

PERBANDINGAN METODE PENYETARAAN SKOR TES MENGGUNAKAN BUTIR BERSAMA DAN TANPA BUTIR BERSAMA

Heri Retnawati

Fakultas Matematika Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta

email: heri_retnawati@uny.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui kesetaraan skor 20 perangkat tes ujian akhir SMP dan membandingkan penyetaraan dengan butir bersama (*equating*) dan tanpa butir bersama (*concordance*). Metode penelitian yang digunakan adalah metode rerata dan rerata, metode rerata dan sigma, dan Haebara, Stocking dan Lord. Objek penelitian ini adalah perangkat tes ujian akhir SMP Mata Pelajaran Matematika Tahun 2014 dan 46.313 respons siswa. Estimasi parameter butir dilakukan dengan program QUEST dan penyetaraan dilakukan dengan program IRTEQ. Interpretasi hasil penyetaraan dilakukan dengan membandingkan kurva karakteristik tes dan mengestimasi kesalahan penyetaraan dengan *root mean square of error (RMSE)*. Hasil penelitian menunjukkan bahwa *pertama*, dua puluh paket yang digunakan pada ujian nasional menunjukkan kecenderungan yang setara. *Kedua*, pada *equating* dengan metode grafis, rerata dan sigma menghasilkan skor paling setara. *Ketiga*, metode Haebara dan metode Stocking dan Lord yang menghasilkan skor-skor dengan RMSE yang paling kecil. *Keempat*, *concordance* menghasilkan RMSE yang lebih kecil dibandingkan *equating*.

Kata kunci: *penyetaraan, concordance, equating, RMSE*

THE COMPARISON OF TEST SCORES LINKING METHOD USING EQUATING AND CONCORDANCE

Abstract

This study was aimed at determining the linking score of 20 tests of the national examination and comparing test score linking methods using equating and concordance. This study used mean and mean, mean and sigma, Haebara, and Stocking & Lord methods. The objects of this study were mathematics national examination tests of junior high schools in 2014 and 46,313 students' responses. The estimation of item parameters was done using the QUEST program while the equating used the IRTEQ program. The interpretation of the results was done by comparing the test characteristic curves and estimating the linking error of the Root Mean Square Error (RMSE). The results show that *first*, 20 sets of tests in the national exams show equal tendencies. *Second*, in equating with graphical methods, the means and sigmas produce the most equal scores. *Third*, Haebara and Stocking & Lord methods generate the smallest RMSE scores. *Fourth*, the concordance produces RMSE smaller than equating.

Keywords: *linking, concordance, equating, RMSE*

PENDAHULUAN

Dalam suatu pelaksanaan ujian skala luas, biasanya digunakan instrumen tes dalam jumlah paket yang banyak. Misalnya saja ujian nasional (UN) yang dilaksanakan di Indonesia. Pada pelaksanaan UN, penyelenggaraan ujian menggunakan lebih dari 1 paket. Pada Tahun 2013, pelaksanaan UN di Indonesia menggunakan 4 paket soal, pada Tahun 2014 menggunakan 20 paket soal, dan pada Tahun 2015 kembali menggunakan 4 paket soal. Paket-paket soal tersebut mengukur hal yang sama, karena paket-paket itu dikembangkan berdasarkan kisi-kisi yang sama.

Ujian menggunakan beberapa paket yang paralel memiliki keunggulan dan kelemahan. Keunggulan menggunakan beberapa paket yakni menjaga kerahasiaan soal dan mengurangi kecurangan selama pelaksanaan tes. Sementara itu, tantangan ketika menggunakan beberapa paket adalah adanya jaminan perangkat-perangkat itu merupakan perangkat paralel dan mengukur indikator yang sama. Hal tersebut bermakna bahwa perangkat-perangkat yang digunakan merupakan perangkat yang setara. Kesetaraan ini dapat dibuktikan baik secara teoretis dan dapat dibuktikan pula secara empiris. Bukti-bukti tersebut terkait dengan konsep menghubungkan tes-tes (*linking*), penyetaraan skor tes (*equating* maupun *concordance*) (Dorans, 2004, pp. 219-223).

Istilah baku dalam menghubungkan dua tes, dalam psikometri dikenal istilah *linking*. Ada tiga derajat menghubungkan suatu tes dengan tes lain. Jika kedua tes secara statistik dan konseptual dapat saling menggantikan, disebut dengan penyetaraan (*equating*) (Brenan & Kolen, 2004, pp.2-4). Jika kedua tes mengukur konstruk yang sama disebut dengan *concordance*, dan jika kondisi penyetaraan tidak terpenuhi disebut dengan prediksi skor harapan (Dorans,

2004, pp. 228-230; Retnawati, 2014, pp. 90-95). Adapun hal-hal yang perlu dipertimbangkan dalam menghubungkan skor-skor tes menurut Freur meliputi kesamaan isi, tingkat kesulitan, dan format butir, dapat dibandingkan kesalahan pengukuran-nya, kondisi administrasi tes yang sama, pemanfaatan tes dan konsekuensinya, termasuk akurasi dan stabilitas dari penyetaraan (Brenan & Kolen, 2004, p. 2).

Terdapat beberapa desain yang dapat dipilih untuk menghubungkan tes-tes. Desain tersebut adalah desain grup tunggal, desain grup ekuivalen, dan desain butir bersama (Hambleton & Swaminathan, 1985, p. 198), serta desain *counter balanced* (Dorans, Moses, Eignor, 2010, p.11). Pada desain grup tunggal, tes-tes yang akan dibuktikan kesetaraannya dikerjakan oleh kelompok yang sama. Pada desain grup ekuivalen, tes-tes yang akan disetarakan diberikan pada kelompok-kelompok yang identik dan hampir sama kemampuannya. Selanjutnya, dengan menggunakan respons peserta tes terhadap tes dan skor-skornya dapat dilakukan suatu proses penyetaraan.

Penyetaraan dapat dilakukan dengan pendekatan klasik dan pendekatan modern. Pendapat ini senada dengan pernyataan dari Ryan & Brockmann (2009, pp. 1-4) yang memilah metode penyetaraan menjadi dua pendekatan, pendekatan teori tes klasik dan pendekatan teori respons butir. Pada pendekatan klasik, dikenal metode linear dan metode equipersentil (Hambleton, Swaminathan & Rogers, 1991, p.123) dan juga metode linear sejajar (Kolen & Brennan, 2004, p.31), *presmoothing*, *postsmoothing* (Moses & Liu, 2011, pp. 3-8). Pada pendekatan modern, terdapat metode rerata dan rerata, rerata dan sigma, metode rerata dan sigma tegar (Hambleton, Swaminathan & Rogers, 1991) dan metode kurva karakteristik butir yang meliputi

metode Stocking & Lord (Kolen & Brennan, 2004, pp.168-175).

