PENGARUH METODE DAN UKURAN SAMPEL TERHADAP VARIANSI SKOR HASIL PENYETARAAN

Tri Rijanto
Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Surabaya
hari_tri2001@yahoo.com

Abstrak

Penelitian ini bertujuan untuk memperoleh informasi perbedaan variansi skor hasil penyetaraan (*equating*) metode linear dan metode ekipersentil untuk ukuran sampel 200, 400, dan 800 pada Ujian Akhir Sekolah Berstandar Nasional (UASBN). Metode yang digunakan adalah simulasi dengan variabel metode penyetaraan dan banyaknya responden. Data penelitian berupa respons peserta UASBN SD/MI tahun pelajaran 2008/2009 mata pelajaran IPA di Jakarta Timur, yang ditentukan menggunakan teknik penarikan sampel acak dengan pengembalian. Hipotesis diuji menggunakan uji kesamaan variansi. Hasil penelitian dengan $\alpha = 0{,}05$ menunjukkan: (1) variansi skor penyetaraan metode ekipersentil (σ^2_{ekp200}) tidak berbeda dengan variansi skor penyetaraan metode linear (σ^2_{lin200}) untuk ukuran sampel 200, (2) variansi skor penyetaraan metode ekipersentil (σ^2_{ekp400}) tidak berbeda dengan variansi skor penyetaraan metode linear (σ^2_{lin400}) untuk ukuran sampel 400, dan (3) variansi skor penyetaraan metode ekipersentil (σ^2_{ekp800}) berbeda dengan variansi skor penyetaraan metode linear (σ^2_{lin800}) untuk ukuran sampel 800.

Kata kunci: *variansi skor, equating, metode ekipersentil, metode linear*

THE IMPACT OF METHODS AND SAMPLE SIZE TO THE SCORE VARIANCE OF EQUATING RESULT

Tri Rijanto
Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Surabaya
hari_tri2001@yahoo.com

Abstract

This study was aimed to obtain information on the difference of score variance as a result of equating linear method and equipercentile method for the sample size of 200, 400, and 800 in the Final Examinition of National Standardized Schools. The method used was a simulation of variables equating method and the number of respondents. The population are examinees from the 2008/2009 elementary school final examination for science class in East Jakarta. Random sampling with replacement technique was used. The hypotheses were tested using similarity variance. The results with $\alpha = 0{,}05$ showed that: (1) the equated score variance from equipercentile method (σ^2_{ekp200}) was not different from the equated score variance from linear method (σ^2_{lin200}) for the sample size of 200, (2) the equated score variance from equipercentile method (σ^2_{ekp400}) was not different from the equated score variance from linear method (σ^2_{lin400}) for the sample size of 400, and (3) the equated score variance from equipercentile method (σ^2_{ekp800}) was different from the equated score variance from linear method (σ^2_{lin800}) for the sample size of 800.

Keywords: *score variance, equating, equipercentile method, linear method*

Pendahuluan

Peraturan Pemerintah Republik Indonesia Nomor 19 tahun 2005 tentang Standar Nasional Pendidikan, Pasal 63 Ayat (1) menyebutkan bahwa penilaian pendidikan pada jenjang pendidikan dasar dan menengah terdiri atas: (1) penilaian hasil belajar oleh pendidik, (2) penilaian hasil belajar oleh satuan pendidikan, dan (3) penilaian hasil belajar oleh pemerintah. Penilaian hasil belajar oleh pemerintah bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan teknologi dan dilakukan dalam bentuk ujian nasional (UN).

Ujian nasional untuk sekolah dasar/madrasah ibtidaiyah/sekolah dasar luar biasa (SD/MI/SDLB) dilakukan pertama kali pada 2008. Ujian tersebut bernama Ujian Akhir Sekolah Berstandar Nasional (UASBN) sesuai dengan Peraturan Menteri Pendidikan Nasional RI Nomor 39 tahun 2007. Ujian ini bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam (IPA) serta mendorong tercapainya target wajib belajar pendidikan dasar yang bermutu (Depdiknas, 2007:6).

Ujian Akhir Sekolah Berstandar Nasional SD/MI/SDLB dilaksanakan secara terintegrasi dengan ujian sekolah/madrasah. Artinya setiap paket soal UASBN terdiri dari 25% soal yang ditetapkan oleh Badan Standar Nasional Pendidikan (BSNP) dan berlaku secara nasional, serta 75% soal yang ditetapkan oleh penyelenggara UASBN tingkat provinsi berdasarkan spesifikasi yang ditetapkan oleh BSNP. Dengan kata lain dalam paket soal terdiri dari 25% butir *anchor items* dan sisanya butir soal yang dibuat oleh masing-masing provinsi.

Instrumen penilaian yang digunakan oleh pemerintah dalam bentuk ujian nasional (UN) menurut Peraturan Menteri Pendidikan Nasional RI Nomor 20 tahun 2007 tentang Standar Penilaian Pendidikan memenuhi persyaratan substansi, konstruksi, bahasa, dan memiliki bukti validitas empirik serta menghasilkan skor yang dapat diperbandingkan antarsekolah, antardaerah, dan antartahun (Depdiknas, 2007:9). Keterbandingan skor antarsekolah, kabupaten/kota, provinsi, dan antartahun dapat diperoleh

jika semua peserta tes mengerjakan soal-soal (paket tes) yang sama. Perbedaan skor antarmereka menunjukkan perbedaan tingkat kemampuannya. Dalam praktiknya, pengadministrasian soal-soal sama antartahun, merugikan peserta tes yang mengerjakan pada tahun-tahun pertama dan menguntungkan mereka yang ikut tes pada tahun-tahun terakhir. Juga, pengadministrasian soal-soal yang sama di setiap sekolah sangat beresiko terhadap kebocoran.

