**PENGARUH UKURAN SAMPEL TERHADAP HASIL PENYETARAAN TES**

**UJIAN NASIONAL MATEMATIKA SMP BERDASARKAN TEORI RESPON BUTIR**

**Oleh : Nuralang**

**SMP Negeri 1 Pallangga**

**E-mail : nuralang5@gmail.com**

*Abstrak*

*Penelitian ini bertujuan untuk mengetahui perbedaan hasil penyetaraan tes berdasarkan perbedaan ukuran sampel dan untuk mengetahui prosedur mendapatkan penyetaraan tes berdasarkan equateIRT. Jenis penelitian ini adalah exploratif, yaitu akan mengungkap kesetaraan skor tes berdasarkan teori respons-butir. Dalam penelitian ini yang diungkap adalah pengaruh ukuran sampel terhadap hasil penyetaraan tes. Subjek dalam penelitian ini adalah butir perangkat soal dan seluruh lembar jawaban peserta tes di Kabupaten Gowa. Jumlah subjek adalah 2600 siswa.*

*Prosedur penyetaraan tes, tahap pertama pemilihan jenis penyetaraan yaitu penyetaraan horizontal, kedua pemilihan desain penyetaraan Common-Item Nonequivalent Group, ketiga dilanjutkan dengan estimasi parameter dengan menggunakan model 2PL dan keempat dilakukan proses penyetaraan dengan menggunakan equateIRT.*

*Berdasarkan hasil penyetaraan tes antara paket 01 dan paket 02 (ukuran sampel 200, 400, dan 700 ) dengan program R equateIRT menyatakan bahwa terdapat perbedaan hasil penyetaraan tes antara ukuran sampel 200 dan 400 dimana paket soal dengan ukuran sampel 400 menghasilkan penyetaraan yang lebih akurat. Terdapat perbedaan hasil penyetaraan tes antara ukuran sampel 200 dan 700, dimana paket soal dengan ukuran sampel 700 menghasilkan penyetaraan yang lebih akurat. Terdapat perbedaan hasil penyetaraan antara ukuran sampel 400 dan 700, dimana paket soal ukuran sampel 700 menghasilkan penyetaraan yang lebih akurat. Hal ini menunjukkan bahwa ukran sampel 700 menghasilkan penyetaraan yang lebih akurat dibandingkan ukuran sampel 200 dan 400.*

*Kata kunci : Teori Respon Butir, Penyetaraan Tes, Ukuran Sampel*

***ABSTRACT***

*This study aims to determine differences equivalency test results based on the differences in sample size based equateIRT. This type of research is explorative, with descriptive approach that will reveal the equality of test scores by-item response theory. This study would reveal the effect of sample size on the results of equivalency test. Subjects in this study is about the item and all the answer sheets of students of junior high school mathematics examination test results in Gowa. Subject total is 2600 student, with the number of men in 1167 and the number of women in 1433.*

*Procedure of equivalency test, the first stage is the equating type selection of horizontal equating; the second is the design selection of equating Common-Item Nonequivalent Group; followed by the third parameter estimation using a model 2PL; and the fourth is equating process which is carried out by using equateIRT.*

*Based on the results of equating between the package 01 and the package 02 with a sample size of 200, equating between the package 01 and the package 02 with a sample size of 400, and the equating of packages 01 and packages 02 with a sample size of 700; shows that equating coefficients α and β is generated by the package obout with a sample size of 700 α = 1, and β = 0. Theoretically standard errors generated by a package by the sample size 700 smaller than the package of a sample size of 200 and 400, this means that a package of the sample size 700 produces more accurate equating*

*Keyword: Item Response Theory, Equating Test, Sample Size*

Ujian nasional yang dilaksanakan terdiri atas beberapa paket soal yang menjadi masalah apakah paket-paket soal tersebut dapat mengukur secara obyektif terhadap peserta didik. Agar peserta didik yang satu dengan peserta didik yang lainnya tidak ada yang dirugikan atau diuntungkan dari paket soal yang berbeda maka dalam evaluasi pembelajaran aspek keadilan merupakan salah satu prinsip yang penting, artinya peserta didik yang menghadapi ujian nasional di berbagai wilayah Indonesia dan waktu yang berbeda harusnya mendapatkan perlakuan yang adil. Dengan demikian apakah soal ujian nasional mempunyai tingkat keseteraan yang sama antara paket yang satu dengan yang lainnya.

Menurut Suryabrata (1987: 132) bahwa “Dalam pelaksanaan evaluasi yang menggunakan beberapa perangkat tes yang

berbeda dan mengukur hal yang sama perlu dilakukan penyetaraan dari beberapa perangkat tes tersebut, karena dengan adanya penyetaraan dapat dijamin keadilan bagi peserta tes”. Agar peserta didik mendapatkan penilaian yang adil dalam menghadapi soal tes yang berbeda paket. Menurut Miyatun & Mardapi (2000: 2-3) bahwa “tingkat kesetaraan perangkat tes yang berbeda akan dapat diketahui melalui proses penyetaraan”. Selanjutnya Hambleton & Swaminathan (1985) dalam Salasi (1996) mengingatkan bahwa sekalipun perangkat tes yang disusun berdasarkan kisi-kisi yang sama, jarang sekali bahkan hampir tidak ditemukan perangkat tes yang benar-benar setara dalam sebarang serta peringkat indeks kesukaran.

