

AKURASI METODE CONCORDANCE BERDASARKAN PANJANG TES DAN UKURAN SAMPEL

Imbuh Yuwono
Pusdiklat BMKG Jakarta

Wardani Rahayu
FMIPA Universitas Negeri Jakarta,
Jakarta Timur

ABSTRACT

This research used two sets of test instruments, namely the Final Examinations of Mathematics of seventh grade, which were respectively developed by SMPN 223 Pasar Rebo East Jakarta and SMPN 196 Cipayung East Jakarta. These two sets of test instruments were developed based on different construct on its indicator. The method used was an experimental method that was aimed to examine the accuracy of concordance method between linear and equipercetile methods based on test length of 20 and 30 items with sample size of 100 and 150. Samples are taken by using multiple random sampling with as many as 20 replications. Research hypotheses were tested by using t-test analysis. The study concluded that the RMSD score of the equipercetile method was greater than the same score of the linear method. This indicated that linear method was more accurate compared to equipercetile method.

Keywords

Accuracy, concordance method, linear method, equipercetile method

ABSTRAK

Penelitian ini menggunakan dua perangkat tes, yaitu: perangkat tes Ulangan Akhir Semester mata pelajaran Matematika kelas VII yang dikembangkan oleh dua sekolah yang berbeda, yaitu: SMPN 223 Pasar Rebo Jakarta Timur dan SMPN 196 Cipayung Jakarta Timur. Kedua perangkat tes tersebut disusun berdasarkan kisi-kisi yang berbeda pada indikatornya. Metode yang digunakan adalah eksperimen yang bertujuan untuk meneliti akurasi metode *concordance* antara metode linier, dan metode ekipersentil berdasarkan pada panjang tes 20 dan 30 serta ukuran sampel 100 dan 150. Pengambilan sampel dilakukan melalui metode acak berulang dengan pengembalian sebanyak 20 replikasi. Hipotesis penelitian diuji dengan uji-t. Hasil penelitian disimpulkan bahwa nilai RMSD metode ekipersentil lebih besar dari metode linier. Hal ini menunjukkan bahwa metode linier lebih akurat dibandingkan metode ekipersentil.

Kata Kunci

Akurasi, metode *concordance*, metode linier, metode ekipersentil

Alamat Korespondensi

Jl. Pemuda 2, No. 36, Rawamangun,
Jakarta Timur, DKI Jakarta
Indonesia

e-mail:

imbuhyuwono@gmail.com

I. Pendahuluan

Kita sering menghadapi beberapa tes berbeda, yang sekornya perlu diperbandingkan dalam dunia pendidikan. Sekor tes digunakan sebagai salah satu informasi yang sangat penting untuk pengambilan keputusan kepada seseorang atau program, misalnya: keputusan untuk kenaikan kelas, kelulusan, penerimaan, dan syarat untuk sertifikasi profesi, atau digunakan sebagai salah satu standar penilaian dalam evaluasi program.

Pemberian tes yang berbeda menimbulkan tuntutan tentang kesetaraan tes. Karakteristik tes tersebut dipertanyakan, baik tingkat kesukaran atau daya bedanya, karena sekor tertentu di satu tes akan berbeda nilainya pada tes lain, jika keduanya langsung diperbandingkan maka akan

sangat merugikan peserta tes yang mendapat tes yang lebih sulit atau sebaliknya. Komponen pendidikan yang tidak merata di setiap sekolah menambah masalah dalam pemantauan mutu pendidikan antar sekolah atau daerah dengan menggunakan sekor hasil belajar.

Tes UAS pada tingkat SMP di wilayah Jakarta Timur dibangun oleh masing-masing sekolah, sehingga walaupun berasal dari kompetensi dasar yang sama namun indikator soal kadang berbeda. Sekor hasil UAS pada beberapa SMP di wilayah tersebut tidak dapat langsung diperbandingkan untuk mengetahui kemajuan hasil belajar di wilayah Jakarta Timur.