Sebagai jalan keluar agar keadilan *(fairness)* dan kerahasiaan *(test security)* soal-soal ujian terjaga mengharuskan pengadministrasian paket-paket tes berbeda antartahun, daerah, dan tempat tes. Tetapi masalah lain muncul, dengan mengadministrasian paket-paket yang berbeda, perbedaan skor antarpeserta tes tidak dapat langsung disimpulkan adanya perbedaan kemampuan antarmereka, karena tingkat kesukaran paket yang digunakan mempengaruhi perbedaan tersebut.

Untuk menanggulangi ketidakadilan tersebut dilakukan penyamaan atau penyetaraan matriks skor *(equating)*. Penyetaraan matriks skor merupakan cara untuk memperoleh suatu konversi nilai dari skor mentah suatu paket, ke skor mentah paket yang lain. Jadi, melalui penyetaraan matriks skor dimungkinkan siswa menjawab benar 32 soal di Paket A, misalnya, mendapat nilai sama dengan siswa menjawab benar 30 soal di Paket B, karena Paket B lebih sukar dua soal dari Paket A.

Akibat lebih serius dengan tidak adanya penyetaraan *(equating)* adalah ketidakadilan pada kelulusan. Batas lulus 5,5 misalnya, pada paket-paket sukar akan merugikan siswa-siswa yang mengerjakannya dan akan menguntungkan mereka yang mendapatkan paket-paket mudah. Oleh karena itu kemungkinan besar ada siswa yang seharusnya lulus tetapi karena mendapatkan paket sukar menjadi tidak lulus. Dengan demikian masalah penyetaraan menjadi penting untuk dikaji melalui penelitian yang komprehensif.

Untuk dapat membandingkan atau menyetarakan skor ujian nasional diperlukan desain penyetaraan yang tepat. Hal ini sejalan dengan pendapat Marco, *et al.,* (1983:148) bahwa karena penyetaraan merupakan prosedur empirik, maka dibutuhkan desain untuk mengumpulkan data dan suatu aturan untuk mentransformasikan skor dari suatu tes pada skor tes yang

lain. Menurut Hambleton dan Swaminathan (1985:198) ada tiga desain dasar penyetaraan, yaitu metode kelompok tunggal *(single group method)*, metode kelompok ekivalen *(equivalent group method)*, dan metode tes jangkar *(anchor test design)*.

Terdapat dua cara penyamaan skor pada teori klasik yaitu penyamaan metode linear dan metode ekipersentil. Menurut Hambleton, Swaminathan, dan Rogers (1991:124) asumsi penyetaraan metode linear adalah kedua skor tes distribusinya berbeda, distribusi tersebut terkait dengan rerata dan simpangan bakunya. Dengan demikian pada metode linear membutuhkan asumsi rerata dan simpangan baku atau variansinya berbeda.

Pada teori klasik, terdapat jumlah butir tertentu serta jumlah peserta tertentu. Dalam waktu yang bersamaan, semua peserta menjawab semua butir ujian. Jawaban setiap peserta terhadap setiap butir ujian dapat berupa jawaban benar atau berupa jawaban salah. Jawaban benar dan salah menjadi skor pada ujian tersebut. Ada kalanya jawaban benar dan salah tersebut dikaitkan dengan sukar dan tidak sukarnya butir ujian tersebut dirasakan oleh peserta ujian. Akan tetapi sukar dan tidak sukarnya butir ujian sangat bergantung kepada para peserta ujian. Butir yang dirasakan tidak sukar oleh satu peserta mungkin saja dirasakan sukar oleh peserta lain. Dengan demikian dapat saja terjadi peserta tertentu merasa semua butir ujian adalah sukar sementara ada pula peserta lainnya yang merasa semua butir ujian adalah tidak sukar.

Taraf sukar dan daya beda butir merupakan ciri butir pada ujian klasik. Taraf sukar butir pada tes yang menggunakan skor dikotomi, rerata skor butir yang diperoleh berkaitan dengan proporsi peserta tes yang menjawab benar butir tersebut. Proporsi ini untuk butir *ke-i* biasanya dinyatakan sebagai p_i, pada analisis butir klasik p_i disebut sebagai taraf sukar butir *(item difficulty)*. Nilai optimum p tergantung pada tujuan penyelenggaraan tes. Misalnya, untuk tes yang diadakan untuk seleksi beasiswa atau tes penempatan, nilai optimum p rendah, tetapi tes yang dirancang untuk menyeleksi siswa yang perlu remidial nilai optimum relatif tinggi. Untuk menyeleksi siswa diperlukan rentangan kemampuan yang cukup besar maka nilai optimum p mendekati 0,50 (Croker dan Algina, 1986:69). Jika kemampuan peserta lebih dari indeks taraf sukar butir, maka

probabilitas menjawab benar akan lebih dari probabilitas menjawab salah sehingga p>0,50. Sebaliknya, jika kemampuan peserta tes kurang dari indeks taraf sukar butir, maka probabilitas jawaban benar akan kurang dari probabilitas jawaban salah sehingga $p < 0{,}50$.

