

Penyetaraan tes soal potensi akademik berdasarkan teori respon butir  
menggunakan *Software R* di STMIK AKBA

Akbar Iskandar<sup>1</sup>, Ahmad Rudi<sup>2</sup>

<sup>1</sup>STMIK AKBA,

e-mail: akbar.iskandar06@gmail.com

<sup>2</sup>SMA Addaraen Makassar

e-mail: ahmadrudi726@gmail.com

(Received: Juni-2017; Reviewed: Juli-2017; Revised: Juli-2017; Accepted: Agustus-2017; Published: Agustus-2017)



©2017 –EST Program Pascasarjana Universitas Negeri Makassar. Ini adalah artikel dengan akses terbuka dibawah licensi CC BY-NC-4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>).

**ABSTRACT**

*This study aims to determine the quality of the package about the Academic Potential Test, to know the procedure of equalizing the package of test of Academic Potential Test and to know how to get two sets of tests that have equality with the method of testEquatingData .. The type of research is explorative, which will reveal the equivalence of test scores based on response- Items. Tithe equivalence procedure, horizontal equalizer selection, common item nonequivalent group equalization selection, third parameter estimation using 2P model (different power and difficulty level) and four equalization process using testEquatingData. Results Equivalence tests with the EQuestData method based on the grain response theory using the 2 parameter model, indicating that the average difficulty level on both packages differs from the packet value of 01 (-1.32934) and the packet value 02 (-2.63645). While the difference between the two packages is different and shows that it can not distinguish the ability of the participants, between high ability and low participant ability, based on the value of package 01 (-0.26831) and package 02 (-0.37774).*

**Keywords:** *Software R, Test Equating, Item Respons Theory.*

**ABSTRAK**

*Penelitian ini bertujuan untuk mengetahui kualitas paket soal Tes Potensi Akademik, mengetahui prosedur penyetaraan paket soal Tes Potensi Akademik dan mengetahui cara mendapatkan dua perangkat tes yang memiliki kesetaraan dengan metode tesEquatingData.. Jenis penelitian adalah eksploratif, yaitu akan mengungkap kesetaraan skor tes berdasarkan teori respons-butir. Prosedur penyetaraan tes, pemilihan jenis penyetaraan horizontal, pemilihan desain penyetaraan common item nonequivalent group, ketiga estimasi parameter dengan menggunakan model 2P (daya beda dan tingkat kesulitan) dan keempat proses penyetaraan dengan menggunakan testEquatingData. Hasil Penyetaraan tes dengan Metode tesEquatingData berdasarkan teori respon butir menggunakan model 2 parameter, menunjukkan bahwa rata-rata tingkat kesulitan pada kedua paket berbeda dengan nilai paket 01 (-1.32934) dan nilai paket 02 (-2.63645). Sedangkan ditinjau dari daya beda kedua paket juga berbeda dan menunjukkan bahwa tidak dapat membedakan kemampuan peserta, antara kemampuan tinggi dengan kemampuan peserta rendah, berdasarkan nilai paket 01 (-0.26831) dan paket 02 (-0.37774).*

**Kata Kunci:** *Software R, Penyetaraan Tes, Teori Respon Butir*

## **PENDAHULUAN**

Software R merupakan program aplikasi yang banyak digunakan dalam menganalisis kualitas butir soal baik pada tingkat nasional maupun pada tingkat internasional namun kurang dimanfaatkan oleh sebagian pendidik baik dosen ataupun guru dalam pengambilan kebijakan penentuan kualitas soal yang dapat digunakan dalam tes seleksi maupun mengukur potensi akademik yang dimiliki oleh mahasiswa. Sedangkan prestasi akademik dipengaruhi oleh 3 faktor utama, yaitu: konteks, input mahasiswa serta proses perkuliahan.

Salah satu kegiatan yang banyak dilakukan dalam dunia pendidikan adalah melakukan tes untuk menguji kemampuan peserta didik. Apabila hasil tes tersebut dijadikan dasar untuk menentukan 'nasib' peserta ujian (misalnya lulus/tidak lulus, diterima/titolak), maka sudah seharusnya kualitas tes ujian memenuhi syarat sebagaimana yang ditentukan. Beberapa kriteria yang dikelompokkan pendekatan tradisional untuk mengukur kualitas tes diantaranya adalah validitas, reliabilitas, tingkat kesukaran daya beda item soal.

