; ci* = 1	ai* = 1; ci* = 1	ai* = 1; ci* = 1	ai* = 1; ci* = 1
2PL	1 ke 2	bi*= 1,57bi + 6,99	bi*= 9,08b - 3,01	bi*= 1,26bi + 1,39	bi*= 1,45bi + 1,25
		ai* = ai/1,57 ; ci* = 0	ai* = ai/9,08 ; ci* = 0	ai* = ai/1,26 ; ci* = 0	ai* = ai/1,45 ; ci* = 0
	2 ke 1	bi*= 0,64bi -4,46	bi*= 0,11bi 0,33	bi*= 0,92bi -0,85	bi*= 0,66bi -0,84
		ai* = ai/0,64 ; ci* = 0	ai* = ai/0,11 ; ci* = 0	ai* = ai/0,92 ; ci* = 0	ai* = ai/0,66 ; ci* = 0
3PL	1 ke 2	bi*= 0,27bi + 6,99	bi*= 24,28bi - 30,08	bi*= 1,29bi + 1,12	bi*= 0,63bi + 1,28
		ai* = ai/0,64 ; ci* = ci	ai* = ai/0,11 ; ci* = ci	ai* = ai/0,92 ; ci* = ci	ai* = ai/0,66 ; ci* = ci
	2 ke 1	bi*= 3,67bi -25,68	bi*= 0,04bi - 1,24	bi*= 1,33bi - 0,87	bi*= 2,13bi -2,94
		ai* = ai/0,27 ; ci* = ci	ai* = ai/24,28 ; ci* = ci	ai* = ai/1,29 ; ci* = ci	ai* = ai/0,63 , ci* = ci

Sumber: Data Primer, **Tahun** : 2021

Berdasarkan tabel 4, penyetaraan perangkat soal *try out* UN matematika paket 1 dan paket 2 dapat dilakukan mengacu pada persamaan baru untuk menyetarakan tes. Dengan menggunakan bantuan MS. Excel, maka dapat dilakukan estimasi parameter butir baik pada model 1 PL, 2 PL, maupun 3 PL untuk mengetahui kesetaraan paket 1 dan paket 2.

b. Keakuratan Estimasi Penyetaraan

Tabel 5. Perbandingan RMSD Berbagai Model Penyetaraan

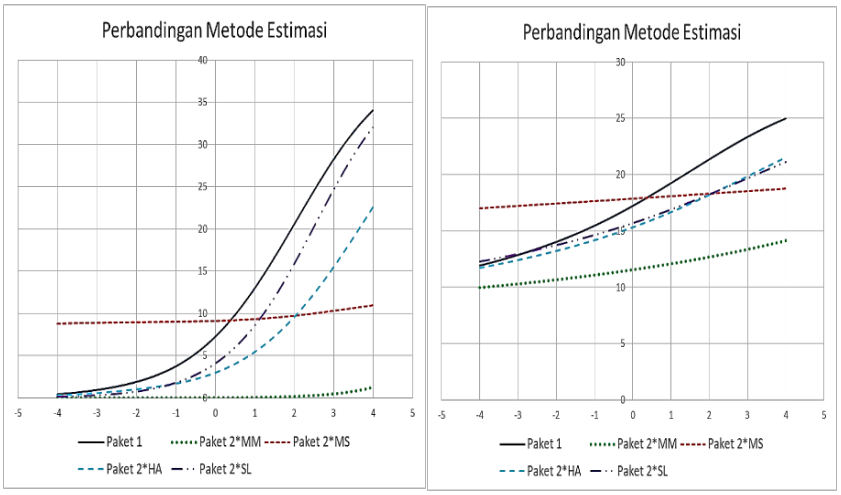
Model	Arah	Karakteristik	RMSD			
			MM	MS	HA	SL
1PL	**1 ke 2**	**b**	5,72	32,03	1,70	0,75
	2 ke 1	**b**	17,18	2,37	0,51	1,76
2 PL	**1 ke 2**	**b**	17,09	229,36	7,43	12,67
		a	0,15	0,37	0,09	0,13
	2 ke 1	**b**	87,17	215,10	18,35	81,73
		a	0,24	3,34	0,03	0,21
3PL	**1 ke 2**	**b**	13,09	394,98	5,34	6,21
		a	9,04	3,25	0,76	1,98
	2 ke 1	**b**	55,20	20,41	6,90	23,60
		a	5,72	32,03	1,70	0,75
Rata-rata		**b**	32,574	17,18	2,37	0,51
		a	3,493	17,09	229,36	7,43