Berdasarkan permasalahan tersebut maka peneliti akan mengadakan penelitian tentang penyetaraan tes. Dengan harapan soal yang setara bisa dijadikan sebagai bahan tes, misalnya dalam pelaksanaan ulangan harian, sehingga tingkat kesadaran untuk menyontek tidak terjadi lagi. Karena peserta didik tidak merasa ada diskriminan dalam menyelesaikan tes yang berbeda. Hal yang berkaitan dengan penyetaraan seperti panjang tes, karakteristik dari sekumpulan anchor item, ukuransampel minimal, variasi kemampuan peserta tes, metode penyetaraan, kesalahan penyetaraan sangat penting untuk diteliti dalam kerangka teoritis dan keperluan pengukuran di Indonesia.

Selanjutnya Syaifuddin (2005: 10) menggunakan ukuran sampel 300, 500, dan 800 menyatakan bahwa “ukuran sampel memiliki pengaruh langsung pada keakuratan estimasi parameter item tes, untuk itu sampel representatif dan besar bisa diharapkan memberikan hasil penyetaraan yang lebih akurat”. Pada penelitian ini akan mengkaji tentang pengaruh ukuran sampel terhadap hasil penyetaraan tes ujian nasioanl matematika SMP berdasarkan teori respon butir

Berdasarkan beberapa penelitian tersebut maka penelitian ini menggunakan ukuran sampel 200, 400, dan 700 sebagai perbandingan apakah ukuran sampel kecil (200), sudah nampak hasil estimasi penyetaraan, dan sampel 400 lebih nampak dari hasil estimasi penyetaraan serta ukuran sampel besar (700) akan lebih nampak lagi hasil penyetaraan dua perangkat tes.

Berdasarkan latar belakang permasalahan di atas, maka rumusan masalah dalam penelitain ini adalah (1) Apakah terdapat perbedaan hasil penyetaraan tes ujian nasional matematika berdasarakan ukuran sampel 200 dan 400 ?

(2) Apakah terdapat perbedaan hasil penyetaraan tes ujian nasional matematika berdasarkan ukuran sampel 200 dan 700 ? (3) Apakah terdapat perbedaan hasil penyetaraan tes ujian nasional matematika berdasarkan ukuran sampel 400 dan 700 ? (4) Bagaimana prosedur untuk mendapatkan hasil penyetaraan tes berdasarkan metode *equateIRT ?*

Tujuan penelitian ini adalah untuk (1) Mengetahui perbedaan hasil penyetaraan tes berdasarkan ukuran sampel 200 dan 400.

(2) Mengetahui perbedaan hasil penyetaraan tes berdasarkan ukuran sampel 200 dan 700.(3) Mengetahui perbedaan hasil penyetaraan tes berdasrakan ukuran sampel 400 dan 700. Mengetahui prosedur mendapatkan hasil penyetaraan tes berdasrakan metode *equateIRT.*

Teori tes modern yang berkembang saat ini untuk teori pengukuran adalah teori respon butir. Asumsi utama teori ini adalah peluang peserta tes menjawab benar suatu butir tidak ditentukan oleh peluang menjawab benar butir yang lain, dikenal dengan asumsi independen. Muncul teori respon butir karena teori tes klasik memiliki berbagai keterbatasan seperti adanya sifat group dependent dan item dependent maka munculnya teori respon butir menjadi sangat berguna dan terus dikembangkan karena mampu mengatasi keterbatasan tersebut.

Selanjutnya teori respon butir juga dikembangkan atas dasar dua postulat yaitu: 1) performansi subyek pada suatu butir dapat diprediksikan oleh seperangkat faktor yang disebut yang disebut *latent trait* atau kemampuan dan 2) hubungan performansi subyek pada suatu butir dan perangkat kemampuan laten yang mendasarinya digambarkan oleh fungsi naik monoton yang disebut *Item* *Charascteristic Curve* (ICC), selain itu bahwa asumsi- asumsi yang melandasi teori respon butir adalah unidimensi, independensi lokal, dan fungsi karakteristik butir atau kurva karakteristik butir. (Hambleton dkk, 1991: 9)

Ujian Nasional pada dasarnya bertujuan untuk motivasi peningkatan kompetensi guru, motivasi guru agar melaksanakan pembelajaran dengan baik dan menggunakan sistem asesmen yang tepat. “Apabila kompetensi minimal guru telah tercapai proses pembelajaran dan asesmen di kelas telah dilaksanakan sesuai dengan standar, maka tidak ada kehawatiran terhadap Ujian Nasional” (Mardapi, 2012: 230). Penilaian pada jenjang Sekolah Menengah Pertama/Madrasa sederajat pada tahun ajaran 2014/2015 tidak hanya dilakukan oleh guru pada sekolah yang bersangkutan saja. Penilaian juga dilakukan oleh pihak independen yang ditunjuk oleh pemerintah yaitu Badan Standar Nasional Pendidikan (BSNP). BSNP menyelenggarakan penilaian yaitu Ujian Nasional. UN diberlakukan pada empat mata pelajaran, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, Ilmu Pengetahuan Alam (IPA).