Pendekatan ini walaupun tergolong tradisional, namun masih banyak dipakai untuk menguji perangkat evaluasi pembelajaran. Pendekatan yang tergolong modern, salah satunya menggunakan pendekatan logistik disebut dikenal dengan IRT dimana software tersebut berbasis open source, sehingga mudah diakses dan tidak menyalahi peraturan tentang hak cipta. Analisis modern atau lebih dikenal dengan teori respon butir lebih sempurna dalam mengakuisisi hasil analisis. Kelemahan yang ada di teori klasik seperti ketidakmampuan mengestimasi parameter butir dapat diperhatikan lebih lanjut.

Mardapi, (2012:223) menguraikan bahwa ujian merupakan salah satu bentuk asesmen formatif yang tujuannya untuk mengetahui pencapaian standar pendidikan dan Hasil ujian nasional dapat dijadikan masukan terhadap perbaikan proses pembelajaran di satuan pendidikan serta mendorong pendidik untuk selalu menyempurnakan strategi pembelajaran yang digunakan. Penyempurnaan atau perbaikan ini diharapkan akan meningkatkan prestasi belajar peserta didik baik pada aspek kognitif, psikomotor, dan afektif. Semua aspek ini penting dan harus dinilai untuk mengetahui pencapaiannya. Tamsyani (2016:11) menambahkan bahwa Kualitas sumber daya

manusia merupakan prasyarat bagi kemajuan suatu bangsa khususnya sekolah, sedangkan untuk mendapatkan kualitas SDM yang baik maka harus memiliki instrument tes yang baik pula untuk menyeleksinya.

Selanjutnya Daeng (2015:29) menambahkan bahwa dalam proses belajar mengajar guru harus memegang prinsip yang telah dikeluarkan oleh depdiknas terkait dengan penentuan materi pembelajaran seperti kesesuaian (relevansi), keajegan (konsistensi), dan kecukupan (adequacy). Selanjutnya, dalam penentuan cakupan materi pembelajaran harus pula diperhatikan apakah materi pembelajaran yang disajikan berupa aspek kognitif (fakta, konsep, prinsip, prosedur) aspek afektif, ataukah aspek psikomotor. Selain ketiga prinsip di atas, prinsip lain yang perlu digunakan dalam menentukan cakupan materi pembelajaran adalah keluasan, kedalaman dan, kecukupan materi.

Dalam membuat perangkat tes yang sempurna paralel tidaklah mudah. Masih saja terdapat kekurangan pada masalah keadilan. Bisa jadi, perangkat antara satu dengan yang lainnya memiliki kualitas tes yang berbeda seperti paket 1 lebih mudah atau lebih sulit daripada perangkat tes 2 atau bahkan sebaliknya. Hal tersebut menjadi kendala dalam setiap penilaian ketika hasil ujian menggunakan pilhan ganda dijadikan suatu patokan tertentu untuk mengukur prestasi calon peserta tes.

Inteni, Candiasa, & Suarni, (2013:9) Penilaian dapat digunakan sebagai salah satu faktor dalam menentukan keberhasilan proses pembelajaran, bukan hanya sebagai cara untuk menilai keberhasilan siswa dalam menguasai materi pembelajaran. Guru sebagai salah satu faktor utama dalam meningkatkan mutu pendidikan harus terampil menilai proses dan hasil belajar peserta didik seperti membuat alat-alat penilaian, mengolah data dan hasil penilaian, mendiagnosis kesulitan belajar, serta memanfaatkan hasil penilaian.

Namun yang terjadi sampai saat ini banyak instrumen penilaian hasil belajar, baik yang digunakan oleh untuk ulangan harian maupun yang digunakan dalam seleksi calon mahasiswa baru oleh kampus untuk ujian masuk dan soal ujian nasional yang sudah lama dikenal di Indonesia yang mana belum memenuhi persyaratan ideal. Sementara instrumen penilaian hasil belajar yang mengandung bias butir akan merugikan siswa yang memiliki kemampuan sama dengan siswa yang menjawab benar hanya dikarenakan kelompoknya berbeda.

