

JURNAL STATISTIKA DAN SAINS DATA
Volume 4 Edisi 1, Oktober, 2026, 173-188
https://journal.student.uny.ac.id/index.php/jssd
https://dx.doi.org/10.21831/jssd.v4i1.21888

PENYETARAAN PERANGKAT TES MENGGUNAKAN METODE REGRESI DAN RERATA DAN SIGMA PADA TEORI RESPON BUTIR

Nurrokhim*, Universitas Negeri Yogyakarta
Heri Retnawati, Universitas Negeri Yogyakarta
*e-mail: Nurrokhim.2019@student.uny.ac.id, Heriretnawati7@uny.ac.id

Abstrak. Penelitian ini bertujuan untuk menganalisis karakteristik instrument tes UN bidang Bahasa Indonesia 2016, menganalisis kesetaraan instrumen tes UN bidang Bahasa Indonesia menggunakan metode regresi dan rerata dan sigma. Data yang digunakan adalah data UN Bahasa Indonesia tahun 2016 seluruh DIY. Adapun sumber data berasal dari Universitas Negeri Yogyakarta. Data diolah dengan *R studio* dan dianalisis dengan langkah sebagai berikut: menyiapkan data, uji asumsi, melakukan pemodelan ndata, membandingkan kecocokan butir, menentukan model terbaik. Melakukan estimasi parameter butir, dan melakukan penyetaraan. Hasil temuan menunjukkan bahwa hasil jumlah butir fit untuk paket 1 berjumlah 49, paket 2 berjumlah 45, paket 3 berjumlah 47, paket 4 berjumlah 48, dan paket 5 berjumlah 48. Adapun dengan menggunakan analisis IRT diperoleh model 3PL merupakan model yang terbaik. Hasil penyetaraan menggunakan metode rerata dan sigma yang dimodelkan dengan regresi ialah $\theta_{21} *= 0{,}78\theta - 0{,}39$ untuk persamaan kesetaraan skor paket 2 ke paket 1, $\theta_{31} *= 0{,}57\theta - 0{,}93$ untuk persamaan kesetaraan skor paket 3 ke paket 1, selanjutnya $\theta_{31} *= 0{,}7\theta - 0{,}49$ untuk persamaan kesetaraan skor paket 4 ke paket 1. Dalam penyetaraan juga didapatkan nilai RMSE untuk paket 2 ke 1 sebesar 3,548796, paket 3 ke 1 sebesar 7,759861 dan paket 4 ke 1 sebesar 4,582123.
Kata kunci: Penyertaan perangkat tes, metode regresi, rerata dan sigma, teori respons butir.

Abstract. *This research aims to analyze the characteristics of the 2016 Indonesian Language UN test instrument, analyze the equivalence of the Indonesian Language UN test instrument using the regression method and the mean and sigma. The data used is the 2016 Indonesian Language UN data throughout Yogyakarta. The data source comes from Yogyakarta State University. The data was processed with R studio and analyzed with the following steps: preparing data, testing assumptions, modeling data, comparing item suitability, determining the best model. Perform item parameter estimation, and perform equating. The findings showed that the results of the number of fit items for package 1 amounted to 49, package 2 amounted to 45, package 3 amounted to 47, package 4 amounted to 48, and package 5 amounted to 48. As for using IRT analysis, the 3PL model was obtained as the best model. The results of equating using the mean and sigma method modeled by regression are* $\theta_{21} *= 0{,}78\theta - 0{,}39$ *for the equality of package 2 scores to package 1,* $\theta_{31} *= 0{,}57\theta - 0{,}93$ *for the equality of package 3 scores to package 1, then* $\theta_{41} *= 0{,}7\theta - 0{,}49$ *for the equality of package 4 scores to package 1. In the equation, the RMSE value for package 2 to 1 is 3.548796, package 3 to 1 is 7.759861 and package 4 to 1 is 4.582123.*
Keywords: Test set inclusion, regression method, mean and sigma, item response theory.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

PENDAHULUAN

Dalam dunia pendidikan, kegiatan penilaian hasil belajar peserta didik merupakan salah satu tugas penting yang harus dilakukan oleh pendidik (Zainal, 2020). Evaluasi hasil belajar dilakukan untuk mengetahui keberhasilan dari sebuah proses pembelajaran. Evaluasi juga berfungsi sebagai pertimbangan bagi pendidik dalam memperbaiki dan menyempurnakan program dan kegiatan pembelajaran. Istilah yang digunakan dalam sistem evaluasi adalah pengukuran dan penilaian, yang meskipun berbeda, namun saling berhubungan. Pengukuran adalah membandingkan sesuatu dengan satu ukuran (kuantitatif), sedangkan penilaian adalah mengambil suatu keputusan terhadap sesuatu dengan ukuran baik atau buruk (kualitatif).

Salah satu komponen wajib dalam evaluasi adalah instrumen tes (Andayani & Ramalis, 2019). Instrumen tes adalah teknik penilaian dengan butir-butir pertanyaan atau pernyataan yang dikerjakan oleh peserta didik untuk mengetahui kemampuan siswa. Setelah tahap validasi isi, alat instrumen akan memasuki tahap validasi empiris dengan menguji beberapa siswa. Analisis karakteristik tes dapat mengidentifikasi butir tes yang baik dan buruk. Analisis butir soal dalam proses belajar mengajar dapat menggunakan tes yang telah distandardisasikan maupun tes buatan pendidik. Tes yang telah distandarisasikan telah mengalami proses validitas dan reliabilitas, sehingga tes tersebut valid dan reliabel untuk tujuan tertentu.

Perangkat soal yang baik seharusnya melalui tahap telaah kualitatif dan empiris sebelum digunakan untuk mengukur kemampuan siswa (Haryani, 2015). Analisis butir soal secara kualitatif menilai butir soal dari aspek materi, konstruksi, dan bahasa berdasarkan kaidah penulisan soal. Penelaahan soal secara empiris didasarkan pada bukti empiris untuk mengetahui sejauh mana butir soal membedakan antara mereka yang tinggi dan rendah kemampuannya. Analisis soal secara empiris bisa menggunakan pendekatan teori tes klasik dan teori respons butir.

Ujian Nasional (UN) adalah salah satu bentuk evaluasi standar pendidikan dasar dan menengah secara nasional. Berdasarkan Undang-Undang No. 21 Tahun 2003, evaluasi pendidikan adalah kegiatan pengendalian, penjaminan, dan penetapan mutu pendidikan terhadap berbagai komponen pendidikan. Pelaksanaan UN menggunakan lebih dari satu paket soal. Pada tahun 2013, UN menggunakan 4 paket soal; tahun 2014 menggunakan 20 paket soal; dan tahun 2015 kembali menggunakan 4 paket soal. Meskipun paket soal ini dibuat berdasarkan kisi-kisi soal yang sama, perbedaan paket soal memungkinkan terjadinya perbedaan karakteristik, sifat, dan tingkat kesukaran yang diukur (Hambleton & Swaminathan, 1985).