Penyetaraan dengan pendekatan modern pada dasarnya menghitung tingkat kesulitan dan kemampuan peserta ke suatu skor dengan suatu persamaan linear. Pendekatan modern untuk mengestimasi parameter butir ini dilakukan untuk mengetahui karakteristik butir (Mardapi, 1998). Estimasi koefisien dan konstantanya saja yang berbeda. Jika parameter daya pembeda butir (a), tingkat kesulitan (b), dan c parameter tebakan semu, pada metode rerata dan rerata, menurut Hambleton, Swaminathan, Rogers (1991, pp. 123-128) disajikan dengan $b_2 = \alpha b_1 + \beta$, $a_2 = \frac{a_1}{\alpha}$.

Selanjutnya diperoleh:

$$\bar{b}_2 = \alpha \bar{b}_1 + \beta \quad (1)$$

$$\beta = \bar{b}_2 - \alpha \bar{b}_1 \quad (2), (2)$$

Keterangan:

$$\bar{a}_2 = \frac{\bar{a}_1}{\alpha} \text{ atau } \alpha = \frac{\bar{a}_1}{\bar{a}_2}$$

- \bar{b}_1 : rerata indeks kesukaran butir bersama tes 1
- \bar{b}_2 : rerata indeks kesukaran butir bersama tes 2
- \bar{a}_1 : rerata indeks daya beda butir bersama tes 1
- \bar{a}_2 : rerata indeks daya beda butir bersama tes 2
- α, β : konstanta penyetaraan

Persamaan (1) dan (2) digunakan untuk menghitung konstanta penyetaraan tes dengan berdasarkan metode rerata dan rerata. Konstanta penyetaraan α dan β pada metode rerata dan sigma menggunakan rerata dan simpangan baku dari parameter indeks kesulitan yang dapat dijelaskan sebagai berikut (Hambleton,

Swaminathan, & Rogers, 1991, pp. 123-128). Misal skor tes 1 disetarakan ke skor tes 2, hubungan parameter indeks kesulitan butir berhubungan linear sebagai berikut. $b_2 = \alpha b_1 + \beta$, sehingga diperoleh $\bar{b}_2 = \alpha \bar{b}_1 + \beta$, dan $S_2 = \alpha S_1$. Dengan persamaan tersebut, dapat diperoleh

$$\alpha = \frac{S_2}{S_1} \quad (3)$$

$$\beta = \bar{b}_2 - \alpha \bar{b}_1 \quad (4)$$

Keterangan:

- \bar{b}_1 : rerata indeks kesukaran butir tes 1
- \bar{b}_2 : rerata indeks kesukaran butir tes 2
- S_1 : simpangan baku indeks kesukaran butir tes 1
- S_2 : simpangan baku indeks kesukaran butir tes 2
- α, β : konstanta penyetaraan

Menentukan konstanta penyetaraan pada metode rerata dan sigma maupun rerata dan rerata tidak melibatkan semua parameter butirnya secara simultan. Alternatifnya, dapat digunakan metode penyetaraan yang melibatkan semua parameter butirnya secara simultan, yaitu metode kurva karakteristik dari Haebara dan Stocking & Lord (Kolen & Brennan, 1995). Pada metode Haebara, penyetaraan parameter butirnya didasarkan pada fungsi karakteristik butir. Prosedur komputasinya menggunakan variasi yang pertama, yang dapat dijelaskan sebagai berikut (Kolen, & Brennan, 2004, pp.168-175; Retnawati, 2014, p.107). 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:

$$H(\theta_i) = \sum_{j=1}^n (T_j - T_j^*)^2 \quad (5)$$

dengan $T_j = P_j(\theta_i)$, $T_j^* = P_j^*(\theta_i)$, dengan n banyaknya butir *anchor* $P_j(\theta_i)$ probabilitas menjawab benar butir j oleh peserta berkemampuan θ_i , $P_j^*(\theta_i)$: Probabilitas hasil transformasinya serta transformasi pada butir *anchor*, $b_j^* = \alpha b_j + \beta$, $a_j^* = \frac{a_j}{\alpha}$, dan $c_j^* = c_j$.

Didefinisikan fungsi yang persamaannya sebagai berikut.

$$F = \frac{1}{N} \sum_{i=1}^N H(\theta_i) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^n (T_j - T_j^*)^2 \quad (6)$$

dengan N sebarang bilangan asli menyatakan banyaknya titik pada skala θ . Fungsi F pada persamaan (8) merupakan fungsi dalam α dan β . Selanjutnya konstanta penyetaraan α dan β dipilih sedemikian rupa sehingga fungsi F minimum. Fungsi F mencapai nilai minimum bila

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = 0 \quad (7)$$

Persamaan (7) nonlinear dan mempunyai solusi numerik sehingga persamaan tersebut dapat diselesaikan dengan menggunakan prosedur numerik. Salah satu metode yang dapat digunakan untuk menyelesaikan persamaan tersebut adalah metode numerik Newton Raphson. Pada metode Stocking dan Lord, formula komputasinya menggunakan variasi yang kedua, prosedur komputasinya disajikan sebagai berikut (Kolen, & Brennan, 2004, p. 170, Retnawati, 2014, p.109). Kuadrat dari selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteris tes dari dua skala yang sudah disetarakan dinyatakan dengan $SL(\theta_i)$ yaitu:

$$SL(\theta_i) = (T_i - T_i^*)^2 \quad (8)$$

Keterangan:

$$T_i = \sum_{j=1}^n P_j(\theta_i)$$

$$T_i^* = \sum_{i=1}^n P_j^*(\theta_i)$$

n : panjang tes-*anchor*

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

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

T_i : skor murni peserta berkemampuan θ_i pada tes dasar

T_i^* : skor murni hasil transformasi

Dengan transformasi pada tes dengan butir bersama, $b_j^* = \alpha b_j + \beta$, $a_j^* = \frac{a_j}{\alpha}$, dan $c_j^* = c_j$.

Selanjutnya definisikan fungsi:

$$F = \frac{1}{N} \sum_{i=1}^N (T_i - T_i^*)^2 \quad (9)$$

dengan N adalah sebarang bilangan asli menyatakan banyaknya titik pada skala θ . Selanjutnya, konstanta penyetaraan α dan β dipilih sehingga fungsi F minimum. Fungsi F pada persamaan (9) mencapai minimum bila

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = 0 \quad (10)$$

Persamaan (10) *nonlinear* dan mempunyai solusi numerik sehingga persamaan tersebut hanya dapat diselesaikan dengan menggunakan prosedur numerik yang dapat diselesaikan di antaranya menggunakan metode Newton Raphson.

