

## KEAKURASIAN PENYETARAAN VERTIKAL DENGAN METODE EKUIPRESENTIL PADA KEMAMPUAN BERPIKIR TINGKAT TINGGI (HOTS)

Riyan Arthur  
Fakultas Teknik, Universitas Negeri Jakarta  
[arthur@unj.ac.id](mailto:arthur@unj.ac.id)

### *Abstract*

*This research is aimed to know the differences in the accuracy of vertical equivalence with the representative method based on the proportion of items (Anchor Item) expressed by the value of Root Mean Square Error (RMSE) on High Order Thinking Skills (HOTS) of Vocational students on Building Engineering. This research used the experimental method. The experiments were not carried out in the learning process but were carried out after obtaining the answer patterns and scores of the test participants through the implementation of a High Order Thinking Skills test. The sampling technique was carried out with the Stratified Random Sampling with data analysis techniques using the T-test. The results of the study stated that there were differences in the accuracy of vertical equivalence of the equilibrium method in High Order Thinking Skills (HOTS) between students of class X with Class XI SMK Building Engineering for the anchor proportion 10%, 15 and 20%.*

*Keyword: Vertical Equivalence, Equipresentil Method, Anchor item*

### *Abstrak*

*Penelitian ini bertujuan untuk mengetahui perbedaan keakurasian penyetaraan vertikal dengan metode ekuiresentil berdasarkan proporsi butir pengait (Anchor Item) yang dinyatakan oleh nilai Root Mean Square Error (RMSE) pada kemampuan berpikir tingkat tinggi siswa SMK Teknik Bangunan. Metode yang digunakan pada penelitian ini adalah metode eksperimen. Eksperimen tidak dilakukan pada proses pembelajaran, tetapi dilakukan setelah memperoleh pola jawaban dan skor peserta tes melalui pelaksanaan tes kemampuan berpikir tingkat tinggi. Teknik Pengambilan sampel dilakukan dengan Stratified Random Sampling dengan Teknik analisis data menggunakan uji t. Hasil penelitian menyatakan bahwa terdapat perbedaan keakurasian penyetaraan vertikal metode ekuiresentil pada kemampuan berpikir tingkat tinggi (HOTS) antara siswa kelas X dengan kelas XI SMK Teknik Bangunan untuk proporsi jumlah butir pengait 10%, 15 dan 20%.*

*Kata Kunci: Penyetaraan Vertikal, Metode Ekuiresentil, Anchor item.*

### **PENDAHULUAN**

Salah satu peningkatan kualitas pembelajaran yang harus dilakukan adalah penilaian pada kemampuan berpikir tingkat tinggi siswa yang biasa disebut *High Order Thinking Skill* (HOTS). Menurut Brookhart (2007:5) “*Higher order thinking is approached as the “top end” of Bloom’s taxonomy: Analyze, Evaluate, and Create*”.

Artinya kemampuan analisa, sintesa dan mencipta menjadi penting dalam kemampuan ini (Basito et al., 2018; 2013: 7)

Naga mengungkapkan (2012:342) pada Teori Klasik, skor tes adalah relatif, bergantung kepada kelompok butir yang digunakan serta kepada kelompok peserta yang mengerjakannya. Oleh karena itu,

perbedaan kemampuan siswa yang ekstrim dalam suatu kelas akan membuat guru menemui masalah dalam pengukuran. Siswa yang memiliki kemampuan yang tinggi kemungkinan akan mampu menjawab semua butir soal yang berada pada tingkatnya. Bahkan ada kemungkinan dapat menjawab butir soal untuk materi yang berada pada tingkat di atasnya. Untuk siswa pada kelompok tersebut cocok jika diberi tes untuk tingkat lanjut. Tetapi, untuk siswa dengan kemampuan yang rendah maka siswa tersebut tidak dapat menjawab dengan benar hampir pada semua butir tes. Akibatnya, informasi tentang kemampuan siswa tersebut sangatlah minim.