Atau dengan kata lain, instrumen yang mengandung bias butir tidak memberi peluang sama untuk menjawab benar pada peserta tes yang memiliki kemampuan sama hanya karena berasal dari kelompok yang berbeda (Mistiani, 2015:284).

Dalam melaksanakan kegiatan evaluasi, berhasil atau tidaknya sangat ditentukan oleh tepat atau tidaknya pelaksanaan ujian. Untuk melaksanakan ujian ini memerlukan alat-alat. Bagi ujian tertulis maka alatnya adalah butir-butir soal tertulis. Bagi ujian lisan maka alatnya adalah butir soal tertulis yang disediakan bagi setiap testi, atau sekurang-kurangnya pokok pertanyaan yang sudah tertulis dan dipersiapkan sebelumnya. Bagi ujian praktek, maka alatnya adalah lembar pengamatan yang berisi segi-segi yang diamati beserta rentang skor masing-masing.

Ruslan (2005:132) dalam evaluasi pembelajaran aspek keadilan merupakan salah satu prinsip yang penting. Artinya, peserta didik yang menghadapi ujian di berbagai wilayah dan waktu yang berbeda mesti mendapat perlakuan yang adil. Mereka mesti diberi kesempatan untuk menjawab butir-butir tes yang memiliki isi (*content*) dan tingkat kesukaran yang sama. Persoalan akan muncul apabila peserta didik diuji dengan tes yang sama dalam waktu yang berbeda. Seperti yang terjadi pada kelas-kelas parallel, peserta didik yang mengambil ujian pada waktu terakhir akan diuntungkan dengan adanya soal ujian yang bocor sedangkan seleksi pada umumnya dilakukan untuk 2 tujuan, yaitu untuk menyesuaikan jumlah calon mahasiswa dengan tempat yang tersedia dan untuk mendapatkan mahasiswa yang dapat mengikuti program pendidikan yang sulit dengan sukses dan kemudian dapat menjadi anggota profesi yang efektif (Permatasari, Prabandari, & Kristina, 2016:81).

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". Hambleton & Swaminathan, 1985 (Bintarawati, 2012:112) "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 sebaran serta peringkat indeks kesukaran. Dengan demikian, menjadi suatu keharusan bagi pengembang tes atau lembaga pengembang tes

untuk melakukan penyetaraan terhadap perangkat tes yang digunakan.

Satu tantangan besar bagi pengembang asesmen pendidikan adalah mengatasi persoalan yaitu bagaimana membandingkan prestasi dan berbagai kelompok peserta tes, berdasarkan tahun ajaran yang berbeda maupun berbagai sekolah yang berbeda. Meskipun teori klasik masih digunakan secara luas dalam pengukuran pendidikan di Indonesia, namun teknik tersebut memiliki keterbatasan yaitu ketika tes-tes yang berbeda diberikan kepada para peserta tes yang berbeda dan hasilnya diungkapkan dengan skor mentah, kemampuan peserta tes tidak dapat dibandingkan berdasarkan pada skor total. Dalam hal ini, yang menjadi persoalan adalah seorang peserta tes yang lebih rendah kemampuannya, mungkin telah mendapatkan skor yang lebih tinggi pada sebuah tes yang mudah, dibandingkan dengan peserta tes lain yang lebih tinggi kemampuannya yang mungkin memperoleh nilai yang lebih rendah pada tes yang lebih sulit. Keadaan ini menggambarkan bahwa skor-skor tersebut tidak dapat diperbandingkan karena dimungkinkan kedua tes tersebut dibuat pada skala yang berbeda. Oleh karena itu perlu disusun tes-tes yang setara, agar supaya skor tes yang diperoleh dapat saling dipertukarkan (Syaifuddin, 2005:4).