Pada kasus perbandingan metode *linking* skor tes, untuk mengetahui metode terbaik dapat digunakan berbagai cara. Cara pertama yang mudah dilakukan adalah

dengan memanfaatkan kurva karakteristik tes. Kurva ini dapat digambarkan dengan grafis dengan memanfaatkan jumlahan dari nilai fungsi informasi butir untuk keseluruhan butir penyusun tes tersebut (Hambleton & Swaminathan, 1991; Retnawati, 2014, p.19). Salah satu di antaranya adalah dengan menggunakan *root mean square of error* (RMSE) atau disebut juga RMSD. Cara ini telah digunakan oleh Kim & Cohen (2002, pp. 25-61), Antara & Bastari (2015, pp.13-24), dan Uysal & Kilmen (2016, pp.1-11). Formula RMSE sebagai berikut.

$$RMSE(\theta_r) = \sqrt{\frac{\sum_{i=1}^n (\theta_r - \theta_E)^2}{n}} \quad (11)$$

Beberapa penelitian telah dilakukan terkait dengan penyetaraan skor. Aşiret & Sünbul (2016) menggunakan metode identitas, rerata, linear, *circle-arc*, dan metode *equipercentile pre-smoothed* dengan 2 dan 3 momen pada berbagai ukuran sampel. Hasil equating dievaluasi menggunakan RMSE. Hasil studi menunjukkan bahwa metode *circle-arc* menghasilkan kesalahan equating yang lebih kecil dibandingkan dengan metode yang lain. Hasil studi pada 7 lintas asesmen menunjukkan bahwa penyetaraan dengan butir bersama yang fiks, metode penyetaraan kurva karakteristik jenis Stocking and Lord, dan metode rerata dan sigma sama bagusnya dan lebih bagus dibandingkan kalibrasi bersama (*concurrent calibration*) (Pang, Madera, Radwan, & Zhang, 2010). Sementara, itu Yu & Popp (2005, pp.1-19) memberikan hasil yang berbeda pada penelitiannya dengan menyatakan bahwa tidak ada metode tunggal yang paling baik untuk penyetaraan skor tes dan konteks yang berbeda memberikan hasil yang berbeda.

Beberapa *software* yang dapat digunakan untuk melakukan *equating*.

Dengan menggunakan parameter butir dan kemampuan berbagai program, beberapa peneliti mengembangkan *software*. Han (2009, pp. 491-493) mengembangkan program untuk melakukan penyetaraan dengan teori respons butir. Untuk metode Kernel dalam *equating*, (Anderson, Braunberg, & Wiberg, 2013, pp. 1-25) mengembangkan program berbasis program-R. Pemanfaatan program ini membantu peneliti mempermudah mengestimasi persamaan equating dan estimasi parameter hasil *equating*, sesuai kebutuhan peneliti.

Terkait dengan perangkat tes yang digunakan di Indonesia ada 20 paket soal dan beberapa metode penyetaraan perlu dibandingkan untuk melihat metode yang menghasilkan kesalahan yang paling kecil serta diperlukan suatu penelitian terkait dengan hal tersebut. Penelitian ini bertujuan untuk mendeskripsikan kesetaraan perangkat ujian akhir dengan metode grafik kurva karakteristik tes, mengetahui perbandingan kesetaraan perangkat dengan metode *concordance*, mengetahui kesetaraan perangkat menggunakan metode *equating*, dan membandingkan metode penyetaraan perangkat dengan metode *concordance* dan *equating*.

METODE

Studi ini bersifat deskriptif eksploratif, untuk mendeskripsikan kesetaraan perangkat ujian akhir dengan metode grafik kurva karakteristik tes, mengetahui perbandingan kesetaraan perangkat dengan metode *concordance*, mengetahui kesetaraan perangkat menggunakan metode *equating*, dan membandingkan metode penyetaraan perangkat dengan metode *concordance* dan *equating*.

Objek studi ini adalah perangkat ujian akhir Mata Pelajaran Matematika SMP Tahun 2014 sebanyak 20 paket berikut data

keseluruhan respons peserta terhadap tes tersebut, khusus di wilayah DI Yogyakarta. Peserta tes ini berjumlah 46.313 siswa dari seluruh siswa SMP di 4 kabupaten dan 1 kota di Provinsi Daerah Istimewa Yogyakarta yang meliputi Kota Yogyakarta, Kabupaten Bantul, Kabupaten Sleman, Kabupaten Gunungkidul, dan Kabupaten Kulon Progo. Data dikumpulkan dengan metode dokumentasi, yang diperoleh dari kantor Kementerian Pendidikan dan Kebudayaan Republik Indonesia pada bulan September 2015. Dengan menggunakan perangkat tes yang ada, butir-butir

kemudian dipetakan berdasarkan indikator ketercapaian kompetensi lulusan. Pemetaan ini dilakukan untuk memeroleh informasi butir bersama (*anchor/common items*) dengan Paket 1. Hasil pemetaan butir bersama berikut siswa yang mengerjakan tiap paket disajikan pada Tabel 1.

Setelah dipetakan, dilakukan analisis data. Langkah pertama yang dilakukan adalah memilah-milah data tiap paket, kemudian disimpan dalam file yang berbeda-beda. Dengan menggunakan *software* QUEST, tingkat kesulitan butir soal dan parameter kemampuan peserta diestimasi

Tabel 1

Butir Bersama Paket 2-20 dengan Paket 1 dan Banyaknya Peserta Tes tiap Paket

Paket	Nomor Butir bersama (dengan paket 1)	Banyaknya Peserta Tes
1	-	2310
2	1, 6, 8, 14, 15, 21, 22, 29, 36, 37, 40 (11 butir)	2359
3	1, 6, 8, 10, 14, 15, 21, 24, 29, 37, 38	2388
4	6, 8, 13, 14, 20, 21, 27, 34, 35, 37	2402
5	4, 6, 8, 9, 14, 18, 21, 23, 32, 37	2382
6	6, 8, 14, 21, 37	2362
7	6, 8, 13, 14, 21, 22, 27, 31, 36, 37	2165
8	4, 5, 6, 8, 12, 14, 21, 37	2583
9	4, 5, 6, 8, 12, 14, 21, 33, 37, 40	2242
10	3, 6, 8, 12, 14, 17, 21, 26, 37, 40	2305
11	4, 6, 8, 11, 14, 18, 21, 26, 37, 40	2341
12	6, 8, 14, 21, 37, 40	2347
13	4, 6, 8, 14, 21, 37, 40	2334
14	2, 6, 8, 9, 14, 16, 21, 23, 30, 37	2324
15	6, 8, 11, 14, 20, 21, 24, 25, 34, 37, 39	2310
16	3, 6, 8, 10, 14, 17, 21, 24, 31, 37, 38	2384
17	4, 5, 6, 8, 14, 21, 33, 37, 40	2297
18	2, 6, 7, 8, 14, 16, 21, 30, 35, 37	2246
19	6, 8, 14, 21, 37	2179
20	3, 6, 7, 8, 14, 21, 28, 35, 37	2053
Total Peserta		46313

pada setiap paket soal ujian. Parameter yang dihasilkan digunakan untuk menggambar kurva karakteristik tes semua perangkat dalam satu bidang gambar untuk mengetahui kesetaraan sebelum dilakukan penyetaraan.

Langkah selanjutnya adalah mengestimasi konstanta penyetaraan dengan butir bersama (*equating*) dengan metode rerata dan rerata, rerata dan sigma, Haebara, dan Lord & Stocking dengan menggunakan *software* IRTEQ, kemudian *output*-nya diinterpretasikan untuk mengestimasi konstanta penyetaraan tanpa butir bersama (*concordance*) dengan metode rerata dan rerata, rerata dan sigma, Haebara, dan Lord & Stocking. Kurva karakteristik tes semua perangkat kemudian digambarkan dalam satu bidang gambar untuk mengetahui kesetaraan setelah dilakukan penyetaraan untuk tiap metode penyetaraan, baik dengan butir bersama maupun tanpa butir bersama. Kemampuan siswa dengan menggunakan konstanta penyetaraan dengan butir bersama dan butir bersama selanjutnya dihitung.