Untuk mengetahui kemampuan berpikir tingkat tinggi pada tingkatan kelas tertentu dapat dilakukan dengan penyesuaian skor pada tingkat di bawahnya. Tujuannya adalah agar hasil tes yang berasal dari tingkat dan skala skor yang berbeda dapat diperbandingkan dalam skala yang sama atau setara. Proses penyesuaian skor hasil tes yang berbeda agar dapat diperbandingkan disebut penyetaraan atau equating.

Crocker dan Algina (2008: 473) mengatakan bahwa penyetaraan skor pada tes yang memiliki tingkat kesulitan dan tingkat kelas yang berbeda tetapi mengukur trait yang sama disebut penyetaraan vertikal. Kolen dan Brenan (2014: 13) mengemukakan bahwa secara umum, terdapat 3 (tiga) buah desain penyetaraan yang dapat digunakan untuk mengumpulkan data, yaitu *random group design*, *single group design*, dan *common-item non-equivalent group design*.

Untuk memperoleh data yang akurat, kesalahan dalam penyetaraan skor harus diminimalisir. Besarnya kesalahan dalam penyetaraan dapat diminimalkan dengan memilih cara dan desain penyetaraan yang sesuai. Desain penyetaraan ini digunakan untuk

merancang cara pengumpulan data untuk penyetaraan tes. Sementara metode digunakan untuk menentukan cara pengolahan data yang telah diperoleh.

Penyetaraan linier menurut Crocker dan Algina (2008: 458), adalah penyetaraan yang dilakukan berdasarkan asumsi bahwa selain dari rerata dan varians yang berbeda, distribusi skor X dan Y adalah sama. Hal ini sejalan dengan pendapat Kolen dan Brenann (2014: 35) bahwa suatu fungsi penyetaraan dapat dikatakan sebagai fungsi penyetaraan ekuipersentil jika distribusi skor pada tes X yang dikonversi ke skala bentuk tes Y adalah sama dengan distribusi skor pada populasi tes Y

Dengan demikian, penyetaraan ekuipersentil adalah penyetaraan antara dua tes yang berbeda berdasarkan kesamaan peringkat persentil. Sehingga kesamaan peringkat persentil yang berasal dari distribusi skor yang berbeda dapat dianggap bahwa hal tersebut menunjukkan kemampuan yang sama.

## **METODE PENELITIAN**

Metode yang digunakan pada penelitian ini adalah metode eksperimen. Eksperimen tidak dilakukan pada proses pembelajaran, tetapi dilakukan setelah memperoleh pola jawaban dan skor peserta tes melalui pelaksanaan tes kemampuan berpikir tingkat tinggi.

Penelitian ini menggunakan dua jenis variabel, yaitu variabel bebas dan variabel terikat. Variabel bebas yang digunakan adalah proporsi butir pengait. Sementara variabel terikatnya adalah keakurasian penyetaraan vertikal dengan metode ekuipersentil yang dinyatakan oleh besarnya nilai *Root Mean Square Error* (RMSE). Populasi yang digunakan pada penelitian ini adalah siswa SMK Tahun Ajaran 2016-2017 di wilayah DKI Jakarta. Dari populasi ini kemudian diambil sampel untuk 2 (dua) level kelas yang berbeda,

yaitu kelas X dan XI dengan teknik stratified random sampling adapun

disainnya sebagai berikut:

**Tabel 1. Disain penelitian**

Proporsi Butir Pengait (B)	Metode Penghalusan (A)	
	Kelas X (A <sub>1</sub> )	Kelas XI (A <sub>2</sub> )
10 (B <sub>1</sub> )	$\overline{RMSE}_{A_1B_1}$	$\overline{RMSE}_{A_2B_1}$
15 (B <sub>2</sub> )	$\overline{RMSE}_{A_1B_2}$	$\overline{RMSE}_{A_2B_2}$
20 (B <sub>3</sub> )	$\overline{RMSE}_{A_1B_3}$	$\overline{RMSE}_{A_2B_3}$

## HASIL DAN PEMBAHASAN

Proses penyetaraan vertikal pada Kelas X dan Kelas XI dengan menggunakan metode ekuipersentil yang menggunakan prepenghalusan log linier dapat dilakukan setelah diperoleh data acak sebanyak 30 kali untuk setiap kelas dan proporsi jumlah butir pengait.