Penyetaraan dilakukan untuk mengatasi hal ini, sesuai dengan jenis perbedaan atau hubungan antar tesnya. Jika tes tersebut adalah paket tes

yang secara statistik dan konseptual dapat saling menggantikan (*interchangeable*), maka penyetaraan dapat diketahui dengan *equating*, jika sama distribusinya dan mengukur konstruk yang sama, namun tidak dapat saling menggantikan karena tes berasal dari kisi-kisi yang berbeda, maka hubungan antar tes disebut dengan *concordance*, dan jika kondisi untuk *equating* dan *concordance* tidak terpenuhi, digunakan prediksi skor harapan dengan regresi (Neil J. Dorans, 2004)

Desain pengumpulan data dalam penyetaraan *concordance* hanya cocok dengan rancangan A atau *single group* karena tidak ada butir bersama/jangkar (*common/anchor item*). Berdasarkan teori tes klasik, ada 3 metode yang dapat digunakan, yakni: metode linier, metode paralel linier, dan metode ekipersentil sedangkan dengan menggunakan teori respons butir, ada 4 metode yang dapat ditempuh, yakni: metode regresi, metode rerata dan sigma, metode tegar dan sigma, dan metode kurva karakteristik (Heri Retnawati dan Kana Hidayati, 2010; Hambleton, R. K., dan H. Swaminathan, 1885; Dali S. Naga, 2013).

Concordance dapat menghasilkan informasi yang akurat jika kemampuan siswa dapat terukur dengan baik, makin tinggi akurasi berarti makin kecil kesalahan dalam pengukuran maupun kesalahan metode *concordance*. Dikenal dua macam kesalahan atau kekeliruan dalam penyetaraan, yaitu: kekeliruan acak dan kekeliruan sistematis (Naga, 1992).

Kekeliruan acak atau kekeliruan sampel adalah perbedaan antara keadaan sebenarnya yang ada pada populasi yang disebabkan hasil ukuran pada sampel tersebut hanya merupakan salah satu dari sekian banyak kemungkinan hasil pencuplikan berulang dari suatu populasi, kesalahan sampel disebut kesalahan pengukuran standar (*standard error*). Adapun kesalahan sistematis biasanya terjadi pada kesalahan memilih atau menerapkan cara pengumpulan data (*equating design*). Makin kecil kesalahan sampel maka makin stabil pengukuran tersebut, sehingga metode *concordance* akan makin akurat (Tri Rijanto, 2012).

Selain pemilihan sampel, panjang tes juga berpengaruh pada validitas, reliabilitas, dan keakuratan metode *equating* dengan teori respons butir. Makin banyak butir tes maka

makin tinggi koefisien validitas dan reliabilitasnya. Penetapan panjang tes berdasarkan karakteristik butir dan kebutuhan waktu penyelesaian tes, di sekolah biasanya menggunakan 20 sampai 50 butir pada satu tes mata pelajaran tertentu. Maka jumlah atau banyaknya butir soal yang tepat dalam penyusunan tes akan mengurangi kesalahan pengukuran. Panjang tes berpengaruh pada validitas dan reliabilitas, makin banyak butir tes maka makin tinggi koefisien validitas dan reliabilitasnya, demikian pula dengan jumlah sampel. Panjang tes dan jumlah sampel juga berpengaruh pada reliabilitas dan keakuratan metode penyetaraan (*equating*) dengan teori respons butir (Ade Lyana, 2016; Dali S. Naga, 2013; Anne R. Fitzpatrick dan WM. Yen, 2001)

Metode yang sering digunakan untuk mengestimasi besarnya kesalahan pengukuran adalah *root mean square error* (RMSE) atau biasa disebut *root mean square difference* (RMSD). Rumusan RMSE atau RMSD mengukur perbedaan antara nilai yang diharapkan atau diprediksi dengan nilai hasil pengamatan. Rerata RMSD menunjukkan akurasi penyetaraan, di mana rata-rata RMSD yang kecil menunjukkan tingkat akurasi yang lebih tinggi dari pada rata-rata RMSD yang besar. Rerata RMSD didapat dengan menjumlahkan nilai RMSD dari sampel dibagi jumlah sampel yang disetarakan. Akurasi relatif penyetaraan adalah perbandingan rerata RMSD suatu metode penyetaraan dengan metode penyetaraan yang lainnya (Cohen, Allan S. dan Seock-Ho Kim, 2002).

Penelitian yang berkaitan dengan *concordance* telah dilakukan Heri dan Kana (2012) menggunakan tes Ujian Akhir Sekolah pada dua SMA di Yogyakarta dengan panjang tes 45 butir tanpa replikasi. Adapun Dwining (2012) menyetarakan tes Ulangan Akhir Semester Kimia pada dua SMA di kabupaten Bekasi dengan sampel 40, panjang tes 35 dan 31 kali replikasi, keduanya pada teori tes klasik. Sedangkan Wardani Rahayu (2015) *concordance* pada teori tes modern menggunakan data hasil tes *try out* SMP berdasarkan ukuran sampel 500 dan 800, panjang tes 40 butir dengan 21 replikasi.

