

Perbandingan Kemampuan Siswa Dengan *Test Score Equating*

Zainal Abidin Koemadji[†]

Abstrak

Ability parameters in item response theory will invariant in group of items, in other words, two different respondent with two test with known parameters will have ability estimation on the same scale. Equating of these different tests will give the equivalence of the tests. This paper gives examples on how to perform the equating.

Keywords: Ability, equating, true score.

1. Pendahuluan

Parameter ability (θ) dari seorang responden menurut *item response theory* (IRT) akan invariant pada sekelompok item, artinya, terlepas dari kesalahan pengukuran, dugaan parameter ability juga akan invariant pada sekelompok item. Sehingga, jika dua responden yang mengerjakan dua test berbeda yang diketahui nilai parameter itemnya akan memiliki dugaan ability pada skala yang sama. Tetapi hal sebaliknya terjadi jika parameter item dan ability tidak diketahui, untuk itu perlu dilakukan penskalaan atau *equating* dari dua test yang berbeda agar diperoleh skor yang ekuivalen.

Tulisan ini membahas tentang *test equating* dengan beberapa gugus data yang berbeda.

2. Test Score Equating (TSE)

Metode equating yang menggunakan teori *latent trait* disebut *true score equating* (TSE). Dengan metode TSE dapat diperoleh skor yang ekuivalen dari dua jenis test. Hubungan antara *latent trait score* (LTS) dengan *true score* (TS) diberikan oleh persamaan

$$T = \sum_{g=1}^G P_g(\theta)$$

dimana G adalah banyaknya item, θ adalah ability dan $P_g(\theta)$ adalah model logistik tiga parameter. Dari nilai T dapat kita peroleh kurva karakteristik test untuk setiap test yang dilakukan, dan dari kurva ini, skor untuk dua test dapat dikatakan ekuivalen jika dihubungkan pada LTS yang sama.

Misalkan terdapat dua jenis test, yaitu X dan Y , dan dimungkinkan untuk memperoleh skor seorang responden dari test X , dengan plot TS yang ekuivalen dapat diduga skor responden tersebut pada test Y walaupun tidak dilakukan pengukuran dengan menggunakan test Y , sehingga plot TS yang ekuivalen menyatakan equating dari TS .

Lebih lanjut dapat dilakukan pendugaan skor yang ekuivalen dengan menggunakan plot TS yang ekuivalen sebagai fungsi $Y^* = f(x)$, yang akan mentransformasi skor X ke skor yang ekuivalen pada skor dari test Y . Penentuan fungsi tersebut dapat dilakukan dengan menggunakan analisis regresi.

[†] Staf Pengajar pada Jurusan Matematika Universitas Negeri Gorontalo

3. Data

Data yang digunakan adalah data simulasi dan data real, dengan tiga buah prosedur yang diuraikan berikut :

- A. Untuk Prosedur A, data yang digunakan adalah data simulasi dengan membangkitkan tiga gugus data (X, Y, Z) dengan menganggap bahwa ketiga gugus ini memiliki kemampuan (*ability*) yang setara, tetapi dengan tingkat kesulitan soal yang berbeda. Data dibangkitkan dengan perbedaan interval pada sebaran untuk parameter tingkat kesulitan (*difficulty*). Tingkat kesulitan untuk gugus data X , $U(-2, 0)$, untuk gugus data Y adalah $U(0, 1)$ dan $U(1, 2)$ untuk gugus data Z . Hal ini didasarkan pada asumsi bahwa rentang nilai tingkat kesulitan berada pada selang $(-2, 2)$. Jumlah item adalah 60 dengan 200 responden setiap gugus data.
- B. Prosedur B merupakan modifikasi dari prosedur A, dimana selain terdapat perbedaan pada sebaran tingkat kesulitan, juga diasumsikan bahwa antara setiap gugus memiliki kemampuan yang berbeda. Sebaran kemampuan dibangkitkan dengan sebaran Normal dengan parameter yang berbeda, gugus data X dianggap memiliki kemampuan terendah dibangkitkan dengan sebaran $N(0, 0.25)$, gugus data Y dianggap memiliki kemampuan yang lebih baik, dibangkitkan dengan sebaran $N(0.25, 0.5)$, dan gugus data Z dibangkitkan dengan sebaran $N(0.5, 1)$.
- C. Prosedur C dilakukan dengan menggunakan data real, data terdiri atas dua gugus (X dan Y), dengan 200 responden untuk setiap gugus, dan 60 item dengan 17 *anchor* item. Pendugaan parameter dilakukan dengan menganggap keseluruhan responden (400 responden) mengikuti test dengan jumlah item $43 + 43 + 17 = 103$, dimana $n_X = 43$, $n_Y = 43$ adalah *unique* item untuk kedua test, dan $n_a = 17$ adalah *anchor* item. Untuk gugus data X , n_Y dianggap sebagai banyaknya item yang tidak dikerjakan karena tidak cukup waktu (*not reached*), demikian juga sebaliknya.