Pada kebanyakan program tes dalam skala besar, penyusunan tes-tes yang setara merupakan kegiatan yang sangat penting, untuk penanganan yang cepat, apabila terjadi kebocoran tes, dan untuk membandingkan hasil tes dari peserta yang menggunakan tes-tes yang berbeda itu. Hal yang sama juga dialami oleh sekolah sebagai suatu institusi pengelola pendidikan yang senantiasa berurusan dengan program tes, meskipun umumnya tidak dalam skala besar (Tumilisar, 2006:1-2).

Aminah (2012:93) mengungkapkan bahwa terdapat dua cara penyetaraan tes yaitu penyetaraan tes vertical dan horizontal dimana penyetaraan tes vertical adalah suatu usaha untuk menyetarakan skor tes pada dua tes atau lebih yang dirancang berbeda tingkat kesulitan tetapi mengukur isi dan jenis kemampuan yang sama. Penyetaraan vertical dirancang untuk kontinuitas tes. Kontinuitas tes diartikan sebagai keberlanjutan tes yang digunakan untuk mengukur perkembangan atau perubahan tingkat kemampuan siswa. Cara penyetaraan horizontal dilakukan pada tes paralel yang memiliki kesamaan isi dan tingkat kesulitan, kemudian

diberikan pada kelompok siswa yang memiliki tingkat kemampuan setara.

Jika hasil tes itu digunakan untuk menentukan kenaikan kelas atau penjurusan program, tentunya hal itu menjadi tidak adil karena tidak dilakukan ekivalensi skor untuk tes yang berbeda itu. Masalah tersebut dapat diatasi dengan melakukan penyetaraan skor yang diperoleh dari peserta yang mengambil tes-tes itu dan kesetaraan antarskor dapat dilakukan secara statistic dimana proses statistik, dikenal sebagai metode penyetaraan (*equating*), telah dikembangkan untuk menangani masalah ini Kolen & Brennan, 1995 (Antara & Bastari, 2015:15).

Penyetaraan tes sangat dirasakan kegunaan mengingat mutu pendidikan di Indonesia belum merata dan keadaan geografis Indonesia sebagai negara kepulauan yang cukup luas. Hal ini mengakibatkan pengukuran secara serentak dalam waktu yang sama tidak mudah dilakukan. Adanya kebijaksanaan otonomi pendidikan pelaksanaan pendidikan sudah menjadi wewenang daerah setempat, namun pemerintah pusat tetap berkewajiban untuk mengendalikan kualitas pendidikan nasional. Misalnya, dengan penetapan kurikulum nasional dan kemampuan standar pada tingkat tertentu. Pengendalian akan mudah dilakukan jika perangkat tes di tiap-tiap daerah diketahui tingkat kesetaraan. Tingkat kesetaraan perangkat tes yang berbeda akan dapat diketahui melalui proses penyetaraan (Miyatun & Mardapi, 2000 :2-3).

Berdasarkan uraian di atas, memberikan gambaran bahwa penyetaraan tes sebagai salah satu upaya pengembangan alat ukur dalam mengevaluasi hasil belajar siswa perlu dikembangkan dalam berbagai bidang ilmu, terutama pada tingkat pendidikan dasar dan menengah hingga perguruan tinggi. Menurut Aminah (2012:92) "Penyetaraan skor adalah membandingkan skor yang diperoleh dari perangkat tes yang satu (X) dan perangkat tes lainnya (Y) yang dilakukan melalui proses penyetaraan skor pada kedua perangkat tes tersebut. serta dapat digunakan untuk mengatur skor pada format-format tes sehingga skor pada format tersebut dapat diperbandingkan.

Mensetarakan perangkat-perangkat tes yaitu mengkonversikan sistem unit-unit sesuatu perangkat ke sistem unit-unit perangkat yang lain, sehingga skor-skor yang dijabarkan dari dua perangkat tes setelah konversi secara langsung akan setara. Jika penyetaraan itu dilakukan dengan baik, maka akan mungkin

menyatakan bahwa terjadi penurunan atau peningkatan rata-rata kelompok. Dengan proses penyetaraan yang benar-benar memenuhi persyaratan, maka selanjutnya orang akan dapat mengukur perkembangan, memetakan kecenderungan-kecenderungan, menggabungkan data atau informasi yang berasal dari perangkat-perangkat-perangkat tes yang berbeda.