Dalam penelitian ini, dua metode *concordance* digunakan untuk menyetarakan dua kelompok sekor tes hasil UAS Matematika SMP kelas VII, berdasarkan pemilihan ukuran sampel 100 dan

150 dari 420 sekor responden serta memilih panjang tes 20 dan 30 dari 40 butir tes, kemudian direplikasi sebanyak 20 kali.

2. Metode Penelitian

Penelitian ini merupakan penelitian eksperimen. Variabel bebas dalam penelitian ini adalah metode *concordance*, panjang tes, dan ukuran sampel, sedangkan variabel terikatnya adalah akurasi yang dihitung dengan RMSD.

Populasi terjangkau dalam penelitian ini adalah siswa kelas VII SMPN 223 Jakarta Timur dan siswa kelas VII SMPN 196 Jakarta Timur. Pengambilan sampel dilakukan dengan metode acak berulang dengan pengembalian (*random sampling with replacement*).

Instrumen dalam penelitian ini menggunakan dua instrumen tes Ulangan Akhir Semester (UAS) SMP mata pelajaran Matematika kelas VII semester I tahun ajaran 2016/2017 yang dikembangkan oleh dua sekolah yang berbeda berdasarkan kisi-kisi yang berbeda pada indikatornya. Setelah uji coba instrumen nantinya hanya diambil 34 butir soal valid dari 40 butir soal

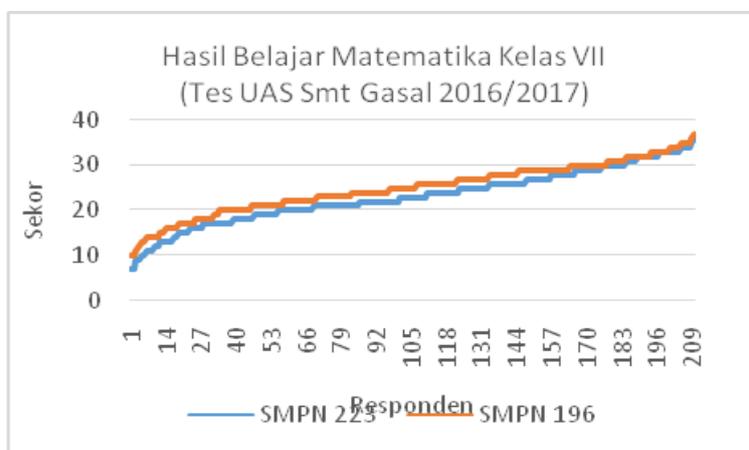
pada SMPN 223, dan 32 butir soal valid dari 40 butir soal pada SMPN 196. Reliabilitas instrumen tes UAS I memiliki koefisien 0,784, sedangkan reliabilitas instrumen tes UAS II memiliki koefisien 0,760, sehingga kedua instrumen tes dinyatakan reliabel untuk dipergunakan.

Seluruh kelompok sekor hasil UAS dengan butir valid diambil sebanyak 20 dan 30 butir saja sehingga terdapat dua kelompok sekor responden, kemudian dipilih 100 dan 150 sekor untuk *concordance* dengan metode linier dan ekipersentil. Delapan kelompok sekor tersebut dihitung RMSD untuk mengetahui akurasinya.

Uji normalitas untuk setiap kelompok data dan uji homogenitas seluruh data RMSD dilakukan dengan SPSS.

3. Hasil Penelitian dan Pembahasan

Hasil belajar Matematika kelas VII dari dua sekolah pada tes UAS tahun ajaran 2016/2017 menunjukkan bahwa tes UAS pada SMPN 223 lebih sulit dari pada SMPN 196, hal ini sesuai hasil uji taraf sukar pada uji coba sebelumnya.



Gambar I. Hasil Belajar Matematika Kelas VII

Gambar I. menunjukkan bahwa sekor perolehan hasil tes siswa menunjukkan bahwa rerata hasil belajar siswa pada mata pelajaran Matematika pada SMPN 223 lebih rendah dibandingkan dengan SMPN 196.

