

**METODE ESTIMASI PARAMETER DAN METODE EQUATING PADA UKURAN SAMPEL KECIL BERDASARKAN ITEM RESPONDS THEORY
(PARAMETER ESTIMATION AND EQUATING METHOD ON SMALL SAMPLE SIZE BASED ON ITEM RESPONSE THEORY)**

Wardani Rahayu

FMIPA Universitas Negeri Jakarta
wardani.rahayu@unj.ac.id

ABSTRACT

The purpose of this study is to determine the accuracy of parameter estimation method of item parameter and equating method on small sample size based on Item Response Theory. Parameter estimation methods used are Joint Maximum Likelihood estimation and Bayesian estimation method, equating method using Mean and Sigma and Robust Mean and Sigma method. Estimated parameter using two parameter logistic model on small sample size of 300. Data Source using junior high school National Math Exam in 2011 data. Accuracy of parameter estimation and equating method can be seen from Root Mean Square Error (RMSE). The results of this study is the same equating method generate the same accuracy between Joint Maximum Likelihood and Bayesian method, while at the same parameter estimation method, resulting different accuracy between and Mean and Sigma and Robust Mean and Sigma methods.

Key words: parameter estimation method, equating method, small sample size, RMSE

ABSTRAK

Tujuan penelitian ini adalah untuk mengetahui keakurasian metode estimasi parameter butir parameter dan metode equating pada ukuran sampel kecil berdasarkan Item Response theory. Metode estimasi parameter yang digunakan adalah metode estimasi Joint Maximum Likelihood dan metode estimasi Bayesian, metode equating yang digunakan adalah metode Rerata dan Sigma dan Tegar Rerata dan Sigma. Estimasi parameter menggunakan model logistik dua parameter pada ukuran sampel kecil yakni 300. Sumber data menggunakan data Ujian Nasional Matematika SMP tahun 2011. Keakurasian metode estimasi parameter dan metode equating dilihat dari nilai Root Mean Square Error (RMSE). Hasil penelitian ini adalah pada ukuran sampel kecil, metode equating yang sama menghasilkan keakurasian yang sama antara metode Joint Maximum Likelihood dan metode Bayesian, sementara pada metode estimasi parameter sama, menghasilkan keakurasian yang berbeda antara metode Rerata dan Sigma dan Tegar Rerata dan Sigma.

Katakunci: metode estimasi parameter, metode equating, ukuran sampel kecil, RMSE

1. PENDAHULUAN

Ujian Nasional merupakan kegiatan pengukuran pencapaian kompetensi peserta didik yang dilakukan pada akhir satuan pendidikan yang bersifat nasional. Penyusunan butir soal ujian dilakukan oleh Pemerintah Pusat dan Pemerintah Daerah berdasarkan kisi-kisi yang dikembangkan oleh Kementerian Pendidikan dan Kebudayaan Indonesia. Kisi-kisi ini disusun berdasarkan standar kompetensi kelulusan pendidikan sesuai tingkat satuan pendidikan.

Pelaksanaan Ujian Nasional mata pelajaran matematika tahun 2011 menggunakan lima paket dan tahun 2013 menggunakan dua puluh paket berbeda berisikan butir soal yang berbeda dengan beberapa butir gandeng (*anchor*). Butir soal ini dibuat berdasarkan kisi-kisi soal yang sama sehingga menghasilkan lima paket soal yang berbeda untuk satu kelas pada tahun 2012 dan dua puluh paket soal yang berbeda pada tahun 2013. Butir-butir pada Ujian Nasional hanya mengukur kemampuan kognitif peserta didik dan tidak mengukur kemampuan psikomotorik maupun afektif.

Paket soal yang dikembangkan oleh Kementerian Pendidikan dan Kebudayaan Indonesia. untuk Ujian Nasional mata pelajaran Matematika diharapkan setara. Kesetaraan yang dimaksud adalah taraf sukar setiap butir soal antar paket sama. Oleh karena itu amat penting dilakukan validasi isi perangkat soal Ujian Nasional oleh ahli materi matematika dan pendidikan matematika untuk menentukan kesesuaian butir soal dengan indikator, bahasa, konstruksi soal, kebenaran materi soal dan kesetaraan butir soal.