Berdasarkan data yang didapat, data mentah langsung dihitung. Adapun deskripsi data mentah dibagi atas 2 kelompok berdasarkan kelas, yaitu kelas X dan XI. Berikutnya tiap-tiap kelas dibagi atas proporsi jumlah butir soal pengait yaitu sebesar 10%, 15% dan 20% yang dapat ditampilkan sebagai berikut:

**Tabel 2. Deskripsi Nilai RMSE**

	Kelas X			Kelas XI		
	10%	15%	20%	10%	15%	20%
Rata-rata	2,01	1,94	1,923	1,545	1,577	1,574
Std Dev	0,138	0,169	0,153	0,073	0,100	0,119
Minimum	1,744	1,662	1,632	1,353	1,289	1,295
Maksimum	2,339	2,213	2,274	1,663	1,795	1,838

Berdasarkan nilai pada Tabel 4.1 di atas, dapat diketahui secara deskriptif bahwa rata-rata RMSE terbesar ada pada kelas X dengan proporsi butir pengait sebanyak 10% sebesar 2,01 dan rata-rata RMSE terkecil ada pada kelas XI dengan proporsi jumlah butir pengait sebanyak 10% sebesar 1,545. Begitu pun untuk nilai RMSE terbesar ada pada kelas X dengan proporsi sebanyak 10% sebesar 2,339 dan nilai RMSE terkecil ada pada kelas XI

dengan proporsi butir sebanyak 10% sebesar 1,663.

Sebelum melakukan pengujian hipotesis, terlebih dahulu dilakukan pengujian persyaratan analisis yang terdiri dari uji normalitas dan uji homogenitas data. Hasil dari kedua pengujian ini akan menentukan jenis uji hipotesis akan dilakukan dengan uji statistik parametrik atau non parametrik.

Uji normalitas keempat kelompok data dilakukan dengan menggunakan Uji Liliefors.

Kesimpulan hasil Uji Normalitas yang dilakukan pada keempat data RMSE ditunjukkan oleh Tabel berikut:

**Tabel 3 Kesimpulan Hasil Uji Normalitas**

	Kelas X			Kelas XI		
	10%	15%	20%	10%	15%	20%
$L_{hitung}$	0,121	0,110	0,100	0,117	0,104	0,104
$L_{tabel}$	0,161	0,161	0,161	0,161	0,161	0,161
Kesimpulan	normal	normal	normal	normal	normal	normal

Selain Uji Normalitas, uji persyaratan analisis lain yang harus dilakukan adalah Uji Homogenitas dengan Uji F(Fisher). Hipotesis yang akan diuji adalah:

Dengan kriteria pengujian yang digunakan adalah:

Kesimpulan hasil Uji Homogenitas yang dilakukan pada keempat data RMSE ditunjukkan oleh Tabel berikut:

**Tabel 4 Kesimpulan Hasil Uji Homogenitas**

	10 Persen Butir pengait	
	X	XI
$F_{hitung}$	7,394	
$F_{tabel}$	1,861	
Kesimpulan	Tidak homogen	
	15 Persen Butir pengait	
	X	XI
$F_{hitung}$	15,023	
$F_{tabel}$	1,861	
Kesimpulan	Tidak homogen	
	20 Persen Butir pengait	
	X	XI
$F_{hitung}$	1,127	
$F_{tabel}$	1,861	
Kesimpulan	homogen	

**Pengujian Hipotesis**

Berdasarkan hasil uji persyaratan analisis, keempat kelompok data berdistribusi normal, seluruh kelompok data homogen. Sehingga, uji hipotesis dilakukan dengan 2 (dua) cara, yakni uji t untuk sampel bebas untuk data yang berdistribusi

normal dan homogen serta uji t' untuk sampel bebas yang berdistribusi normal namun tidak homogen. Tabel 4 berikut menunjukkan jenis uji hipotesis yang dilakukan berdasarkan kelompok sampel yang akan diuji.