Pada tiap kasus, kemudian diestimasi *root mean square of error (RMSE)* antara skala kemampuan hasil estimasi terhadap skala kemampuan setelah dilakukan penyetaraan. Grafik RMSE tiap kasus, selanjutnya digambarkan pada satu bidang gambar dan kemudian melakukan perbandingan RMSE antara penyetaraan dengan butir bersama (*equating*) dan penyetaraan tanpa butir bersama (*concordance*) dengan menggunakan grafik dan melakukan interpretasi kesamaannya.

Hasil RMSE digunakan untuk membandingkan metode-metode penyetaraan. Metode yang terbaik ditentukan berdasarkan RMSE. Semakin kecil RMSE, semakin akurat metode penyetaraan yang digunakan.

HASIL PENELITIAN DAN PEMBAHASAN

Berdasarkan informasi dari Balitbang Kemdikbud Indonesia sebagai pengembang

tes, dasar utama penyusunan soal adalah standar kompetensi lulusan, dikembangkan dengan indikator tertentu, dan tingkat kesulitan distribusinya 25% mudah, 50% sedang, dan 25% sulit. Menggarisbawahi pernyataan ini dikaitkan dengan parameter butir, hanya parameter tingkat kesulitan butir yang dijadikan asumsi penyusunan soal, bukan parameter yang lain seperti daya pembeda atau tebakan semu. Terkait dengan hal tersebut, model estimasi parameter yang digunakan adalah parameter tingkat kesulitan saja (satu parameter). Dengan menggunakan respons peserta, estimasi parameter butir dan kemampuan peserta dilakukan dengan menggunakan QUEST. Hasil statistik deskriptif estimasi parameter butir dari 20 paket perangkat UN Mata Pelajaran Matematika disajikan pada Tabel 2.

Mencermati hasil pada Tabel 2, diperoleh bahwa rerata tingkat kesulitan sama, demikian pula halnya standar deviasinya berkisar 1,00 atau distribusi tingkat kesulitan tiap paket mendekati distribusi normal baku. Dengan mengecek kembali butir yang memiliki tingkat kesulitan tertinggi tiap paket, butir nomor 3 terkait dengan operasi perpangkatan bilangan merupakan butir yang memiliki indeks tingkat kesulitan yang paling tinggi.

Dengan memanfaatkan parameter butir tersebut, kemudian dilakukan analisis baik *equating* maupun *concordance*. Hasil analisis ini dilakukan dengan bantuan program IRTEQ dengan masukan berupa parameter butir hasil estimasi. Metode yang digunakan meliputi empat metode, yaitu rerata dan rerata, rerata dan sigma, TCC Haebara, dan TCC Stocking & Lord. Hasil analisis untuk *Equating* disajikan pada Tabel 3, dan untuk *Concordance* disajikan pada Tabel 4.

Tabel 2

Statistik Deskriptif Estimasi Parameter Butir dari 20 Paket Perangkat UN Mata Pelajaran Matematika

Paket	Rerata	Sd	Min	Maks	Paket	Rerata	Sd	Min	Maks
1	0	1,12	-1,68	4,11	11	0	1,17	-3,08	4,23
2	0	1,18	-2,81	4,07	12	0	1,03	-1,96	4,02
3	0	1,01	-2,04	3,92	13	0	1,15	-3,08	3,94
4	0	1,07	-1,62	4,06	14	0	1,16	-2,99	4,04
5	0	1,11	-2,17	3,78	15	0	1,32	-2,16	5,19
6	0	1,05	-1,93	4,01	16	0	1,2	-2,7	4,03
7	0	1,15	-2,83	3,83	17	0	0,96	-1,59	3,72
8	0	1,18	-2,79	4,02	18	0	1,11	-3,03	3,86
9	0	1	-1,9	3,66	19	0	1,01	-2,09	3,74
10	0	1,09	-2,18	3,9	20	0	1,15	-1,31	4,17

Tabel 3

Hasil Equating dengan Menggunakan 4 Metode

	Rerata dan Rerata	Rerata dan Sigma	TCC Haebara	TCC Stocking & Lord
2 ke 1	$b^* = b_2 + 0,14$	$b^* = 1,34b_2 + 0,09$	$b^* = 1,064b_2 + 0,07$	$b^* = 1,08b_2 + 0,04$
3 ke 1	$b^* = b_2 + 0,085$	$b^* = 1,05b_2 + 0,04$	$b^* = 1,014b_2 + 0,03$	$b^* = 1,014b_2 + 0,03$
4 ke 1	$b^* = b_2 + 0,04$	$b^* = 0,96b_2 + 0,05$	$b^* = 0,96b_2 + 0,04$	$b^* = 0,98b_2 + 0,03$
5 ke 1	$b^* = b_2$	$b^* = 1,09b_2$	$b^* = 0,97b_2 + 0,01$	$b^* = 1,02b_2 + 0,01$
6 ke 1	$b^* = b_2 - 0,13$	$b^* = 1,26b_2 - 0,17$	$b^* = 1,10b_2 - 0,11$	$b^* = 1,07b_2 - 0,09$
7 ke 1	$b^* = b_2 + 0,44$	$b^* = 1,53b_2 + 0,31$	$b^* = 1,14b_2 + 0,34$	$b^* = 1,15b_2 + 0,36$
8 ke 1	$b^* = b_2 + 0,04$	$b^* = 0,99b_2 + 0,05$	$b^* = 0,98b_2 + 0,04$	$b^* = 0,99b_2 + 0,04$
9 ke 1	$b^* = b_2 - 0,07$	$b^* = 0,99b_2 - 0,11$	$b^* = 1,02b_2 - 0,08$	$b^* = 1,05b_2 - 0,09$
10 ke 1	$b^* = b_2 + 0,30$	$b^* = 0,93b_2 + 0,28$	$b^* = 0,88b_2 + 0,10$	$b^* = 0,97b_2 + 0,22$
11 ke 1	$b^* = b_2 + 0,03$	$b^* = 1,06b_2 + 0,01$	$b^* = 1,01b_2 + 0,02$	$b^* = 1,01b_2 + 0,03$
12 ke 1	$b^* = b_2 - 0,03$	$b^* = 1,18b_2 - 0,09$	$b^* = 1,07b_2 - 0,06$	$b^* = 1,04b_2 - 0,05$
13 ke 1	$b^* = b_2 + 0,16$	$b^* = 1,00b_2 + 0,16$	$b^* = 0,91b_2 + 0,00$	$b^* = 1,00b_2 + 0,10$
14 ke 1	$b^* = b_2 - 0,06$	$b^* = 0,87b_2 - 0,02$	$b^* = 0,85b_2 - 0,03$	$b^* = 0,94b_2 - 0,06$
15 ke 1	$b^* = b_2 + 0,50$	$b^* = 0,99b_2 + 0,50$	$b^* = 0,76b_2 + 0,41$	$b^* = 0,97b_2 + 0,45$
16 ke 1	$b^* = b_2 + 1,00$	$b^* = 1,39b_2 + 0,13$	$b^* = 0,88b_2 + 0,84$	$b^* = 1,06b_2 + 0,88$
17 ke 1	$b^* = b_2 + 0,02$	$b^* = 1,22b_2 + 0,01$	$b^* = 1,05b_2 + 0,02$	$b^* = 1,05b_2 + 0,01$
18 ke 1	$b^* = b_2 + 0,29$	$b^* = 0,60b_2 + 0,27$	$b^* = 0,73b_2 + 0,18$	$b^* = 0,87b_2 + 0,17$
19 ke 1	$b^* = b_2 + 0,13$	$b^* = 1,09b_2 + 0,09$	$b^* = 1,03b_2 + 0,11$	$b^* = 1,02b_2 + 0,11$
20 ke 1	$b^* = b_2 + 1,07$	$b^* = 2,18b_2 + 1,58$	$b^* = b_2 + 0,95$	$b^* = 1,22b_2 + 1,05$