Penyetaraan bukan untuk memprediksi skor pada salah satu tes ke skor tes yang lain, melainkan untuk perbandingan antara skor yang disetarakan. Tujuan penyetaraan adalah untuk menghasilkan skor yang ekuivalen yang berasal dari dua atau lebih perangkat tes. "Pada dasarnya penyetaraan (*equating*) bertujuan untuk menyetarakan skor dengan cara membandingkan skor yang diperoleh dari mengerjakan perangkat tes yang satu dengan skor yang diperoleh dari mengerjakan perangkat tes lainnya yang dilakukan melalui proses penyetaraan skor pada kedua perangkat tes tersebut". Jadi, Melalui penyetaraan penyetaraan dapat diperoleh skor yang dapat diperbandingkan.

Berdasarkan beberapa pendapat tentang penyetaraan tes, dapat dikatakan bahwa penyetaraan atau *equating* merupakan suatu prosedur dalam membandingkan perangkat tes yang satu (X) ke perangkat tes yang lain (Y), sehingga hasil-hasilnya dapat diperbandingkan atau dipertukarkan peggunaannya.

## METODE

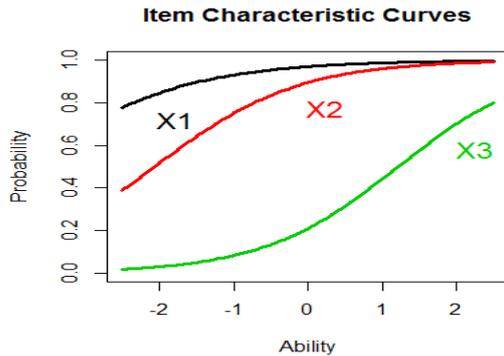
Jenis penelitian ini adalah penelitian yang bersifat *exploratif*, yaitu mencari, mengungkap, menggali secara cermat dan lengkap fakta-fakta yang terkandung dalam suatu permasalahan yang bersifat spesifik. Dalam penelitian ini, peneliti mengungkap karakteristik soal TPA tahun ajaran 2016-2017 model 2 parameter, kesetaraan tes dengan metode *tesEquating Data*. Penelitian dilaksanakan di STMIK AKBA. Subjek dalam penelitian ini adalah butir soal dan seluruh lembar jawaban peserta tes. Desain penyetaraan yang digunakan adalah desain *common item nonequivalent groups*. Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi. Teknik Analisis Data (1) Proses Validasi Isi, 2. (2) pengumpulan data dengan menggunakan desain *common item nonequivalent group*. (3) Estimasi parameter butir dengan analisis *R Program V.3.2.0*

**HASIL DAN PEMBAHASAN**

**Hasil**

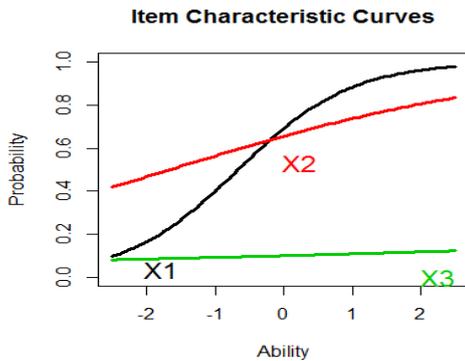
**a. Tingkat Kesulitan Paket 01 dan 02**

**Gambar 1.** Kurva Karakteristik Butir Soal Paket 01 Model 2P Berdasarkan Tingkat Kesulitan



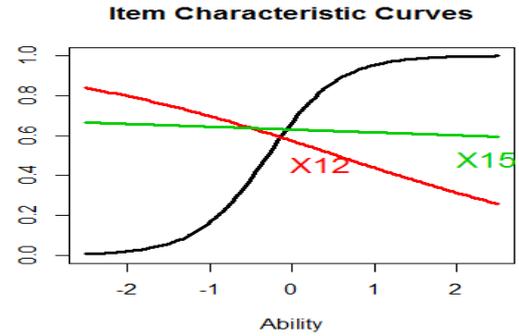
Gambar 1. Menunjukkan bahwa dari Paket 01 untuk butir 1, 2 dan 3 berada pada kategori sukar dan sangat sukar. Berdasarkan grafik di atas butir 3 dengan koefisien (3,220), butir 2 (-2,057) dan butir 1 (-3,913).

