



## Bagaimana hasil penyetaraan paket tes USBN pada mata pelajaran matematika dengan teori respons butir?

Eri Yusron<sup>1, a \*</sup>, Heri Retnawati<sup>2, b</sup>, Ibnu Rafi<sup>3, c</sup>

<sup>1</sup> Program Studi Penelitian dan Evaluasi Pendidikan, Pascasarjana, Universitas Negeri Yogyakarta.

Jalan Colombo No. 1, Karangmalang, Yogyakarta 55281, Indonesia

<sup>2</sup> Jurusan Pendidikan Matematika, Fakultas Matematika, dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta.

Jalan Colombo No. 1, Karangmalang, Yogyakarta 55281, Indonesia.

<sup>3</sup> Program Studi Pendidikan Matematika, Pascasarjana, Universitas Negeri Yogyakarta.

Jalan Colombo No. 1, Karangmalang, Yogyakarta 55281, Indonesia.

E-mail: <sup>a</sup> [eyusron98@gmail.com](mailto:eyusron98@gmail.com); <sup>b</sup> [heri\\_retnawati@uny.ac.id](mailto:heri_retnawati@uny.ac.id); <sup>c</sup> [ibnurafi789@gmail.com](mailto:ibnurafi789@gmail.com)

\* Corresponding Author

### ARTICLE INFO

#### Article history

Received: 16 April 2020

Revised: 17 June 2020

Accepted: 1 July 2020

#### Keywords

penyetaraan paket tes USBN; perangkat tes matematika; teori respons butir; *equating of USBN test packages*; *mathematics test packages*; *item response theory*

### ABSTRACT

Penelitian deskriptif eksploratif dengan pendekatan kuantitatif ini bertujuan untuk mendeskripsikan kesetaraan perangkat Ujian Sekolah Berstandar Nasional (USBN) tahun 2018/2019 untuk mata pelajaran matematika wajib SMA. Pengumpulan data dilakukan melalui dokumentasi respons siswa pada USBN 2018/2019 untuk mata pelajaran matematika wajib SMA. Respons siswa berasal dari lima paket soal dari empat sekolah menengah atas di Daerah Istimewa Yogyakarta dan Kalimantan Selatan. Data dianalisis menggunakan teknik *equating* berdasarkan teori respons butir dengan metode *mean-mean*, *mean-sigma*, *Haebara*, dan *Stocking Lord*. Estimasi parameter butir dan penyetaraan dilakukan dengan bantuan Program R. Hasil penyetaraan menggunakan empat metode menunjukkan bahwa lima paket tes USBN 2018/2019 untuk mata pelajaran matematika wajib cenderung setara satu sama lain dan penyetaraan menggunakan metode *Haebara* menghasilkan kesetaraan yang lebih baik dibandingkan tiga metode lainnya. Dalam artikel ini juga disajikan contoh butir sulit beserta peluangnya untuk digunakan sebagai referensi dalam meningkatkan kualitas pembelajaran matematika

*This exploratory descriptive research with a quantitative approach was focused on describing the equality of test packages of USBN year 2018/2019 for compulsory mathematics subjects in Senior High School. Data collection was done through documentation of students' responses to the USBN 2018/2019 for compulsory mathematics subjects. These students' responses were collected from five test packages from four senior high schools in Province of Special Region of Yogyakarta and South Kalimantan, Indonesia. The data were analyzed by using an equating technique based on the item response theory with the methods of mean-mean, mean-sigma, Haebara, and Stocking Lord. The item parameter estimation and equating were conducted with the aid of the R program. The results of equating showed that the five test packages of the USBN 2018/2019 for compulsory mathematics subjects tend to be equal to each other and the equating through Haebara method yields better equality than the other methods. This article also presents the example of a difficult item as well as its opportunity to be used as a reference for enhancing the quality of mathematics learning.*



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



**How to Cite:** Yusron, E., Retnawati, H., & Rafi, I. (2020). Bagaimana hasil penyetaraan paket tes USBN pada mata pelajaran matematika dengan teori respon butir?. *Jurnal Riset Pendidikan Matematika*, 7(1), 1-12. doi:<https://doi.org/10.21831/jrpm.v7i1.31221>

## PENDAHULUAN

Ujian sekolah dilakukan satuan pendidikan untuk mengukur kompetensi peserta didik. Proses pengukuran tersebut merupakan wujud dari penilaian hasil belajar yang diselenggarakan oleh satuan pendidikan. Pada proses penyelenggaraannya, instrumen yang digunakan untuk ujian sekolah harus memenuhi persyaratan substansi, konstruksi dan bahasa sesuai dengan jenjang dan kompetensi peserta didik. Sejalan dengan perkembangan sistem pendidikan di Indonesia, ujian sekolah mengalami perkembangan. Untuk melakukan standarisasi ujian sekolah secara nasional di Indonesia, maka diadakan Ujian Sekolah Berstandar Nasional (USBN).

USBN merupakan kegiatan pengukuran kompetensi peserta didik dalam satuan pendidikan. Instrumen yang digunakan dalam kegiatan USBN berdasarkan POS 2018/2019 ([Badan Standar Nasional Pendidikan, 2018](#)) terdiri dari 20%-25% soal dari pusat dan 75%-80% soal disusun oleh guru. Hal ini yang menjadikan setiap instrumen yang digunakan untuk USBN oleh setiap sekolah berbeda-beda. Pada kasus ini akan muncul kemungkinan kalau kesetaraan setiap instrumen USBN yang digunakan oleh setiap sekolah tidak setara terutama dalam tingkat kesukaran. Hal ini dapat diatasi dengan melakukan penyetaraan (*equating*) skor antar instrumen tes yang berbeda-beda tersebut.

*Equating* merupakan suatu proses statistik yang digunakan untuk menyetarakan skor pada kegiatan tes ([Kolen & Brennan, 2014](#)). [Herkusumo \(2011\)](#) menjelaskan bahwa *equating* merupakan prosedur dalam menyetarakan skor dari perangkat tes satu ke perangkat tes lainnya, sehingga dapat diketahui perbandingan secara langsung dari skor individu yang menggunakan perangkat tes tersebut. [Hambleton et al. \(1991\)](#) menyatakan bahwa *equating* adalah perbandingan skor tes satu dengan skor tes lain dengan penyetaraan skor pada kedua tes. Dapat disimpulkan bahwa *equating* adalah proses statistik untuk menyetarakan skor tes yang satu dengan skor tes yang lainnya sehingga dapat diketahui konversi secara langsung dari instrumen yang memiliki spesifikasi yang sama. Adapun tujuan dari *equating* adalah menghasilkan skor yang dapat saling menggantikan antar tes satu sama lainnya ([Retnawati, 2014](#)). Untuk melakukan penyetaraan tes dalam bentuk ujian dapat dilakukan proses statistik dengan menyetarakan skor tes satu dengan skor tes lainnya.

*Equating* terdiri dari dua jenis, yaitu horizontal dan vertikal ([Retnawati, 2014](#)). Pada *equating* horizontal, instrumen tes disetarakan dengan mengukur tingkat atau kelas yang sama, tes-tes yang dibandingkan diberikan pada kelompok peserta tes yang memiliki kemampuan yang sama ([Hambleton et al., 1991](#)). Pada *equating* vertikal, proses penyetaraan yaitu pada tes-tes yang mengukur tingkat atau kelas yang berbeda ([Retnawati, 2014](#)). Jenis-jenis *equating* ini merupakan acuan dalam menentukan proses penyetaraan yang dilakukan. Kemudian, beberapa hal yang perlu diperhatikan dalam melakukan penyetaraan pada jenis *equating* vertikal adalah materi tes, dimensi pengukuran, kemampuan, dan tingkat kesulitan ([Crocker & Algina, 2006](#)).