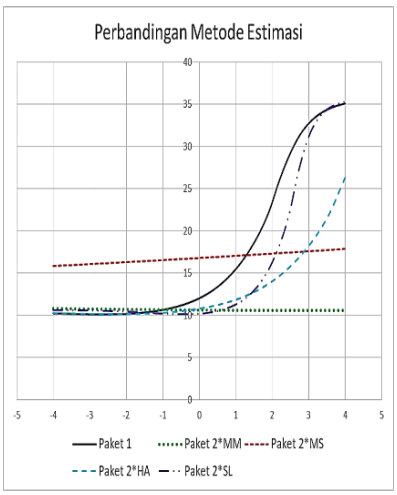
Sumber: Data Primer, **Tahun** : 2021

Setelah menentukan persamaan regresi dari masing-masing arah dan model IRT yang digunakan, selanjutnya adalah menentukan keakuratan estimasi berdasarkan metode yang penyetaraan yang digunakan dengan menghitung nilai RMSD yang terkecil dari keempat metode yang digunakan menggunakan formula (17) dan (18). Tabel berikut menyajikan perbandingan nilai RMSD dari masing-masing model dam metode estimasi penyetaraan yang digunakan.

c. Hasil Penyetaraan Paket 1 ke Paket 2

Hasil regresi pada tabel 4 kemudian digunakan untuk menentukan karakteristik baru untuk seluruh butir pada paket 2 dengan menggunakan berbagai metode penyetaraan. Secara visual, perbandingan keempat metode penyetaraan berdasarkan model IRT yang digunakan dapat dilihat pada gambar 2. Berdasarkan gambar 2, dapat diketahui bahwa untuk penyetaraan paket 1 ke paket 2*, untuk semua model IRT, pada model 1 PL metode penyetaraan yang paling presisi adalah metode Stocking-Lord, untuk 2 PL yang paling dekat adalah metode Haebara, sedangkan untuk model 3 PL adalah model Stocking-Lord.



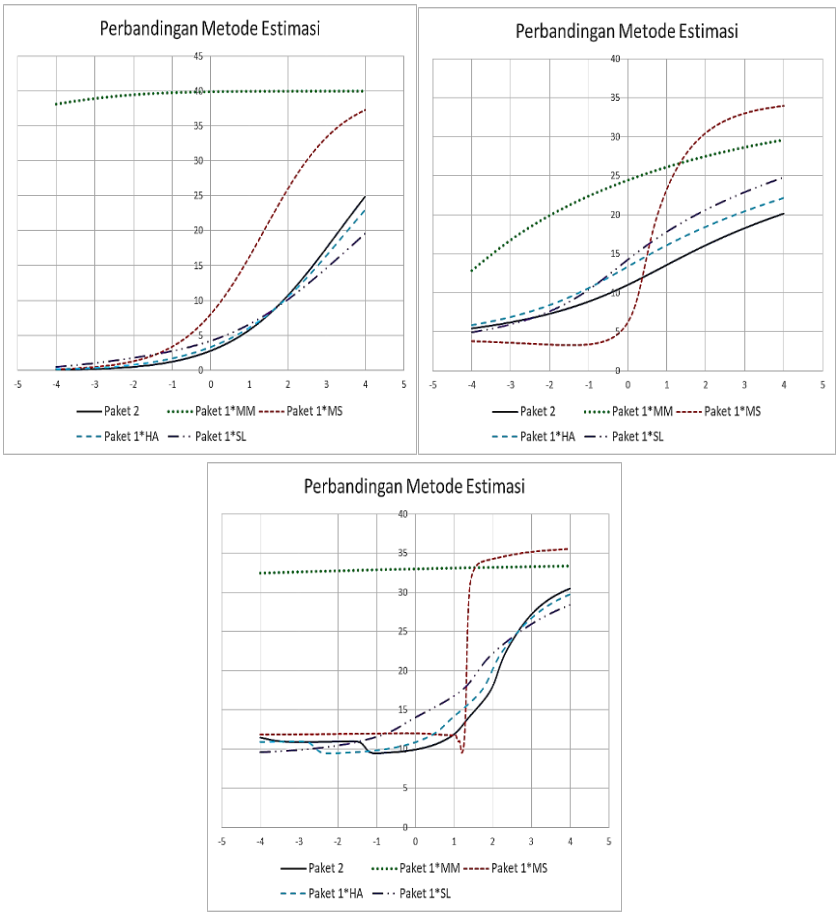


Gambar 1. Grafik Perbandingan Berbagai Metode Estimasi Penyetaraan Paket 1 ke Paket 2