Kegiatan penyataraan tes dilakukan dengan mengembangkan konversi suatu system unit tes yang lain sehingga setelah dikonversi skor yang berasal dari dua atau lebih perangkat tes menjadi setara dan dapat dipertukarkan. Menurut Dorans, Moses & Eignor (2010: 3) “Penyetaraan adalah bentuk kuat untuk menghubungkan antara skor pada dua tes. Tujuan penyetaraan adalah untuk menghasilkan skor pada dua bentuk tes, sehingga skor dari setiap tes dapat diperbandingkan dari pengujian yang sama”. Dengan demikian, menjadi suatu keharusan bagi pengembang tes atau lembaga tes untuk menyetarakan perangkat tes tersebut. Menurut Croker & Algina (1986) dalam Naga (1992: 346) dua skor satu di alat ukur X dan lainnya di alat ukur Y mengukur ciri yang sama dengan reliabilitas yang sama serta peringkat persentil yang sepadan dengan skor itu adalah sama. Sedangkan Hambleton & Swaminathan (1985: 123) menyatakan “bahwa penyetaraan merupakan prosedur statistik untuk menetapkan hubungan antara skor dari dua tes atau lebih”.

Sampel menurut Steel &Torrie (1980) dalam Rijanto (2012) adalah “bagian dari populasi, kadang-kadang mencakup seluruh populasi dan umumnya informasi dari sampel digunakan untuk penarikan kesimpulan tentang populasi itu”. Ukuran sampel merupakan salah satu faktor yang mempengaruhi penyetaraan, keakuratan estimasi parameter item tes, demikian juga pada tingkat kesalahan penyetaraan. Semakin besar ukuran sampel, kesalahan penyetaraan semakin kecil. Kolen dan Brennan (1995: 264) telah menunjukkan bahwa meningkatnya ukuran sampel, *standard error of equating*  *(SEE)* yang dihasilkan semakin kecil. Untuk itu sampel yang representatif dan besar bisa diharapkan memberikan hasil penyetaraan yang lebih akurat. Namun demikian ukuran sampel minimum yang diperlukan pada penyetaraan perlu diteliti.

Penelitian ini adalah jenis penelitian yang bersifat *eksploratif,* dengan pendekatan deskriptif berdasarkan teori respon butir yaitu mencari, mengunkap, menggali, mengolah data secara cermat dan lengkap fakta-fakta yang terkandung dalam suatu permasalahan yang bersifat spesifik. Dalam penelitian ini peneliti mengunkap karakteristik soal matematika Ujian Nasional SMP Tahun Pelajaran 2015/2016 model dua parameter, kesetaraan tes dengan metode *library equate IRT.*

Penelitian ini dilaksanakan di Kabupaten Gowa dengan Sembilan sekolah SMP yaitu: SMPN 1 Sungguminasa, SMPN 2 Sungguminasa, SMPN 3 Sungguminasa, SMPN 4 Sungguminasa, SMPN 1 Pallangga, SMPN 3 Pallangga, SMPN 5 Pallangga, SMPN 1 Barombong, dan SMPN 2 Barombong. Kegaitan penelitian dilaksanakan pada bulan Pebruari sampai bulan April tahun 2016.

Data penelitian ini adalah respon peserta didik ujian nasional SMP paket 01 dan paket 02 soal matematika tahun pelajaran 2015/2016. Sumber data berupa lembar jawaban peserta didik yang telah didokumentasikan.

Subyek dalam penelitian ini adalah butir soal dan seluruh lembar jawaban siswa dari hasil tes ujian Nasional matematika di Kabupaten Gowa. Pemilihan subjek dalam penelitian ini dengan menggunakan teknik *purposif sampling* berdasarkan tujuan. Artinya peneliti menentukan sendiri sampel yang diambil karena ada pertimbangan tertentu untuk memenuhi sampel yang dibutuhkan. Peneliti berupaya untuk memperoleh imformasi yang relevan dengan tujuan penelitian yaitu informasi tentang perbedaan hasil penyetaraan tes berdasarkan ukuran sampel 200, 400 dan 700 pada dua perangkat tes. Jumlah subyek 2600 peserta didik

Desain penyetaraan yang digunakan pada penelitian ini adalah desain *commom item non equivalent groups*. Paket soal ujian nasional SMP tahun pelajaran 2014/2015 yang terdiri dari 2 paket soal. Paket soal yang digunakan dalam penelitian ini adalah paket 01 dan paket 02. Jumlah total soal pada setiap paket 40 butir, terdiri dari 16 butir anchor 40% (butir Z) dan 24 butir soal yang berbeda pada masing-masing paket. Paket (01 + Z) dan paket (02 + Z) dengan ukuran sampel 200, 400 dan 700.

Dalam penelitian ini teknik Pengumpulan data yang dilakukan dengan menggunakan teknik dokumentasi dengan mengutip respon peserta didik dari hasil tes ujian nasional SMP pada mata pelajaran matematika di Kabupaten Gowa Tahun Pelajaran 2014/2015.