Ciri kedua dalam analisis butir klasik adalah daya beda butir untuk tes berbentuk dikotomi. Daya beda butir adalah kemampuan suatu butir untuk membedakan peserta tes dari kelompok kemampuan tinggi dengan peserta tes kelompok kemampuan rendah. Daya beda butir dinyatakan dalam bentuk indeks daya beda *(discrimination index)* yang biasanya diberi lambang *D*. Jika $D>0$ berarti peserta kelompok kemampuan tinggi menjawab benar lebih banyak dibandingkan peserta kelompok kemampuan rendah. Sebaliknya jika harga $D < 0$ berarti peserta kelompok kemampuan rendah menjawab lebih banyak dari kelompok kemampuan tinggi.

Daya beda butir sebenarnya diestimasi menggunakan metode korelasi butir total. Kaplan dan Saccuzzo (1982:147) menyebut metode ini sebagai *point biserial method* karena menggunakan korelasi biserial titik *(point biserial correlation)*. Harga koefisien beserial titik antara $-1 \leq r_{pbis} \leq 1$. Indeks daya beda minimum 0,2 menurut Nunnaly (1970), Croker dan Algina (1986), Mehren dan Lehman (1991), dan Aiken (1994), (dalam Naga, 2008:89) dikatakan memadai dan indeks daya beda < 0,2 diangap tidak memadai. Indeks taraf sukar butir dan daya beda butir pada teori klasik sangat tergantung pada kelompok peserta tes. Artinya, harga kedua indeks akan berubah jika peserta tes juga berubah yang dikenal dengan istilah tidak invarian.

Pengukuran dalam pendidikan mengenal dua macam kekeliruan, yaitu kekeliruan acak atau kekeliruan sampel *(sampling error)* dan kekeliruan sistematik *(systematic error)* (Naga, 1992:116). Kekeliruan sampel adalah perbedaan antara keadaan sebenarnya yang ada pada populasi *(true score)*. Hal ini disebabkan oleh karena hasil ukuran pada sampel tersebut hanya merupakan salah satu dari sekian banyak kemungkinan hasil pengukuran yang dapat dicuplik secara berulang-ulang dari suatu populasi. Kekeliruan sampel tetap saja muncul meskipun alat ukur yang dipakai, situasi dan kondisi pengukuran, maupun jenis kemampuan yang diukur tetap sama. Dengan demikian ukuran sampel berpengaruh terhadap hasil pengukuran.

Sampel menurut Steel dan Torrie (1980:3) adalah bagian dari populasi, kadang-kadang mencakup seluruh populasi dan umumnya informasi dari sampel digunakan untuk penarikan kesimpulan tentang populasi itu. Pada sisi lain, dikenal juga pengertian sampel butir. Lord dan Novick (1968:235-236), secara eksplisit membedakan antara sampel butir *(sample of items)* yang ditarik secara acak dari suatu populasi butir dan sampel peserta tes *(sampel of examinees)* yang ditarik secara acak dari suatu populasi peserta tes. Sedangkan Thordike *et al.* (1991:197) menyebutkan pengertian butir sampel *(sampel item)* dan bukan sampel butir, mereka mendefinisikan butir sampel sebagai contoh-contoh yang serupa dengan apa yang akan muncul pada tes, terkait dengan (1) atribut stimulus untuk membatasi kompleksitas dari stimulus yang dihadapi peserta tes dan (2) atribut respons untuk menggambarkan dengan cepat bagaimana konstruk itu akan diukur.

Naiman, Resenfeld, dan Zirkel (1993:156) menyatakan bahwa untuk sampel besar dengan ukuran sampel *(sample size)* >30 secara statistik akan menghasilkan suatu distribusi rerata cukup dekat pada distribusi normal, sehingga kalkulasi berdasarkan kurva normal masuk akal. Eid meneliti tentang pengaruh ukuran sampel *(sample size)* pada penyetaraan butir tes mengusulkan untuk menggunakan ukuran sampel 200, 400, dan 800. Setiadi (1997:7) dalam penelitiannya terhadap estimasi parameter butir menyatakan bahwa sampel yang relatif kecil berukutan 100 atau 200, sedangkan Livingston dan Feryok (1987:9-10) melakukan penelitian pada penyetaraan ekipersentil estimasi frekuensi dengan penghalusan pada sampel berukuran 100 sampai dengan 3000 dan akurasi penyetaraan terjadi pada sampel berukuran 300. Dengan demikian penelitian ini menggunakan ukuran sampel 200, 400, dan 800.

Tujuan penyetaraan skor adalah untuk mambandingkan skor yang diperoleh dari perangkat tes yang satu (X) dan perangkat tes lainnya (Y) yang dilakukan melalui proses penyetaraan skor pada kedua perangkat tersebut (Hambleton dan Swaminathan, 1990:199). Selanjutnya Croker dan Algina (1986:4) mengemukakan bahwa dua skor, hasil pengukuran yang menggunakan instrumen X dan instrumen Y dapat disetarakan skornya jika keduanya mengukur kemampuan atau trait yang sama. Kemudian Aiken (1997:82) juga mengemukakan bahwa skor hasil pengukuran dua perangkat

tes yang paralel dapat disetarakan untuk mengkonversi skor dari perangkat tes satu terhadap skor dari perangkat tes lainnya.

Menurut Wiersma dan Jurs (1990:148) penyetaraan skor adalah suatu prosedur empiris yang diperlukan untuk mentransformasikan skor suatu tes ke skor tes yang lain. Karena merupakan prosedur empiris maka penyetaraan skor didasarkan pada data skor tes. Pendapat lain dikemukakan oleh Brown (1976:216) yang mengatakan bahwa skor dua tes dapat disetarakan jika kedua tes tersebut diperoleh dari populasi yang sama. Senada dengan Brown, Kolen menyatakan bahwa penyetaraan skor dapat dilakukan jika kelompok peserta setara, karena ketidaksetaraan yang ekstrim akan berpengaruh dalam perhitungan (Keeves, 1997:733).