Tabel 5. Teknik Pengujian Hipotesis

	10 Persen Butir pengait	
	X	XI
Uji Normalitas	Normal	
Uji Homogenitas	Tidak Homogen	
Uji Hipotesis	Uji t	
	15 Persen Butir pengait	
	X	XI
Uji Normalitas	Normal	
Uji Homogenitas	Tidak Homogen	
Uji Hipotesis	Uji t	
	20 Persen Butir pengait	
	X	XI
Uji Normalitas	Normal	
Uji Homogenitas	Homogen	
Uji Hipotesis	Uji t	

Setelah melalui uji persyaratan analisis, dilakukan analisis dan perhitungan

dengan menggunakan analisa uji T sebagai berikut:

Tabel 6. Hasil Analisa Uji t

No.	Kelompok	Hasil Uji T	Jumlah Replikasi	Keterangan
1	A1 – B1	15,690	30	Signifikan
2	A2 – B2	9,788	30	Signifikan
3	A3 – B3	10,701	30	Signifikan

### Pengujian Hipotesis 1

Hipotesis yang pertama menyatakan bahwa keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas X lebih kecil dari keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas XI untuk proporsi butir pengait 10%, dengan kata lain, rata-rata RMSE pada kelas X lebih besar dari rata-rata RMSE pada kelas XI untuk proporsi butir pengait 10%.

Berdasarkan hasil kesimpulan uji prasyarat analisis yang ditunjukkan oleh Tabel 4.5, maka pengujian hipotesis yang pertama dilakukan dengan menggunakan Uji t pada taraf signifikansi 5%.

Dari hasil pengujian hipotesis Uji t untuk sampel bebas diperoleh nilai  $t = 15,690$  dan sig. (2-tailed) atau  $p_{\text{value}} = 0,000$ :  $2 = 0,000$ . Dengan demikian, Sig.  $< 0,05$  sehingga  $H_0$  ditolak. Artinya, rata-rata RMSE antara kelas X lebih besar dari rata-rata RMSE pada kelas XI untuk jumlah butir pengait 10%. Berdasarkan hasil pengujian hipotesis tersebut maka dapat disimpulkan bahwa keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas X lebih rendah dibanding keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas XI untuk proporsi jumlah butir pengait 10%.

### Pengujian Hipotesis 2

Hipotesis yang kedua menyatakan bahwa keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas X lebih kecil dari keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas XI untuk proporsi butir pengait 15%, dengan kata lain, rata-rata RMSE pada kelas X lebih besar dari rata-rata RMSE pada kelas XI untuk proporsi butir pengait 15%.

Berdasarkan hasil kesimpulan uji prasyarat analisis yang ditunjukkan oleh Tabel 4, maka pengujian hipotesis yang pertama dilakukan dengan menggunakan Uji t pada taraf signifikansi 5%.

Dari hasil pengujian hipotesis Uji t untuk sampel bebas diperoleh nilai  $t = 9,778$  dan sig. (2-tailed) atau  $p_{\text{value}} = 0,000: 2 = 0,000$ . Dengan demikian, Sig.  $< 0,05$  sehingga  $H_0$  ditolak. Artinya, rata-rata RMSE antara kelas X lebih besar dari rata-rata RMSE pada kelas XI untuk jumlah butir pengait 15%. Berdasarkan hasil pengujian hipotesis tersebut maka dapat disimpulkan bahwa keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas X lebih rendah dibanding keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas XI untuk proporsi jumlah butir pengait 15%.

### Pengujian Hipotesis 3

Hipotesis yang ketiga menyatakan bahwa keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas X lebih kecil dari keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas XI untuk proporsi butir pengait 20%, dengan kata lain, rata-rata RMSE pada kelas X lebih besar dari rata-rata RMSE pada kelas XI untuk proporsi butir pengait 20%.

Berdasarkan hasil kesimpulan uji prasyarat analisis yang ditunjukkan oleh Tabel 4.5, maka pengujian hipotesis yang pertama dilakukan dengan menggunakan Uji t pada taraf signifikansi 5%.