Hal ini memperkuat asumsi bahwa *concordance* harus dilakukan pada kedua kelompok sekor agar kita dapat mengetahui perbandingannya, dan

untuk mengetahui akurasi kedua metode dilakukan perhitungan RMSD. Data nilai RMSD secara deskriptif disajikan dalam Tabel I.

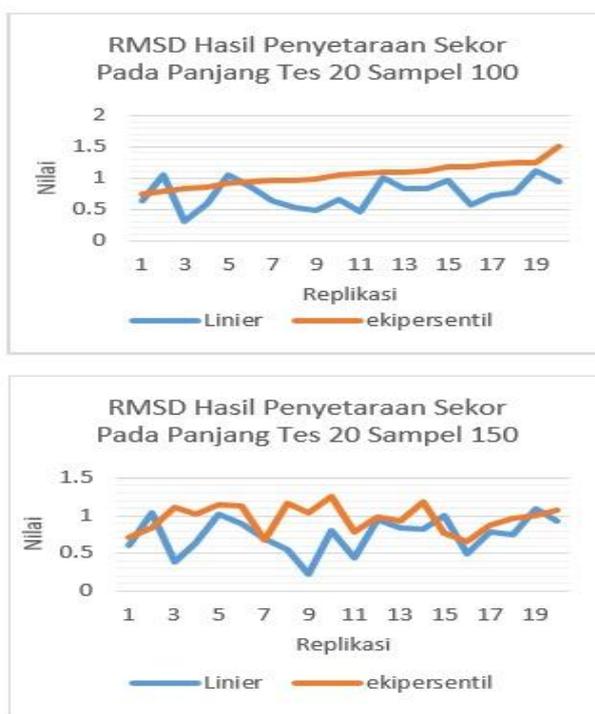
Pada Tabel I, secara deskriptif rerata RMSD paling tinggi setelah 20 kali replikasi ditunjukkan oleh kelompok metode linier panjang tes 30 sampel 100, dan paling rendah pada kelompok metode linier panjang tes 20 sampel 150.

Tabel 1. Nilai RMSD Hasil Concordance pada 20 Replikasi

Metode Concordance	Panjang tes	Ukuran Sampel	Max	Min
Linier	20	100	1,111	0,301
		150	1,097	0,227
	30	100	1,506	0,748
		150	1,250	0,649
Ekipersentil	20	100	1,142	0,415
		150	1,139	0,367
	30	100	1,325	0,718
		150	1,345	0,800

Gambar 2. berikut ini menunjukkan bahwa pada panjang tes 20 dan 100 sampel direplikasi kedua dan kelima nilai RMSD metode linier lebih tinggi dari pada metode ekipersentil. Nilai RMSD pada panjang tes 20 dan sampel 150 direplikasi

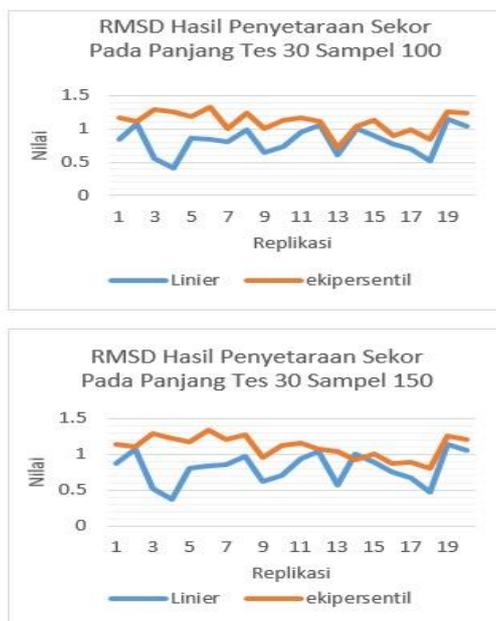
kedua, lima belas dan sembilan belas nilai RMSD metode linier lebih tinggi dari pada metode ekipersentil, sedangkan pada replikasi ketujuh dan dua belas bernilai sama.



Gambar 2. RMSD Hasil Metode Linier dan Ekipersentil Pada Panjang Tes 20

Gambar 3. menunjukkan bahwa pada panjang tes 30 dan 150 sampel direplikasi kelima belas nilai RMSD metode linier sedikit lebih tinggi daripada ekipersentil, sedangkan pada replikasi kedua bernilai sama.