Kesetaraan menurut *judgment* ahli materi matematika dan pendidikan matematika tidak bisa langsung dapat digunakan ketika akan membandingkan hasil ujian nasional sekelompok peserta didik dari dua paket yang berbeda. Oleh karena itu perlu dilakukan penyetaraan skor dari respon peserta didik. Terdapat tiga macam penyetaraan skor yaitu *equating*, *concordance* dan *prediction*. [1] *Equating* digunakan untuk penyetaraan skor antar paket soal yang berbeda yang mengukur konstruk yang sama. *Equating* dapat digunakan dengan pendekatan teori tes klasik dan teori responsi butir. Penyetaraan dengan menggunakan teori tes klasik diantaranya adalah metode Linier, metode Paralel Linier dan metode Ekipersentil [2], sedangkan untuk teori responsi butir adalah Rerata dan Sigma,

Tegar Rerata dan Sigma, Kurva Karakteristik dan metode Minimum Khi-Kuadrat.
[3]

Equating berdasarkan *item respons theory* dengan menggunakan Rerata dan Sigma, dan Tegar Rerata dan Sigma melibatkan taraf sukar butir. Lord memberikan persamaan transformasi linear untuk parameter kemampuan peserta tes. Parameter kemampuan peserta tes kedua kelompok disamakan skalanya dengan model logistik dua parameter (L2P) menggunakan rumus

$$\theta_{j2}^* = A\theta_{j2} + K \cdot [4] .$$

Penentuan nilai parameter kemampuan peserta tes dan taraf sukar butir dalam estimasi dalam pengukuran dinamakan estimasi parameter. Terdapat tiga metode mengestimasi parameter peserta tes diantaranya metode Likelihood maksimum [5], [6] dan metode Bayesian. [7], [8]. Hasil estimasi kemampuan peserta tes juga dipengaruhi oleh ukuran sampel. Semakin besar ukuran sampel maka semakin akurat hasil estimasi kemampuan peserta tes. Pada penelitian akan dibatasi pada ukuran sampel kecil dan permasalahan yang timbul jika estimasi parameter butir dilakukan dengan metode estimasi berbeda maka jenis metode *equating* dan metode estimasi butir manakah yang paling akurat pada ukuran sampel kecil.

2. Tinjauan Pustaka

2.1. Equating

Parameter butir kelompok dua disamakan skalanya dengan parameter butir kelompok satu melalui transformasi linear dengan model logistik dua parameter (L2P). Rumus transformasi linearnya untuk kemampuan peserta tes. [9] adalah

$$\theta_{j2}^* = A\theta_{j1} + K$$

* menyatakan nilai transformasi, sehingga estimasi parameter kemampuan peserta tes ke-j dari kelompok satu ke kelompok dua.

2.1.1. Metode Rerata dan Sigma (*Rerata dan Sigma Method*)

Marco mengemukakan metode Rerata Dan Sigma (RS) untuk perangkat tes yang berbentuk dikotomi. [10] Metode ini menggunakan rerata dan simpangan baku dari estimasi taraf sukar butir dari ke dua kelompok peserta tes. Menentukan koefisien A dan K pada persamaan (1) dengan menggunakan rumus

$$K = \mu_{b_2} - A\mu_{b_1} \text{ dan } A = \frac{\sigma_{b_1}}{\sigma_{b_2}} \text{ (Hambleton, Swaminathan, Roger; 1991: 131)}$$

2.1.2. Metode Tegar Rerata dan Sigma (*Tegar Rerata dan Sigma Method*)

Metode ini menggunakan rerata dan simpangan baku dari estimasi taraf sukar butir dari ke dua kelompok peserta tes. Menentukan koefisien A dan K pada persamaan (1) dengan menggunakan rumus

$$K = \mu_{b_R^w} - A\mu_{b_F^w}, \text{ dan } A = \frac{\sigma_{b_R^w}}{\sigma_{b_F^w}} \text{ (Hambleton, Swaminathan, Roger; 1991: 133)}$$