Dalam menghubungkan tes terdapat beberapa desain yang bisa digunakan, seperti desain grup tunggal, desain grup ekuivalen, dan desain butir bersama atau *anchor* ([Hambleton et al., 1991](#)). Desain grup tunggal yaitu dua tes yang akan disetarakan diberikan kepada grup yang sama ([Retnawati, 2014](#)). Desain ini memerlukan perhatian khusus karena pada pelaksanaannya akan ada dua atau lebih instrumen yang akan digunakan untuk tes secara bergantian yang akan dilakukan pada grup yang sama, sehingga akan memunculkan faktor latihan, kelelahan, dan kepenatan ([Retnawati, 2014](#)). Kemudian kemungkinan yang muncul lainnya yaitu kesukaran tes akhir dibanding dengan tes awal berbeda (*fatigue effect*). Untuk mengatasi faktor latihan, kelelahan, dan kepenatan bisa dilakukan tes pada waktu dan tempat yang berbeda ([Retnawati, 2014](#)). Selain itu, dalam mengatasi *fatigue effect* bisa dilakukan dengan cara *counter balancing* yaitu sampel dibagi menjadi beberapa sub dan setiap tes dilaksanakan secara bersilangan pada setiap sub sampel ([Herkusumo, 2011](#)). Hal ini yang membuat desain penyetaraan ini jarang digunakan karena tidak praktis dan banyak faktor yang dapat berpengaruh pada skor yang dihasilkan.

Pada desain grup ekuivalen, dua instrumen tes yang diberikan kepada dua kelompok yang mempunyai kemampuan yang sama atau ekuivalen. Jadi, secara operasional desain ini menyiapkan dua instrumen tes yang akan disetarakan, terdiri dari instrumen tes 1 dan instrumen tes 2, kemudian kedua tes tersebut diberikan kepada dua kelompok, kelompok A diberi instrumen tes 1 dan kelompok B diberi instrumen tes 2. Kelompok A dan B tersebut harus mempunyai kemampuan yang sama. Desain ini sama dengan desain sebelumnya, yaitu mempunyai kelebihan dan kelemahan. Kelebihan dari desain grup ekuivalen merupakan kelemahan dari desain grup tunggal, di mana kelebihan desain ini yaitu lebih

praktis dan menghilangkan pengaruh kelelahan, latihan, dan kepenatan. Kelemahan dari desain ini yaitu memerlukan sampel yang besar karena desain ini dapat memunculkan bias dalam hasil penyetaraan. Bias tersebut timbul karena kelompok-kelompok yang digunakan sebagai sampel belum tentu mempunyai kemampuan yang sama.

Desain tes dengan butir *anchor* merupakan dua tes atau lebih yang diberikan pada dua atau grup yang berbeda, di mana instrumen tes tersebut memiliki *common item* atau butir bersama (*anchor*) (Kolen & Brennan, 2014; Retnawati, 2014). Instrumen tes dengan *anchor* terdiri dari dua jenis, yaitu tes dengan butir *anchor* eksternal dan internal. Instrumen tes dengan butir *anchor* eksternal memiliki butir *anchor* yang berdiri sendiri. Sementara instrumen tes dengan butir *anchor* internal memiliki butir *anchor* yang berada pada tes tersebut. Keuntungan dari instrumen tes dengan butir *anchor* internal yaitu responden hanya menerima satu instrumen, waktu pelaksanaan menjadi singkat karena hanya mengerjakan satu instrumen dan skor butir *anchor* dapat diperhitungkan dalam skor total tes (Retnawati, 2014).

Dalam melakukan penyetaraan ada beberapa hal yang perlu diperhatikan. Instrumen tes yang digunakan untuk mengukur kemampuan dan sifat yang berbeda tidak dapat disetarakan, skor tes yang mempunyai karakteristik reliabilitas tidak sama hendaknya tidak disetarakan, skor tes mentah yang tingkat kesukarannya tidak bervariasi tidak dapat disetarakan, skor antar perangkat tidak dapat disetarakan jika tidak paralel dan skor yang berasal dari tes yang beda materi tidak dapat disetarakan. Untuk melakukan penyetaraan dapat dilakukan pembuktian secara statistik menggunakan teori tes klasik atau teori respons butir (Hambleton & Swaminathan, 1985).

Pada teori respons butir atau *item response theory* (IRT), *equating* merupakan proses menemukan skor dari dua atau lebih bentuk tes paralel ke skala skor umum (Hambleton et al., 1991). Pada penerapannya, ada dua kategori metode yang dapat digunakan untuk melakukan penyetaraan, yaitu metode momen dan metode grafik (Retnawati, 2014). Pada kategori metode momen ada tiga metode penyetaraan, yaitu metode rerata-rerata, metode rerata-sigma, metode rerata-sigma tegar (Hambleton et al., 1991). Kemudian, kategori metode grafik yaitu metode kurva karakteristik Haebara dan Stocking Lord (Kolen & Brennan, 2014). Metode rerata dan rerata bertujuan untuk menentukan konstanta yang melibatkan dua parameter, yaitu parameter daya pembeda dan tingkat kesukaran butir (Lloyd & Hoover, 1980). Penyetaraan konstanta  $\alpha$  dan  $\beta$  dapat dihitung dengan rerata dari parameter yang terlibat dari daya pembeda dengan tingkat kesukaran butir. Metode kedua yaitu metode rerata dan sigma (Marco, 1977), di mana pada metode ini konstanta penyetaraan  $\alpha$  dan  $\beta$  melibatkan rerata dan simpangan baku dari parameter tingkat kesulitan (Retnawati, 2014). Sedangkan pada metode Haebara maupun Stocking Lord, penyetaraan parameter butirnya dilihat berdasarkan kurva karakteristik tesnya (Haebara, 1980).

Metode-metode *equating* dapat dibandingkan melalui kurva karakteristik tes yang terbentuk. Kurva karakteristik tes dapat digambarkan dengan membandingkan jumlah dari nilai fungsi informasi butir secara keseluruhan pada instrumen tes (Hambleton & Swaminathan, 1985; Retnawati, 2014). Semakin rapat gambar kurva karakteristik setiap instrumen maka tingkat kesetaraannya semakin baik. Selain itu, dalam melakukan perbandingan metode *equating* juga dapat dilakukan menggunakan *root mean square error* (RMSE) seperti yang telah dilakukan pada penelitian-penelitian sebelumnya (misalnya Antara & Bastari, 2015; Aşiret & Sünbül, 2016; Kolen & Brennan, 1995; Retnawati, 2014; Uysal & Kilmen, 2016).