Penyetaraan skor digunakan sebagai teknik untuk membedakan kemampuan peserta didik. Penyetaraan skor memungkinkan penggunaan perangkat tes yang berbeda terhadap kelompok yang berbeda sesuai standar kemampuannya sehingga hasil yang diperoleh dapat dibandingkan dan peserta didik tidak merasa dirugikan atau diuntungkan oleh perbedaan tingkat kesukaran perangkat tes. Menurut Harris (2016), penyetaraan tes memiliki beberapa keuntungan utama: memungkinkan penggunaan perangkat tes berbeda yang hasilnya dapat dibandingkan, mengatasi kebocoran tes dengan mengganti perangkat tes, dan memberikan fleksibilitas lingkungan dan waktu.

Menurut Hambleton dan Swaminathan (1985), penyetaraan tes harus memperhatikan beberapa hal seperti kesamaan sifat dan kemampuan yang diukur, reabilitas tes, variasi tingkat kesukaran butir, dan bukti bahwa perangkat tes tersebut paralel. Penyetaraan dapat dibuktikan secara statistik menggunakan teori tes klasik atau teori respons butir. Teori respons butir menyetarakan skor dari dua atau lebih bentuk tes paralel ke skala skor umum menggunakan metode seperti rerata dan sigma, serta metode kurva karakteristik butir (Hambleton et al., 1991; Kolen & Brenan, 2004).Metode regresi juga digunakan untuk menyetarakan perangkat

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

instrumen tes dengan mengkonversi parameter daya beda dan tingkat kesukaran butir dari satu perangkat tes ke perangkat tes lain. Hal ini bertujuan untuk mengetahui bagaimana parameter daya beda dan tingkat kesukaran setiap butir soal pada kelompok peserta tes tertentu seandainya diberikan perangkat tes lain.

Proses penyetaraan sangat penting mengingat instrumen tes UN yang digunakan adalah instrumen yang berbeda. Meskipun dalam pengembangan instrumen tes UN terdapat butir bersama, hal ini belum tentu mencirikan bahwa instrumen tersebut setara. Oleh karena itu, diperlukan proses penyetaraan instrumen tes UN dan perbandingan penyetaraan untuk hasil yang lebih akurat antara instrumen yang disetarakan dengan instrumen sebagai tolak ukur penyetaraan. Berdasarkan pemaparan masalah di atas, tujuan penelitian ini adalah mengetahui kesetaraan instrumen tes UN Bahasa Indonesia tahun 2016 menggunakan Teori Respon Butir. Metode yang digunakan adalah metode regresi dan rerata serta sigma.

METODE

Penelitian ini merupakan penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Data yang digunakan adalah data UN Bahasa Indonesia tahun 2016 se DIY. Analisis yang digunakan adalah menggunakan CTT dan IRT untuk memperoleh model terbaik. Metode penyetaraan menggunakan metode regresi dan rerata dan sigma. Analisis komparatif model penyetaraan instrument tes bidang Bahasa Indonesia dilakukan melalui beberapa tahap analisis kuantitatif dengan memanfaatkan software R-studio.

TEKNIK ANALISIS DATA

Berikut ini langkah-langkah Analisis data yang dilakukan setelah memperoleh data:

1. Melakukan uji CTT untuk menyeleksi butir.

$$R_{pbis} = [\frac{\bar{x}_i - \bar{x}}{s_x}] \sqrt{\frac{p_1}{1-p_1}}$$

dengan,
$Rpbis$: Koefisien korelasi titik biserial
$\bar{xi}$: variabel kontinu
$\bar{x}$: skor rata-rata x
sx : standar deviasu skor x
$p1$: Proporsi peserta tes menjawab benar

2. Melakukan uji asumsi unidimensi, invariansi parameter, dan independensi lokal untuk setiap masing-masing paket.
3. Melakukan Pemodelan (1PL, 2PL, dan 3PL) pada data dengan menggunakan Pendekatan Teori Respons Butir
Model 1PL dengan rumus:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{\theta - b_i}}$$

Keterangan:
i : 1,2,3,...,n
$Pi(\theta)$: Probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir i dengan benar
θ : tingkat kemampuan subyek (sebagai variabel bebas)

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

bi : parameter kesukaran butir ke-i
e : bilangan natural yang nilainya mendekati 2,718
n : banyaknya butir dalam tes

Model 2PL dengan rumus:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Keterangan:
i : 1,2,3,...,n
$Pi(\theta)$: Probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir i dengan benar
θ : tingkat kemampuan peserta tes (sebagai variabel bebas)
bi : parameter kesukaran butir ke-i
ai : Parameter daya pembeda ke-i
e : bilangan natural yang nilainya mendekati 2,718
n : banyaknya butir dalam tes
D : faktor penskalaan yang harganya 1,7.

Model 3PL dengan rumus:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Keterangan:
i : 1,2,3,...,n
$Pi(\theta)$: Probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir i dengan benar
θ : tingkat kemampuan peserta tes (sebagai variabel bebas)
bi : parameter kesukaran butir ke-i
ai : Parameter daya pembeda ke-i
ci : parameter tebakan semu butir ke-i
e : bilangan natural yang nilainya mendekati 2,718
n : banyaknya butir dalam tes
D : faktor penskalaan yang harganya 1,7.

4. Membandingkan tingkat kecocokan butir dari pemodelan 1PL, 2PL, dan 3PL.
5. Menentukan model yang paling cocok yang digunakan pada data penelitian.
6. Melakukan estimasi parameter butir.
7. Melakukan penyetaraan dengan metode rerata dan sigma.

Pada dasarnya, metode rerata dan sigma bekerja dengan menghitung rerata dan simpangan baku dari skor tes pada dua kelompok tes yang berbeda. Berdasarkan model penyetaraannya, hubungan parameter parameter kesulitan butir berhubungan linear,

$$b_2 = \alpha b_1 + \beta,$$

Sehingga diperoleh $\bar{b_2} = \alpha\bar{b_1} + \beta,$
Dan $s_2 = \alpha s_1$
Jadi $\boldsymbol{\alpha = \frac{s2}{s1}}$

Dan $\beta = b_2 - \alpha b_1$

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Keterangan :
$\bar{b_1}$ dan $\bar{b_2}$: berturut-turut rerata parameter kesukaran butir tes 1 dan tes 2,
s_1 dan s_2 : berturut-turut simpangan baku parameter kesukaran butir tes 1 dan 2,
α dan β : konstanta penyetaraan