d. Penyetaraan Paket 2 ke Paket 1

Secara visual, perbandingan keempat metode penyetaraan berdasarkan model IRT yang digunakan dapat dilihat pada ambar 2. Berdasarkan gambar 2, dapat diketahui bahwa untuk penyetaraan paket 2 ke paket 1*, untuk semua model IRT, pada model 1 PL, 2PL, dan 3PL metode penyetaraan yang paling presisi adalah metode Haebara. hal ini sejalan dengan estimasi RMSD dari metode Haebara yang memiliki indeks yang paling kecil di antara seluruh metode penyetaraan yang lain.

Meskipun demikian, tidak ada satupun metode penyetaraan yang dapat memberikan hasil yang sangat presisi untuk paket 1 dan paket 2 dari soal *try out* UN SMP tahun 2014 pada mata pelajaran matematika, hal ini kelihatan pada grafik nampak bahwa pakel 2 memiliki kesukaran yang lebih dibandingkan dengan paket 1. Grafik *Test Charasteristics Curve*/ TCC, kelihatan bahwa paket 2 cenderung lebih ke kanan dibandingkan dengan paket 1 untuk semua metode penyetaraan yang digunakan.



Gambar 2. Grafik Perbandingan Berbagai Metode Estimasi Penyetaraan Paket 2 ke Paket 1

2. Pembahasan

Karakteristik butir paket 1 dan paket 2 soal *try out* UN matematika menunjukkan masih banyak item yang belum fit yang artinya item tersebut tidak mampu menggambarkan kemampuan siswa (Sainuddin, 2014). Item/butir ini perlu direvisi agar dapat digunakan pada tes yang akan datang. Item dengan karakteristik yang baik dapat dijadikan bank soal yang dapat sewaktu-waktu digunakan untuk tes yang relevan seperti ujian semester dan ujian kompetensi. Informasi Tes menunjukkan bahwa Model 3PL memiliki informasi yang cukup menggambarkan kemampuan peserta tes (Syafii et al., 2021) dibandingkan dua model lainnya dalam hal ini siswa pada rentang kemampuan -4 sampai 4 skala logit. Hal ini disebabkan model 3PL mempertimbangkan 3 karakteristik yakni kesukaran, daya beda dan tebakan semu peserta tes (Meng et al., 2016).

Berdasarkan tabel 3, dapat diketahui bahwa paket 2 cenderung lebih sulit dibandingkan dengan paket 1 hal ini terlihat bahwa koefisien dari 1 ke 2 menghasilkan nilai A (α) dan B (β) yang positif sehingga jika dilakukan penyetaraan terhadap parameter butir akan menghasilkan angka yang lebih besar dan positif (Papanastasiou, 2015; Uyar & Gübeş, 2020). Hasil skoring tes dengan menggunakan *row score* akan menghasilkan penilaian yang tidak adil (Ketterlin-Geller et al., 2018) jika menggunakan skor asli dari paket 2 yang memiliki tingkat kesulitan yang lebih dibandingkan paket 1. Sehingga solusi yang paling tepat adalah dengan menggunakan penyetaraan terlebih dahulu untuk paket 1 ke paket 2 atau sebaliknya sehingga skor yang diperoleh lebih adil. Berbagai metode penyetaraan (Elvira & Sainuddin, 2021; Retnawati, 2014) dapat dipertimbangkan dalam penyetaraan skor kemampuan siswa ke depannya, namun perlu memperhatikan keakuratan metode tersebut.

Khusus pada kasus soal *try out* UN matematika ini ternyata menunjukkan bahwa estimasi dengan menggunakan metode Haebara dan Stocking-Lord tingkat presisi yang memadai (0.608) (Elvira & Sainuddin, 2021). Selain itu, hal ini didukung pula pada hasil penyetaraan pada gambar 1 dan gambar 2 yang menunjukkan hampir semua model IRT yang digunakan menunjukkan metode penyetaraan Haebara dan Stocking-Lord mendekati skor paket soal yang disetarakan (Nisa & Retnawati, 2018). Kedua metode ini memiliki tingkat presisi yang mumpuni karena keduanya memiliki kebaruan dan pendekatan statistik yang lebih baik dibandingkan dua metode lainnya (Kolen & Brennan, 2014).