**Gambar 2.** Kurva Karakteristik Butir Soal Paket 02 Model 2P Berdasarkan Tingkat Kesulitan



Gambar 2. Menunjukkan bahwa dari Paket 01 untuk butir 1, 2 dan 3 berada pada kategori sukar dan sangat sukar. Berdasarkan grafik di atas butir 3 dengan koefisien (23,787), butir 2 (-1,648) dan butir 1 (-0,661).

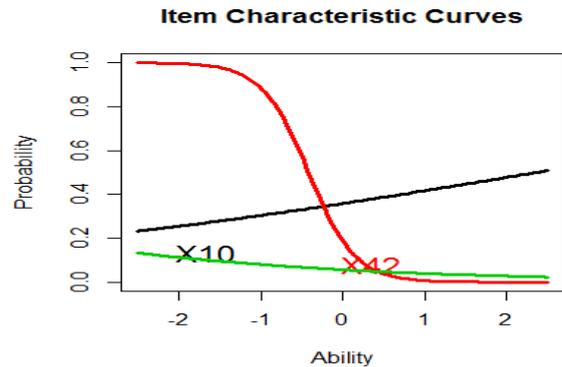
**b. Daya Beda Paket 01 dan Paket 02**



**Gambar 3.** Kurva Karakteristik Butir Soal Paket 01 Model 2P Berdasarkan Daya Beda

Gambar 3. Menunjukkan bahwa dari Paket 01 untuk butir 10, 12 dan 15 berada pada kategori daya beda baik dan tidak baik. Berdasarkan grafik di atas butir 10 dengan koefisien (2,336), butir 12 (-0,541) dan butir 15 (-0,062).

**Gambar 4.** Kurva Karakteristik Butir Soal Paket 02 Model 2P Berdasarkan Daya Beda



Gambar 4. Menunjukkan bahwa dari Paket 01 untuk butir 10, 12 dan 15 berada pada kategori daya beda baik dan tidak baik. Berdasarkan grafik di atas butir 10 dengan koefisien (0,246), butir 12 (-3,517) dan butir 15 (-0,376).

**3. Proses Penyetaraan**

**Tabel 1.** Hasil Analisis Penyetaran Paket 01 dan Paket 02 Berdasarkan Tingkat Kesulitan

Item	Tingkat Kesulitan Paket 01	Item	Tingkat Kesulitan Paket 02
X1 *	-1.2285	X1 *	-1.2285
X4 *	-1.3191	X4 *	-1.3191
X5 *	1.8568	X5 *	1.8568
X9 *	-0.0121	X9 *	-0.0121
X11 *	1.76	X11 *	1.76

Item	Tingkat Kesulitan Paket 01	Item	Tingkat Kesulitan Paket 02
X13 *	-3.813	X13 *	-3.813
X17 *	-2.6885	X17 *	-2.6885
X19 *	-3.4957	X19 *	-3.4957
X22 *	-64.7993	X22 *	-64.7993
X23 *	0.1564	X23 *	0.1564
X2	0.792	X2	49.0468
X3	0.8956	X3	-0.8029
X6	-5.5107	X6	-7.968
X7	5.428	X7	-7.239
X8	21.0161	X8	-7.4211
X10	0.9299	X10	-1.3613
X12	-1.3145	X12	-4.6144
X14	-5.4764	X14	2.7797
X15	2.9942	X15	-3.9942
X16	-0.5261	X16	-2.949
X18	-41.6036	X18	-2.1295
X20	3.5077	X20	-5.943
X21	-16.6343	X21	-7.4089
X24	5.7664	X24	-3.7301
X25	-1.1956	X25	-11.6686
X26	0.2582	X26	-0.8061
X27	-1.78	X27	-3.647
X28	-0.9063	X28	-11.9799
X29	0.9001	X29	-10.4849
X30	3.5058	X30	-15.8527
X31	0.5322	X31	-1.0979
X32	-3.7073	X32	-5.0422
X33	-0.2995	X33	-11.0177
X34	-0.5566	X34	-2.3956
X35	-2.3107	X35	-3.7131
X36	-0.1692	X36	15.7942
X37	-2.3627	X37	4.6689
X38	0.274	X38	-9.3939
X39	-1.266	X39	-6.53
X40	-1.0608	X40	-2.192
Rata-rata	-1.32934	Rata-rata	-2.63645