Teknik Analisis Data yang digunakan Proses Validasi Isi, Proses pengumpulan data Estimasi parameter butir dengan analisis R Program V.3.2.2*,*Proses penyetaraan,

**Proses Penyetaraan**

Berdasarkan hasil validasi isi instrumen soal ujian nasional matematika 2014/2015 dinyatakan valid dengan koefisien internal sebesar 0,95 soal paket 01 dan 0,93 soal paket 02. Hal ini berarti bahwa hasil penilaian kedua validator memiliki “relevansi yang kuat” . Ini sesuai dengan pendapat Ruslan (2009: 19), bahwa koefisien validasi isi tinggi (>75%) maka dapat dinyatakan pengukuran atau intervensi yang dilakukan adalah valid.

Hasil analisis parameter soal paket 01 dan paket 02 dapat dilihat dengan menggunakan teori respon butir model 2 parameter (2P), sehingga kualitas butir soal dilihat dari tingkat kesulitan, daya beda dan fungsi informasi butir. Analisis parameter soal paket 01 dan paket 02 berdasarkan *output* program R, dengan ukuran sampel 200, diketahui bahwa tingkat kesulitan paket 01 terdapat 6 butir dalam kategori sangat mudah, 7 butir dalam kategori mudah, 9 butir dalam kategori sedang, 9 butir dalam kategori sukar, dan 9 butir dalam kategori sangat sukar. Sedang tingkat kesulitan paket 02 terdapat 9 butir dalam kategori sangat mudah, 4 butir dalam kategori mudah, 16 butir dalam kategori sedang, 8 butir dalam kategori sukar dan 3 butir dalam kategori sangat sukar.

Analisis parameter soal paket 01 dan paket 02 berdasarkan *output* program R, dengan ukuran sampel 400, diketahui bahwa tingkat kesulitan paket 01 terdapat 9 butir dalam kategori sangat mudah, 1 butir dalam kategori mudah, 4 butir dalam kategori sedang, 12 butir dalam kategori sukar, dan 14 butir dalam kategori sangat sukar. Sedang tingkat kesulitan paket 02 terdapat 9 butir dalam kategori sangat mudah, 1 butir dalam kategori mudah, 3 butir dalam kategori sedang, 19 butir dalam kategori sukar dan 8 butir dalam kategori sangat sukar.

Analisis parameter soal paket 01 dan paket 02 berdasarkan *output* program R, dengan ukuran sampel 700, diketahui bahwa tingkat kesulitan paket 01 terdapat 9 butir dalam kategori sangat mudah, 1 butir dalam kategori mudah, 6 butir dalam kategori sedang, 19 butir dalam kategori sukar, dan 5 butir dalam kategori sangat sukar. Sedang tingkat kesulitan paket 02 terdapat 7 butir dalam kategori sangat mudah, 0 butir dalam kategori mudah, 5 butir dalam kategori sedang, 15 butir dalam kategori sukar dan 13 butir dalam kategori sangat sukar.

Analisis parameter soal paket 01 dan paket 02 berdasarkan *output* program R, dengan ukuran sampel 200, diketahui bahwa daya beda paket 01 terdapat 16 butir dalam kategori sangat baik, 8 butir dalam kategori baik, 12 butir dalam kategori buruk, dan 4 butir tidak terdapat dalam kategori. Sedang daya beda pada paket 02 terdapat 9 butir dalam kategori sangat baik, 15 butir dalam baik, 12 butir dalam kategori buruk, dan 4 butir tidak terdapat dalam kategori

Analisis parameter soal paket 01 dan paket 02 berdasarkan *output* program R, dengan ukuran sampel 400, diketahui bahwa daya beda paket 01 terdapat 29 butir dalam kategori sangat baik, 1 butir dalam kategori baik, dan 10 butir dalam kategori buruk. Sedang daya beda pada paket 02 terdapat 29 butir dalam kategori sangat baik, 1 butir dalam baik, dan 10 butir dalam kategori buruk.

Analisis parameter soal paket 01 dan paket 02 berdasarkan *output* program R, dengan ukuran sampel 700, diketahui bahwa daya beda paket 01 terdapat 19 butir dalam kategori sangat baik, 12 butir dalam kategori baik, dan 9 butir dalam kategori buruk. Sedang daya beda pada paket 02 terdapat 23 butir dalam kategori sangat baik, 11 butir dalam baik, dan 6 butir dalam kategori buruk.

Berdasarkan hasil analisis dapat dinyatakan bahwa kualitas butir yang baik berdasarkan tingkat kesulitan dengan ukuran sampel 200 pada paket 01 terdapat 26 butir, paket 02 terdapat 28 butir. Ukuran sampel 400 kualitas butir yang baik berdasarkan tingkat kesulitan pada paket 01 terdapat 17 butir, pada paket 02 terdapat 23 butir. Ukuran sampel 700 kualitas butir yang baik berdasarkan tingkat kesulitan pada paket 01 terdapat 26 butir, pada paket 02 terdapat 20 butir. Hal ini sesuai dengan pendapat (Hambleton & Swaminathan, 1985: 107). Suatu butir dikatakan baik jika nilai berkisar antara -2 dan +2. Nilai bi mendekati -2 maka indeks kesukaran butir sangat rendah sedang nilai bi mendekati +2 maka indeks kesukaran butir sangat tinggi untuk kelompok peserta tes.