8. Memodelkan hasil penyetaraan dengan regresi

Metode regresi dalam penyetaraan skor tes didasarkan pada konsep bahwa hubungan antara skor pada dua tes dapat dimodelkan menggunakan persamaan regresi linier. Menurut Angoff (1971), metode ini melibatkan penyesuaian skor berdasarkan perkiraan hubungan linier antara skor tes. Proses ini mencakup estimasi parameter regresi yang paling sesuai untuk data yang ada, yang kemudian digunakan untuk mengkonversi skor dari satu tes ke tes lainnya. Estimasi parameter butir memenuhi persamaan model regresi linier sebagai berikut:

$$y = \alpha x + \beta + \varepsilon$$

Dengan model penduga

$$\hat{y} = \hat{\alpha} x + \hat{\beta}$$

$$\hat{\alpha} = \frac{rxy \cdot sy}{sx}$$

$$\hat{\beta} = \bar{y} - \alpha \bar{x}$$

Keterangan:
$\hat{y}$: estimasi kemampuan atau estimasi parameter butir pada perangkat tes Y
$\hat{x}$: estimasi kemampuan atau estimasi parameter butir pada perangkat tes X
r_{xy} : Koefisien korelasi antara X dan Y
$\bar{y}, \bar{x}$: rerata dari y dan x
$s_x s_y$: simpangan baku dari x dan y

HASIL DAN PEMBAHASAN

Hasil

Data yang digunakan adalah data UN Bahasa Indonesia tahun 2016 tingkat SMP/MTS sederajat yang terdiri dari 5 paket. Setiap paket akan diestimasi menggunakan model Rasch, 2PL, dan 3PL. Sebelum melakukan estimasi dilakukan Analisis CTT terlebih dahulu. Analisis CTT digunakan untuk menyeleksi butir yang dapat digunakan untuk pemilihan model dan analisis lainya. Model terbaik dipilih berdasarkan jumlah butir yang cocok pada masing-masing model pada setiap paket. Setelah didapatkan model yang terbaik langkah selanjutnya adalah pengujian asumsi untuk masing-masing paket : unidimensi, invariansi parameter, dan independensi lokal.

1. Analisis CTT

Sebelum melakukan analisis-analisis lainnya, langkah pertama dilakukan adalah seleksi butir pada 5 paket yang terdiri dari 50 butir. Butir-butir tersebut dianalisis menggunakan program R studio dengan bantuan library CTT dan library MIRT. Suatu butir dikatakan lolos jika nilai pBis nya adalah positif. Hasil analisis CTT terdapat pada Tabel 1

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Tabel 1. Hasil Analisis CTT 5 Paket

Paket	Butir yang tidak lolos	Total butir yang lolos
Paket 1	36	49
Paket 2	5,22,27,32,44	45
Paket 3	5,28,44	47
Paket 4	21,38	48
Paket 5	110,36	48

Berdasarkan Tabel 1, Butir-butir yang lolos akan dilanjutkan keanalisis berikutnya dan butir yang tidak lolos dieliminasi.

2. Kecocokan model

Data perangkat tes UN Bahasa Indonesia dilakukan uji kecocokan model pada masing-masing paket untuk menentukan model terbaik atau yang paling cocok untuk data tersebut. Pada setiap paket akan diestimasi dengan model rasch, 2PL, 3PL. Selanjutnya pada tiap paket akan dipilih 1 model yang cocok. Penentuan kecocokan model dipilih dari total kecocokan model terbanyak dari setiap paket. Suatu butir dikatakan cocok jika nilai p-value > 0,05.

3. Model terbaik

Berdasarkan hasil analisis menggunakan model Rasch, 2PL, dan 3PL akan ditentukan model terbaik untuk memperoleh hasil analisis yang diinginkan. Pada penelitian ini, model terbaik ditentukan berdasarkan model yang memiliki paling banyak butir yang cocok dan dilihat dari nilai AIC dan BIC yang paling kecil.

Tabel 2. Ringkasan butir cocok pada model paket 1-5

Paket	**Model**	**Kecocokan Model**	**AIC**	**BIC**
	Model Rasch	15 cocok	92793,65	93100,65
Paket1	Model 2PL	42 cocok	90451,95	**91053,67**
	Model 3PL	**47 cocok**	**90385,73**	91288,32
	Model Rasch	7 cocok	102275,8	102557,4
Paket2	Model 2PL	35 cocok	99466,19	**100017,2**
	Model 3PL	**38 cocok**	**99446,83**	100273,3
	Model Rasch	8 cocok	117637,6	117931,3
Paket3	Model 2PL	35 cocok	115195,2	**115770,4**
	Model 3PL	**42 cocok**	**115098,4**	115961,4
	Model Rasch	13 cocok	91175,82	91474,06
Paket4	Model 2PL	44 cocok	88779,61	**89365,86**
	Model 3PL	**46 cocok**	**88719,24**	89598,62
	Model Rasch	4 cocok	131664,3	131807,5
Paket5	Model 2PL	20 cocok	129529,2	130114,7
	Model 3PL	**26 cocok**	**127840,5**	**128718,8**

Berdasarkan tabel 2, dihasilkan bahwa model 3PL memiliki model yang cocok untuk data perangkat tes UN Bahasa Indonesia tahun 2016, karena memiliki butir cocok yang terbanyak dan memiliki nilai AIC dan BIC terendah terbanyak.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

4. Asumsi Teori Respon Butir

Model teori respons butir memiliki tiga asumsi yang harus terpenuhi. Tiga asumsi itu adalah unidimensi, invariansi parameter, dan independensi lokal.

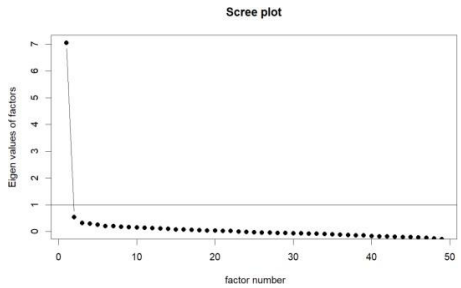
a. Unidimensi

Asumsi unidimensi dapat dibuktikan salah satu diantaranya dengan menggunakan analisis faktor, untuk melihat nilai eigen pada matris varians kovarians inter-butir. Analisis data dengan analisis faktor didahului dengan analisis kecukupan sampel. Pada tulisan ini akan dibuktikan asumsi unidimensi pada data respons peserta tes terhadap ujian nasional mata pelajaran Bahasa Indonesia tahun 2016 yang berjumlah 5 paket.