Penyetaraan skor di antara ujian yang berbeda dilakukan melalui rancangan tertentu. Menurut Hambleton (1990:199) ada tiga rancangan penyetaraan skor yang umum digunakan, yaitu rancangan kelompok tunggal *(single group design)*, rancangan kelompok acak *(random group design),* dan rancangan butir bersama *(anchor items design)*. Pada rancangan kelompok tunggal, secara acak, peserta dibagi ke dalam dua kelompok, yaitu kelompok K1 dan kelompok K2. Dengan keacakan itu dimaksudkan agar K1 dan K2 adalah setara, yakni mereka memiliki kemampuan yang sama dengan variansi yang sama pula. Kemudian kelompok K1 mengerjakan tes X dan kelompok K2 mengerjakan tes Y. Dari kedua hasil tes tersebut kemudian dicari patokan konversinya.

Rancangan kelompok acak, secara acak peserta dibagi ke dalam dua kelompok, yaitu kelompok K1 dan kelompok K2. Cara acak ini dimaksudkan agar kedua kelompok peserta adalah setara. Selanjutnya kelompok K1 mengerjakan tes X dan kemudian tes Y, serta kelompok K2 mengerjakan tes Y dan kemudian tes X. Dengan demikian setiap kelompok mengerjakan dua tes dalam urutan yang terbalik dan selanjutnya dari hasil tes tersebut disetarakan.

Pada rancangan butir bersama tidak perlu secara acak membagi dua kelompok peserta seperti halnya pada rancangan kelompok tunggal dan rancangan kelompok acak. Sudah cukup terdapat kelompok peserta K1 dan K2. Namun demikian, rancangan ini tidak menolak penggunaan kelompok yang setara, yaitu kelompok peserta seperti pada rancangan kelompok

tunggal dan rancangan kelompk acak. Selanjutnya, tes X ditambah dengan sejumlah butir sebagai tes Z serta tes Z yang sama ditambahkan juga ke tes Y. Jadi bersama tambahan tersebut terdapat perangkat tes (X+Z) dan perangkat tes (Y+Z). Butir ujian Z dikenal sebagai tes jangkar *(anchor test)*.

Selanjutnya kelompok K1 mengerjakan tes (X+Z) dan kelompok K2 mengerjakan tes (Y+Z). Dari kedua tes tersebut, khususnya tes Z yang dikerjakan oleh kedua kelompok kemudian dicari patokan konversinya. Dilihat dari penyiapan peserta tes, rancangan butir bersama adalah yang paling praktis karena tidak mensyaratkan kesetaraan di antara kelompok peserta. Oleh karena itu tulisan ini menggunakan rancangan butir bersama dan rancangan ini pula yang digunakan dalam UASBN. Panjang tes pada UASBN 2009 mata pelajaran IPA Paket 01 dan Paket 02 sebanyak 40 butir, yang terdiri dari 30 butir (75%) merupakan butir daerah dan 10 butir (25%) merupakan butir pusat atau butir bersama *(anchor items)*. Hal ini sejalan dengan hasil penelitian Yahya Umar (1987:105) bahwa dengan panjang tes 40 butir dibutuhkan minimal 10 butir (25%) *anchor items* agar diperoleh hasil yang baik. Demikian pula hasil penelitian Yetti Supriyati (2003) yang merekomendasikan proporsi minimal butir *anchor items* sebesar 20% dari butir total.

Pada rancangan butir bersama kelompok peserta K1 mengerjakan tes X dan Z sedangkan kelompok peserta K2 mengerjakan tes Y dan Z. Dengan demikian, tes Z dikerjakan oleh kedua kelompok. Untuk membedakan digunakan notasi berbeda pada hasil tes tersebut. Tes Z yang dikerjakan oleh kelompok K1 menghasilkan skor Z_1 dengan rerata skor μ_{Z1} dan simpangan baku σ_{Z1}. Tes Z yang dikerjakan oleh kelompok peserta K2 menghasilkan skor Z_2 dengan rerata skor μ_{Z2} dan simpangan baku σ_{Z2}. Hasil gabungan ujian Z yang dikerjakan oleh kelompok peserta K1 dan K2 adalah skor Z dengan rerata μ_Z dan simpangan baku σ_Z.

Tes X hanya dikerjakan oleh kelompok peserta K1 dan menghasilkan skor X dengan rerata μ_X dan simpangan baku σ_X, sedangkan tes Y hanya dikerjakan oleh kelompok K2 dan menghasilkan skor Y dengan rerata μ_Y dan simpangan baku σ_Y. Dari sini yang diperhatikan adalah regresi skor X terhadap Z yang dikerjakan oleh kelompok peserta K1, serta regresi skor Y

terhadap skor Z yang dikerjakan oleh kelompok peserta K2. Kedua regresi tersebut masing-masing menghasilkan koefisien arah b_{XZ1} dan b_{YZ2}.

Selisih nilai X yang disebabkan oleh selisih $\mu_{\tilde{z}}$ dan $\mu_{\tilde{z}1}$ adalah sebesar $b_{XZ1}(\mu_{\tilde{z}} - \mu_{\tilde{z}1})$ sehingga rerata μ_x bergeser dengan nilai itu. Demikian pula, dengan jalan yang sama, selisih nilai Y yang ditimbulkan oleh selisih $\mu_{\tilde{z}}$ dan $\mu_{\tilde{z}2}$ adalah sebesar $b_{XZ2}(\mu_{\tilde{z}} - \mu_{\tilde{z}2})$ sehingga rerata μ_Y juga bergeser dengan nilai itu. Hal ini juga terjadi pada simpangan baku.