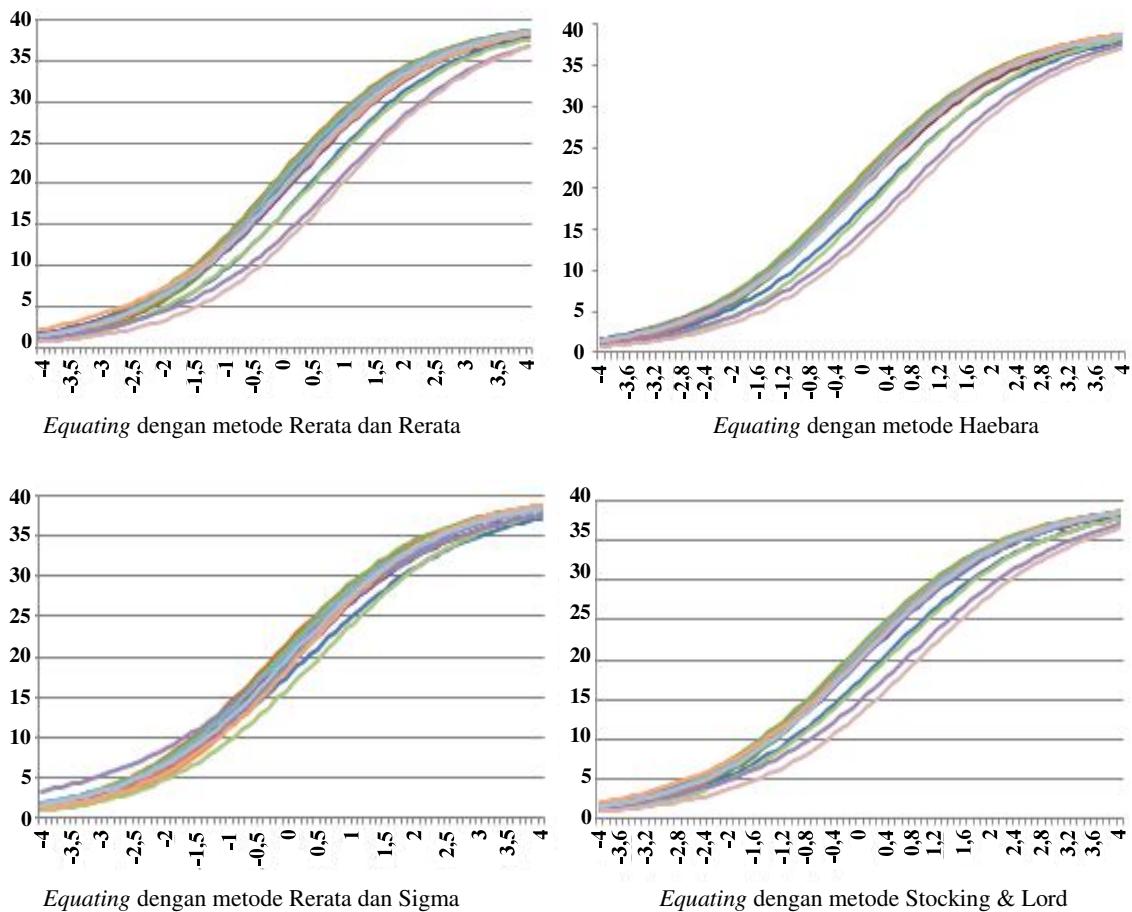
Tabel 4
Hasil Concordance dengan Menggunakan 4 Metode

	Rerata dan Rerata	Rerata dan Sigma	TCC Haebara	TCC Stocking & Lord
2 ke 1	$b^* = b_2 + 0,10$	$b^* = 1,07b_2 + 0,10$	$b^* = 0,964b_2 + 0,07$	$b^* = 1,01b_2 + 0,07$
3 ke 1	$b^* = b_2 + 0,06$	$b^* = 1,29b_2 + 0,07$	$b^* = 1,05b_2 + 0,03$	$b^* = 1,074b_2 + 0,04$
4 ke 1	$b^* = b_2 + 0,18$	$b^* = 1,05b_2 + 0,19$	$b^* = 0,95b_2 + 0,03$	$b^* = 1,02b_2 + 0,15$
5 ke 1	$b^* = b_2 + 0,06$	$b^* = 1,18b_2 + 0,07$	$b^* = 0,96b_2 + 0,02$	$b^* = 1,03b_2 + 0,04$
6 ke 1	$b^* = b_2 + 0,09$	$b^* = 1,22b_2 + 0,09$	$b^* = 1,06b_2 + 0,08$	$b^* = 1,05b_2 + 0,08$
7 ke 1	$b^* = b_2 + 0,18$	$b^* = 0,97b_2 + 0,18$	$b^* = 0,93b_2 + 0,08$	$b^* = 0,98b_2 + 0,13$
8 ke 1	$b^* = b_2 + 0,18$	$b^* = 0,97b_2 + 0,18$	$b^* = 0,93b_2 + 0,08$	$b^* = 0,98b_2 + 0,13$
9 ke 1	$b^* = b_2 + 0,20$	$b^* = 1,08b_2 + 0,21$	$b^* = 1,00b_2 + 0,13$	$b^* = 1,02b_2 + 0,15$
10 ke 1	$b^* = b_2 + 0,30$	$b^* = 0,93b_2 + 0,28$	$b^* = 0,88b_2 + 0,10$	$b^* = 0,97b_2 + 0,22$
11 ke 1	$b^* = b_2 + 0,06$	$b^* = 1,12b_2 + 0,06$	$b^* = 1,01b_2 + 0,02$	$b^* = 1,03b_2 + 0,03$
12 ke 1	$b^* = b_2 + 0,08$	$b^* = 1,26b_2 + 0,08$	$b^* = 0,96b_2 + 0,02$	$b^* = 1,03b_2 + 0,04$
13 ke 1	$b^* = b_2 + 0,06$	$b^* = 1,18b_2 + 0,07$	$b^* = 0,96b_2 + 0,02$	$b^* = 1,03b_2 + 0,04$
14 ke 1	$b^* = b_2 + 0,07$	$b^* = 1,12b_2 + 0,07$	$b^* = 0,96b_2 + 0,03$	$b^* = 1,03b_2 + 0,05$
15 ke 1	$b^* = b_2 + 0,03$	$b^* = 0,98b_2 + 0,04$	$b^* = 0,66b_2 + 0,02$	$b^* = 0,99b_2 + 0,01$
16 ke 1	$b^* = b_2 + 0,23$	$b^* = 0,88b_2 + 0,21$	$b^* = 0,58b_2 + 0,13$	$b^* = 0,95b_2 + 0,14$
17 ke 1	$b^* = b_2 + 0,16$	$b^* = 1,15b_2 + 0,18$	$b^* = 1,03b_2 + 0,08$	$b^* = 1,05b_2 + 1,12$
18 ke 1	$b^* = b_2 + 0,39$	$b^* = 0,85b_2 + 0,34$	$b^* = 0,82b_2 + 0,14$	$b^* = 0,94b_2 + 0,28$
19 ke 1	$b^* = b_2 + 0,18$	$b^* = 1,10b_2 + 0,19$	$b^* = 0,97b_2 + 0,01$	$b^* = 1,03b_2 + 0,13$
20 ke 1	$b^* = b_2 + 0,14$	$b^* = b_2 + 0,14$	$b^* = 0,67b_2 + 0,07$	$b^* = 1,00b_2 + 0,10$