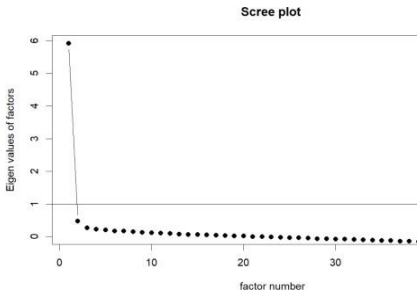
Tabel 3. Hasil KMO dan Sig. Bartlett paket 1-5

	Paket 1	Paket 2	Paket 3	Paket 4	Paket 5
KMO	0,95	0,95	0,94	0,94	0,93
Sig. Bartlett	0,00	0,00	0,00	0,00	0,00

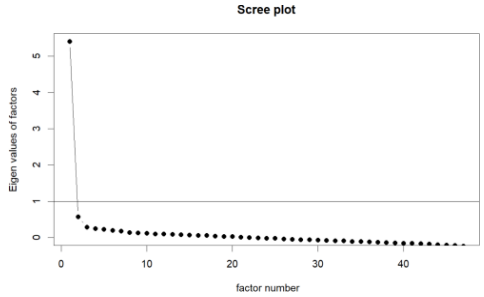
Berdasarkan analisis kecukupan sampel menunjukkan nilai MSA pada masing-masing paket lebih dari 0.05 dan nilai signifikansi bartlett kurang dari 0.05 yang bearti kecukupan sampel untuk semua paket terpenuhi. Selanjutnya asumsi unidimensi akan dilihat dari gambar scree plotnya. Berikut gambar scree plot untuk paket 1-5



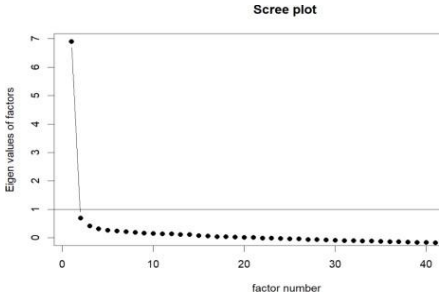
Gambar 1. Scree Plot Paket 1



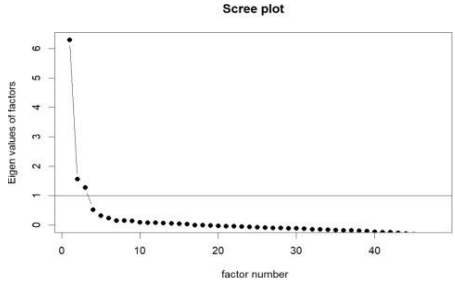
Gambar 2. Scree Plot Paket 2



Gambar 3. Scree Plot Paket 3



Gambar 4. Scree Plot Paket 4



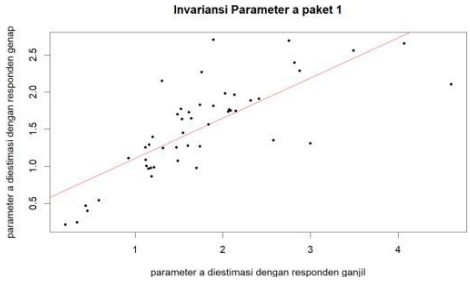
Gambar 5. Scree Plot Paket 5

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

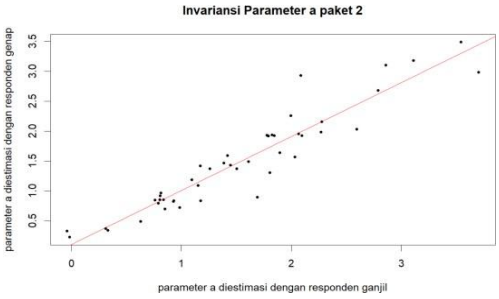
Gambar diatas menunjukkan scree plot dari analisis faktor mengalami penurunan yang sangat tajam antara faktor 1 dan faktor 2, nilai eigen kemudian mulai miring pada faktor ke-2 sehingga scree plot hampir membentuk sudut siku kanan. Ini menunjukkan bahwa hanya ada 1 faktor dominan dalam perangkat tes UN Bahasa Indonesia untuk paket1-4. Artinya Paket 5 tidak dilanjutkan analisis berikutnya.

b. Invariansi Parameter

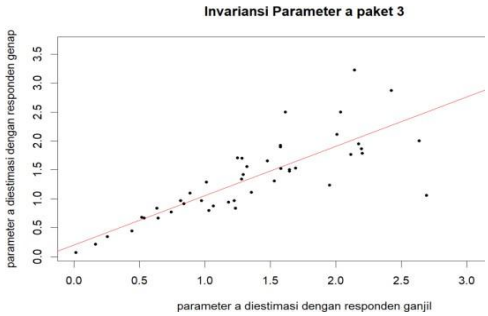
Pengujian asumsi invariansi parameter dilakukan dengan cara membagi data menjadi: data pertama berisi peserta yang berada pada urutan ganjil dan data kedua berisi peserta yang berada pada urutan genap untuk masing-masing paket. Setelah itu data pada setiap paket diestimasi dengan menggunakan model 3PL dan akan diperoleh estimasi parameter a dan b. Plot invariansi parameter a akan ditunjukkan pada Gambar 6-9.



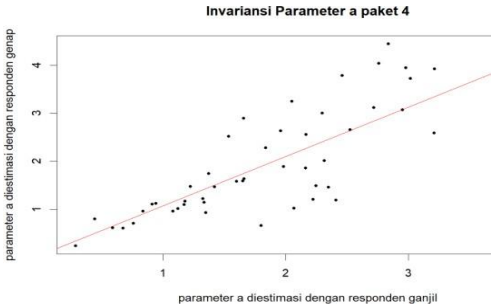
Gambar 6. Invariansi Parameter a Paket 1



Gambar 7. Invariansi Parameter a Paket 2

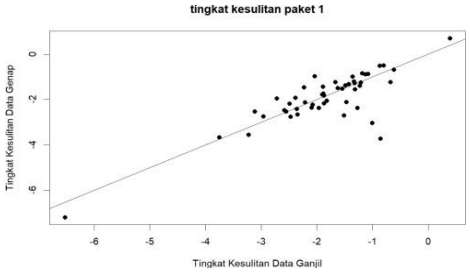


Gambar 8. Invariansi Parameter a Paket 3

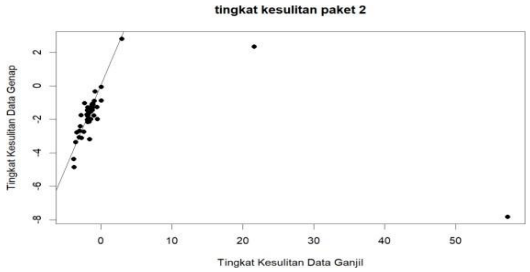


Gambar 9. Invariansi Parameter a Paket 4

Pada Gambar 6-9 terlihat bahwa titik-titik mengikuti garis lurus, sehingga dapat disimpulkan asumsi parameter diskriminan (a) paket 1 sampai 4 terpenuhi. Selanjutnya plot invariansi parameter b akan ditunjukkan pada Gambar 10-13.