4. Hasil dan Pembahasan

Prosedur A.

Analisis deskriptif dari parameter ketiga gugus adalah sebagai berikut.

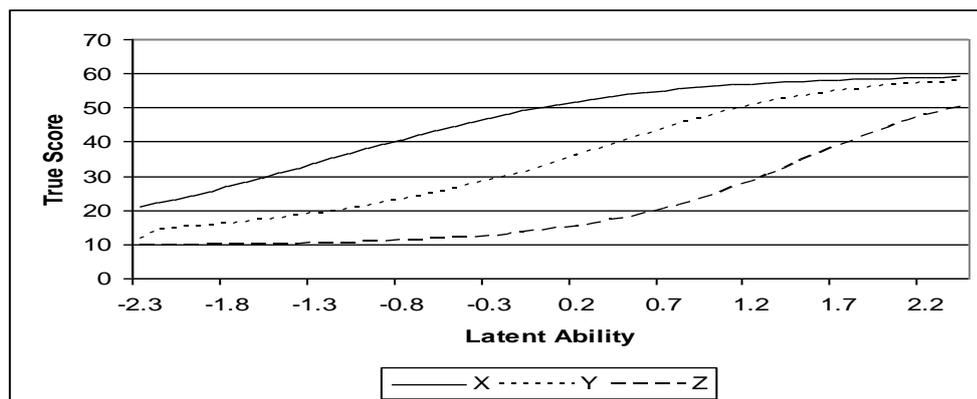
Tabel 1. Dugaan Parameter Model (A)

Statistik	X			Y			Z		
	a	b	c	a	b	c	a	b	c
Mean	0.98224	- 1.10048	0.20218	1.10433	0.54497	0.17666	1.06214	1.49703	0.15592
SE Mean	0.05007	0.10103	0.00445	0.05429	0.04904	0.00861	0.04426	0.05024	0.00862
S. Baku	0.38782	0.78254	0.03446	0.42054	0.37987	0.06666	0.34282	0.38919	0.06674
Min	0.36150	- 2.83107	0.08410	0.42562	- 1.88040	0.05036	0.57971	0.58540	0.03259
Maks	2.05934	0.54500	0.28740	2.26614	1.56418	0.36178	1.81867	2.17371	0.29281

Analisis deskriptif dari ability dan plot TS untuk Prosedur A diberikan pada tabel dan gambar berikut.

Tabel 2. Ability (A)

	X	Y	Z
Mean	0.00421	-0.01178	-0.00548
SE Mean	0.07008	0.07086	0.06403
Simpangan baku	0.99102	1.00205	0.90552
Minimum	-2.90729	-1.80435	-1.31647
Maksimum	1.89964	2.17238	2.37557



Gambar 1. Plot True Score Prosedur A.

Hasil di atas menunjukkan bahwa gugus data Z dengan tingkat kesulitan tertinggi akan menghasilkan skor yang terendah dari ketiga gugus untuk ability yang sama. Dari plot tersebut dapat diduga skor pada test Y dan Z jika diketahui skor pada test X, contohnya jika diketahui skor pada test X untuk ability $\theta = 1.1$ adalah 56.68975, maka skor pada test Y untuk ability yang sama adalah 49.33207, dan skor pada test Z adalah 26.52985.

Prosedur B.

Analisis deskriptif dari parameter ketiga gugus adalah sebagai berikut.

Tabel 3. Dugaan Parameter Model (B)