Koefisien pada metode rerata dan rerata pada hasil analisis *Equating* maupun *Concordance* bernilai 1. Hal ini menunjukkan bahwa hubungan antarpaket berupa transformasi linear dari tingkat kesulitan paket 1. Hal yang berbeda adalah koefisiennya. Koefisien yang perbedaannya menonjol (lebih dari 0,10) pada beberapa paket pada *equating* dan pada *concordance*. Hal ini menunjukkan paket-paket ini memiliki tingkat kesulitan yang sedikit lebih tinggi dibandingkan paket 1.

Hasil analisis *equating* dan *concordance* menghasilkan persamaan-persamaan transformasi indeks tingkat kesulitan.

Hasil transformasi tingkat kesulitan berupa tingkat kesulitan suatu paket dengan skala tingkat kesulitan paket 1. Tingkat kesulitan yang sudah setara ini digunakan untuk menggambarkan kurva karakteristik tes dari tiap metode. Gambar ini kemudian dibandingkan dengan metode lain. Semakin berimpit dengan kurva karakteristik tes paket 1, semakin setara paket tersebut dengan paket 1. Dengan kata lain, jika kesembilan belas kurva karakteristik tes semakin dekat, kedua puluh paket tersebut semakin setara. Kurva karakteristik dengan *equating* disajikan pada Gambar 1 dan dengan *concordance* disajikan pada Gambar 2.

Gambar 1. Kurva karakteristik Tes Paket 2-20 setelah Disetarkan ke Paket 1 dengan *Equating*

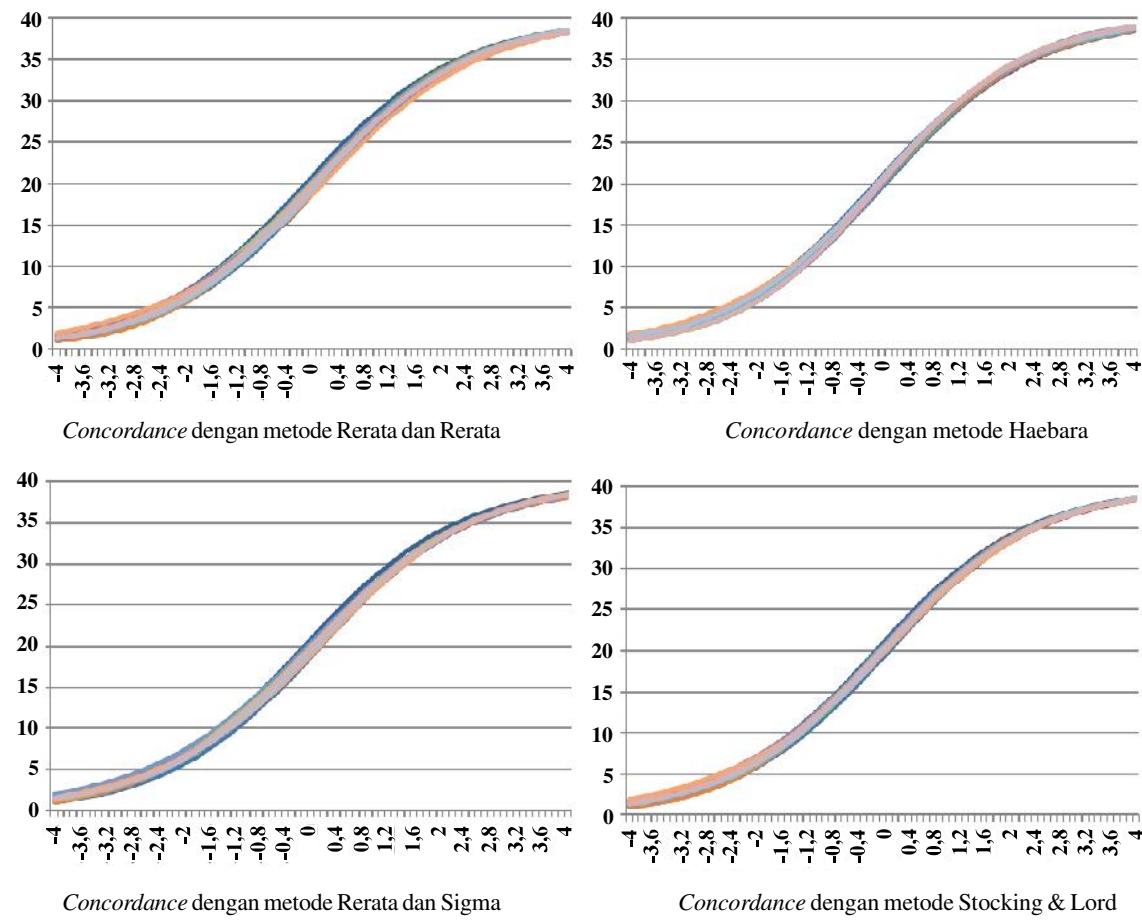
Mencermati Gambar 1, dapat diperoleh bahwa *Equating* dengan metode rerata dan sigma menghasilkan kurva yang saling mendekati, mengindikasikan bahwa metode ini menghasilkan skor-skor tes yang mendekati setara. Pada *concordance*, keempat metode menghasilkan gambar yang relative sama. Namun secara grafis, metode yang menghasilkan grafik yang paling rapat metode Haebara dan Stocking & Lord. Hasil selengkapnya disajikan pada Gambar 2.