Dengan $b_{jF}^w = w_j^* b_{jF}$ dan $b_{jR}^w = w_j^* b_{jR}$

$$w_j^* = \frac{w_j}{\sum_{i=1}^n w_i}$$

$$w_j = [\text{maks}\{\sigma_{b_{j1}}^2, \sigma_{b_{j2}}^2\}]^{-1}$$

2.2. Metode Estimasi Parameter

2.2.1. Estimasi Joint Maximum Likelihood

Estimasi Maximum Likelihood dengan kemampuan peserta tes diperlukan fungsi Likelihood. Fungsi Likelihood untuk peserta tes dengan kemampuan θ (Naga, 2012: 473) yaitu

$$L(X_1, X_2, X_3, \dots, X_n | \theta) = \prod_{j=1}^n P_j(\theta)^{X_j} Q_j(\theta)^{1-X_j} \dots\dots\dots(1)$$

Logaritma dari fungsi likelihood (1) adalah

$$\begin{aligned} \ln L(X_1, X_2, X_3, \dots, X_n | \theta) &= \ln \prod_{j=1}^n P_j(\theta)^{X_j} Q_j(\theta)^{1-X_j} \\ &= \sum_{j=1}^n [X_j \ln P_j(\theta) + (1 - X_j) \ln Q_j(\theta)] \end{aligned}$$

Likelihood maksimum dapat diperoleh melalui

$$\frac{d}{d\theta} [\ln L(X_1, X_2, X_3, \dots, X_n | \theta)] = 0$$

2.2.2. Estimasi Bayesian

Pada prosedur Bayesian (Hambleton dan Swaminathan, 1990: 94) dapat diasumsikan bahwa distribusi kemampuan peserta tes berdistribusi normal baku

yaitu $\theta \sim N(0,1)$ atau fungsi kepadatannya $f(\theta) \propto e^{-\frac{1}{2}\theta^2}$. Jika kemampuan awal peserta tes independen maka berdistribusi posterior adalah

$$\begin{aligned} f(\theta_1, \theta_2, \theta_3, \dots, \theta_N | X) &\propto L(X | \theta_1, \theta_2, \theta_3, \dots, \theta_N) f(\theta_1, \theta_2, \theta_3, \dots, \theta_N) \\ &\propto L(X | \theta_1, \theta_2, \theta_3, \dots, \theta_N) \cdot (f(\theta_1) \cdot f(\theta_2) \cdot f(\theta_3) \dots f(\theta_N)) \\ &\propto L(X | \theta_1, \theta_2, \theta_3, \dots, \theta_N) \cdot \left(e^{-\frac{1}{2}\theta_1^2} \cdot e^{-\frac{1}{2}\theta_2^2} \cdot e^{-\frac{1}{2}\theta_3^2} \dots e^{-\frac{1}{2}\theta_N^2} \right) \\ &\propto L(X | \theta_1, \theta_2, \theta_3, \dots, \theta_N) \cdot \left(e^{-\frac{1}{2}\sum_{j=1}^N \theta_j^2} \right) \dots\dots\dots(2) \end{aligned}$$

Logaritma dari fungsi likelihood (2) adalah

$$\ln f(\theta_1, \theta_2, \theta_3, \dots, \theta_N | X) = C + \ln L(X | \theta_1, \theta_2, \theta_3, \dots, \theta_N) - \frac{1}{2} \sum_{j=1}^N \theta_j^2, \quad C = \text{konstanta} \quad \dots\dots (3)$$

Solusi persamaan (3) adalah estimasi bayes's modal $\theta_1, \theta_2, \theta_3, \dots, \theta_N$ dapat diperoleh melalui

$$\frac{\partial}{\partial \theta_j} \ln f(\theta_1, \theta_2, \theta_3, \dots, \theta_N | X) = 0, \quad j = 1, 2, \dots, N \quad \dots\dots\dots (4)$$

2.3. Root Standar Measurement Error (RSME)

Keakuratan metode *equating* dan metode estimasi dilihat dari RMSE, semakin kecil RMSE maka disimpulkan bahwa metode *equating* dan metode estimasi yang diujicobakan semakin akurat. RSME atau RMSD ditentukan dengan menggunakan rumus sebagai berikut

$$RMSE(\theta) = \sqrt{\frac{\sum_{j=1}^n (\bar{\theta}_i - \theta_i)^2}{n}} \quad (\text{Cohen dan Seock; 1998, Kartono; 2008})$$