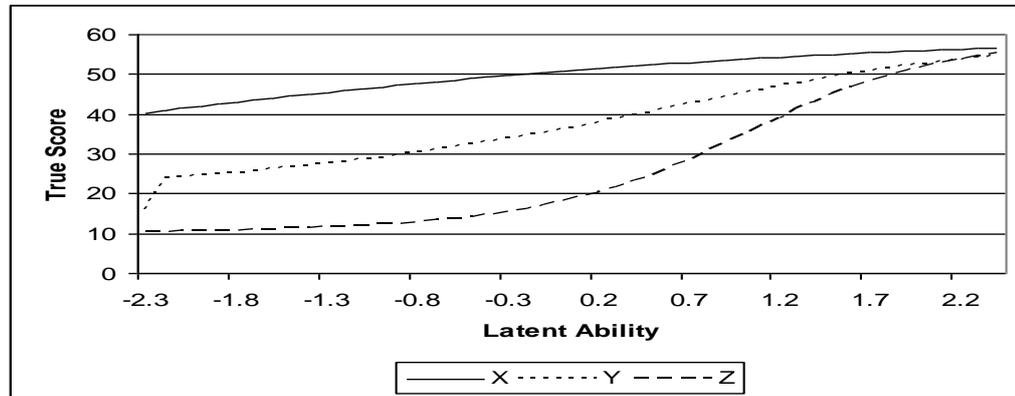
Statistik	X			Y			Z		
	a	b	c	a	b	c	a	b	c
Mean	0.39969	-	0.21069	0.68828	0.59652	0.22007	1.12066	1.00955	0.16474
SE Mean	0.01397	0.25131	0.00189	0.03079	0.08905	0.00833	0.05477	0.04849	0.00893
S. Baku	0.10819	1.94664	0.01463	0.23851	0.68976	0.06449	0.42424	0.37565	0.06913
Min	0.22576	-	0.19514	0.31527	-	0.06184	0.57499	0.21866	0.03485
Maks	0.61267	0.73662	0.25893	1.37951	2.16611	0.47472	2.22561	1.93586	0.32691

Analisis deskriptif dari ability dan plot TS untuk Prosedur B diberikan pada tabel dan gambar berikut.

Tabel 4. Ability (B)

	X	Y	Z
--	---	---	---

Mean	-0.03531	0.02188	-0.00614
SE Mean	0.04719	0.06468	0.06909
Simpangan baku	0.66732	0.91469	0.97702
Minimum	-1.70122	-2.03632	-1.61397
Maksimum	1.65227	2.15844	2.25994



Gambar 2. Plot True Score Prosedur B.

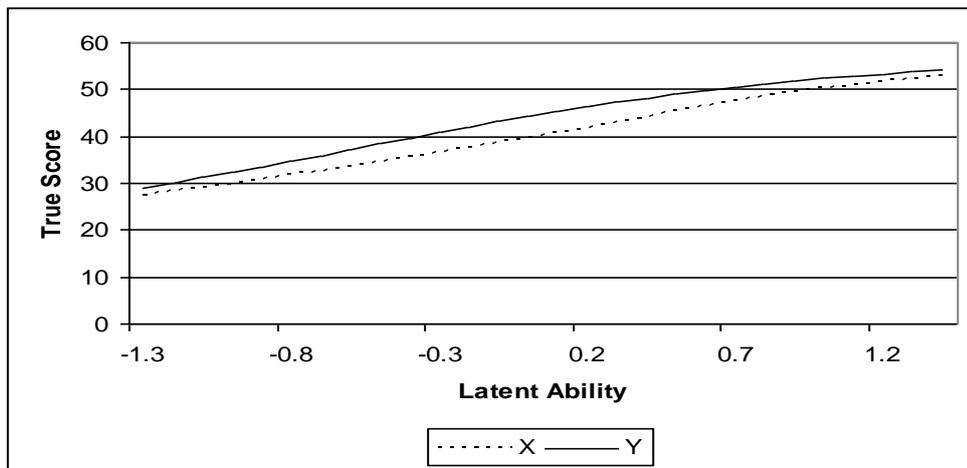
Gambar 2 menunjukkan bahwa dengan tingkat kesulitan yang berbeda dan ability yang berbeda, pola yang dihasilkan tidak jauh berbeda dengan hasil yang diperoleh pada prosedur A (Gambar 1), walaupun pada ability tertentu, gugus data Z akan memiliki skor yang lebih baik dari gugus data Y, dan mendekati skor gugus data X. Dari plot tersebut dapat diduga skor pada test Y dan Z jika diketahui skor pada test X, contohnya jika diketahui skor pada test X untuk ability $\theta = 1.1$ adalah 54.07910, maka skor pada test Y untuk ability yang sama adalah 46.09114, dan skor pada test Z adalah 36.90527.

Prosedur C.

Analisis deskriptif dari parameter ketiga gugus adalah sebagai berikut.