Banyak *software* yang dapat digunakan untuk melakukan analisis *equating* di antaranya yaitu Program R. Analisis *equating* pada Program R bisa dilakukan dengan menggunakan *package equate IRT* (Battaaz, 2015). Paket ini menyajikan beberapa fasilitas mulai dari *functionalities*, *data preparation*, hingga *data analysis* (*direct equating coefficients*, *chain equating coefficients*, *average equating coefficients*, dan *equated scores*) (Battaaz, 2015). Untuk melakukan penyetaraan dengan Program R terlebih dahulu perlu dicari *parameter logistic* (PL) yang cocok untuk data yang akan digunakan. Mencari *parameter logistic* (1PL, 2PL, 3PL) pada Program R dapat dilakukan dengan menggunakan *package ltm* (Rizopoulos, 2006) dan *package mirt* (Chalmers, 2012). Kemudian hasil estimasi karakteristik butir pada setiap instrumen dengan model yang cocok digunakan sebagai data untuk melakukan estimasi penyetaraan. Selanjutnya setelah data siap, dilakukan *link* terlebih dahulu antar instrumen dan dilakukan analisis *equating* yang akan menghasilkan konstanta pada setiap penyetaraan tes.

Jika kita cermati proses penyetaraan sangatlah penting, terlebih instrumen tes yang digunakan pada USBN merupakan instrumen tes yang berbeda. Walaupun dalam pengembangan instrumen tes USBN di dalamnya terdapat butir *anchor* atau butir bersama, hal ini tidak dapat mencirikan bahwa instrumen tersebut setara. Sehingga untuk mengatasi hal tersebut perlu dilakukan suatu proses penyetara-

raan instrumen tes USBN. Selain itu, perlu juga dilakukan perbandingan metode penyetaraan agar hasil penyetaraan lebih presisi antara instrumen yang disetarakan dengan instrumen sebagai tolak ukur penyetaraan.

Penelitian-penelitian terdahulu telah dilakukan untuk mengonfirmasi kesetaraan perangkat tes yang digunakan dalam ujian akhir (Herkusumo, 2011; Iriyadi et al., 2019; Kartowagiran et al., 2018; Rahayu, 2015; Retnawati, 2016; Rijanto, 2011; Setiawan, 2019). Akan tetapi kebanyakan dari penelitian tersebut lebih difokuskan pada penyetaraan perangkat tes untuk Ujian Nasional (UN). Lebih lanjut, dari beberapa penelitian tersebut, penelitian yang berfokus pada USBN adalah penelitian yang dilakukan oleh Herkusumo (2011), Retnawati (2016), dan Rijanto (2011). Namun demikian, Herkusumo (2011) melakukan penyetaraan perangkat tes USBN untuk sekolah dasar berdasarkan teori tes klasik, Retnawati (2016) melakukan penyetaraan pada perangkat tes ujian akhir mata pelajaran matematika SMP, dan Rijanto (2011) melakukan penyetaraan perangkat tes USBN mata pelajaran IPA untuk sekolah dasar. Dengan kata lain, belum banyak penelitian yang mengeksplorasi penyetaraan perangkat tes mata pelajaran matematika wajib yang digunakan dalam USBN, terlebih untuk jenjang Sekolah Menengah Atas (SMA) atau yang sederajat. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut. Berdasarkan hal tersebut, penelitian ini bertujuan untuk mendeskripsikan kesetaraan perangkat USBN tahun 2018/2019 pada mata pelajaran matematika wajib SMA.

### METODE

Penelitian ini merupakan penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Pengumpulan data dilakukan melalui dokumentasi respons siswa pada USBN 2018/2019 untuk mata pelajaran matematika wajib SMA. Respons siswa berasal dari lima paket tes dari empat Sekolah Menengah Atas (SMA) di Daerah Istimewa Yogyakarta dan Kalimantan Selatan. Jumlah responden dari Daerah Istimewa Yogyakarta sebanyak 340 siswa yang berasal dari dua sekolah. Kemudian jumlah responden dari Provinsi Kalimantan Selatan sebanyak 379 siswa yang berasal dari dua sekolah. Sehingga, total ada 719 siswa yang terlibat sebagai responden dalam penelitian ini. Sebagian besar butir soal pada kelima paket tes tersebut berbeda karena soal terdiri dari 20-25% butir soal berstandar nasional (*anchor*) dan 75-80% butir soal yang dikembangkan oleh guru di bawah koordinasi Dinas Pendidikan Provinsi (Badan Standar Nasional Pendidikan, 2018). Setiap paket tersebut memiliki butir berstandar nasional (butir bersama atau *anchor*). Rincian sebaran butir bersama, jumlah butir soal, dan jumlah responden untuk masing-masing paket tes disajikan pada Tabel 1.

**Tabel 1.** Sebaran Butir Bersama dan Jumlah Responden

Paket tes	Butir bersama	Jumlah butir soal	Jumlah responden
A	3,10,15,16,19,22,27,29	40	109
B	3,10,15,16,19,22,27,28	40	108
C	10,15,16,19,27,28	40	123
D	3,10,15,16,19,22,27,29	40	112
E	3,10,15,16,19,22,27,29	40	267

Pada penelitian ini, analisis dilakukan dalam beberapa tahap. Pada tahap pertama dilakukan analisis karakteristik tes pada setiap paket tes berdasarkan teori respons butir dan dianalisis kecocokan model parameter logistiknya. Proses analisis ini menggunakan bantuan Program R. Pada tahap kedua, dilakukan analisis penyetaraan dengan Program R berdasarkan teori respon butir untuk mengetahui konstanta hasil penyetaraan. Analisis ini menggunakan empat metode penyetaraan yaitu metode rerata-rerata (*mean-mean*), rerata-sigma (*mean-sigma*), Haebara, dan Stocking Lord. Pada tahap ketiga dilakukan penyetaraan, dimana paket tes B, C, D, dan E disetarakan ke paket tes A dan dibuat kurva hasil penyetaraannya. Proses penyetaraan ini dilakukan dengan menyetarakan setiap paket ke paket A melalui perhitungan hasil konstanta dari hasil analisis dengan metode rerata-rerata, rerata-sigma, Haebara, dan Stocking Lord. Selain dilakukan perbandingan hasil penyetaraan dengan kurva karakteristik tes, dilakukan juga perbandingan hasil penyetaraan dengan melihat nilai RMSE pada setiap metode. Semakin kecil nilai RMSE semakin presisi tingkat pengukuran, termasuk pada penyetaraan (Retnawati, 2016).

HASIL DAN PEMBAHASAN

Hasil Estimasi Parameter Butir Perangkat Tes USBN

Hasil penelitian menunjukkan bahwa model IPL merupakan model parameter logistik yang paling cocok dengan data yang diestimasi. Model IPL menghasilkan estimasi parameter tingkat kesulitan butir seperti disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Tingkat Kesulitan Lima Paket Tes USBN Matematika

Deskripsi	Paket Tes A	Paket Tes B	Paket Tes C	Paket Tes D	Paket Tes E
Rata-rata	0,534	0,211	-0,117	1,075	0,436
Simpangan baku	2,037	1,451	1,959	2,396	1,998
Minimum	-2,542	-2,822	-4,657	-0,227	-0,726
Maksimum	6,448	3,506	4,675	6,555	4,663

Hasil estimasi parameter butir menunjukkan bahwa rata-rata tingkat kesulitan dari lima paket tes berkisar pada rentang -0,1 hingga 1 (Tabel 2). Dari lima paket tes tersebut, paket tes D merupakan paket tes yang memiliki rata-rata tingkat kesulitan yang paling tinggi dibandingkan dengan empat paket tes lainnya. Pada paket tes D tersebut juga terdapat butir soal dengan tingkat kesulitan paling tinggi, di mana indeks kesulitannya mencapai 6,555. Butir soal tersebut adalah butir yang mengukur capaian kompetensi transformasi geometri siswa (Gambar 1).