Penyetaraan ekipersentil menurut Braun & Holland (1982) dan Lord (1950) didasarkan pada definisi bahwa skala skor untuk kedua tes sebanding dan distribusi skor kedua tes identik (Linn, 1987:247). Sementara itu menurut Croker dan Algina (1986:346), secara umum diterima bahwa skor pada dua tes dianggap ekivalen jika skor-skor tersebut memiliki tara persentil *(percentile rank)* yang sepadan.

Menurut Braun dan Holland (1982) dan Lord (1950) penyetaraan ekipersentil didasarkan pada definisi bahwa skala skor untuk kedua tes sebanding dan distribusi skor ke dua tes identik (Linn, 1987:247). Cook dan Petersen menyatakan penyetaraan ekipersentil tidak seperti metode penyetaraan lainnya, tidak membutuhkan asumsi-asumsi terhadap tes-tes yang akan disetarakan (Naga, 1992:226). Dengan demikian metode penyetaraan ekipersentil mengasumsikan bahwa skor pada tes X dan Y adalah identik atau ekivalen, bertujuan agar distribusi dari skor tes X yang diubah sama dengan distribusi skor tes Y, dan tidak membutuhkan asumsi-asumsi seperti pada metode linear. Dengan demikian penelitian ini bertujuan untuk mengetahui (1) perbedaan variansi skor hasil penyetaraan antara metode menyetaraan linear dan metode penyetaraan ekipersentil untuk ukuran sampel 200, (2) perbedaan variansi skor hasil penyetaraan antara metode menyetaraan linear dan metode penyetaraan ekipersentil untuk ukuran sampel 400, dan (3) perbedaan variansi skor hasil penyetaraan antara metode menyetaraan linear dan metode penyetaraan ekipersentil untuk ukuran sampel 800.

Metode Penelitian

Penelitian ini menggunakan simulasi dengan dua variabel yang diteliti mengggunakan pola jawaban peserta tes diperoleh melalui pelaksanaan UASBN SD/MI negeri dan swasta tahun pelajaran 2008/2009 mata pelajaran IPA. Variabel bebas dalam penelitian ini adalah metode penyetaraan dan jumlah responden. Metode penyetaraan pada penelitian ini adalah metode penyetaraan linear dan metode penyetaraan ekipersentil, sedangkan variabel jumlah responden berturut-turut adalah 200, 400, dan 800. Variabel terikatnya adalah variansi skor hasil penyetaraan.

Populasi peserta tes adalah peserta tes UASBN tahun pelajaran 2008/2009 mata pelajaran IPA se-Jakarta Timur. Populasi tersebut sebanyak 44.401 siswa yang terbagi menjadi dua bagian yaitu sebanyak 22.201 siswa yang mengerjakan Paket 01 dan sebanyak 22.200 siswa yang mengerjakan Paket 02. Pengambilan sampel menggunakan metode acak berulang dengan pengembalian *(random sampling with replacement),* yaitu cara pengambilan sampel secara acak dengan jumlah tertentu yang dikembalikan lagi menjadi populasi untuk mendapat peluang menjadi sampel berikutnya, sehingga sampel tersebut mempunyai peluang untuk dipilih kembali pada saat pengambilan sampel berikutnya.

Data diperoleh melalui Pusat Penilaian Pendidikan (Puspendik) Badan Penelitian dan Pengembangan (Balitbang) Kementerian Pendidikan Nasional. Teknik analisis data menggunakan uji kesamaan variansi dengan analisis Uji F. Analisis data menggunakan bantuan perangkat lunak MINITAB, *ITEMAN for Windows Version 3.50d*, SPSS, dan Excel. Hipotesis statistik dalam penelitian ini terdiri dari tiga hipotesis nol dan tiga hipotesis alternatif, yaitu (1) H_o: $\frac{\sigma^2_{eks200}}{\sigma^2_{lin200}} = 1$ H_1: $\frac{\sigma^2_{ekp200}}{\sigma^2_{lin200}} > 1$; (2) H_o: $\frac{\sigma^2_{ekp400}}{\sigma^2_{lin400}} = 1$ H_1: $\frac{\sigma^2_{ekp400}}{\sigma^2_{lin400}} > 1$; dan (3) H_o: $\frac{\sigma^2_{ekp800}}{\sigma^2_{lin800}} = 1$ H_1: $\frac{\sigma^2_{ekp800}}{\sigma^2_{lin800}} > 1$.

Hasil Penelitian dan Pembahasan

Analisis variansi skor hasil penyetaraan dilakukan terhadap masing-masing cuplikan dengan ukuran sampel 200, 400, dan 800 baik untuk Paket 01 maupun Paket 02, masing-masing sebanyak 41 kali cuplikan. Jadi, seluruhnya terdapat 246 analisis dan menghasilkan variansi skor hasil penyetaraan dengan metode linear dan metode ekipersentil seperti disajikan pada Tabel 1 dan Tabel 2.