Penelitian yang dilakukan ini menunjukkan hasil estimasi koefisien penyetaraan dari berbagai model yang digunakan umumnya masih memiliki nilai presisi yang sangat jauh dari harapan hal ini dapat dilihat pada gambar 1 dan 2. Menurut hemat penulis, hal ini terjadi karena beberapa faktor sebagai berikut: 1). Peserta tes masih kurang serius mengerjakan soal *try out* ini dikarenakan tidak adanya dampak bagi mereka sehingga beberapa orang cenderung mengerjakan tidak serius (Zhang & Yang, 2014), 2). Item tes yang digunakan masih dalam tahap pengembangan, sehingga ada beberapa item/butir yang sangat sulit dan belum dipahami oleh peserta tes (Metsämuuronen, 2012; Runnels, 2013), 3). Kesiapan peserta tes mengerjakan soal (Hamid et al., 2019; Tavakoli & Samian, 2014) *try out*

ini sehingga masih banyak item yang tidak dapat digambarkan, dan 4). Distribusi item/butir tes yang kurang proporsional (Feng et al., 2019; Wang et al., 2017)sehingga ditemukan bahwa paket 2 lebih sulit dibandingkan dengan paket 1. Soal Bersama (*anchor)* yang digunakan pada paket 1 dan paket dua hanya terdiri atas tiga item dari 40 item (8%). Idealnya item bersama yang digunakan dalam analisis penyetaraan model *anchor* adalah 20% sampai 40% untuk mendapatkan hasil penyetaraan yang lebih akurat (Kolen & Brennan, 2014; Syahrul et al., 2016), hal ini juga yang menjadi penyebab beberapa metode penyetaraan kurang presisi.

D. Kesimpulan

Kesimpulan yang dapat diambil mengenai hubungan soal UN paket 1 dan paket 2, tidak dapat dikatakan setara, hal ini diketahui setelah melakukan analisis dan menghitung persamaan regresi, diketahui bahwa paket 2 soal *try out* UN memiliki tingkat kesukaran yang lebih tinggi dibandingkan paket 1. penyetaraan paket 1 ke paket 2*, untuk semua model IRT, pada model 1 PL metode penyetaraan yang paling presisi adalah metode Stocking-Lord, untuk 2 PL yang paling dekat adalah metode Haebara, sedangkan untuk model 3 PL adalah model Stocking-Lord. penyetaraan paket 2 ke paket 1*, untuk semua model IRT, pada model 1 PL, 2PL, dan 3PL. Penggunaan butir bersama sangat kecil yakni berkisar 8% saja sehingga mempengaruhi hasil estimasi koefisien penyetaraan dan tingkat presisi hasil penyetaraan tes paket 1 dan paket 2 *try out* UN.

Saran

Berdasarkan penelitian ini bahwa metode yang tepat digunakan untuk estimasi koefisien penyetaraan tes adalah metode Haebara dan Stocking -Lord. Penggunaan item/butir bersama ada tes paralel hendaknya berkisar antara 20 – 40 %. Hal ini dapat meningkatkan keakuratan estimasi model penyetaraan. Penyetaraan tes paralel sangat penting bagi kalangan pendidik untuk agar skor/kemampuan siswa yang diperoleh tidak tumpang tindih atau lebih adil.

Daftar Pustaka

Crocker, L., Alglna, J., Staudt, M., Mercurio, S., Hintz, K., & Walker, R. A. (2008). I ntroduction to Classical and Modern Test Theory. In *Cengage Learning*.

Elvira, M., & Sainuddin, S. (2021). Equating Test Instruments Using Anchor to Map Student Abilities Through the R Program Analysis. *Proceedings of the International Conference on Engineering, Technology and Social Science (ICONETOS 2020)*, *529*(Iconetos 2020), 651–657. https://doi.org/10.2991/assehr.k.210421.095

Feng, Y., Qiao, Y., Zhao, X., & Li, J. (2019). Study on Sample Size of Candidates Oriented to Online Test. *In 2019 14th International Conference on Computer Science & Education (ICCSE)*, 1006–1010.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Apanese Psychological Research*, *22*(3), 144–149.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Priciple and Applications* (1 st editi). Springer Science+Business Media, LLC.