\* = *Anchor Item* (Butir yang sama dari kedua perangkat)

Berdasarkan hasil pengolahan data Tabel 1. Menunjukkan bahwa paket 01 dan paket 02 ditinjau dari tingkat kesulitan, bahwa kedua paket soal berbeda dalam tingkat kesulitan, paket 02 lebih mudah dari pada paket 01. Berdasarkan rata-rata tingkat kesulitan paket 01 (-1.32934) dan rata-rata paket 02 (-2.63645).

**Tabel 2.** Hasil Analisis Penyetaran Paket 01 dan Paket 02 Berdasarkan Daya Beda

Item	Daya Pembeda Paket 01	Item	Daya Pembeda Paket 02
X1 *	1.6519	X1 *	1.6519
X4 *	1.031	X4 *	1.031
X5 *	0.9785	X5 *	0.9785
X9 *	0.6023	X9 *	0.6023
X11 *	0.7629	X11 *	0.7629
X13 *	0.8504	X13 *	0.8504
X17 *	1.3544	X17 *	1.3544
X19 *	0.6776	X19 *	0.6776
X22 *	0.0109	X22 *	0.0109
X23 *	0.9352	X23 *	0.9352
X2	1.0369	X2	-0.0268
X3	-0.5462	X3	-2.0241
X6	0.617	X6	-0.4264
X7	-0.5024	X7	-0.3418
X8	-0.0257	X8	-0.3976
X10	1.1779	X10	-1.3373
X12	1.294	X12	0.2483
X14	0.4065	X14	-0.4343
X15	-0.3282	X15	0.1475
X16	1.5127	X16	-0.9583
X18	-0.0249	X18	1.3471
X20	-0.771	X20	-0.3643
X21	0.1022	X21	0.2378
X24	0.1311	X24	-0.3165
X25	-0.948	X25	-0.1613
X26	-0.4544	X26	-0.9535
X27	-2.3311	X27	-1.1006
X28	-2.9037	X28	-0.1044
X29	-1.7252	X29	0.024
X30	-2.5688	X30	-0.226
X31	0.669	X31	-1.1841
X32	-0.333	X32	-0.3978
X33	1.6604	X33	-0.0461
X34	-1.297	X34	-0.4061
X35	-0.7037	X35	-0.6314
X36	-0.3996	X36	0.0748
X37	-0.5643	X37	0.3518
X38	1.0028	X38	-0.128
X39	-2.3076	X39	-0.4294
X40	1.0751	X40	-1.3675
Rata-rata	-0.26831	Rata-rata	-0.37774

\* = *Anchor Item* (Butir yang sama dari kedua perangkat)

Berdasarkan hasil pengolahan data Tabel 2. Menunjukkan bahwa paket 01 dan paket 02 ditinjau dari daya beda, bahwa kedua paket soal tidak mampu membedakan antara kemampuan

peserta dengan kemampuan tinggi dan kemampuan rendah. Berdasarkan rata-rata daya beda paket 01 (-0.26831) dan rata-rata paket 02 (-0.37774).

## SIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan sebelumnya dapat disimpulkan bahwa Kualitas soal paket 01 Tes Potensi Akademik berdasarkan tingkat kesulitan terdapat 23 butir kategori baik dan paket 02 terdapat 10 butir kategori baik. Sedangkan Kualitas soal paket 01 Tes Potensi Akademik berdasarkan daya beda terdapat 22 butir kategori baik dan paket 02 terdapat 17 butir kategori baik.