Bayangan garis  $x-2y+4=0$  oleh rotasi dengan pusat  $O(0,0)$  sejauh  $90^\circ$  berlawanan arah dengan arah putar jarum jam dilanjutkan dengan pencerminan terhadap garis  $y=x$  adalah ....

A.  $x+2y+4=0$    B.  $x+2y-4=0$    C.  $2x+y+4=0$    D.  $2x-y+4=0$    E.  $2x+y-4=0$

Gambar 1. Contoh Butir Sulit dalam Perangkat Tes USBN Matematika 2018/2019

Untuk menyelesaikan soal pada Gambar 1, tentu siswa harus memiliki pemahaman terhadap konsep, prinsip, dan prosedur yang baik terkait transformasi, rotasi, pencerminan terhadap suatu garis berbentuk  $y=mx+c$ , operasi pada matriks, trigonometri, dan komposisi transformasi geometri. Tentu dalam hal ini terdapat *element interactivity* yang sangat tinggi dalam mempelajari atau memecahkan masalah transformasi geometri pada soal tersebut. Tingginya *element interactivity* inilah yang diduga menyebabkan soal tersebut memiliki indeks kesulitan yang tinggi. Selain itu, [Albab et al. \(2014\)](#) juga menyatakan bahwa siswa masih kesulitan dalam memahami konsep transformasi geometri. Padahal, pemahaman terhadap konsep transformasi geometri merupakan hal penting dalam memecahkan masalah tersebut. Oleh karena itu, mengingat kesulitan siswa dalam memahami konsep transformasi geometri dan tingginya *element interactivity* pada materi transformasi geometri, guru dalam pembelajaran perlu memfasilitasi siswa untuk mengembangkan kemampuan koneksi matematisnya pada kompetensi transformasi geometri tersebut. Kemampuan koneksi matematis ini berkaitan dengan kemampuan siswa dalam memahami keterhubungan antara gagasan, konsep, representasi ([Eli et al., 2013](#); [García-García & Dolores-Flores, 2018](#)), definisi, teorema, dan prosedur matematika, serta keterkaitan antara matematika dengan disiplin ilmu lain atau kehidupan nyata ([García-García & Dolores-Flores, 2018](#)). Kemampuan ini penting untuk membangun pemahaman matematis siswa, dan sebaliknya, pemahaman matematis siswa juga perlu dikembangkan agar siswa memahami koneksi dalam matematika ([García-García & Dolores-Flores, 2018](#)). Dengan kata lain, kedua hal tersebut merupakan dua hal penting bagi siswa dalam memecahkan masalah, karena ketika siswa memiliki pemahaman konsep atau pemahaman matematis dan kemampuan koneksi yang rendah, mereka akan kesulitan dan melakukan kesalahan dalam menerapkan rumus, melakukan perhitungan, dan melakukan operasi dan manipulasi aljabar ([Hadi et al., 2018](#)). Untuk mengembangkan koneksi matematis siswa, guru dapat menerapkan model, metode, atau strategi pembelajaran, seperti *Relating, Experiencing, Applying, Cooperating, dan Transferring* (REACT) ([Putri & Santosa, 2015](#)) dan *Problem-Based Learning* (PBL) ([Kartikasari & Widjajanti, 2017](#); [Malasari et al., 2017](#)) atau dengan memanfaatkan teknologi berupa program *GeoGebra* ([Zengin, 2019](#)).

Keberhasilan siswa dalam memecahkan masalah maupun prestasi belajarnya tentu tidak hanya ditentukan oleh pemahaman matematis dan kemampuan koneksi matematis siswa. Ada banyak faktor lain yang dapat mempengaruhi kemampuan pemecahan masalah atau prestasi belajar siswa. Salah satu faktor tersebut adalah motivasi siswa ([Fane & Sugito, 2019](#); [Kurniawan & Wustqa, 2014](#); [Shores &](#)

Shannon, 2007). Motivasi ini dapat berupa motivasi intrinsik dan ekstrinsik. Menurut Pintrich et al. (1991) motivasi intrinsik berkaitan dengan dorongan dalam diri untuk mengerjakan suatu tugas sebagai akibat adanya tantangan, rasa ingin tahu, dan keinginan untuk menjadi lebih berpengetahuan dan terampil. Adapun motivasi ekstrinsik berkaitan dengan keinginan untuk melaksanakan suatu tugas sebagai akibat adanya dorongan dari luar diri seseorang yang dapat berupa nilai, penghargaan, kompetisi, dan pengakuan dari orang lain. Bagaimanapun juga, kedua jenis motivasi ini penting bagi siswa dalam menghadapi ujian, seperti ujian nasional ataupun ujian sekolah berstandar nasional, karena dengan motivasi yang tinggi siswa akan belajar lebih giat sehingga siap dalam menghadapi ujian tersebut (Retnawati et al., 2019).

### Hasil Penyetaraan Perangkat Tes USBN

Hasil estimasi parameter butir yang telah diperoleh, selanjutnya dijadikan dasar dalam melakukan analisis *equating* atau penyetaraan. Hasil estimasi penyetaraan perangkat tes USBN 2018/2019 untuk mata pelajaran matematika wajib disajikan pada Tabel 3. Setelah hasil estimasi penyetaraan antar-tes didapatkan kemudian dilakukan penyetaraan tes dengan memasukkan nilai konstanta  $\alpha$  dan  $\beta$  pada persamaan setiap metode. Penyetaraan dilakukan hanya pada paket tes B, C, D, dan E yang disetarakan ke tes A. Hasil penyetaraan paket tes B, C, D, dan E ke paket tes A disajikan pada Tabel 4. Hasil penyetaraan yang telah disajikan pada Tabel 4 kemudian digunakan untuk menggambar kurva karakteristik tes seperti pada Gambar 1, Gambar 2, Gambar 3, dan Gambar 4.

Gambar 1 menunjukkan bahwa penyetaraan dengan metode *mean-mean* menghasilkan kurva yang saling berdekatan. Hal ini mengindikasikan bahwa instrumen paket tes B, C, D, dan E yang diujikan kepada siswa pada USBN hampir memiliki karakteristik tingkat kesulitan yang setara satu sama lain. Ketika *equating* dilakukan dengan menggunakan metode *mean-sigma*, kurva karakteristik dari empat tes tersebut juga saling berdekatan satu sama lain (Gambar 2). Hal tersebut menunjukkan bahwa keempat paket tes tersebut memiliki karakteristik yang hampir sama. Lebih lanjut, bila kita cermati secara mendalam, dapat diperoleh informasi bahwa kurva karakteristik tes paket B cenderung mendekati berhimpit dengan kurva karakteristik tes paket C. Adapun kurva karakteristik tes paket D memiliki kecenderungan untuk berhimpit dengan kurva karakteristik tes paket E. Hasil ini mengindikasikan bahwa paket tes B cenderung lebih setara dengan paket tes C, sedangkan paket tes D cenderung setara dengan paket tes E.

Selanjutnya, ketika penyetaraan dilakukan dengan menggunakan metode Haebara diperoleh kurva karakteristik tes sebagaimana yang tersaji dalam Gambar 3. Gambar 3 menunjukkan bahwa kurva karakteristik tes keempat paket tes yang diujikan kepada siswa pada USBN mendekati berhimpit. Bahkan jarak antara kurva karakteristik tes paket yang satu dengan yang lainnya yang dihasilkan dengan metode Haebara ini lebih rapat dari pada yang dihasilkan dengan menggunakan metode *mean-mean*. Hubungan keempat kurva karakteristik tes yang demikian menunjukkan bahwa keempat paket tes hampir setara satu sama lainnya. Lebih lanjut, pola kurva karakteristik yang dihasilkan dengan metode Haebara ini memiliki kemiripan dengan yang dihasilkan dengan metode *mean-sigma*, di mana terdapat kecenderungan bahwa kurva karakteristik tes paket tes B sangat mendekati berhimpit dengan kurva karakteristik tes paket tes C dan kurva karakteristik tes paket tes D sangat mendekati berhimpit dengan kurva karakteristik tes paket tes E.