Dari hasil pengujian hipotesis Uji t untuk sampel bebas diperoleh nilai  $t = 10,701$  dan sig. (2-tailed) atau  $p_{\text{value}} = 0,000: 2 = 0,000$ . Dengan demikian, Sig.  $< 0,05$  sehingga  $H_0$  ditolak. Artinya, rata-rata RMSE antara kelas X lebih besar dari rata-rata RMSE pada kelas XI untuk jumlah butir pengait 20%. Berdasarkan hasil pengujian hipotesis tersebut maka dapat disimpulkan bahwa keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas X lebih rendah dibanding keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kelas XI untuk proporsi jumlah butir pengait 20%.

Hasil penelitian tersebut dapat diartikan bahwa adanya butir gandeng (*Anchor Item*) di tiap kelas pada ujian dapat dijadikan sebuah acuan dalam menyetarakan kualitas tes yang akan diberikan kepada siswa.

Dalam pelaksanaan evaluasi yang menggunakan beberapa paket tes yang berbeda (bertingkat) dan mengukur hal yang sama, perlu dilakukan penyetaraan (*Equating*) (Herkusumo, 2008:1).

Hasil penelitian tersebut diperkuat oleh Dewi yang mengatakan Penyetaraan vertikal metode ekuipersentil dengan proporsi kelompok gandeng 20% lebih akurat dibandingkan dengan proporsi gandeng 10% dan 30%. (2017, p. 1)

### SIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah Terdapat perbedaan keakurasian penyetaraan vertikal metode ekuipersentil pada kemampuan berpikir tingkat tinggi (HOTS) antara siswa kelas X dengan kelas XI SMK Teknik Bangunan untuk proporsi jumlah butir pengait 10%.

Terdapat perbedaan keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kemampuan berpikir

tingkat tinggi (HOTS) antara siswa kelas X dengan kelas XI SMK Teknik Bangunan untuk proporsi jumlah butir pengait 15%.

Terdapat perbedaan keakurasian penyetaraan vertikal dengan metode ekuipersentil pada kemampuan berpikir tingkat tinggi (HOTS) antara siswa kelas X dengan kelas XI SMK Teknik Bangunan untuk proporsi jumlah butir pengait 20%.

#### DAFTAR RUJUKAN

- Basito, M. D., Arthur, R., & Daryati, D. (2018). Hubungan Efikasi Diri terhadap Kemampuan Berpikir Tingkat Tinggi Siswa SMK Program Keahlian Teknik Bangunan pada Mata Pelajaran Mekanika Teknik. *Jurnal Pensil*, 7(1), 1–14.
- Brookhart, S. M. (2007). *Assess HOTS In Your Classroom*. Virginia: ASCD Alexandria.
- Crocker, L., Algina, J., Staudt, M., Mercurio, S., Hintz, K., & Walker, R. A. (2008). *Introduction to Classical and Modern Test Theory*. Ohio: Cengage Learning.
- Dewi, E. P. (2017). Kestabilan Penyetaraan Vertikal dengan Metode Ekuipersentil berdasarkan Proporsi Kelompok Gandeng. *Jurnal Evaluasi Pendidikan*, 8(2), 63–66.
- Herkusumo, A. P. (2011). Penyetaraan (Equating) Ujian Akhir Sekolah Berstandar Nasional (UASBN) Dengan Teori Tes Klasik. *Jurnal Pendidikan Dan Kebudayaan*, 17(4), 455–471.
- Kolen, M. J., & Brennan, R. L. (2014). *Statistics for Social and Behavioral Sciences Test Equating, Scaling, and Linking*. New York: Springer International Publishing.
- Naga, D. S. (2012). *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Jakarta: Nagarani Citrayasa.
- S., J. L. R., Dolipas, B. B., & Villamor, B. B. (2013). Higher Order Thinking Skills and Academic Performance in Physics of College Students: A Regression Analysis. *International Journal of Innovative Interdisciplinary Research*, (4), 7–48.