Tabel 1. Variansi Skor Hasil Penyetaraan Metode Linear

Cuplikan Ke	Variansi Skor			Cuplikan Ke	Variansi Skor		
	s^2_{lin200}	s^2_{lin400}	s^2_{lin800}		s^2_{lin200}	s^2_{lin400}	s^2_{lin800}
1	25,064	28,625	25,983	22	24,495	25,035	27,130
2	25,237	26,277	24,257	23	29,022	23,641	25,383
3	21,936	27,998	27,538	24	25,459	23,062	23,032
4	24,242	24,417	25,349	25	24,874	25,579	23,420
5	20,485	25,177	23,256	26	27,721	26,519	25,128
6	23,112	28,163	24,750	27	27,420	22,658	25,284
7	26,317	23,150	23,665	28	23,673	22,169	24,004
8	24,373	27,485	26,838	29	25,249	25,525	22,833
9	25,285	24,518	25,575	30	24,257	24,460	25,386
10	24,644	28,006	25,055	31	22,896	23,741	25,969
11	27,669	24,281	23,592	32	23,406	26,971	25,015
12	25,250	24,867	24,373	33	20,842	22,833	25,364
13	28,172	24,663	25,480	34	26,415	21,903	24,816
14	22,533	22,732	25,685	35	28,342	28,106	24,988
15	26,620	29,690	26,312	36	28,256	25,483	26,331
16	27,942	23,645	23,761	37	24,354	21,502	25,544
17	24,316	23,972	23,763	38	24,883	23,815	26,394
18	21,786	22,707	23,286	39	28,987	24,498	25,297
19	22,232	24,167	24,185	40	28,083	23,828	24,764
20	24,573	25,782	25,393	41	21,455	24,977	23,709
21	24,322	22,560	25,268	Rerata	25,119	24,855	24,986
				Simpangan Baku	2,240	2,042	1,139

Tabel 2. Variansi Skor Hasil Penyetaraan Metode Ekipersentil

Cuplikan Ke	Variansi Skor			Cuplikan Ke	Variansi Skor		
	s^2_{ekp200}	s^2_{ekp400}	s^2_{ekp800}		s^2_{ekp200}	s^2_{ekp400}	s^2_{ekp800}
1	28,072	28,847	30,449	22	27,466	29,408	30,367
2	31,286	29,049	30,633	23	32,481	32,132	27,923
3	29,072	29,050	27,034	24	32,107	26,371	26,819
4	31,865	30,475	27,476	25	29,591	29,336	29,125
5	30,125	33,465	27,098	26	29,955	28,635	28,883
6	28,981	27,222	28,470	27	29,788	29,758	29,276
7	28,961	28,658	28,449	28	25,413	28,185	29,439
8	30,842	29,248	29,516	29	29,373	29,632	31,484
9	27,022	27,818	28,715	30	27,373	27,639	30,363
10	25,874	30,538	26,344	31	28,647	27,684	31,088
11	30,535	30,652	29,363	32	28,353	31,567	29,976
12	28,472	27,727	29,382	33	28,397	30,558	26,645
13	33,138	29,510	30,830	34	29,180	28,994	26,152
14	34,047	32,860	31,069	35	32,580	28,738	29,924
15	29,739	29,425	30,797	36	27,163	29,724	30,557
16	30,956	27,411	27,635	37	28,351	28,845	31,057
17	28,215	33,959	31,499	38	31,454	31,665	31,261
18	24,765	28,215	28,161	39	31,618	31,967	27,015
19	30,952	29,194	29,155	40	31,408	33,618	29,165
20	30,391	29,815	27,205	41	31,348	30,207	27,336
21	31,790	28,087	30,927	Rerata	29,645	29,642	29,168
				Simpangan Baku	2,122	1,832	1,601

Sebelum teknik analisis statistik uji kesamaan variansi digunakan, terlebih dahulu perlu diperiksa apakah data penelitian telah memenuhi persyaratan. Uji persyaratan analisis yang digunakan adalah uji normalitas Kolmogorov-Smirnov, rangkuman hasil uji normalitas dapat dilihat pada Tabel 3. Berdasarkan Tabel 3 dapat dinyatakan bahwa keenam data variansi skor penyetaraan berasal dari populasi yang berdistribusi normal dengan taraf signifikasi 5%.

Tabel 3. Hasil Uji Normalitas Varians Skor Penyetaraan

No	Varians	Uji Normalitas Kolmogorov-Smirnov			
		Statistik	db	Sgn.	Sebutan
1	s^2_{lin200}	0,114	41	0,200	Normal
2	s^2_{lin400}	0,089	41	0,200	Normal
3	s^2_{lin800}	0,073	41	0,200	Normal
4	s^2_{ekp200}	0,071	41	0,200	Normal
5	s^2_{ekp400}	0,108	41	0,199	Normal
6	s^2_{ekp800}	0,121	41	0,136	Normal

Ada tiga hipotesis yang diuji dalam penelitian ini, yaitu hipotesis yang membandingkan variansi skor penyetaraan s^2_{ekp200} dengan $s^2_{lin200,}$ variansi skor penyetaraan s^2_{ekp400} dengan s^2_{lin400}, dan variansi skor penyetaraan s^2_{ekp800} dengan s^2_{lin800}. Setelah dilakukan analisis terhadap data variansi skor penyetaraan diperoleh hasil yang selengkapnya dapat dilihat pada Tabel 4.

Tabel 4. Tabel 4. Statistik Variansi Skor Penyetaraan

No	Variansi	Statistik		
		Rerata	SD	Variansi
1	s^2_{ekp200}	29,645	2,122	4,460
2	s^2_{lin200}	25,119	2,240	5,218
3	s^2_{ekp400}	29,642	1,832	3,281
4	s^2_{lin400}	24,855	2,042	4,065
5	s^2_{ekp800}	29,168	1,601	2,582
6	s^2_{lin800}	24,986	1,139	1,305

Hipotesis pertama dinyatakan bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp200}) lebih besar daripada metode

menyetaraan linear (s^2_{lin200}) untuk ukuran sampel 200. Hasil analisis yang terdapat pada Tabel 4 menunjukkan bahwa rerata variansi skor hasil penyetaraan yang dihasilkan dari penyetaraan menggunakan metode ekipersentil (s^2_{ekp200}) adalah 4,460 dan metode linear (s^2_{lin200}) sebesar 5,218. Untuk memastikan apakah kedua variansi skor ini berbeda secara signifikan atau tidak, maka dilakukan analisis uji-F sebagai berikut: $F_{h1} = s^2_{ekp200} / s^2_{lin200}$ = 4,460/5,218= 0,855.