Terakhir, penyetaraan dengan menggunakan metode Stocking Lord juga menghasilkan kurva karakteristik tes yang hampir berhimpit satu sama lain (Gambar 4). Karakteristik kurva tes yang demikian menunjukkan bahwa keempat paket tes tersebut memiliki karakteristik tingkat kesulitan yang hampir setara. Dari keempat metode yang telah dilakukan dalam analisis *equating*, secara grafis, metode Haebara merupakan metode yang menghasilkan grafik yang paling rapat dibandingkan tiga metode lainnya.

Hasil *equating* berdasarkan kurva karakteristik tes selanjutnya dikonfirmasi dengan menggunakan RMSE. Dari perhitungan terhadap RMSE untuk keempat metode *equating*, metode Haebara menghasilkan rata-rata nilai RMSE paling kecil dibandingkan dengan rata-rata RMSE pada tiga metode lainnya (lihat Tabel 5). Hal tersebut membuktikan bahwa penyetaraan dengan metode Haebara merupakan metode yang paling akurat untuk mengestimasi kesetaraan keempat paket tes yang diujikan dalam USBN. Temuan ini tentunya berbeda dengan temuan yang diperoleh Retnawati (2016) ketika meneliti kesetaraan skor 20 perangkat tes ujian akhir SMP mata pelajaran matematika. Hasil penelitiannya me-

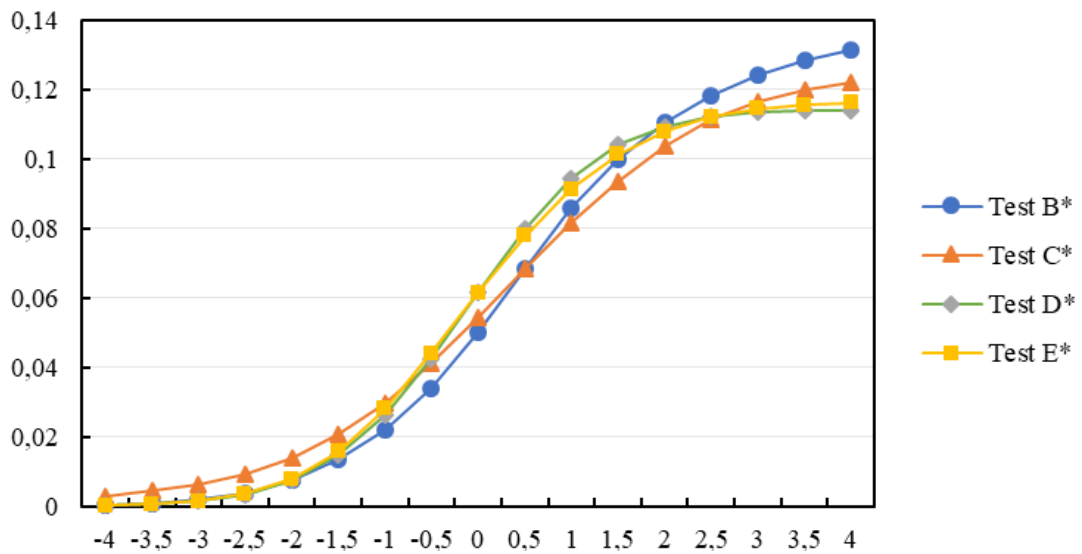
nunjukkan bahwa ketika membandingkan penyeteraan dengan butir bersama, RMSE yang dihasilkan pada metode Stocking Lord merupakan yang paling kecil.

Tabel 3. Hasil Estimasi Penyeteraan Antar Paket Tes dengan Empat Metode

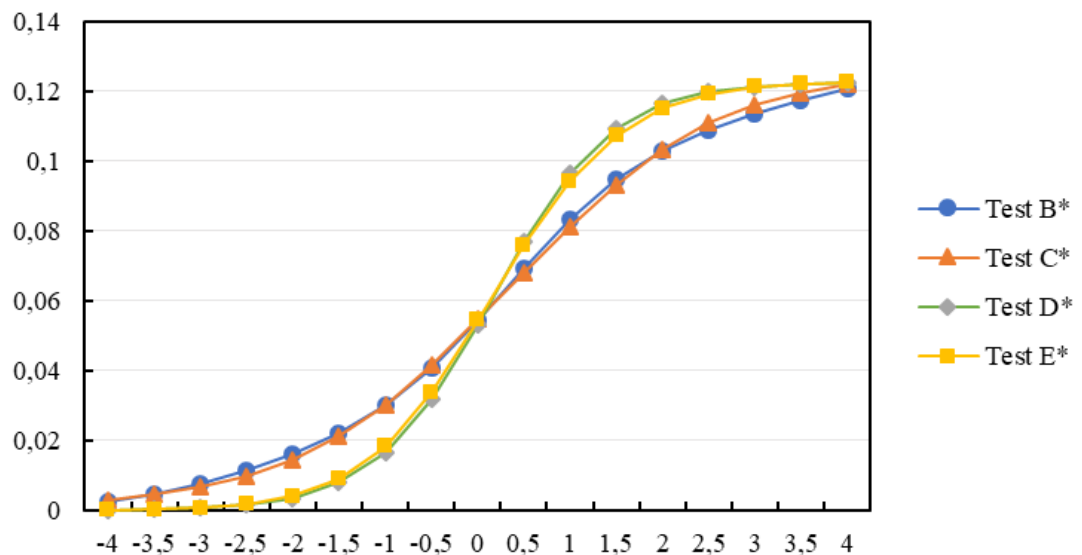
Link test	Const.	Metode							
		Mean-Mean		Mean-Sigma		Haebara		Stocking-Lord	
		Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
A ke B	$\alpha$	1,146	0,158	0,712	0,114	0,930	0,131	1,052	0,143
	$\beta$	-0,401	0,188	-0,170	0,140	0,056	0,156	-0,167	0,162
A ke C	$\alpha$	0,977	0,127	0,962	0,149	0,800	0,111	1,011	0,129
	$\beta$	-0,639	0,173	-0,630	0,161	-0,279	0,141	-0,525	0,157
A ke D	$\alpha$	0,653	0,082	1,177	12,267	0,468	0,064	0,566	0,071
	$\beta$	0,727	4,797	0,447	1,758	0,619	0,126	0,520	0,123
A ke E	$\alpha$	0,610	0,066	0,981	2,149	0,413	0,049	0,541	0,058
	$\beta$	0,110	0,862	-0,088	0,316	0,037	0,083	-0,045	0,090
B ke A	$\alpha$	0,872	0,120	1,404	0,225	0,832	0,116	0,948	0,128
	$\beta$	0,350	0,162	0,238	0,198	0,127	0,147	0,159	0,154
B ke C	$\alpha$	0,853	0,114	1,350	0,187	0,797	0,110	0,971	0,128
	$\beta$	-0,297	0,148	-0,402	0,190	-0,302	0,141	-0,366	0,154
B ke D	$\alpha$	0,569	0,074	1,652	17,220	0,421	0,057	0,532	0,068
	$\beta$	0,955	4,797	0,727	1,189	0,650	0,121	0,616	0,123
B ke E	$\alpha$	0,533	0,060	1,377	3,014	0,413	0,047	0,519	0,058
	$\beta$	0,324	0,861	0,146	0,280	0,092	0,082	0,050	0,089
C ke A	$\alpha$	1,023	0,133	1,040	0,161	0,739	0,103	0,979	0,125
	$\beta$	0,654	0,177	0,655	0,182	0,457	0,142	0,517	0,155
C ke B	$\alpha$	1,173	0,157	0,741	0,102	0,801	0,116	1,032	0,136
	$\beta$	0,348	0,174	0,297	0,143	0,403	0,150	0,377	0,160
C ke D	$\alpha$	0,668	0,081	1,223	12,756	0,378	0,051	0,549	0,066
	$\beta$	1,153	4,798	1,218	6,293	0,839	0,127	0,815	0,129
C ke E	$\alpha$	0,625	0,065	1,020	2,233	0,376	0,042	0,533	0,054
	$\beta$	0,509	0,861	0,555	1,125	0,264	0,082	0,243	0,088
D ke A	$\alpha$	1,532	0,193	0,850	8,863	1,863	0,231	1,787	0,222
	$\beta$	-1,113	7,351	-0,380	5,451	-0,813	0,228	-0,938	0,224
D ke B	$\alpha$	1,756	0,228	0,606	6,313	1,799	0,236	1,885	0,244
	$\beta$	-1,677	8,427	-0,440	3,884	-0,948	0,231	-1,157	0,242
D ke C	$\alpha$	1,497	0,182	0,817	8,523	1,701	0,210	1,825	0,220
	$\beta$	-1,727	7,184	-0,996	5,243	-1,348	0,231	-1,486	0,241
D ke E	$\alpha$	0,935	0,092	0,834	8,883	0,973	0,096	0,970	0,095
	$\beta$	-0,569	4,568	-0,460	5,459	-0,549	0,119	-0,548	0,119
E ke A	$\alpha$	1,638	0,177	1,019	2,233	1,805	0,190	1,842	0,196
	$\beta$	-0,180	1,412	0,089	0,165	0,266	0,165	0,076	0,165
E ke B	$\alpha$	1,878	0,212	0,726	1,590	1,800	0,204	1,940	0,217
	$\beta$	-0,608	1,617	-0,106	0,143	0,082	0,166	-0,087	0,172
E ke C	$\alpha$	1,601	0,166	0,980	2,146	1,711	0,177	1,874	0,191
	$\beta$	-0,815	1,380	-0,544	0,164	-0,307	0,157	-0,449	0,167
E ke D	$\alpha$	1,069	0,105	1,199	12,777	0,996	0,099	1,036	0,102
	$\beta$	0,609	4,884	0,552	0,678	0,578	0,127	0,566	0,127

