



## KALIBRASI TES CAMPURAN DIKOTOMUS 2PLM DAN POLITOMUS GRM MENGUNAKAN PROSEDUR GRM DAN GPCM

Samritin

Universitas Muhammadiyah Buton

samritin75@gmail.com

### ABSTRACT

Item parameter estimation is an important step in test development. Estimation of test parameters must pay attention to the suitability of the response characteristics with the calibration procedure to be used. This study has simulated mixed test response data of 35 dichotomous items of 2PLM with 5 polytomous items of GRM. Response data was generated using a sample size of 1500 examinees. Response data were analyzed descriptively from the aspect of correlation coefficient, root mean square difference (RMSD), mean absolute difference (MAD), variance, and bias. The results of this study indicate that the estimation of the slope and threshold parameters of the dichotomous test model 2 logistic parameters using GRM calibration produces correlation coefficients, RMSD, MAD, variance, and estimation bias that are almost the same as those of the GPCM calibration; parameter estimation of the slope and threshold of the GRM model polytomous test using GRM calibration produces a smaller correlation coefficient, RMSD, MAD, variance, and estimation bias compared to the GPCM calibration results; and estimation of the slope parameters and threshold of the mixed dichotomous 2 parameter logistic model (2PLM) test and the GRM polytomous test calibrated using the GRM procedure produced a higher correlation coefficient but had lower RMSD, MAD, variance, and bias compared to the results of calibration using the GPCM procedure. Based on the results of data analysis, this study concluded that the parameter estimation of the 2PLM dichotomous and polytomous GRM mixed test is more precise when calibrated using the GRM procedure.

**Keywords:** mixed tests, parameter estimation, GRM, GPCM.

### ABSTRAK (Indonesia)

Estimasi parameter butir merupakan salah tahapan penting dalam pengembang tes. Estimasi parameter tes harus memperhatikan kesesuaian karakteristik respon dengan prosedur kalibrasi yang hendak digunakan. Penelitian ini telah mensimulasikan data respon tes campuran 35 item dikotomus 2PLM dengan 5 butir politomus GRM. Data respon dibangkitkan menggunakan sample size 1500 examinees. Data respon dianalisis secara deskriptif dari aspek correlation coefficient, root mean square difference (RMSD), mean absolute difference (MAD), varians, dan bias. Hasil penelitian ini menunjukkan bahwa estimasi parameter slope dan threshold tes dikotomus model 2 parameter logistik menggunakan kalibrasi GRM menghasilkan koefisien korelasi, RMSD, MAD, varians, dan bias estimasi yang hampir sama dengan hasil kalibrasi GPCM; estimasi parameter slope dan threshold tes politomus model GRM menggunakan kalibrasi GRM menghasilkan koefisien korelasi, RMSD, MAD, varians, dan bias estimasi yang lebih kecil dibandingkan dengan hasil kalibrasi GPCM; dan estimasi parameter slope dan threshold tes campuran dikotomus 2 parameter logistik model (2PLM) dan tes politomus GRM yang dikalibrasi menggunakan prosedur GRM menghasilkan koefisien korelasi yang lebih tinggi namun memiliki RMSD, MAD, varians, dan bias lebih rendah dibandingkan dengan hasil kalibrasi menggunakan prosedur GPCM. Berdasarkan hasil analisis data penelitian ini menyimpulkan bahwa estimasi parameter tes campuran dikotomus 2PLM dan politomus GRM lebih tepat bila dikalibrasi menggunakan prosedur GRM.

**Kata kunci:** tes campuran, estimasi parameter, GRM, GPCM.

## A. PENDAHULUAN

Pengembangan instrumen penilaian pendidikan dapat didasarkan pada salah satu dari dua teori tes yaitu teori tes klasik (*Classical Test Theory-CTT*) dan teori respons butir (*Item Response Theory-IRT*). Kedua teori tes tersebut dapat diaplikasikan untuk menganalisis *item* tes yang bersifat dikotomis maupun politomis. Teori tes klasik sangat umum digunakan di Indonesia. Namun, teori tes klasik memiliki banyak kelemahan Retnawati (2014). Teori respon butir hadir untuk mengatasi kelemahan-kelemahan yang terdapat pada teori tes klasik dalam penilaian kualitas tes, butir tes, dan interpretasinya.

Teori respon butir dapat digunakan untuk menganalisis parameter kemampuan peserta tes dan parameter butir semua bentuk tes baik tes dikotomis maupun politomis. Tes bentuk dikotomis dapat dianalisis menggunakan model IRT dikotomis dan tes dalam bentuk uraian dapat dianalisis dengan model IRT politomis. Ada beberapa model politomis yang telah diusulkan untuk model-model unidimensi antara lain: *nominal response model (NRM)*, *response model for multiple-choice items*, *rating scale model (RSM)*, *partial credit model (PCM)*, *graded response model (GRM)*, *sequential model for ordered response*, dan *generalized partial credit model (GPCM)* (Van der Linden & Hambleton, 1997; DeMars, 2010). Kelima model tersebut masing-masing memiliki karakteristik yang berbeda.

Pada dasarnya model IRT politomis dapat dikelompokkan ke dalam dua macam, yaitu model respons nominal dan model respons ordinal. Pengelompokan ini ditentukan oleh asumsi tentang karakteristik data. Jika suatu butir memiliki alternatif jawaban yang tidak terurut serta adanya karakteristik tingkat kemampuan yang diukur maka model yang digunakan adalah model ordinal. Model respons nominal dapat diterapkan pada butir yang mempunyai alternatif jawaban yang tidak terurut dengan adanya berbagai tingkat kemampuan yang diukur (Retnawati, 2014). Model GPCM, GRM, RSM dan PCM termasuk dalam model respons ordinal.