Dengan menggunakan parameter kemampuan, hasil estimasi dan persamaan hasil analisis dengan *equating* dan *concordance* dapat diestimasi skala kemampuan hasil *equating* dan *concordance*. Hasil ini digunakan

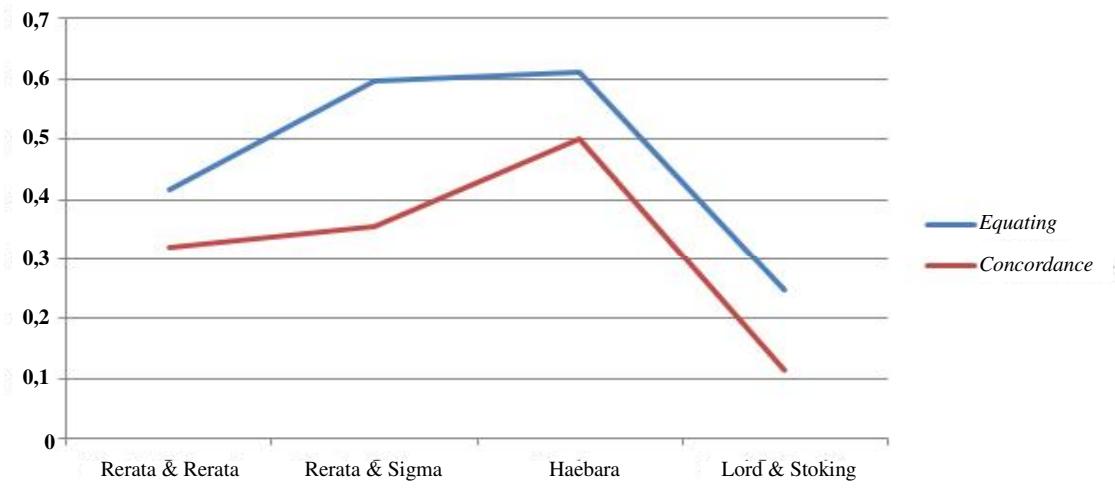
untuk menghitung RMSE. Hasilnya disajikan pada Tabel 5 untuk *equating* dan Tabel 6 untuk *concordance*. Dengan menggunakan tabel tersebut, dapat dibuat grafik untuk melihat kecenderungannya. Grafik perbandingan dengan *equating* dan *concordance* disajikan pada Gambar 3.

Berdasarkan hasil analisis data dari 20 paket soal matematika ujian nasional, ada butir yang sangat sulit. Indeks kesulitan ini berada di atas 3 diestimasi dengan model Rasch. Setelah dicermati kembali, butir yang sangat sulit ini adalah butir 3. Butir ini mengukur indikator operasi perpangkatan bilangan. Naskah butir soal tersebut dari salah satu paket sebagai berikut.

Gambar 2. Kurva karakteristik Tes setelah disetarakan ke Paket 1 dengan *Concordance*



Gambar 3. Perbandingan RMSE dengan *Equating* dan *Concordance*



Tabel 5. RMSE Equating

Paket	Metode			
	Rerata & Rerata	Rerata & Sigma	Haebara	Lord & Stocking
P2	0,140	0,499	0,122	0,133
P3	0,085	0,086	0,035	0,036
P4	0,040	0,072	0,067	0,038
P5	1,638	1,954	1,682	0,133
P6	1,875	2,200	1,920	0,133
P7	0,440	0,881	0,435	0,522
P8	0,040	0,048	0,043	0,038
P9	0,070	0,112	0,082	0,110
P10	0,300	0,499	0,183	0,209
P11	0,030	0,499	0,028	0,037
P12	0,030	0,499	0,104	0,067
P13	0,160	0,499	0,099	0,100
P14	0,060	0,499	0,226	0,076
P15	0,500	0,499	2,351	0,477
P16	1,000	0,499	0,806	0,968
P17	0,020	0,499	1,952	0,071
P18	0,290	0,499	0,372	0,193
P19	0,130	0,499	0,125	0,121
P20	1,070	0,499	0,950	1,217
Rerata	0,417	0,597	0,610	0,246
Varians	0,319	0,316	0,601	0,109

Hasil dari $125^{\frac{2}{3}}$ adalah

- A. 5
- B. 15
- C. 25
- D. 50

Untuk mengerjakan butir ini, peserta tes harus mengubah dahulu 125 menjadi bilangan dengan pangkat tiga yaitu 5^3 . Selanjutnya, dioperasikan $[5^{\frac{2}{3}}]^3$ sehingga menghasilkan $5^2 = 25$. Memanipulasikan 125 menjadi 5^3

kemudian mengoperasikannya dengan pangkat $\frac{2}{3}$ menyebabkan butir ini menjadi butir yang sangat sulit.

Mencermati Gambar 3 tersebut dapat diperoleh bahwa metode-metode pada *concordance* menghasilkan RMSE yang lebih rendah secara seragam dibandingkan metode-metode pada *equating*. Pada *equating* maupun *concordance* ada kecenderungan metode rerata dan rerata lebih rendah dibandingkan metode rerata dan sigma dan metode Haebara, namun yang paling rendah adalah metode Lord &

Stocking. Hal ini mengindikasikan bahwa metode Lord & Stocking menghasilkan RMSE terkecil baik pada *equating* maupun *concordance* (Tabel 6).

Parameter butir hasil estimasi kemandian digunakan untuk melakukan *equating* dan *concordance* dengan berbagai metode. Dengan menggunakan grafik kurva karakteristik tes, metode rerata, dan sigma menghasilkan grafik yang paling setara pada *equating*. Pada *concordance*, keempat metode menghasilkan gambar yang relatif sama, yang keempat-empatnya relatif setara. Hal ini disebabkan karena hanya satu parameter yang menentukan persamaan

penyetaraan, yaitu tingkat kesulitan, dan perbandingan tingkat kesulitan bernali satu.

Hasil tersebut sedikit berbeda dengan perbandingan dengan menggunakan RMSE. Pada RMSE dihitung dengan memanfaatkan parameter kemampuan hasil estimasi, dan parameter hasil *equating*. Hasil ini ditentukan oleh parameter kemampuan hasil estimasi, yang diskalakan dengan persamaan penyetaraan. Persamaan penyetaraan ini yang memberikan kontribusi besarnya RMSE. Semakin kecil RMSE, semakin akurat metode penyetaraan yang digunakan. Pada studi ini, metode yang menghasilkan

Tabel 6. RMSE pada *Concordance*

Paket	Metode			
	Rerata & Rerata	Rerata & Sigma	Haebara	Lord & Stocking
P2	0,100	0,157	0,076	0,076
P3	0,060	0,419	0,080	0,113
P4	0,180	0,216	0,077	0,161
P5	1,633	1,694	1,596	0,076
P6	1,870	1,933	1,831	0,076
P7	0,180	0,174	0,111	0,122
P8	0,180	0,173	0,114	0,122
P9	0,200	0,239	0,130	0,158
P10	0,300	0,157	0,183	0,209
P11	0,060	0,157	0,028	0,061
P12	0,080	0,157	0,056	0,067
P13	0,060	0,157	0,056	0,066
P14	0,070	0,157	0,059	0,075
P15	0,030	0,157	1,919	0,024
P16	0,230	0,157	0,562	0,127
P17	0,160	0,157	1,864	0,152
P18	0,390	0,157	0,253	0,247
P19	0,180	0,157	0,041	0,149
P20	0,140	0,157	0,485	0,100
Rerata	0,321	0,354	0,501	0,115
Varians	0,264	0,270	0,501	0,003

RMSE terkecil adalah metode Lord & Stocking.