Dengan:

- N = ukuran sampel
- $\bar{\theta}_i$ = kemampuan peserta ke-i setelah disetarakan
- θ_i = kemampuan peserta ke-i sebelum disetarakan

3. METODE PENELITIAN

Metode penelitian yang digunakan adalah metode eksperimen. Data yang digunakan dalam penelitian ini berupa skor Ujian Nasional Matematika SMP tahun

2011 pada paket A17 dan B29 untuk wilayah Jakarta. Sekor hasil pekerjaan siswa ini berbentuk option jawaban A, B, C, D, dan E Panjang perangkat tes Matematika SMP adalah 40. Ukuran sampel yang digunakan dalam penelitian ini adalah 300 dan dinamakan ukuran sampel kecil. Metode *equating* yang digunakan adalah metode Rerata dan Sigma, dan metode Tegar Rerata dan Sigma. Metode estimasi yang digunakan adalah Joint Maximum Likelihood dan Bayesian

Tabel 1. Desain Penelitian Estimasi parameter dan Metode Equating Pada Ukuran Sampel Kecil

Estimasi Joint Maximum Likelihood (A1)		Estimasi Bayesian (A2)	
Rerata dan Sigma (B1)	Tegar Rerata dan Sigma (B2)	Rerata dan Sigma (B1)	Tegar Rerata dan Sigma (B2)
$Y_{1.1.1}$	$Y_{1.2.1}$	$Y_{2.1.1}$	$Y_{2.2.1}$
$Y_{1.1.1}$	$Y_{1.2.1}$	$Y_{2.1.1}$	$Y_{2.2.1}$
.	.	.	.
.	.	.	.
$Y_{1.1.30}$	$Y_{1.2.30}$	$Y_{2.1.30}$	$Y_{2.2.30}$

Keterangan

- A1B1 : kelompok *equating* dengan metode Rerata dan Sigma dengan menggunakan metode Joint Maximum Likelihood
- A1B2 : kelompok *equating* dengan metode Rerata dan Sigma dengan menggunakan metode Joint Maximum Bayesian
- A2B1 : kelompok *equating* dengan metode Tegar Rerata dan Sigma dengan menggunakan metode Joint Maximum Likelihood
- A2B2 : kelompok *equating* dengan metode Tegar Rerata dan Sigma dengan menggunakan metode Bayesian

Prosedur penelitian ini adalah

- 1) Menentukan butir yang cocok model dengan model logistik dua parameter (L2P),
- 2) Mengambil secara acak 300 skor peserta tes untuk kelompok pertama dan kelompok kedua, masing-masing dilakukan sebanyak 30 kali replikasi
- 3) Melakukan estimasi kemampuan peserta tes dan taraf sukar dan daya beda dengan menggunakan metode Joint *Maximum* Likelihood dan Estimasi Bayesian dengan *BILOG MG3*
- 4) Menentukan nilai *A* dan *K* dengan menggunakan metode RS dan TRS dengan software *BLINK* (Wardani Rahayu: 2010)

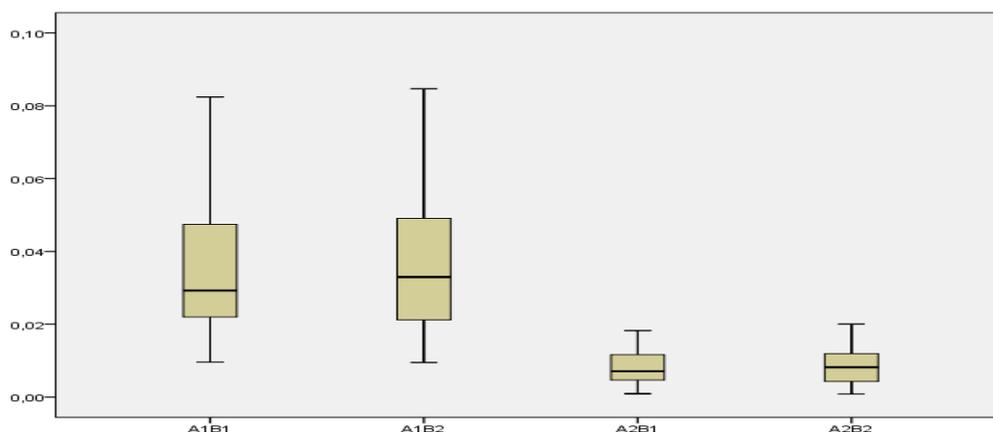
- 5) Melakukan equating parameter butir
- 6) Menghitung nilai RSME dan rata-rata RSME setiap replikasi
- 7) Pengujian hipotesis