Tabel 4. Penyeteraan Tes dengan Tes A dengan Empat Metode

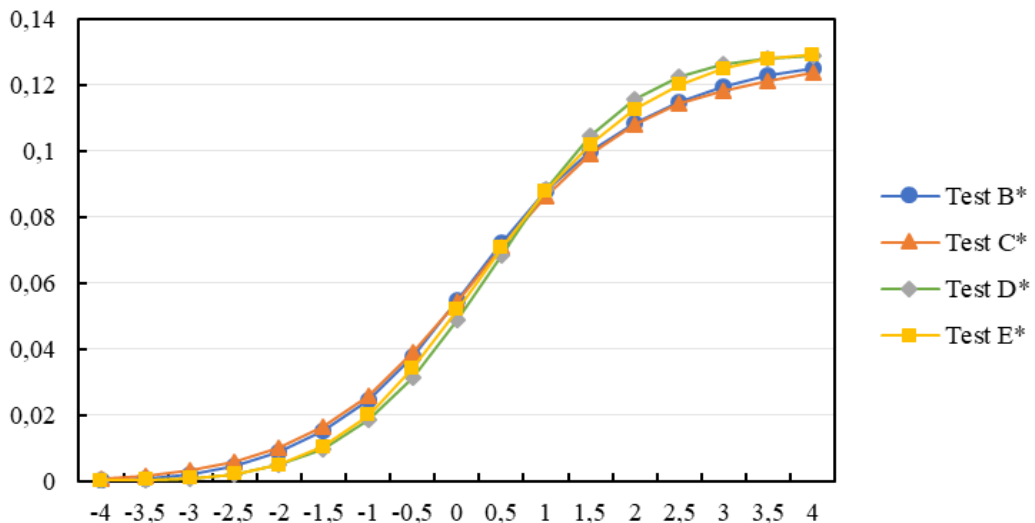
Link	Metode			
	Mean-Mean	Mean-Sigma	Haebara	Stocking-Lord
B ke A	$b^* = b1 + 0,350$	$b^* = 1,404 b1 + 0,238$	$b^* = 0,832 b1 + 0,127$	$b^* = 0,948 b1 + 0,159$
C ke A	$b^* = b1 + 0,654$	$b^* = 1,040 b1 + 0,655$	$b^* = 0,739 b1 + 0,457$	$b^* = 0,979 b1 + 0,517$
D ke A	$b^* = b1 - 1,113$	$b^* = 0,850 b1 - 0,380$	$b^* = 1,863 b1 - 0,813$	$b^* = 1,787 b1 - 0,938$
E ke A	$b^* = b1 - 0,180$	$b^* = 1,019 b1 + 0,089$	$b^* = 1,805 b1 + 0,266$	$b^* = 1,842 b1 + 0,076$



Gambar 1. Kurva Karakteristik Tes yang Disetarakan ke Tes A dengan Metode *Mean-Mean*.

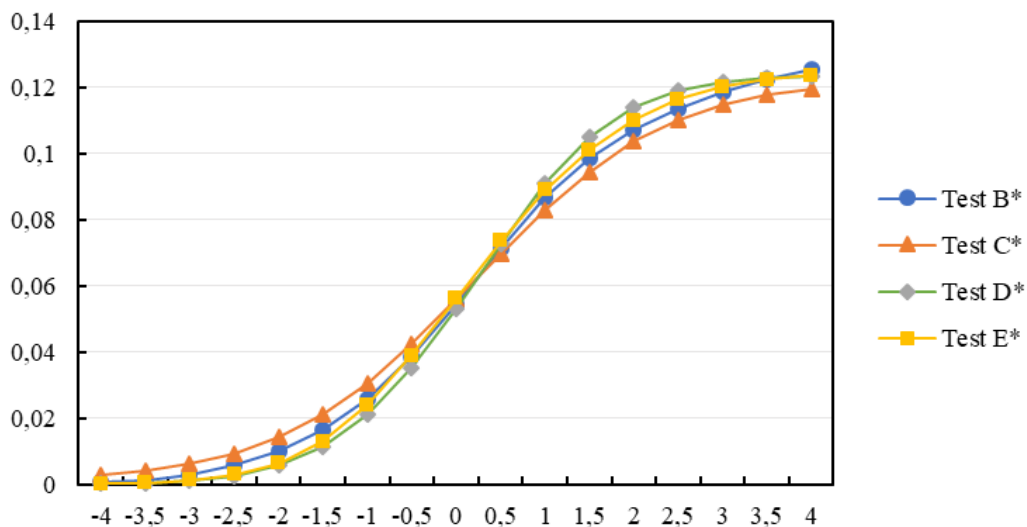


Gambar 2. Kurva Karakteristik Tes yang disetarakan ke Tes A dengan Metode *Mean-Sigma*