Dengan menggunakan *concordance*, diperoleh grafis yang lebih rapat dan RMSE yang lebih kecil. Hal ini disebabkan karena jumlah butir bersama yang digunakan untuk *equating* pada tiap paket berbeda jumlahnya dan tidak sampai 40 butir. Sementara itu, pada *concordance*, keseluruhan butir digunakan sebagai butir bersama dalam perhitungan rerata maupun standar deviasi. Hasil ini memberikan hasil yang lebih stabil untuk menentukan koefisien α dan konstanta β . Hal ini sesuai dengan penelitian Pang, Madera, Radwan, dan Zhang (2010) yang menyatakan bahwa penggunaan penyetaraan skor lebih baik dibandingkan dengan *concurrent calibration*. Terkait dengan perbandingan metode penyetaraan skor, Yu & Popp (2005, pp.1-19) mengatakan bahwa tidak ada metode penyetaraan yang paling baik. Hal ini menunjukkan bahwa perlunya penelitian lanjutan mengenai perbandingan metode *equating* maupun *concordance*.

Dalam menyetaraan skor, ada banyak faktor yang mempengaruhi *equating*. Pada prosesnya perlu estimasi parameter terlebih dahulu, kemudian menggunakan berbagai metode penyetaraan. Ketika mengestimasi parameter, berbagai variabel mempengaruhi hasil estimasi. Variabel-variabel tersebut di antaranya model yang digunakan untuk estimasi, banyaknya peserta tes, banyaknya butir dalam tes, metode estimasi baik parameter butir maupun parameter kemampuan. Pada penyetaraan, variabel yang perlu diteliti adalah distribusi parameter butir, distribusi parameter kemampuan, perbandingan metode estimasi, banyaknya butir bersama dalam tes, dan *software* yang digunakan. Variabel-variabel tersebut dapat diteliti pengaruh dan sifat-sifatnya menggunakan studi simulasi dengan data riil untuk memodelkannya.

Analisis kesetaraan perangkat tes maupun analisis butir seperti ini sangat diperlukan dalam pendidikan. Hal ini terkait dengan pemanfaatan skor tes skala luas yang berdampak pada kebijakan yang lain, misalnya seleksi masuk siswa untuk melanjutkan ke jenjang yang lebih tinggi maupun pemetaan mutu pendidikan. Analisis butir seperti yang dilakukan Lumapow (2012, pp. 61-75) dapat digunakan pula untuk memperbaiki kualitas pembelajaran di sekolah. Jika penilaian hasil pembelajaran menggunakan model alternatif dengan penyekoran politomi seperti portofolio yang dilakukan Masruri & Nurhadi (2007, pp. 167-186) analisis penyetaraan skor hasil penskoran data ini juga perlu dilakukan.

SIMPULAN

Hasil studi menunjukkan bahwa dua puluh paket yang digunakan pada ujian nasional menunjukkan kecenderungan yang setara. Pada *equating* dengan metode grafis menggunakan kedekatan kurva karakteristik tes, rerata dan sigma menghasilkan skor yang paling setara. Pada metode kurva karakteristik, metode Haebara dan metode Stocking dan Lord menghasilkan skor-skor dengan RMSE yang paling kecil. Dengan menggunakan 20 perangkat ujian nasional, menghubungkan skor hasil tes dengan desain *concordance* menghasilkan RMSE lebih kecil dibandingkan *equating*.

DAFTAR PUSTAKA

Anderson, B., Braunberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating with the package *kequate*. *Journal of Statistical Software*, 55(6), 1-25.

Antara, A. A. P., & Bastari. (2015). Penyetaraan vertikal dengan pendekatan klasik dan item response theory pada siswa sekolah dasar. *Jurnal*

Penelitian dan Evaluasi Pendidikan, 19(1), 13-24.

Aşiret, S., & Sünbül, S. Ö. (2016). Investigating test equating methods in smallsamples through variousfactors. *Educational Sciences: Theory & Practice*, 16, 647-668.

Brennan, R. L. & Kolen, M. J. (2004). Concordance between ACT and ITED scores from different population. *Jurnal Applied Psichological Measurement*, 28(4), 219-226.

Dorans, N. J. (2004). Equating, concordance and expectation. *Jurnal Applied Psichological Measurement*, 28(4), 219-226.

Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). *Principles and practices of test score equating research report*. Diunduh dari <http://www.ets.org/research/contact.html>.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.

Han, K. T. (2009). IRTEQ: Windows application that implements IRT scaling and equating [computer program]. *Applied Psychological Measurement*, 33(6), 491-493.

Kim S. H. & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under graded response model. *Applied Psychological Measurement*, 26(25-61).

Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practices*. New York: Springer.

Lumapow, H. (2012). Identifikasi materi sulit ujian nasional bahasa inggris pada siswa jurusan bahasa. *Jurnal Kependidikan*, 42(1), 61-75.

Mardapi, D. (1998). analisis butir dengan teori tes klasik dan teori respons butir. *Jurnal Kependidikan*, 28(2).

Masruri, M. S., & Nurhadi. (2007). Peningkatan kualitas pembelajaran mata kuliah penilaian dan pencapaian belajar geografi melalui penerapan model portofolio. *Jurnal Kependidikan*, 37(2), 167-186.

Moses, T. P., & Liu, J. (2011). Smoothing and equating methods applied to different types of test score distributions and evaluated with respect to multiple equating criteria. *ETS Research Report Series*, 2011(1), i-25. Diunduh dari https://www.ets.org/research/policy_research_reports/publications/report/2011/isez.

Pang, X., Madera, E., Radwan, N., & Zhang, S. (2010). A comparison of four test equating. *Methods Research Report*. Diunduh dari <http://www.eqao.com>.

Retnawati, H., & Hidayati, K. (2007). *Perbandingan metode concordance berdasarkan teori tes klasik* (Laporan penelitian). Yogyakarta: Lembaga Penelitian UNY.

Retnawati, H. (2014). *Teori respons butir dan penerapannya*. Yogyakarta: Parama.

Ryan, J., & Brockmann, F. (2009). A practitioner's introduction to equating with primers on classical theory and item respons theory. *Council of Chief State School Officers*.

Uysal, İ., & Kilmen, S. (2016). Comparison of item response theory test equating methods for mixed format tests. *International Online Journal of Educational Sciences*, 8(2), 1-11.

Yu, C. H., & Popp, S. E. O. (2005). Test equating by common items and common subjects: Concepts and applications. *Practical Assessment, Research & Evaluation*, 10(4), 1-19.

INDEKS SUB-JEK

Symbols

A

B

BUTIR BERSAMA, [164](#)

C

concordance, [164](#)

D

desain butir bersama, [165](#)
desain grup ekuivalen, [165](#)
desain grup tunggal, [165](#)

E

equating, [164](#)

F

G

H

I

J

K

L

M

metode graf k kurva karakteristik tes, [168](#)
metode Haebara, [164](#)
METODE PENYETARAAN SKOR TES, [164](#)
metode Stocking dan Lord, [164](#)

N

O

P

program RTEQ, [164](#)
program QUEST, [164](#)

Q

R

root mean square of error (RMSE), [164](#)

S

T

U

V

W

X

Y

Z