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 (daya beda dan tingkat kesulitan) dan keempat dilakukan proses penyetaraan dengan menggunakan *testEquatingData*.

Penyetaraan tes dengan Metode *testEquatingData* berdasarkan teori respon butir menggunakan model 2 parameter, menunjukkan bahwa rata-rata tingkat kesulitan pada kedua paket berbeda dengan nilai paket 01 (-1.32934) dan nilai paket 02 (-2.63645). Sedangkan ditinjau dari daya beda kedua paket juga berbeda dan menunjukkan bahwa tidak dapat membedakan kemampuan peserta, antara kemampuan tinggi dengan kemampuan peserta rendah, berdasarkan nilai paket 01 (-0.26831) dan paket 02 (-0.37774).

## SARAN

Untuk penelitian selanjutnya disarankan dengan penyetaraan tes dengan beberapa metode penyetaraan dan model parameter yang berbeda dengan subyek yang sama dengan dengan tes yang berbeda. Selanjutnya dengan Melakukan penyetaraan tes, sudah seharusnya dilakukan oleh perguruan tinggi swasta, dengan upaya itu, setelah beberapa tahun baik dosen ataupun perguruan tinggi, memiliki bank tes yang setara.

## DAFTAR PUSTAKA

- Aminah, N. S. (2012). Karakteristik metode penyetaraan skor tes untuk data dikotomos. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 16, 88–101.
- Antara, A. A. P., & Bastari. (2015). Penyetaraan Vertikal Dengan Pendekatan Klasik Dan Item Response Theory Pada Siswa Sekolah Dasar, 19(1), 13–24
- Bintarawati, D. (2012). The accuracy of concordance method. *Jurnal Evaluasi Pendidikan*, 3(2), 107–117. Retrieved from <http://journal.unj.ac.id/unj/index.php/jep/article/download/2683/2032>
- Daeng, K. (2015). Pengembangan materi pembelajaran makassar bagi siswa SMP/MTS. di sulawesi selatan *Journal of Educational Science and Technology*, 1, 27–38.
- Inteni, K. A. S., Candiasa, I. M., & Suarni, N. K. (2013). Pengembangan Instrumen Tes Objektif Pilihan Ganda Yang Diperluas Berbasis Web Untuk Mata Pelajaran Tik Kelas Xi Sman Di Kabupaten Karangasem. *E-Journal Program Pascasarjana Universitas Pendidikan Ganesha Program Studi Penelitian Dan Evaluasi Pendidikan*, 3(5).
- Miyatun, E., & Mardapi, D. (2000). Komparasi Metode Penyetaraan Tes Menurut Teori Respon Butir. *Jurnal Penelitian dan Evaluasi UNY (online). Nomor 3, Tahun II. 2000*
- Mardapi, Djemari. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta. PT. Nuha Litera.
- Mistiani, W. (2015). Keadilan gender dalam penilaian hasil belajar. *Musawa*, 7(2), 283–302.
- Permatasari, T. O., Prabandari, Y. S., & Kristina, T. N. (2016). Faktor Kognitif dan Non-Kognitif Pada Seleksi Mahasiswa Baru Sebagai Prediktor Terhadap Prestasi Akademik. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1).
- Ruslan. 2005. Penyetaraan Kemampuan Peserta Tes. *Jurnal Pendidikan Matematika dan Matematika UNM*. Vol. 5, No 2
- Tumilisar, A. J. V. 2006. Akurasi Relatif Penyetaraan Skor Tes untuk Sampel Berukuran 300 Ditinjau dari Metode Penyetaraan dan Teknik Penghalusan. *Jurnal Pendidikan Penabur*. No.06/Th.V/Juni 2006
- Tamsyani, W. (2016). Pengaruh Model Pembelajaran dan Kesadaran Metakognitif. *Journal of Educational Science and Technology*, 2(1), 10–26.
- Syaifuddin, Mohammad. 2005. *Penyetaraan Tes Model Respons Berjenjang*. Disertasi.

Tidak diterbitkan. Program Pascasarjana  
Univeritas Negeri Yogyakarta.