Harga F tabel untuk taraf signifikansi (α) = 0,05 dengan pembilang (n_1 - 1) = 40 dan penyebut (n_2 - 1) = 40 sebesar 1,684. Angka F_{h1} yang diperoleh lebih kecil dibandingkan dengan F tabel. Dengan demikian hipotesis nol diterima. Hal ini berarti bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp200}) tidak berbeda secara signifikan dengan variansi skor hasil penyetaraan menggunakan metode linear (s^2_{lin200}) untuk ukuran sampel 200.

Hipotesis kedua dinyatakan bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp400}) lebih besar daripada metode menyetaraan linear (s^2_{lin400}) untuk ukuran sampel 400. Hasil analisis yang terdapat pada Tabel 4 menunjukkan bahwa rerata variansi skor hasil penyetaraan yang dihasilkan dari penyetaraan menggunakan metode ekipersentil (s^2_{ekp400}) adalah 3,281 dan metode linear (s^2_{lin400}) sebesar 4,065. Untuk memastikan apakah kedua variansi skor ini berbeda secara signifikan atau tidak, maka dilakukan analisis uji-F sebagai berikut: $F_{h2} = s^2_{ekp400} / s^2_{lin400}$ = 3,281/4,065 = 0,807.

Harga F tabel untuk taraf signifikansi (α) = 0,05 dengan pembilang (n_1 - 1) = 40 dan penyebut (n_2 - 1) = 40 sebesar 1,684. Angka F_{h2} yang diperoleh lebih kecil dibandingkan dengan F tabel. Dengan demikian hipotesis nol diterima. Artinya, variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp400}) tidak berbeda secara signifikan dengan variansi skor hasil penyetaraan menggunakan metode linear (s^2_{lin400}) untuk ukuran sampel 400.

Hipotesis ketiga dinyatakan bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp800}) lebih besar daripada metode menyetaraan linear (s^2_{lin800}) untuk ukuran sampel 800. Hasil analisis yang terdapat pada Tabel 4 menunjukkan bahwa rerata variansi skor hasil

penyetaraan yang dihasilkan dari penyetaraan menggunakan metode ekipersentil (s^2_{ekp800}) adalah 2,582 dan metode linear (s^2_{lin800}) sebesar 1,305. Untuk memastikan apakah kedua variansi skor ini berbeda secara signifikan atau tidak, maka dilakukan analisis uji-F sebagai berikut: $F_{h3} = s^2_{ekp800} / s^2_{lin800} = 2{,}582/1{,}305 = 1{,}979$.

Harga F tabel untuk taraf signifikansi (α) = 0,05 dengan pembilang (n_1-1) = 40 dan penyebut (n_2 - 1) = 40 sebesar 1,684. Angka F_{h3} yang diperoleh lebih besar dibandingkan dengan F tabel. Jadi, hipotesis nol ditolak. Artinya, variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp800}) berbeda secara signifikan dengan variansi skor hasil penyetaraan menggunakan metode linear (s^2_{lin800}) untuk ukuran sampel 800.

Berdasarkan hasil pengujian hipotesis pertama dan kedua di atas dapat dinyatakan bahwa hipotesis nol diterima untuk ukuran sampel 200 dan 400. Artinya tidak ada perbedaan variansi antara metode linear dan ekipersentil dengan ukuran sampel 200 dan 400. Kenyataan data dalam menerima hipotesis nol dapat disebabkan oleh berbagai kemungkinan, antara lain adalah asumsi penggunaan penyetaraan metode linear tidak terpenuhi. Hal ini sesuai dengan pendapat Hambleton dan Rogers bahwa asumsi penyetaraan dengan cara linear adalah kedua skor tes distribusinya berbeda, distribusi tersebut terkait dengan rerata dan simpangan bakunya. Jika terdapat kasus yang demikian skor kedua kelompok akan sama. Sebaliknya bila asumsi tersebut benar, maka penyetaraan linear menjadi penyetaraan ekipersentil (Hambleton, Swaminathan, dan Rogers, 1991:124-125). Dengan demikian distribusi skor kedua kelompok pada penelitian mempunyai distribusi yang tidak berbeda, sehingga penyetaraan linear sama dengan penyetaraan ekipersentil.

Pengujian hipotesis ke tiga dengan ukuran sampel 800 menunjukkan hasil yang berbeda, yaitu menolak hipotesis nol. Artinya ada perbedaan variansi antara metode linear dan ekipersentil dengan ukuran sampel 800. Ini dipengaruhi oleh variansi metode linear yang makin kecil untuk ukuran sampel 800. Makin kecilnya variansi metode linear dengan ukuran sampel yang makin besar sesuai dengan penelitian yang dilakukan Hanson *et. al.* (1994) bahwa rerata kuadrat kesalahan penyetaraan *(mean-squared equating*

error) makin kecil untuk ukuran sampel yang makin besar (Kolen dan Brennan, 1995:102).