Hamid, M. O., Hardy, I., & Reyes, V. (2019). Test-takers' perspectives on a global test of English: questions of fairness, justice and validity. *Language Testing in Asia*, *9*(1). https://doi.org/10.1186/s40468-019-0092-9

Ketterlin-Geller, L., Perry, L., Platas, L., & Sitbakhan, Y. (2018). Aligning Test Scoring Procedures with Test Uses of the Early Grade Mathematics Assessment: A Balancing Act. *Global Education Review*, *5*(3), 143–164.

Kilmen, S., & Demirtasli, N. (2012). Comparison of Test Equating Methods Based on Item Response Theory According to the Sample Size and Ability Distribution. *Procedia - Social and Behavioral Sciences*, *46*(1980), 130–134. https://doi.org/10.1016/j.sbspro.2012.05.081

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling and linking. Methods and Practice* (Third). Springer Science+Business Media, LLC. https://doi.org/10.1016/s0140-6736(66)90249-2

Loyd, B. H., & Hoover, H. D. (2016). *Vertical Equating Using the Rasch Model*. *17*(3), 179–193.

Marco, G. (1977). Item Characteristic Curve Solution. *ETS Research Bulletin Series*, i–41.

Meng, H., Guo, F., & Han, K. C. T. (2016). Conquer the IRT Hurdle for Adaptive Test Designs. *ITC 2016 Conference*.

Metsämuuronen, J. (2012). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies*,

4(3), 1–22. https://doi.org/10.5539/ijps.v4n3p1

Nisa, C., & Retnawati, H. (2018). Comparing the methods of vertical equating for the math learning achievement tests for junior high school students. *Research and Evaluation in Education*, *4*(2), 164–174. https://doi.org/10.21831/reid.v4i2.19291

Papanastasiou, E. C. (2015). Psychometric changes on item difficulty due to item review by examinees. *Practical Assessment, Research and Evaluation*, *20*(3), 1–10.

Retnawati, H. (2014). Perbandingan Metode Penyetraan Skor Tes Menggunakan Butir Bersama dan Tanpa Butir Bersama. *Jurnal Kependidikan*, *46*(2), 164–178. https://journal.uny.ac.id/index.php/jpep/article/view/4551

Runnels, J. (2013). Measuring differential item and test functioning across academic disciplines. *Language Testing in Asia*, *3*(1), 1–11. https://doi.org/10.1186/2229-0443-3-9

Sainuddin, S. (2014). Analisis Karakteristik Butir Tes Matematika pada Tes Buatan MGMP Matematika Kota Makassar Berdasarkan Teori Modern (Teori Respon Butir). *Jurnal Penelitian Dan Pendidikan Matematika*, *1*(Pendidikan Matematika), 1–12.

Stocking, M. L., & Lord, F. M. (1983). Developing a Common Metric in Item Response Theory. *Applied Psychological Measurement*, *7*(2), 201–210. https://doi.org/10.1177/014662168300700208

Syafii, A., Haryanto, H., Ahmad, I. F., & Fauziah, A. (2021). Analysis of Items with Item Response Theory (IRT) Approach on Final Assessment for Al-Quran Hadith Subjects. *Jurnal Pendidikan Agama Islam*, *18*(1), 167–194. https://doi.org/10.14421/jpai.2021.181-09

Syahrul, Mansyur, & Rosdiyanah. (2016). Pengaruh Jumlah Butir Anchor Terhadap Hasil Penyetaraan Tes Berdasarkan Teori Respon Butir. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, *46*(2), 207–218.

Tavakoli, E., & Samian, S. H. (2014). Test-wiseness Strategies in PBTs and IBTs: The Case of EFL Test Takers, Who Benefits More? *Procedia - Social and Behavioral Sciences*, *98*(Mc), 1876–1884. https://doi.org/10.1016/j.sbspro.2014.03.618

Uyar, Ş., & Gübeş, N. Ö. (2020). Item parameter estimation for dichotomous items based on item response theory: Comparison of BILOG-MG, Mplus and R (ltm). *Journal of Measurement and Evaluation in Education and Psychology*, *11*(1), 27–42. https://doi.org/10.21031/epod.591415

Wang, X., Liu, Y., & Hambleton, R. K. (2017). Detecting Item Preknowledge Using a Predictive Checking Method. *Applied Psychological Measurement*, *41*(4), 243–263. https://doi.org/10.1177/0146621616687285

Zhang, H. J., & Yang, B. (2014). Study On the Fuel-saving Efficiency of Electric Vehicles Under Empirical Test. In *Applied Mechanics and Materials* (pp. 95–99). Trans Tech Publications Ltd.