Namun dapat dilihat bahwa hampir seluruh nilai RMSD pada metode linier lebih kecil dari nilai RMSD pada metode ekipersentil. Pengujian data secara statistik akan dilakukan setelah uji prasyarat terpenuhi.



Gambar 3. RMSD Hasil Metode Linier dan Ekipersentil Pada Panjang Tes 30

Hasil pengujian hipotesis dilakukan dengan uji-t dengan bantuan SPSS setelah uji prasyarat di bawah ini, di mana dihasilkan kesimpulan bahwa

data RMSD berasal dari populasi yang normal dan homogen.

Tabel 2. Kesimpulan Hasil Uji-t

Hipotesis statistik	T	Sig	Kesimpulan
$H_0: \mu_{MeqP20S100} \leq \mu_{MlinP20S100}$ $H_1: \mu_{MeqP20S100} > \mu_{MlinP20S100}$	4,600	0,000	Tolak H_0
$H_0: \mu_{MeqP20S150} \leq \mu_{MlinP20S150}$ $H_1: \mu_{MeqP20S150} > \mu_{MlinP20S150}$	3,250	0,002	Tolak H_0
$H_0: \mu_{MeqP30S100} \leq \mu_{MlinP30S100}$ $H_1: \mu_{MeqP30S100} > \mu_{MlinP30S100}$	4,835	0,000	Tolak H_0
$H_0: \mu_{MeqP30S150} \leq \mu_{MeqP30S150}$ $H_1: \mu_{MeqP30S150} > \mu_{MeqP30S150}$	4,939	0,000	Tolak H_0

Hasil uji-t membuktikan bahwa RMSD hasil dari *concordance* dengan metode linier pada delapan kelompok RMSD, yaitu: pada panjang tes 20 dan 30 serta ukuran sampel 100 dan 150 lebih kecil dari RMSD dengan metode ekipersentil.

4. Kesimpulan

Berdasarkan hasil pengujian hipotesis, maka dikemukakan kesimpulan, yaitu: *concordance*

5. Daftar Pustaka

Ade Lyana, (2016). *Keakurasian Penyetaraan Vertikal dengan Metode Ekipersentil berdasarkan Panjang Tes dan Ukuran Sampel* Jakarta: Tesis UNJ: 42-43.

dengan metode linier lebih akurat daripada metode ekipersentil, sesuai asumsi awal bahwa metode linier mempunyai sensitifitas lebih merata pada sekor bawah, tengah, dan atas. Hal ini disebabkan karena kedua kelompok sekor tes pada penelitian ini memiliki distribusi skor yang berbeda, distribusi tersebut terkait dengan rerata dan simpangan bakunya.

Anne R. Fitzpatrick dan WM. Yen (2001). *The Effects of Test Length and Sample Size on the Reliability and Equating of Tests Composed of Constructed-Response*

- Items. *Journal Applied Psychological Measurement*.
- Bintarawati, D. (2012). Akurasi Metode Concordance. *Jurnal Evaluasi Pendidikan*, 107-117.
- Cohen, Allan S. dan Seock-Ho Kim (2002). Comparison of Linking and Concurrent Calibration Under Item Response Theory. *Journal Applied Psychological Measurement*.
- Dali S. Naga. (2013). *Teori Sekor Pada Pengukuran Mental*. Jakarta: PT. Nagarani Citryasa
- Dorans, N. J. (Juli 2004). Equating, Concordance, and Expectation. *Journal Applied Psychological Measurement*, 28(4): 227-246.
- Hambleton, R. K., dan H. Swaminathan. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Academic Publisher.
- Heri Retnawati dan Kana Hidayati. (2010). Perbandingan Metode Concordance Berdasarkan Teori Tes Klasik. *Jurnal FMIPA UNY*:1-17.
- Kolen, M. (Juli 2004). Linking Assessment: Concept and History. *Journal Applied Psychological Measurement*, 28(4): 219-226.
- Mary Pommerich, BA Hanson, D Harris, dan JA Sconing. (Juli 2004). Issues on Conducting Between Distinct Test. *Journal Applied Psychological Measurement*, 28(4): 247-273.
- Nancy S. Peterson. (Januari 2008). A discussion of Population Invariants in Equating. *Journal Applied Psychological Measurement*, 32(1): 98-101.
- Neil J. Dorans. (July 2004). Equating, Concordance, and Expectation. *Journal Applied Psychological Measurement*, 28(4): 227-246.