Gambar 3. Kurva Karakteristik Tes yang Disetarakan ke Tes A dengan Metode Haebara





Gambar 4. Kurva Karakteristik Tes yang disetarakan ke Tes A dengan Metode Stocking Lord

Tabel 5. RMSE pada Empat Metode Penyetaraan

Paket Tes	Metode			
	Mean-Mean	Mean-Sigma	Haebara	Stocking Lord
B	0,371	0,660	0,257	0,166
C	0,652	0,655	0,700	0,521
D	1,366	0,934	0,524	1,131
E	1,257	1,370	0,603	1,273
Rerata	0,912	0,905	0,521	0,773

Secara umum, berdasarkan *equating* yang dilakukan terhadap lima paket tes USBN tahun 2018/2019 untuk mata pelajaran matematika wajib dari empat sekolah di Daerah Istimewa Yogyakarta dan Kalimantan Selatan melalui perbandingan kurva karakteristik tes, dapat disimpulkan bahwa kelima paket tes yang digunakan mendekati setara. Hal tersebut menunjukkan bahwa kelima paket tes yang diujikan kepada siswa dalam USBN sudah sesuai dengan prosedur operasional standar USBN yang ditetapkan oleh Badan Standar Nasional Pendidikan (2018), yaitu paket soal USBN merupakan variasi naskah soal yang bersifat paralel. Yang dimaksud paralel di sini tentu adalah kesetaraan karakteristik dari paket-paket soal yang digunakan dalam USBN. Penelitian-penelitian terdahulu (misalnya Kartowagiran et al., 2018; Retnawati, 2016) juga mengungkapkan bahwa paket-paket soal yang diujikan dalam ujian berstandar nasional, salah satunya adalah Ujian Nasional (UN), sudah memenuhi syarat kesetaraan karakteristik. Lebih lanjut, adanya temuan bahwa kelima paket soal mata pelajaran matematika wajib yang digunakan dalam USBN memiliki karakteristik yang hampir setara, mengindikasikan bahwa tidak ada kesenjangan yang berarti terkait kemampuan guru di daerah dalam mengembangkan paket soal USBN. Selain itu, hasil penelitian yang menunjukkan bahwa kelima paket soal matematika wajib yang diujikan dalam USBN memiliki karakteristik yang hampir setara ini tentu diharapkan tidak menurunkan motivasi siswa di masa mendatang dalam menghadapi dan mengerjakan ujian tersebut. Dengan adanya kesetaraan perangkat USBN yang digunakan, diharapkan tidak ada siswa yang merasa dirugikan dengan adanya paket soal yang berbeda-beda atau ada anggapan bahwa nilai yang jelek dikarenakan mendapat paket soal yang sangat sulit dan sebaliknya mendapat nilai yang bagus dikarenakan mendapat paket soal yang sangat mudah (Herkusumo, 2011; Sukirno, 2007).

Hasil penyetaraan ini bisa berdampak pada penilaian pembelajaran. Penilaian pembelajaran yang baik didapat dari proses penilaian yang terencana dan didukung dengan instrumen yang tepat dan sesuai (Black & Wiliam, 2018). Proses penilaian memerlukan suatu instrumen yang dapat melakukan penilaian dengan baik dan sesuai tujuan penilaian. Dalam penilaian pembelajaran, kegiatan penilaian perlu diarahkan pada usaha penelusuran terhadap kegiatan dan capaian belajar siswa, pemeriksaan atau penilaian terhadap pemahaman belajar siswa, dan pencarian terhadap apa yang telah diperoleh siswa dan kendala yang mungkin mereka alami dalam belajar (Chittenden, 1991). Tiga usaha tersebut selanjutnya menurut Engel (Chittenden, 1991) diakhiri dengan penyimpulan yang berupa pelaporan atau pertanggungjawab-

an terhadap capaian belajar siswa dalam pembelajaran di sekolah kepada orang tua dan siswa itu sendiri. Dengan demikian, dalam melakukan penilaian instrumen yang digunakan harus mampu menggali informasi tentang siswa, sehingga dapat melakukan penyimpulan yang tepat tentang kemampuan siswa. Kemudian yang perlu diperhatikan pada penilaian pembelajaran yaitu prinsip penilaian pembelajaran, di mana di antaranya adalah penilaian yang adil. Jadi, instrumen yang digunakan harus tepat tujuan dan adil. Pada hal ini tentunya hasil dari penilaian haruslah mempunyai kesetaraan agar prinsip adil dalam penilaian dapat terwujud. Hasil penyetaraan soal USBN ini juga memberikan dampak pada proses penilaian pembelajaran yaitu dapat menjadi dasar dalam melakukan penilaian pembelajaran.

Untuk menghasilkan paket-paket soal yang memiliki karakteristik yang setara atau mendekati setara tentu harus memperhatikan beberapa hal dalam pengembangan instrumen tes. Instrumen tes yang dikembangkan tentunya harus memperhatikan kualitas dan karakteristik instrumen agar tujuan dari instrumen tersebut tercapai (Azwar, 2012). Tentunya ini juga berlaku dalam mengembangkan instrumen tes yang mempunyai butir *anchor*. Menurut Retnawati (2016), ada dua hal yang perlu diperhatikan dalam melakukan penyetaraan, yaitu proses estimasi parameter dan proses estimasi penyetaraan. Hal-hal yang perlu diperhatikan terkait dengan proses estimasi parameter meliputi banyaknya responden, banyak butir, dan metode estimasi. Adapun hal yang perlu diperhatikan terkait dengan proses estimasi penyetaraan meliputi distribusi parameter butir, distribusi parameter kemampuan, metode, banyak butir, dan *software* yang digunakan.

### SIMPULAN

Analisis *penyetaraan* berdasarkan teori respons butir menggunakan metode *mean-mean*, *mean-sigma*, Haebara, dan Stocking Lord menunjukkan bahwa lima paket soal matematika wajib USBN tahun 2018/2019 untuk mata pelajaran matematika wajib dari empat sekolah di Daerah Istimewa Yogyakarta dan Kalimantan Selatan memiliki karakteristik yang cenderung setara. Dengan demikian, kelima paket soal tersebut sudah memenuhi syarat keparalelan atau kesetaraan sebagaimana disebutkan dalam prosedur operasional standar USBN. Hasil penyetaraan paling akurat dihasilkan menggunakan metode Haebara yang menghasilkan rata-rata RMSE paling kecil dibandingkan dengan tiga metode lainnya. Hasil penelitian memberikan implikasi bahwa siswa tidak perlu lagi beranggapan bahwa mereka dirugikan dengan adanya paket-paket soal yang berbeda. Tentu hal ini diharapkan tidak menurunkan motivasi siswa di masa mendatang dalam belajar dan mempersiapkan diri menghadapi ujian yang sejenis. Selain itu, penelitian ini menunjukkan bahwa soal yang memerlukan kemampuan koneksi matematis menjadi soal yang paling sulit. Hal ini berimplikasi agar proses pembelajaran ke depannya perlu memfasilitasi siswa untuk mengembangkan kemampuan koneksi matematisnya melalui penerapan model, metode, atau strategi yang potensial ataupun dengan memanfaatkan teknologi pembelajaran matematika yang relevan.