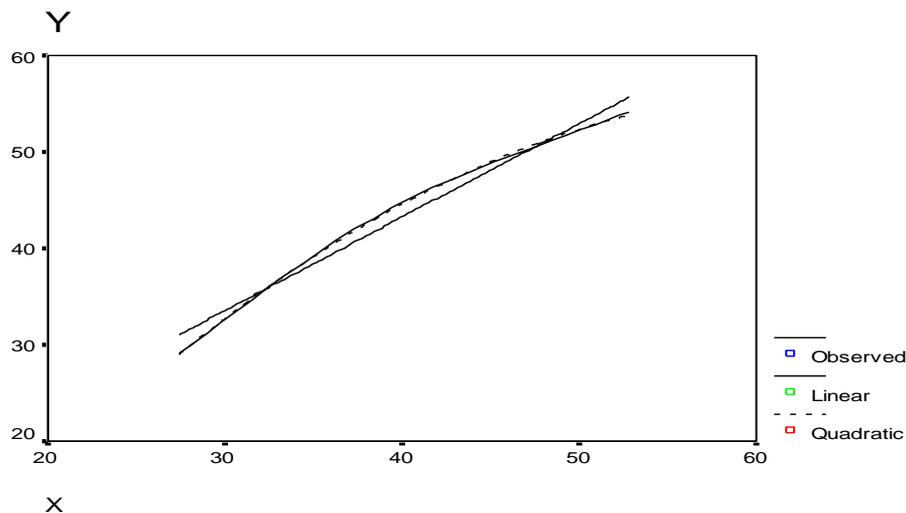
Tabel 5. Dugaan Parameter Model (C)

Statistik	Parameter		
	<i>a</i>	<i>b</i>	<i>c</i>
Mean	0.91324	-0.68573	0.22433
SE Mean	0.05062	0.19751	0.00777
S. Baku	0.51378	2.00450	0.07882
Min	0.09305	-9.21353	0.06026
Maks	2.98441	5.69622	0.50000



Gambar 3. Plot True Score Prosedur C

Hasil di atas menunjukkan bahwa gugus data *X* memiliki *true score* yang lebih rendah dari gugus data *Y* untuk ability yang sama. Dari plot tersebut dapat diduga skor pada test *Y* jika diketahui skor pada test *X*, contohnya jika diketahui skor pada test *X* untuk ability $\theta = 1.1$ adalah 50.93361, maka skor pada test *Y* untuk ability yang sama adalah 52.86667. Pendugaan skor yang ekuivalen juga dilakukan dengan menentukan fungsi $Y^* = f(x)$, dengan analisis regresi diperoleh hasil



Gambar 4. Plot Regresi Prosedur C

Dari plot regresi terlihat model yang sesuai untuk fungsi $Y^* = f(x)$ adalah regresi kuadratik, dengan model :

$$Y = -26.878503 + 2.590686X - 0.020114X^2$$

dengan R^2 model adalah 0.99966. Sebagai perbandingan, untuk $X = 50.93361$ dengan model di atas diperoleh dugaan $Y = 48.85817$. Hasil analisis regresi selengkapnya adalah sebagai berikut :

Multiple R		.99983			
R Square		.99966			
Adjusted R Square		.99963			
Standard Error		.15531			
Analysis of Variance:					
	DF	Sum of Squares		Mean Square	
Regression	2	1759.16327		879.58163	
Residuals	25	.60301		.02412	
F =	36466.59166		Signif F =	.0000	
----- Variables in the Equation -----					
Variable		B	SE B	Beta	T Sig T
X		2.590686	.044639	2.648679	58.036 .0000
X**2		-.020114	.000552	-1.663481	-36.449 .0000
(Constant)		-26.878503	.873312		-30.778 .0000

5. Kesimpulan

Sifat invariant dari parameter item dan *ability* memungkinkan dilakukannya *equating* pada *item response theory*, dimana parameter item dan *ability* memiliki hubungan linier pada dua kelompok. Jika telah diperoleh bentuk hubungan linier tersebut, dugaan parameter item dan *ability* dapat dibuat memiliki skala yang sama. Hasil-hasil di atas menunjukkan bagaimana penggunaan metode *test score equating* digunakan untuk beberapa kelompok data, khususnya prosedur C, perlu dilakukan penelitian lebih lanjut untuk melihat keakuratan prosedur ini.

Daftar Pustaka

- [1] Crocker, L., Algina, J., 1986, "*Introduction to Classical and Modern Test Theory*". Hartcourt Brace Jovanovich College Publishers, Florida.
- [2] Hambleton, R.K. *et al.*, 1991, "*Fundamentals of Item Response Theory, Volume 2*". SAGE Publications, California.
- [3] Yu, Chong Ho & Popp, S.E.O., 2005, "Test equating by common item and common subjects: Concepts and applications". *Practical Assessment Research & Evaluation*, 10(4). [<http://pareonline.net/getvn.asp?v=10&n=4>]