Berdasarkan hasil analisis dapat dinyatakan bahwa kualitas butir yang baik berdasarkan daya beda dengan ukuran sampel 200 pada paket 01 terdapat 24 butir, paket 02 terdapat 24 butir. Ukuran sampel 400 kualitas butir yang baik berdasarkan daya beda pada paket 01 terdapat 30 butir, pada paket 02 terdapat 30 butir. Ukuran sampel 700 kualitas butir yang baik berdasarkan daya beda pada paket 01 terdapat 31 butir, pada paket 02 terdapat 34 butir. Hal ini sesuai dengan pendapat (Hambleton & Swaminathan, 1991: 15). Indeks daya beda yang baik berada diantara 0 - 2.



**Total Information: 83,142**

**Total Information: 46,187**



**Total Information : 19,1**

**Total Information : 18,03**



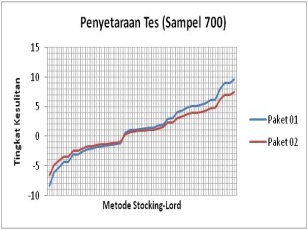
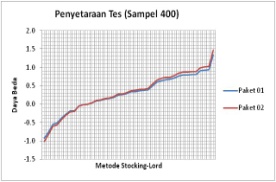
**TotalInformation: 28,653**

**Total Information: 22,991**

Dari hasil informasi butir kedua paket soal untuk sampel 200 menunjukkan bahwa total informasi butir paket 01 sebesar 83,142 sedang total informasi paket 02 sebesar 46,187. Hal ini menunjukkan bahwa paket 01 lebih besar memberikan kekuatan atau sumbangan butir tes dalam mengungkap kemampuan siswa dibandingkan dengan paket 02. Total informasi untuk sampel 400 menunjukan bahwa total informasi butir paket 01 sebesar 28,653 sedang total informasi paket 02 sebesar 22,991. Hal ini menunjukkan bahwa paket 01 lebih besar memberikan kekuatan atau sumbangan butir tes dalam mengunkap kemampuan siswa dibandingkan dengan paket 02. Total informasi butir kedua paket soal ukuran sampel 700 menunjukkan bahwa total informasi paket 01 sebesar 28,653 sedang total informasi paket 02 sebesar 22,991. Hal ini menunjukkan bahwa paket 01 lebih besar memberikan kekuatan atau sumbangan butir tes dalam mengungkap kemampuan siswa dibandingkan dengan paket 02. Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes (Hambleton & Swaminathan, 1985: 94).

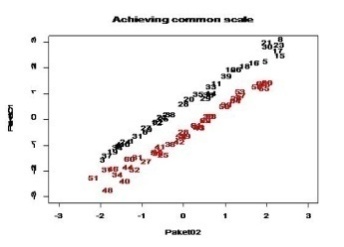
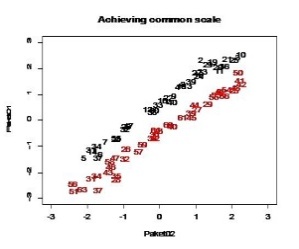
Prosedur dalam penyetaraan tes, tahap pertama pemilihan jenis penyetaraan adapun jenis penyetaraan yang digunakan adalah penyetaraan horizontal yaitu penyetaraan yang dilakukan terhadap jenjang pendidikan yang sama (Kelas IX SMP). Tahap kedua desain penyetaraan, yang digunakan adalah desain *Common-Item Nonequivalent*. Desian ini merupakan desain yang menggunakan dua kelompok peserta tes yang berbeda dan dua perangkat tes yang berbeda yaitu paket 01 dan paket 02, kedua paket tersebut berisi kumpulan butir bersama yang disebut dengan *anchor item,* jumlah *anchor item* yang digunakan adalah 16 butir(3,5,9,10,12,15,17,18,21,22,24,26,28,31,32,38) atau 40% dari 40 butir pada setiap paket soal dan setiap ukuran sampel. Selanjutnya tahap ketiga dilakukan estimasi parameter model dua parameter dengan menggunakan program R (*library ltm*) dengan melihat tingkat kesulitan dan daya beda paket 01 dan paket 02 pada ukuran sampel 200, 400, dan 700. Kemudian tahap keempat, tahap terakhir yaitu proses *equating* atau penyetaraan. Proses penyetaraan dengan cara pemilihan program *exel*  (*input* data), kemudian memilih *library ltm* di program R dilanjutkan dengan *equate IRT*, hasil anlisis dari *equating* ini akan menunjukkan tingkat kesulitan, daya beda dengan butir anchor pada masing-masing paket soal pada setiap ukuran sampel.