Analisis yang dilakukan dalam penelitian ini adalah (1) membandingkan RSME pada A₁B₁ dan A₁B₂, (2) membandingkan RSME pada kelompok A₂B₁ dan A₂B₂, (3) membandingkan RSME pada kelompok A₁B₁ dan A₂B₁ dan (4) membandingkan RSME pada kelompok A₁B₂ dan A₂B₂.

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1. Deskripsi Data

Pada gambar 1 menunjukkan distribusi rata-rata RMSE dari 30 replikasi kelompok A₂B₂ lebih homogen dibandingkan dengan kelompok A₁B₁, A₁B₂ dan A₂B₁. Distribusi rata-rata RMSE A₂B₁ lebih homogen dibandingkan A₁B₁ dan A₁B₂. Distribusi rata-rata RMSE A₁B₂ lebih homogen dibandingkan A₁B₁. Distribusi A₁B₁ dan A₁B₂ berbentuk positif. Ini berarti rata-rata RMSE kelompok A₁B₁ dan A₂B₁ berkumpul pada nilai yang rendah. Berdasarkan eksplorasi data dari gambar 1, nampak rata-rata RMSE A₁B₁ tidak berbeda dengan A₁B₂, rata-rata RMSE A₂B₁ tidak berbeda A₂B₂, sementara rata RMSE A₁B₁ lebih besar daripada A₂B₁, rata-rata RMSE A₁B₂ lebih besar daripada A₂B₂.



Gambar 1. Boxplot Rata-rata RSME dari 30 replikasi

4.2. HASIL PENELITIAN

Hasil uji homogenitas varian diperoleh varians dari rata-rata RMSE A₁B₁ dan A₁B₂ tidak berbeda, varian rata-rata RMSE A₂B₁ dan A₂B₂ tidak berbeda,

sedangkan varian rata-rata RMSE A_1B_1 dan A_2B_1 , varian rata-rata RMSE A_1B_2 dan A_2B_2 berbeda maka pengujian perbandingan rata-rata RSME untuk A_1B_1 dan A_1B_2 , A_2B_1 dan A_2B_2 dengan menggunakan uji t, sementara pengujian perbandingan rata-rata RSME untuk rata-rata RMSE A_1B_1 dan A_2B_1 , rata-rata RMSE A_1B_2 dan A_2B_2 dengan menggunakan statistik non paramterik

Hasil pengujian rata-rata RMSE A_1B_1 dan A_1B_2 diperoleh nilai sig. $0,344 > 0.05$, rata-rata RMSE A_2B_1 dan A_2B_2 diperoleh nilai sig. $0,2545 > 0.05$ maka disimpulkan RMSE A_1B_1 dan A_1B_2 tidak berbeda dan RMSE A_2B_1 dan A_2B_2 juga tidak berbeda. Hasil pengujian rata-rata RMSE A_1B_1 dan A_2B_1 serta A_1B_2 dan A_2B_2 diperoleh nilai sig. $0,00 < 0.05$. Dengan membandingkan nilai rata-rata pada tabel 1 maka disimpulkan RSME A_1B_1 lebih rendah dari RMSE A_1B_2 dan RMSE A_2B_2 lebih rendah dari RMSE A_2B_1 .