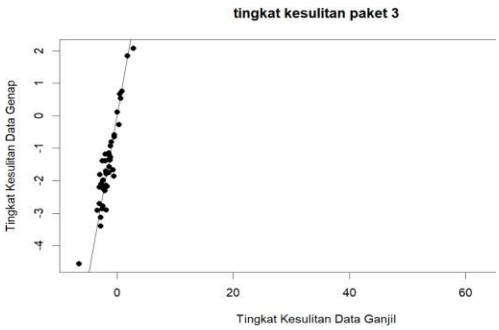


Gambar 10. Invariansi Parameter b Paket 1

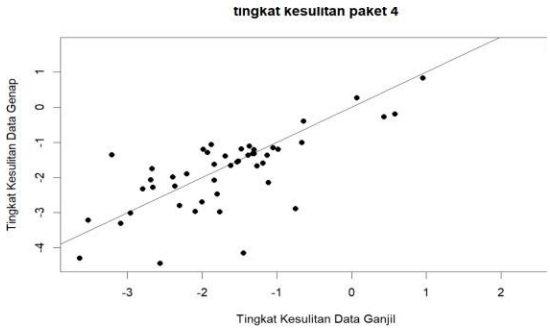


Gambar 11. Invariansi Parameter b Paket 2

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649



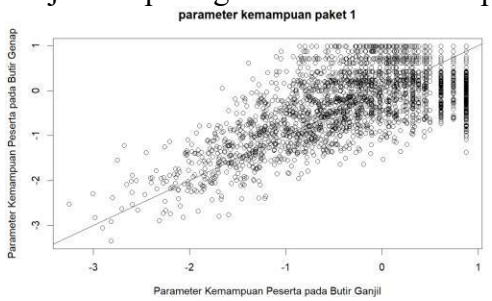
Gambar 12. Invariansi Parameter b Paket 3



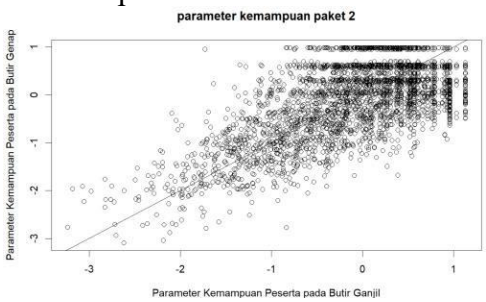
Gambar 13. Invariansi Parameter b Paket 4

Pada Gambar 10-13 diatas terlihat bahwa paket 1, paket 2, paket 3, dan paket 4 terlihat titik-titik mengikuti garis lurus, sehingga dapat disimpulkan asumsi parameter diskriminan (b) terpenuhi.

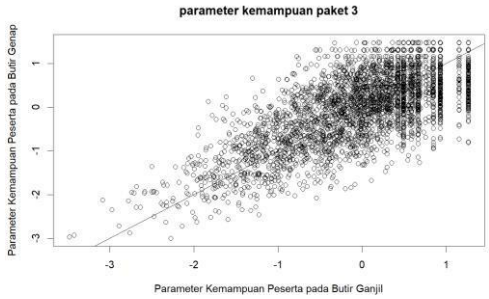
Dalam pengecekkan invariansi parameter kemampuan (θ), jumlah butir soal dari masing-masing paket dikelompokkan menjadi dua data. Butir soal ganjil menjadi data pertama, dan butir soal genap menjadi data kedua yang kemudian masing-masing dianalisis secara terpisah. Kedua kelompok data tersebut digunakan untuk mengestimasi kemampuan peserta (θ). Estimasi kemampuan dilakukan menggunakan model 3PL, kemudian menghasilkan nilai kemampuan (latent-trait) untuk data ganjil dan data genap. Hasilnya kemudian dibuat scater-plot yang ditunjukkan pada gambar 14-17 untuk paket 1 sampai 4.



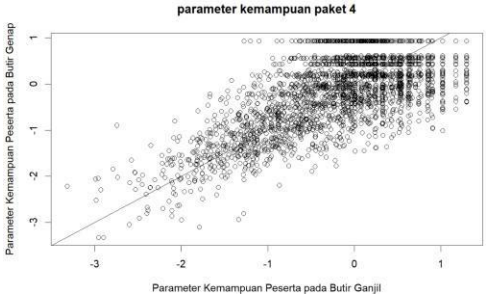
Gambar 14. Invariansi Parameter Kemampuan Paket 1



Gambar 15. Invariansi Parameter Kemampuan Paket 2



Gambar 13. Invariansi Parameter Kemampuan Paket 3



Gambar 14. Invariansi Parameter Kemampuan Paket 4

Pada gambar 14-17, terlihat bahwa titik-titik menyebar disekitar garis dan mengikuti garis lurus, sehingga dapat disimpulkan bahwa asumsi parameter kemampuan pada data perangkat tes UN Bahasa Indonesia tahun 2016 untuk paket 1 sampai 4 terpenuhi.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

c. Independensi lokal

Dalam deteksi asumsi unidimensi diatas, asumsi unidimensi terpenuhi, sehingga dengan terpenuhinya asumsi unidimensi maka asumsi independensi lokal sekaligus terpenuhi (Hambleton dkk, 1991). Hal tersebut juga ditegaskan oleh DeMars yang mengungkapkan bahwa independensi lokal dapat diketahui dengan membuktikan asumsi unidimensional [14]. Ini dapat diartikan bahwa jika asumsi unidimensional terpenuhi, asumsi independensi lokal juga terpenuhi.

d. Ringkasan Hasil asumsi IRT

Berdasarkan poin a, b, dan c dapat disimpulkan bahwa paket yang memenuhi asumsi IRT yaitu unidimensi, invariansi parameter, dan independensi lokal adalah paket 1, paket 2, paket 3, dan paket 4. Selanjutnya keempat paket ini akan dilakukan penyetaraan.

5. Penyetaraan menggunakan metode rerata sigma

Pada metode rerata dan sigma, akan ditentukan konstanta penyetaran α dan β yang melibatkan rerata dan simpangan baku dari parameter kesulitan. Dengan menggunakan fungsi direc pada program R studio, maka didapatkan nilai konstanta penyetaran α dan β menggunakan rerata dan sigma yang terdapapat pada Tabel 10.