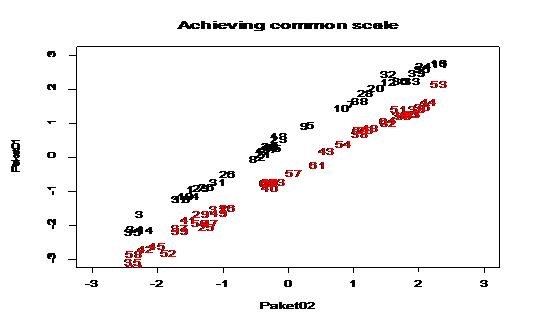
Dari gambar di atas, diperlihatkan hasil penyetaraan tes paket 01 dengan paket 02 dengan ukuran sampel 200. Paket 01 dijelaskan dengan garis biru sedangkan paket 02 dijelaskan dengan garis merah. Tampak garis merah (paket 02) berada diatas garis biru (paket 01), hal ini menunjukkan bahwa paket soal dengan tingkat kesulitan rendah akan berada dibawah nilai kriteria karena proses penyetaraan yang dilakukan dari paket soal yang sulit ke paket soal yang mudah. Sebaliknya paket soal dengan tingkat kesulitan tinggi akan berada diatas nilai yang menjadi kriteria, seperti nampak pada gambar diatas artinya proses penyetaraan yang dilakukan dari paket soal yang mudah ke paket soal sukar. Berdasarkan koefisen penyetaraan tergambar bahwa kedua paket tersebut memiliki tingkat kesulitan yang berbeda. Paket 02 lebih sulit dibandingkan dengan paket 01.





Berdasarkan grafik di atas, diperlihatkan hasil penyetaraan daya beda antara paket 01 dengan paket 02 dengan jumlah sampel 200. Garis biru ditunjukkan dengan paket 01 sedangkan paket 02 dtunjukkan dengan garis merah. Paket 01 lebih mengarah ke arah positif atau mendekati angka nol dibandingkan dengan paket 02. Hal ini menunjukkan bahwa paket 01 lebih mampu dalam hal membedakan antara peserta tes kemampuan tinggi dengan kemampuan rendah, dibandingkan dengan paket 02.





Selanjutnya dilihat dari grafik penyamaan skala paket 01 dan paket 02, untuk sampel 200 dapat dijelaskan bahwa, kemampuan 0 di paket 01 setara dengan kemampuan -0,8 pada paket 02, kemampuan 1 di paket 02 setara dengan kemampuan 0,5 di paket 01, kemampuan 2 pada paket 01 setara dengan kemampuan 1 pada paket 02. Grafik penyamaan skala sampel 400 dapat dijelaskan bahwa kemampuan 1 pada paket 01 setara dengan 0,7 pada paket 02, kemampuan -1 pada paket 01 setara dengan kemampuan -1,5 pada paket 02, kemampuan 2 pada paket 02 setara dengan kemampuan 1,5 pada paket 01. Grafik penyamaan skala sampel 700 dapat dijelaskan bahwa kemampuan 1 pada paket 01 setara dengan kemampuan 0,5 pada paket 02, kemampuan 2 pada paket 01 setara dengan kemampuan 1,5 pada paket 02, kemampuan 1 pada paket 02 setara dengan kemmapuan 0,5 pada paket 01.

Dalam hal proses penyetaraan berdasarkan ukuran sampel menunjukkan bahwa kekuratan penyetaraan dilihat dari jumlah sampel yang digunakan. Semakin banyak ukuran sampel yang digunakan pada proses penyetaraan mempengaruhi nilai koefisien penyetaraan. Secara teori dan beberapa hasil penelitian, bahwa semakin banyak ukuran sampel yang digunakan nilai koefisien dan semakin mendekati 1 dan 0. Begitupula dengan nilai standar error dari koefisien penyetaraan. Semakin banyak ukuran sampel yang digunakan nilai standar error koefisien penyetaraan semakin kecil. Hal ini berarti semakin banyak ukuran sampel maka kualitas penyetaraan semakin nampak.

Tabel 4.19 Hasil analisis perbedaan koefisien penyetaraan berdasarkan ukuran sampel.

|  |  |  |
| --- | --- | --- |
| Ukuran sampel | Koefisien | oefisien |
| 200  400  700 | 1,94983  0,905983  0,77731 | 0,41896  -0,035125  -0,01521 |

Berdasarkan tabel hasil analisis data terdapat perbedaaan hasil penyetaraan paket 01 dan paket 02 dengan ukuran sampel 200 dan 400. Ukuran sampel 400 menghasilkan hasil penyetaraan yang lebih akurat dibandingkan dengan ukuran sampel 200, karena koefisien pada sampel 400 lebih mendekati 1 dan koefisien lebih mendekati 0. Terdapat perbedaan koefisien penyetaraan pada ukuran sampel 200 dan 700. Ukuran sampel 700 menghasilkan koefisien penyetaraan yang lebih akurat dibandingkan dengan ukuran sampel 200, karena koefisien pada ukuran sampel 700 lebih mendekati 1 dan koefisien lebih mendekati 0. Terdapat perbedaan koefisien penyetaraan pada ukuran sampel 400 dan 700. Pada ukuran sampel 700 menghasilkan koefisien penyetaraan yang lebih akurat dibandingkan dengan ukuran sampel 400, karena koefisien pada sampel 700 lebih mendekati 1 dan koefisien lebih mendekati 0

Tabel 4.20 Hasil analisis standar error koefisien penyetaraan berdasarkan ukuran Sampel

|  |  |  |
| --- | --- | --- |
| Ukuran sampel | Standar error koefisien | Standar error koefisien |
| 200  400  700 | 127,7029  0,18068  0,073657 | 7,1967  0,17422  0,068250 |

Berdasarkan tabel hasil analisis data tersebut terdapat perbedaan standar error hasil penyetaraan ukuran sampel 200 dengan 400 paket soal dengan ukuran sampel 400 menghasilkan standar error yang lebih kecil bila dibandingkan dengan paket soal dengan ukuran sampel 200. Hal ini berarti hasil penyetaraan tes ukuran sampel 400 lebih akurat dibandingkan dengan ukuran sampel 200.