Simpulan

Sesuai dengan hasil penelitian dapat dikemukakan beberapa simpulan sebagai berikut: (1) variansi skor hasil penyetaraan menggunakan metode linear (σ^2_{lin200}) terbukti memiliki variansi yang sama dengan variansi skor hasil penyetaraan menggunakan metode ekipersentil (σ^2_{ekp200}) dengan ukuran sampel 200. Artinya, tidak ada perbedaan variansi skor penyetaraan menggunakan metode linear maupun metode ekipersentil untuk ukuran sampel 200, (2) variansi skor hasil penyetaraan menggunakan metode linear (σ^2_{lin400}) terbukti memiliki variansi yang sama dengan variansi skor hasil penyetaraan menggunakan metode ekipersentil (σ^2_{ekp400}) dengan ukuran sampel 400. Artinya, tidak ada perbedaan variansi skor penyetaraan menggunakan metode linear maupun metode ekipersentil untuk ukuran sampel 400, dan (3) variansi skor hasil penyetaraan menggunakan metode linear (σ^2_{lin800}) terbukti memiliki variansi yang berbeda dengan variansi skor hasil penyetaraan menggunakan metode ekipersentil (σ^2_{ekp800}) dengan ukuran sampel 800. Dengan demikian terdapat perbedaan variansi skor penyetaraan menggunakan metode linear dan metode ekipersentil untuk ukuran sampel 800.

Daftar Pustaka

Aiken, Lewis R. 1997. *Psychological Testing and Assemesment.* Boston: Allyn & Bacon.

Brown, Frederick G. 1976. *Principles of Educational and Psychologycal Testing.* New York: Holt, Rinehart and Wiston.

Croker, Linda dan James Algina. 1986. *Introduction to Classical & Modern Test Theory.* New York: Rinehart and Winston Inc.

Departemen Pendidikan Nasional. Peraturan Pemerintah Republik Indonesia Nomor 19 tahun 2005 tentang Standar Nasional Pendidikan. Jakarta: Badan Standar Nasional Pendidikan.

_____. Peraturan Pemerintah Republik Indonesia Nomor 20 tahun 2007 tentang Standar Penilaian Pendidikan. Jakarta: Badan Standar Nasional Pendidikan.

_____. Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 39 tahun 2007 tentang Ujian Akhir sekolah Berstandar Nasional (UASBN) untuk sekolah Dasar/Madrasah Ibtidaiyah/Sekolah Luar Biasa (SD/MI/SDLB) tahun Pelajaran 2007/2008. Jakarta: Badan Standar Nasional Pendidikan.

Eid, Ghada K. "*Effects of Sample Size in the Equating of Test Items*" http://findarticles.com/p/articles/mi_qa3673/is_200510/ai_n15641924 (diakses tanggal 12 Oktober 2008).

Hambleton, Ronald K. dan Hariharan Swaminathan. 1985. *Item Response Theory: Principles and Application.* Boston: Kluwer.

_____. 1990. *Item Respon Theory Principles and Applications.* Boston: Kluwer Nijhoff Publishing.

Hambleton, Ronald K., Hariharan Swaminathan, dan H. Jane Rogers. 1991. *Fundamentals of Item Respons Theory*. California: SAGE Publications, Inc.

Kaplan, Robert M. dan Dennis P. Saccuzzu. 1982. *Psychological Testing, Principle, Application and Issues.* California: Wadsworth, Inc.

Keeves, John (ed). 1997. *An International Hand Book of Measurement*. Oxford: Pergamon.

Kolen, Michael J. dan Robert L. Brennan. 1995. *Test Equating Methods and Practices.* New York: Springer.

Linn, Robert L. 1987. *Educational Measurement*. New York: Macmillan Publishing Company.

Livingston, S. A. dan N. J. Feryok. 1987. "Univariate Versus Bivariate Smooting in Frewuency Evuating" *Laporan Riset* No. 87-36. New Jersey: Educational Testing Service.

Lord, Frederick M. dan Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores.* Massachusetts: Addison-Wesley Publishing Company, Inc.

Marco, Gory L. *et al.* 1983. *A Test of The Adequacy of Curvilinear Score Equating Models* (New Horizon in Testing). Weiss David J. New York: Academic Press.

Naga, Dali S. 1992. *Pengantar Teori Sekor.* Jakarta: Besbats.

_____. 2008. *Probabilitas dan Sekor pada Hipotesis Statistika.* Jakarta: Universitas Tarumanegara.

Naiman, Arnold, Robert Rosenfeld, dan Gene Zirkel. 1993. *Understanding Statistic.* Singapore: McGraw-Hill.

Setiadi, Hari. 1997. Small Sample IRT Item Parameter Estimates. *Disertasi* tidak dipublikasikan.

Steel, Robert G. D. dan James H. Torrie. 1980. *Principles and Procedur of Statistic.* Singapore: Mc.Graw-Hill.

Thordike, Robert M. *et al.* 1991. *Measurement and Evaluation in Psychology and Education.* New York: Macmillan Publishing Company.

Wiersma, William dan Stephen G. Jurs. 1990. *Educational Measurement and Testing.* Boston: Allyn and Bacon.

_____. 2005. *Research Methods in Education.* New York: Pearson Education Inc.

Yahya Umar. 1987. Robusness of the Simple Linking Procedure i Item Banking Using the Rasch Model, *Dissertation*, University of California, p. 105.

Yetti Supriati. 2003. Pengaruh Proporsi *Anchor Items* terhadap Kestabilan Koefisien Penyetaraan Sekor pada Evaluasi Belajar. *Disertasi* tidak dipublikasikan. Jakarta: Pascasarjana Universitas Negeri Jakarta.