### DAFTAR PUSTAKA

- Albab, I. U., Hartono, Y., & Darmawijoyo, D. (2014). Kemajuan belajar siswa pada geometri transformasi menggunakan aktivitas refleksi geometri. *Jurnal Cakrawala Pendidikan*, 3(3). <https://doi.org/10.21831/cp.v3i3.2378>
- Antara, A. A. P., & Bastari, B. (2015). Penyetaraan vertikal dengan pendekatan klasik dan item response theory pada siswa sekolah dasar. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 13–24. <https://doi.org/10.21831/pep.v19i1.4551>
- Aşiret, S., & Sünbül, S. Ö. (2016). Investigating test equating methods in small samples through various factors. *Educational Sciences: Theory & Practice*, 16(2), 647–668. <https://doi.org/10.12738/estp.2016.2.2762>
- Azwar, S. (2012). *Validitas dan reabilitas* (4th ed.). Pustaka Pelajar.
- Badan Standar Nasional Pendidikan. (2018). *Prosedur operasional standar penyelenggaraan ujian sekolah bestandar nasional (POS USBN)*.
- Battauz, M. (2015). equateIRT: An R package for IRT test equating. *Journal of Statistical Software*, 68(7), 1–22. <https://doi.org/10.18637/jss.v068.i07>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 1–25. <https://doi.org/10.1080/0969594X.2018.1441807>

- Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. <https://doi.org/10.18637/jss.v048.i06>
- Chittenden, E. (1991). Authentic assessment, evaluation, and documentation. In V. Perrone (Ed.), *Expanding student assessment* (pp. 22–31). Association for Supervision and Curriculum Development.
- Crocker, L. M., & Algina, J. (2006). *Introduction to classical and modern test theory* (2nd ed.). Cengage Learning.
- Eli, J. A., Mohr-Schroeder, M. J., & Lee, C. W. (2013). Mathematical connections and their relationship to mathematics knowledge for teaching geometry. *School Science and Mathematics*, 113(3), 120–134.
- Fane, A., & Sugito, S. (2019). Pengaruh keterlibatan orang tua, perilaku guru, dan motivasi belajar terhadap prestasi belajar matematika siswa. *Jurnal Riset Pendidikan Matematika*, 6(1), 53–61. <https://doi.org/10.21831/jrpm.v6i1.15246>
- García-García, J., & Dolores-Flores, C. (2018). Intra-mathematical connections made by high school students in performing Calculus tasks. *International Journal of Mathematical Education in Science and Technology*, 49(2), 227–252. <https://doi.org/10.1080/0020739X.2017.1355994>
- Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. *Problems of Education in the 21st Century*, 76(4), 520–532.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144–149. <https://doi.org/10.4992/psycholres1954.22.144>
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer Science+Business Media. <https://doi.org/10.1007/978-94-017-1988-9>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Herkusumo, A. P. (2011). Penyetaraan (equating) Ujian Akhir Sekolah Berstandar Nasional (UASBN) dengan teori tes klasik. *Jurnal Pendidikan Dan Kebudayaan*, 17(4), 455–471. <https://doi.org/10.24832/jpnk.v17i4.41>
- Iriyadi, D., Naga, D. S., & Rahayu, W. (2019). Equating method for prevent discrimination in classroom. *Journal of Educational Science and Technology*, 5(2), 100–109. <https://doi.org/10.26858/est.v5i2.9258>
- Kartikasari, A., & Widjajanti, D. B. (2017). The effectiveness of problem-based learning approach based on multiple intelligences in terms of student's achievement, mathematical connection ability, and self-esteem. *Journal of Physics: Conference Series*, 812(1), 012097. <https://doi.org/10.1088/1742-6596/812/1/012097>
- Kartowagiran, B., Munadi, S., Retnawati, H., & Apino, E. (2018). The equating of battery test packages of Mathematics National Examination 2013-2016. *SHS Web of Conferences*, 42(1), 1–6. <https://doi.org/10.1051/shsconf/20184200022>
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. Springer-Verlag New York. <https://doi.org/10.1007/978-1-4757-2412-7>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. <https://doi.org/10.1007/978-1-4939-0317-7>
- Kurniawan, D., & Wustqa, D. U. (2014). Pengaruh perhatian orangtua, motivasi belajar, dan lingkungan sosial terhadap prestasi belajar matematika siswa SMP. *Jurnal Riset Pendidikan Matematika*, 1(2), 176. <https://doi.org/10.21831/jrpm.v1i2.2674>
- Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17(3), 179–193. <https://doi.org/10.1111/j.1745-3984.1980.tb00825.x>
- Malasari, P. N., Nindiasari, H., & Jaenudin, J. (2017). Preface: International Conference on Recent Trends in Physics (ICRTP 2016). *Journal of Physics: Conference Series*, 812(1), 1–6. <https://doi.org/10.1088/1742-6596/812/1/012025>

- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14(2), 139–160. <https://doi.org/10.1111/j.1745-3984.1977.tb00033.x>
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. National Center for Research to Improve Post secondary Teaching and Learning.
- Rahayu, W. (2015). Metode estimasi parameter dan metode equating pada ukuran sampel kecil berdasarkan item respons theory. *Prosiding Semirata 2015 Bidang MIPA BKS-PTN Barat*, 315–324.
- Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Nuha Medika.
- Retnawati, H. (2016). Perbandingan metode penyetaraan skor tes menggunakan butir bersama dan tanpa butir besama. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, 46(2), 164–178. <https://doi.org/10.21831/jk.v46i2.10383>
- Retnawati, H., Hadi, S., Munadi, S., Hadiana, D., Muhardis, M., Apino, E., Djidu, H., Rafi, I., Yusron, E., & Rosyada, M. N. (2019). When national examination no longer determining graduation, will students accomplish it seriously? *Indonesian Journal of Educational Assesment (IJE)*, 2(2), 40–49. <https://doi.org/10.26499/ijea.v2i2.34>
- Rijanto, T. (2011). Metode penyetaraan skor dan ukuran sampel. *Jurnal Evaluasi Pendidikan*, 2(1), 101–114. <https://doi.org/10.21009/JEP>
- Rizopoulos, D. (2006). ltm : An R package for latent variable modeling. *Journal of Statistical Software*, 17(5). <https://doi.org/10.18637/jss.v017.i05>
- Setiawan, R. (2019). A comparison of score equating conducted using Haebara and Stocking Lord method for polytomous. *European Journal of Educational Research*, 8(4), 1071–1079. <https://doi.org/10.12973/eu-jer.8.4.1071>
- Shores, M. L., & Shannon, D. M. (2007). The effects of self-regulation, motivation, anxiety, and attributions on mathematics achievement for fifth and sixth grade students. *School Science and Mathematics*, 107(6), 225–236. <https://doi.org/10.1111/j.1949-8594.2007.tb18284.x>
- Sukirno, S. (2007). Penyetaraan tes UAN: Mengapa dan bagaimana? *Cakrawala Pendidikan*, 26(3), 305–321. <https://doi.org/10.21831/cp.v3i3.8576>
- Uysal, İ., & Kilmen, S. (2016). Comparison of item response theory test equating methods for mixed format tests. *International Online Journal of Educational Sciences*, 8(2), 1–11. <https://doi.org/10.15345/iojes.2016.02.001>
- Zengin, Y. (2019). Development of mathematical connection skills in a dynamic learning environment. *Education and Information Technologies*, 24(3), 2175–2194. <https://doi.org/10.1007/s10639-019-09870-x>