Terdapat perbedaan standar error hasil penyetaraan ukuran sampel 200 dengan ukuran sampel 700. Paket soal dengan ukuran sampel 700 menghasilkan standar error yang lebih kecil bila dibandingkan paket soal dengan ukuran sampel 200. Hal ini berarti hasil penyetaraan tes dengan ukuran sampel 700 lebih akurat bila dibandingkan dengan ukuran sampel 200.

Terdapat perbedaan standar error hasil penyetaraan antara paket soal ukuran sampel 400 dengan ukuran sampel 700. Paket soal dengan ukuran sampel 700 menghasilkan standar error yang lebih kecil bila dibandingkan dengan paket soal yang dengan ukuran sampel 400. Hal ini berarti hasil penyetaraan paket soal ukuran sampel 700 lebih akurat dibandingkan dengan paket soal ukuran sampel 400.

Ukuran sampel dalam paket soal mempengaruhi kualitas penyetaraan. Semakin banyak ukuran sampel yang digunakan dalam penyetaraan standar error yang dihasilkan semakin kecil. Hal ini sesuai dengan penelitian yang dilakukan sebelumnya yang mengemukakan temuan ini adalah Swediati (1997), Bastari (2000), dan Kim & Cohen (2002) dalam Syaifuddin (2005: 64), pada Penyetaraan Tes Model Respon Berjenjang bahwa keakuratan parameter estimasi yang dihasilkan dari kalibrasi *item* dipengaruhi oleh ukuran sampel. Semakin besar ukuran sampel yang digunakan pada kalibrasi *item,* parameter *item* yang dihasilkan semakin dekat kenilai *true parameter* . Hal ini sesuai dengan kajian pustaka yang dikemukakan oleh Kolen & Brennan (1995: 264) menunjukkan bahwa meningkatnya ukuran sampel, *standar error of equating (SEE)*  yang dihasilkan semakin kecil.

Berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya, maka dapat disimpulkan hal-hal sebagai berikut:

1. Terdapat perbedaan hasil penyetaraan tes

antara ukuran sampel 200 dan 400, dimana

paket soal dengan ukuran sampel 400

menghasilkan penyetaraan yang lebih akurat.

1. Terdapat perbedaan hasil penyetaraan tes antara ukuran sampel 200 dan 700, dimana paket soal dengan ukuran sampel 700 menghasilkan penyetaraan yang lebih akurat.
2. Terdapat perbedaan hasil penyetaraan tes antara ukuran sampel 400 dan 700, dimana paket soal dengan ukuran sampel 700 menghasilkan penyetaraan yang lebih akurat.
3. Prosedur penyetaraan tes, tahap pertama pemilihan jenis penyetaraan horizontal, kedua pemilihan desain penyetaraan *Common-Item Nonequivalent,* ketiga dilanjutkan dengan estimasi parameter dengan menggunakan model 2 parameter (tingkat kesulitan dan daya beda) dan keempat dilakukan proses penyetaraan dengan menggunakan *equateIRT*

Berdasarkan simpulan yang telah dipaparkan sebelumnya, dari penelitian ini diajukan beberapa saran sebagai berikut:

1. Upaya untuk melakukan penyetaraan skor tes, di sekolah-sekolah baik mandiri, maupun secara bersama-sama serumpun mata pelajaran. Dengan tujuan setelah beberapa tahun guru maupun sekolah memiliki beberapa paket soal yang setara.
2. Perlu dilakukan penelitian lebih lanjut terhadap penyetaraan tes dengan ukuran sampel yang lebih besar agar penyetaraan tes lebih akurat, terhadap subyek yang sama dan tes yang berbeda.
3. Penelitian lebih lanjut dengan menggunakan semua model teori respon butir sehingga dapat membandingkan antara Model 1 parameter, 2 parameter, dan 3 parameter.
4. Agar dilakukan penelitian lebih lanjut dengan paket soal yang lain yaitu paket 03, paket 04, dan paket 05. Karena paket soal ujian nasional tahun 2015 terdiri atas lima paket soal dengan setiap paket soal berbeda anchor antara paket yang satu dengan paket soal yang lain

**DAFTAR PUSTAKA**

Allen, Mary.J & Yen, Wendy. M. 1979. *Introduction to Measurement Theory.*Monterey,California: Brooks/Cole Publishing Company.

Azwar, Saifuddin. 2008. *Dasar-Dasar Psikometri.* Yogyakarta: Pustaka Pelajar

Djaali & Pudji Muljono. 2008. *Pengukuran Dalam Bidang Pendidika.* Jakarta :Gramedia Widiasarana Indonesia

Djemari Mardapi, 2008. *Teknik Penyusunan Instrumen Tes dan Non Tes.*Jokjakarta : Mitra Cendekia Press

Djemari Mardapi, 2012. *Pengukuran, Penilaian, Dan Evaluasi Pendidikan*. Yokjakarta : Nuha Litera

Dorans, Neil. J. 2007. Linking Scores from Multiple Health Outcome Instruments. *Journal of Educational Measurement.* 16 : 85-94.