Tabel 4. Hasil Konstanta Penyetaraan Metode Rerata dan Sigma

Konstanta	Paket 2 ke 1	Paket 3 ke 2	Paket 4 ke 1
Aplha	0,7812740	0,5778584	0, 7094957
Betha	-0, 3973212	-0, 9385464	-0, 4997339

6. Penyetaraan menggunakan model Regresi Dari metode rerata dan sigma maka akan diperoleh suatu model Regresi. Pada Tabel 5 akan dijelaskan terkait model regresi untuk pernyetaraan paket 1 dengan paket yang lain,

Tabel 5. Persamaan Regresi pada Penyetaraan Tes Paket 2 ke 1, Paket 3 ke 1 dan Paket 4 ke 1

			Rerata dan Sigma		
Arah	Alfa	Beta	Persamaan Ekivalensi Parameter b	Persamaan Ekivalensi Parameter a	Persamaan Ekivalensi Parameter Kemampuan
2 ke 1	0,781274	-0,3973212	$b* = 0{,}78b - 0{,}3$	$a* = 0{,}7a$	$\theta* = 0.78\theta - 0{,}3$
3 ke 2	0,5778584	-0,9385464	$b* = 0{,}57b - 0{,}9$	$a* = 0.5a$	$\theta* = 0.57\theta - 0{,}9$
4 ke 1	0,7094957	-0,4997339	$b* = 0{,}7b - 0{,}4$	$a* = 0.7a$	$\theta* = 0{,}7\theta - 0{,}4$

7. Nilai RMSE penyetaraan

Setelah disetarakan maka akan dicari nilai RMSE memberikan nilai kesalahan rata-rata dari prediksi model dalam unit yang sama dengan variabel yang diprediksi. Hasil RMSE penyetaraan untuk masing masing paket terdapat pada Tabel 6.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Tabel 6. Nilai RMSE penyetaraan setelah disetarakan

Paket	Nilai RMSE
Paket 2 ke 1	3,548796
Paket 3 ke 1	7,759861
Paket 4 ke 1	4,582123

Berdasarkan analisis terhadap karakteristik butir soal dari paket 1 hingga paket 4 pada Ujian Nasional Matematika 2016, terdapat temuan bahwa sejumlah item masih belum fit dengan kemampuan siswa, yang berarti item-item tersebut perlu direvisi untuk digunakan pada tes masa depan (Sainuddin, 2014). Hal ini menyoroti pentingnya pengembangan bank soal yang inklusif dan adaptif, dimana butir soal yang memiliki karakteristik baik dapat dimasukkan ke dalam bank soal untuk digunakan pada berbagai jenis tes, seperti ujian semester dan ujian kompetensi (Kaplan & Saccuzzo, 2017).

Penelitian ini mengkaji paket 1 sampai paket 4 yang masing-masing terdiri dari 50 butir. Analisis butir dengan pendekatan item response theory (IRT) menggunakan Program R. Komponen psikometri yang menjadi fokus pada bagian ini adalah jumlah item fit untuk masing-masing paket dan parameter logistik yang digunakan. Secara umum untuk melihat perbandingan karakteristik butir dari masing-masing paket berdasarkan jumlah item disajikan pada Tabel 7.

Tabel 7. Jumlah Item Fit Paket 1-4

Paket	Jumlah Item Fit
Paket 1	49
Paket 2	45
Paket 3	47
Paket 4	48

Setelah mengetahui jumlah item fit yang sesuai untuk masing-masing paket, langkah berikutnya adalah melakukan estimasi parameter butir untuk menentukan model yang paling optimal. Analisis karakteristik butir tes, yang disajikan terlebih dahulu dalam statistik terkait, menjadi landasan sebelum menentukan kesetaraan perangkat tes paket menggunakan metode rerata dan sigma. Metode yang terbaik dan digunakan untuk menganalisis karakteristik butir tes adalah model 3PL, yang melibatkan estimasi parameter seperti parameter diskriminan (a), parameter kesukaran (b), dan parameter tebakan semu (c). Hasil estimasi statistik untuk ketiga parameter dari paket 1 sampai paket 4 terdapat pada Tabel 14.

Tabel 8. Statistik Estimasi Parameter diskriminan, kesulitan dan tebakan semu untuk paket 1-4

Paket	Statistik	a	b	c
Paket 1	Mean	1,685	-1,9459	0,2366607
	Min	0,207	-7,2234	0,0005449
	Max	3,848	0,4954	0,8472519

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

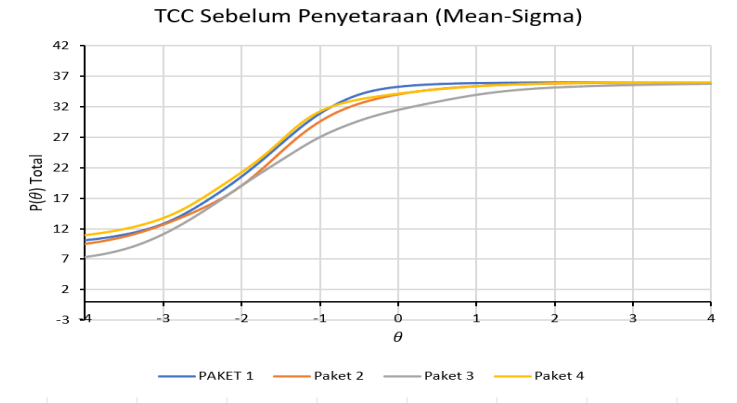
Paket	Statistik	a	b	c
Paket 2	Mean	1,4864	-2,420	0,215183
	Min	0,0476	-18,328	0,003721
	Max	3,4420	2,954	0,731457
Paket 3	Mean	0,03146	-1,3555	0,006166
	Min	1,39528	-1,825	0,183750
	Max	3,59763	2,431	0,659858
Paket 4	Mean	1,8061	-1,712	0,255781
	Min	0,2587	-4,168	0,003462
	Max	3,5000	2,119	0,784430

Berdasarkan Tabel 8. Diketahui nilai parameter a yang bervariasi dari 0,03146 hingga 3,848 dalam data nilai UN B. Indonesia menunjukkan adanya variasi signifikan dalam kemampuan butir soal untuk membedakan antara peserta dengan kemampuan rendah hingga tinggi. Rata-rata nilai parameter a diatas 1, menunjukan bahwa item-item yang termuat didalam paket tersebut sangat baik dalam membedakan antara individu dengan kemampuan yang berbeda.

Selain itu, nilai parameter kesukaran (b) yang ditinjau dalam penelitian ini menunjukkan rentang nilai dari -18,328 hingga 2,954. Menurut Hambleton & Swaminathan (1985), nilai b yang lebih tinggi menandakan tingkat kesulitan soal yang lebih tinggi, sedangkan nilai yang lebih rendah menunjukkan tingkat kesulitan yang lebih rendah.

Selanjutnya, parameter tebakan semu (c), dengan rentang nilai dari 0,0005449 hingga 0,8472519 dan rata-ratanya parameter c mendekati 0, mencerminkan proporsi tebakan benar secara acak tanpa pengetahuan sebelumnya tentang materi soal. Menurut penelitian oleh Embretson & Reise (2000), nilai c yang lebih rendah menunjukkan probabilitas tebakan semu yang lebih rendah, menandakan bahwa soal tersebut lebih efektif dalam membedakan antara respon yang dipengaruhi oleh pengetahuan sejati tentang materi tes.

Meskipun Test Characteristics Curve (TCC) dari keempat paket memiliki karakteristik yang relatif setara, namun perbedaan parameter butir berpotensi menghasilkan skor kemampuan (ability) yang berbeda di seluruh paket soal. Berikut adalah gambaran TCC dari keempat paket soal sebelum dilakukan penyetaraan.