Tabel 1 Rata-rata RSME Kelompok A_1B_1 , A_2B_1 A_2B_1 dan A_2B_2

Kelompok	Rata-rata RMSE
A_1B_1	0,03450
A_1B_2	0,03740
A_2B_1	0,00956
A_2B_2	0,00844

4.3. PEMBAHASAN

Keakuratan metode *equating* dan metode estimasi dilihat RMSE, semakin kecil RMSE maka semakin akurat metode *equating* dan metode estimasi yang menjadi variabel bebas dalam penelitian ini. Hasil pengujian dengan menggunakan uji t menyatakan RSME kelompok A_1B_1 tidak berbeda dengan RSME kelompok A_1B_2 dan RMSE kelompok A_2B_1 tidak berbeda dengan RSME kelompok A_2B_2 , sehingga dapat dinyatakan bahwa (1) pada ukuran sampel kecil dan *equating* metode Rerata dan Sigma, keakurasian estimasi parameter dengan menggunakan Metode Joint Maximum Likelihood sama dengan metode Bayesian, (2) pada ukuran sampel kecil dan *equating* metode Tegar Rerata dan Sigma, keakurasian estimasi parameter dengan menggunakan metode Joint Maximum Likelihood sama dengan metode Bayesian.

Hasil pengujian dengan menggunakan statistik non parametrik diperoleh RSME A_1B_1 rendah dari RMSE A_2B_1 dan RMSE A_2B_2 lebih rendah dari RMSE

A_2B_1 maka disimpulkan (1) pada ukuran sampel kecil dan estimasi parameter dengan menggunakan metode joint maximum likelihood, metode *equating* Tegar Rerata dan Sigma lebih akurat dibanding dengan metode Rerata dan Sigma, (2) pada ukuran sampel kecil dan estimasi parameter dengan menggunakan metode bayesian, metode *equating* Tegar Rerata dan Sigma lebih akurat dibanding dengan metode Rerata dan Sigma

Hasil penelitian ini menunjukan pada ukuran sampel kecil, metode *equating* yang sama menghasilkan keakurasian yang sama antara metode Joint Maximum Likelihood dan metode Bayesian, sementara pada metode estimasi parameter sama, menghasilkan keakurasian yang berbeda yaitu metode Tegar Rerata dan Sigma lebih akurat daripada Rerata dan Sigma. Perbedaan RSME tergantung dari keakurasian metode *equating* karena pada perhitungan RSME menggunakan estimasi kemampuan peserta setelah disetarakan dan sebelum disetarakan dengan menggunakan metode Joint Maximum Likelihood dan metode Bayesian.

5. SIMPULAN

Pada ukuran sampel kecil, metode *equating* yang sama menghasilkan keakurasian yang sama antara metode Joint Maximum Likelihood, dan metode Bayesian dan pada metode estimasi parameter yang sama menghasilkan metode Tegar Rerata dan Sigma lebih akurat dari metode Rerata dan Sigma.

6. PUSTAKA

- [1].Dorrans, Neil J., "Equating, Concordance, and Expectation," *Applied Psychological Measurement*, 28 (4) , 227-246. 2004.
- [2].Baker, Frank B dan Seock-Ho Kim, Item Response Theory Parameter Estimation Techniques. New York, Marcel Dekker Inc., 2004
- [3].Cohen, Allan S. and Seock-Ho Kim, "Effect of Linking Methods on Detection of DIF," *Journal of Educational Measurement*, 29 Issue 1, 51-66. 1992
- [4],[10]. Kolen, Michael and Robert L Brennan, *Test Equating*. New York: Springer, 1995.
- [5]. Ayalo, R. J. De. The Theory and Practice of Item Response Theory. New York, The Guilford Press, 2009
- [6],[7] Hambleton, Ronald K and Hariharan Swaminathan, Item Response Theory

Principles and Application. Boston: Kluwer.Nijhooff Publishing, 1990

[8].Naga, Dali Santun. Teori Sekor Pada Pengukuran Mental. Jakarta, Nagarani Citrayasa, 2012.

[9].Cohen, Allan S. dan Seock-Ho Kim. *Comparison of Linking and Concurenrent Calibration Under Item Rersponse Theory*. Journal Applied Psychological Measurement. Vol 22 No 2 Juni 1998.