Erna Miyatum dan Djemari Mardapi, 2000. *Komparasi Metode Penyetaraan Tes Menurut Teori* *Respon Butir, Jurnal Penelitian dan Evaluasi*, Nomor 3, Tahun II, 2000

Hambleton, R.K, Swaminathan H & Rogers, H.J, 1991. *Item Response theory : Principels and Applications.* Boston, MA : Kluwer Inc.

Hambleton, Ronald K., dan Swaminathan, H. 1985. *Item Response Theory: Principle and Applications.* Boston: Kluwer Nijhoff Publishing.

Handayani Siti. 2014. *Penetaraan Dua Perangkat Tes Pada Tepri Respon Butir, Skripsi,* Tidak diterbitkan. Departemen Statistika FPMIPA Institut Pertanian Bogor.

Herikusumo, A. P. 2011. *Penyetaraan (Equating) Ujian Akhir Sekolah Berstandar* *Nasional (UASBN) dengan Teori Tes Klasik.(online),* *Jurnal Pendidikan dan Kebudayaan,* Vol. 17, Nomor 4, Juli 2011

Hidayati Kana, 2010. *Penerapan Teori Respon Butir Dalam Penyetaraan Tes.* Jurnal Penelitian dan Evaluasi UNY(online).Nomor 3, Tahun II. 2000.

Kartono, SEMNAS Matematika dan Pend Matematika, 2006. *Penyetaraan tes berbentuk Uraian .”*Trend Matematika dan Pembelajaran Matematika di Era ICT” FPMIPA UNNES

Kolen, Michael J. & Robert L. Brennan, 1995. *Test Equating, method and practices*. New York : Springer Verlag New York Inc.

Mardapi, Djemari & Miyatun Erna. 2000 *Komparasi Metode Penyetaraan Tes* *Menurut Teori Respon Butir*. Jurnal Penelitian dan Evaluasi UNY (online), Nomor 3 tahun II, (<http://download>. Portalgaruda.org/, diakses 28 September 2015).

Mansyur., Rasyid, Harun., & Suratno. 2009. *Assesmen Pembelajaran di Sekolah*. Yogyakarta: Multi Pressindo.

Mohammad Syaifuddin, 2005. *Penyetaraan Tes Model Respon Berjenjang.* Disertasi tidak diterbitkan Program Pasca Sarjana Universitas Negeri Yogyakarta.

Mellolo Yonathan. 2014. *Analisis Kualitas Soal Try Out Mata Pelajaran Matematika Pada Jurusan IPA SMA Negeri Rantepao Kabupaten Toraja Utara.* Tesis tidak diterbitkan. Program Pascasarjana Univeritas Negeri Makassar.

Naga, D. S. 1992. *Pengantar Teori Sekor pada Pengukuran Pendidikan.* Jakarta: Gunadarma.

Rakhmad Mulyana,S.Ag.2009. *Pembahasan Mengenai Equating (Penyetaraan) dan Differential Iten functioning (DIF),*Blog Mahasiswa Magister Psikometri Universitas Indonesia

Retnawati Heri. 2014. *Teori Respon Butir dan Penerapannya.* Yogyakarta : Nuha Medika.

Ruslan. 2009. *Validitas Isi*. Buletin Pa’biritta No.10 Tahun VI, 18-19

Rustam, 2000. Laporan Penelitian. *Penyetaraan Perangkat Tes Matematika Program D2 PGSD,* tidak diterbitkan Universitas Terbuka

Salam Sofyan & Bangkona Deri.2012, *Pedoman Penulisan Tesis dan Disertasi.*Program Pasca Sarjana Universitas Negeri Makassar. Badan Penerbit UNM.

Salasi R, 1996. *penyetaraan perangkat tes dalam mata pelajaran matematika pada sekolah lanjutan tingkat pertama (SMP),Tesis,*tidak diterbitkan IKIP Yogyakarta.

Setiadi, Hari (1997). *Small Sample ITR itemParameter Estimates*. Disertasi tidak dipublikasikan

Sukirno DS, 2007. *Penyetaraan Tes UAN: Mengapa dan Bagaimana.* FISE Universitas Negeri Yogyakarta, Article

Sora N, 2015, *Pengertian Populasi dan Sampel Serta Teknik Sampling,* Artikel(<http://www.pengertianku.net/2015/03/pengertian-populasi-dan-sampel-serta-teknik-sampling.html>. Diakses 28 oktober 2015)

Sugiyono, 2011. *Metode Penelitian Kombinasi.* Yogyakarta Alpabeta

Suryabrata, Sumadi. 1987. *Pengembangan Tes hasil Belajar.* Penerbit Rajawali Pers Jakarta.

Try Rijanto, 2012. *Pengaruh Metode dan Ukuran sampel terhadap Variansi Skor* *Hasil Penyetaraan*. Jurnal penelitian dan evaluasi pendidikan tahun 16 nomor 1

Zhu, W. 1998. Test Equating: What, Whay, How?. *Research Quartely for Exercise and Sport, Vol 69, 11-23, Washington, Maret 1998.* (http:proquest.umi. com, Diakses 9 Desember 2014).