Gambar 18. TCC Keempat Paket Soal Sebelum Penyetaraan

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Pembahasan

Langkah awal dalam melakukan penyetaraan adalah menyamakan butir soal setiap paket. Setelah dilakukan analisis CTT dengan menghilangkan butir yang nilai pbisnya kurang dari 0,1, maka diambil butir-butir yang beririsan antara paket 1 sampai paket 4. Didapatkan butir yang beririsan berjumlah 36 butir. Selanjutnya butir-butir ini akan dilakukan penyetaraan untuk masing-masing paket.

Langkah selanjutnya adalah melakukan estimasi parameter butir untuk masing masing paket dengan model 3PL, maka akan didapatkan parameter butir a,b, dan c nya. Setelah dilakukan estimasi maka bisa dilanjutkan untuk penyetaraan menggunakan metode rerata dan sigma.

Analisis selanjutnya adalah mengetahui konstanta hasil penyetaraan menggunakan salah satu metode penyetaraan yakni metode rerata dan sigma. Penyetaraan dilakukan hanya pada paket 2, paket 3, dan paket 4 yang disetarakan ke paket 1. Berikut adalah hasil konstanta penyetaraan yang telah diestimasi menggunakan metode Rerata dan Rerata.

Tabel 9. Hasil Konstanta Penyetaraan Metode Rerata & Rerata

Link Test	Konstanta	Nilai Estimasi
Paket 2 ke 1	α	0, 7812740
	β	-0, 3973212
Paket 3 ke 1	α	0, 5778584
	β	-0, 9385464
Paket 4 ke 1	α	0, 7094957
	β	-0, 4997339

Tabel 9 menunjukkan bahwa konstanta alpha seluruhnya bervariasi mulai dari 0,5778584 sampai 0,781274. Kemudian untuk nilai konstanta betha juga bervariasi dengan interval nilai -0,9385464 hingga -0,3973212. Konstanta penyetaraan yang telah terestimasi kemudian digunakan untuk mentranformasi skor awal menjadi skor baru setelah tes disetarakan. Berikut adalah persamaan regresi kesetaraan skor (θ) metode Rerata & Rerata.

Tabel 10. Persamaan Regresi Kesetaraan Skor

Link Test	Persamaan Kesetaraan Skor
Paket 2 ke 1	$\theta_{21} *= 0{,}78\theta - 0{,}39$
Paket 3 ke 1	$\theta_{31} *= 0{,}57\theta - 0{,}93$
Paket 4 ke 1	$\theta_{31} *= 0{,}7\theta - 0{,}49$

Persamaan kesetaraan skor pada Tabel 10 digunakan untuk mentransformasi ability (θ) peserta tes dengan paket 2, paket 3, dan paket 4 ke ability baru setelah disetarakan ke paket 1. Berikut adalah estimasi ability (θ) setelah ketiga paket disetarakan ke paket 1.

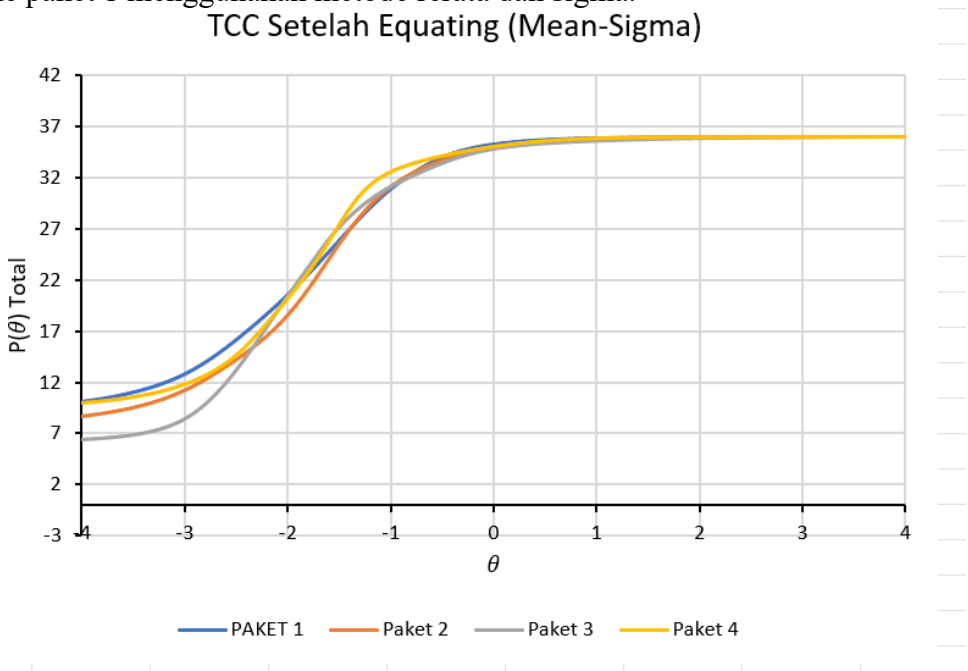
Tabel 11. Hasil Skor Penyetaraan untuk Beberapa Ability

Ability	Paket 2 ke 1	Paket 3 ke 1	Paket 4 ke 1
-4	0,24132654	0,17625113	0,278319
-3	0,31208736	0,23244917	0,328908
-2	0,51599061	0,56665927	0,561433
-1	0,86392714	0,86379248	0,904439
0	0,96789365	0,96630935	0,973181
1	0,99158418	0,98829442	0,995389

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Ability	Paket 2 ke 1	Paket 3 ke 1	Paket 4 ke 1
2	0,99719876	0,99548674	0,998818
3	0,99885558	0,99841822	0,999465
4	0,99948623	0,99946246	0,999720

Tabel 11 menunjukkan bahwa dengan metode Rerata dan sigma, ketika peserta didik berada pada θ = -4 pada paket 2, itu setara dengan θ = 0,24132654 setelah paket 2 disetarakan dengan paket 1. Hal tersebut berlaku ketika paket 3 dan paket 4 disetarakan dengan paket 1 dengan θ = - 4 akan setara dengan 0,17625113 dan 0,278319. Interpretasi tersebut berlaku untuk nilai theta lainnya. Hasil tersebut kemudian diperkuat dengan TCC dari ketiga paket yang telah disetarakan ke paket 1 menggunakan metode rerata dan sigma.



Gambar 19. TCC Setelah Penyetaraan

Gambar 19 menunjukkan bahwa penyetaraan dengan metode Rerata dan sigma menghasilkan kurva yang saling berdekatan untuk paket 2,3, dan 4. Hal tersebut mengindikasikan bahwa untuk paket 2,3,dan 4 hampir memiliki karakteristik tingkat kesukaran yang setara satu sama lain.

Perhitungan RMSE juga dilakukan untuk mengetahui nilai kesalahan rata-rata dari prediksi model dalam unit yang sama dengan variabel yang diprediksi. Nilai RMSE untuk penyetaraan paket 2 ke 1 sebesar 3,548796, paket 3 ke 1 sebesar 7,759861 dan paket 4 ke 1 sebesar 4,582123. Semakin kecil nilai RMSE, semakin baik model dalam melakukan prediksi. Sehingga paket 2 ke 1 merupakan model yang paling baik daripada paket lainya dikarenakan memiliki nilai RMSE yang paling kecil.

Penggunaan row score dalam skoring tes tanpa adanya penyetaraan dapat mengarah pada penilaian yang tidak adil, terutama ketika paket-paket tes memiliki tingkat kesulitan yang bervariasi secara signifikan (KetterlinGeller et al., 2018). Oleh karena itu, proses penyetaraan antar paket tes, baik dari paket 1 ke paket lainnya maupun sebaliknya, diperlukan untuk memastikan bahwa skor yang diperoleh peserta tes mencerminkan kemampuan sebenarnya secara adil.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Hasil dari proses penyetaraan ini memiliki dampak yang signifikan dalam konteks penilaian pembelajaran. Penilaian yang efektif harus didasarkan pada instrumen yang dapat menilai secara akurat dan adil terhadap pencapaian belajar siswa (Stiggins, 2005). Proses ini tidak hanya mengidentifikasi aktivitas dan pemahaman siswa, tetapi juga memberikan laporan yang jelas terkait capaian belajar kepada siswa dan orang tua mereka (Guskey, 2003).

Dalam mengembangkan instrumen tes yang setara atau mendekati setara, perlu memperhatikan berbagai aspek seperti jumlah butir soal, karakteristik peserta tes, dan metode estimasi parameter yang digunakan (Kolen & Brennan, 2014). Proses ini membantu memastikan bahwa instrumen tes tidak hanya valid dan reliabel, tetapi juga mampu mengukur kemampuan siswa dengan akurat di berbagai situasi evaluasi.

Kesimpulannya, penelitian ini memberikan pemahaman yang mendalam tentang pentingnya equating dalam konteks penggunaan Model 3PL pada tes paralel. Dengan menggunakan metode penyetaraan yang tepat, seperti mean-sigma, studi ini menggarisbawahi kebutuhan akan proses penyetaraan yang cermat untuk meminimalkan kesalahan dan memastikan hasil evaluasi pendidikan yang konsisten dan adil bagi semua peserta tes (Madya et al., 2019; Meyer & Zhu, 2013). Penelitian lebih lanjut dapat mengembangkan metode penyetaraan yang lebih inovatif dan responsif terhadap kebutuhan evaluasi pendidikan yang semakin kompleks.

SIMPULAN

Melalui kajian yang dilakukan terhadap penyetaraan menggunakan metode regresi dan rerata sigma, terkait penyetaraan sebagai berikut:

1. Perangkat soal UN Bahasa Indonesia tahun 2016 di Daerah Istimewa Yogyakarta terdiri dari 5 paket yang jumlah item fit setelah dilakukan analisis CTT untuk paket 1 berjumlah 49 butir, paket 2 berjumlah 45 butir, paket 3 berjumlah 47 butir, paket 4 berjumlah 48 butir dan paket 5 berjumlah 48 butir.Dengan menggunakan IRT yang mempertimbangkan kecocokan model dan nilai dari AIC BIC nya untuk masing-masing paket didapatkan model terbaik untuk ke 5 paket yaitu model 3PL.Karakteristik soal untuk masing-masing paket bervariasi mulai dari tingkat mudah, sedang dan sulit dengan dominan rata-rata untuk masing-masing paket dominan sedang.
2. Penyetaraan menggunakan metode rerata dan sigma untuk paket 2 ke 1 diperoleh nilai koefisian a sebesar 0,7812740 dan b sebesar -3,973212. Selanjutnya untuk paket 3 ke 1 nilai a sebesar 0,5778584 dan b sebesar -0,9385464. Kemudian untuk paket 4 ke 1 nilai a sebesar 0,7094957 dan b sebesar -0,4997339. Setelah melakukan penyetaraan dengan rerata dan sigma didapatkan model regresi yakni $\theta 21 *= 0{,}78\theta - 0{,}39$ untuk persamaan kesetaraan skor paket 2 ke paket 1, $\theta 31 *= 0{,}57\theta - 0{,}93$ untuk persamaan kesetaraan skor paket 3 ke paket 1, selanjutnya $\theta 41 *= 0{,}7\theta - 0{,}49$ untuk persamaan kesetaraan skor paket 4 ke paket 1. Dalam penyetaraan juga didapatkan nilai RMSE untuk paket 2 ke 1 sebesar 3,548796, paket 3 ke 1 sebesar 7,759861 dan paket 4 ke 1 sebesar 4,582123.

UCAPAN TERIMAKASIH

Terimakasih kepada Koor Prodi Statistika dan seluruh Dosen Prodi Statistika yang telah memberikan ilmu hingga terselesainya artikel ini.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

DAFTAR PUSTAKA

Andayani, L., & Ramalis, T. (2019). Instrumen tes dalam pendidikan. Bandung: Alfabeta.
Angoff, W. H. (1971). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service.
DeMars, C. E. (2010). Item response theory. New York, NY: Oxford University Press.
Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.
Guskey, T. R. (2003). How's My Kid Doing? A Parent's Guide to Grades, Marks, and Report Cards. San Francisco, CA: Jossey-Bass.
Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer-Nijhoff.
Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage Publications.
Harris, D. J. (2016). Test equating, scaling, and linking: Methods and practices (M. Faizal & I. Santoso, Trans.). Yogyakarta: Pustaka Pelajar. (Original work published 2016).
Haryani, S. (2015). Analisis butir soal dalam pendidikan. Yogyakarta: Pustaka Pelajar.
Kaplan, R. M., & Saccuzzo, D. P. (2017). Psychological testing: Principles, applications, and issues (9th ed.). Boston, MA: Cengage Learning.
Ketterlin-Geller, L. R., Yovanoff, P., & Tindal, G. (2018). Fair and equitable assessment practices. New York, NY: Routledge.
Kolen, M. J., & Brenan, R. L. (2004). Test equating, scaling, and linking: Methods and practices (2nd ed.). New York, NY: Springer.
Madya, S., Wijayanti, I., & Haryanto, E. (2019). Pengembangan instrumen evaluasi pendidikan. Yogyakarta: Universitas Negeri Yogyakarta Press.
Meyer, P., & Zhu, S. (2013). Applying Test Equating Methods: Using Mean-Sigma and IRT-Based Approaches. Journal of Educational Measurement, 50(1), 1-15.
Sainuddin. (2014). Analisis butir soal pada Ujian Nasional. Jakarta: Depdiknas.
Stiggins, R. J. (2005). Student-involved assessment FOR learning. Upper Saddle River, NJ: Pearson.
Undang-Undang No. 21 Tahun 2003 tentang Sistem Pendidikan Nasional [National Education System Act No. 21 of 2003].
Zainal. (2020). Evaluasi hasil belajar. Jakarta: Gramedia.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649