Kesesuaian model IRT dalam kalibrasi tes perlu diperhatikan. Tujuan utama menyesuaikan model IRT adalah untuk memperoleh ketepatan informasi tentang parameter item (misalnya, kesulitan item, diskriminasi item, kemampuan peserta ujian) dari data respon item (Kim dan Bolt, 2007). Ketidakesesuaian model kalibrasi akan menghasilkan penyimpangan terhadap parameter yang sesungguhnya. Dengan pendalaman penelitian, estimasi parameter IRT akan lebih tepat, lebih cepat, dan lebih mudah tersebar untuk berbagai kebutuhan misalnya tes adaptif, sistem penilaian bakat dan sebagainya, dan sistem pengujian akan semakin sempurna (Wang dan Ma, 2010). Beberapa penelitian telah dilakukan untuk mengetahui ketepatan penggunaan model *GRM* dan *GPCM* dalam pengukuran parameter-parameter hasil pengukuran. Dodd melakukan analisis baik dengan *GRM* maupun dengan *GPCM* terhadap pengukuran instrumen sikap (De Ayala, 1993). Disebutkan bahwa De Ayala, Dodd dan Koch juga telah membandingkan penerapan model *GRM* dan *GPCM* pada *CAT* (De Ayala, 1993). Pertanyaan yang muncul kemudian adalah bagaimanakah hasil kalibrasi menggunakan prosedur *GRM* dan prosedur *GPCM* terhadap parameter tes campuran dikotomus dan politomus model *GRM*? Penelitian ini bertujuan untuk mendeskripsikan perbandingan hasil estimasi parameter butir tes campuran multiple choice model 2PL dan uraian model *GRM* yang dikalibrasi menggunakan prosedur *GRM* dan yang dikalibrasi menggunakan prosedur *GPCM*.

## B. METODE

Penelitian ini merupakan penelitian deskriptif yang membandingkan hasil kalibrasi menggunakan prosedur *GRM* dan *GPCM*. Pelaksanaan penelitian ini menggunakan desain simulasi. Simulasi data dibangkitkan dan dikendalikan menggunakan Program Aplikasi Komputer. Pembangkitan data ditentukan menggunakan *sample size* (N) sebanyak 500 *examinees*, paket tes campuran dikotomus

berbentuk *multiple choice* 4 (empat) alternatif jawaban dan tes politomus 4 (empat) jenjang skor. Menurut Mislevy & Bock (1990), paket tes tersebut termasuk dalam tes panjang. Data driven yang diperoleh dianalisis secara deskriptif menggunakan bantuan Program Aplikasi *Microsoft Office Excel*. Aspek yang dianalisis dalam penelitian ini yaitu Koefisien Korelasi, *Mean Square Difference* (MSD), *Mean Absolute Difference* (MAD), Varians, dan Bias Estmasi. Secara detail prosedur penelitian ini adalah sebagai berikut:

1. Membangkitkan data menggunakan Program Aplikasi Komputer WinGen-2 dengan *sample size* sebanyak 1500 *examinees*, *number of item* 35 *multiple choice* dengan model 2 PL pada parameter  $a = 0 - 2$  dan parameter  $b = -2$  sampai 2 (Hambleton & Swaminathan, 1985). Selanjutnya ditambahkan 5 item politomus yang bangkitkan dengan model GRM lalu dikalibrasi menggunakan WinGen sehingga diperoleh *true parameter* dan data respon 30 MC + 5 GRM. Pembangkitan data ini dilakukan 5 replikasi.
2. Membangkitkan data seperti langkah ke-1 tetapi item politomus dibangkitkan dengan model GPCM.
3. Data respon yang telah dibangkitkan pada langkah ke-2 dikalibrasi menggunakan Program Aplikasi Komputer Parscale. Pada langkah ke-2 ini setiap data respon di samping dikalibrasi dengan prosedur GRM, juga dikalibrasi dengan prosedur GPCM.
4. Menghitung koefisien korelasi, *mean absolute difference* (MAD), *mean square difference* (MSD), varians dan bias estimasi parameter butir hasil kalibrasi GRM dan GPCM.
5. Melakukan komparasi koefisien korelasi, *mean absolute difference* (MAD), *mean square difference* (MSD), varians dan bias estimasi parameter item antara hasil kalibrasi menggunakan prosedur GRM dengan hasil kalibrasi menggunakan prosedur GPCM.



### C. HASIL PENELITIAN DAN PEMBAHASAN

Data hasil simulasi yang dibangkitkan dan dikalibrasi secara serentak dianalisis secara parsial dan serentak. Secara parsial, data dipisahkan antara hasil kalibrasi tes dikotomus dan hasil kalibrasi tes politomus. Data politomus yang dihasilkan memiliki empat tingkatan jawaban sehingga *threshold* yang diperoleh sebanyak empat tingkatan yaitu  $b_1$ ,  $b_2$ ,  $b_3$ , dan  $b_4$ . Data parsial hasil simulasi disajikan dalam Tabel 1 dan Tabel 2.

Tabel 1. Data Hasil Kalibrasi Parameter Tes Dikotomus

Aspek Analisi	Kalibrasi	<i>Slope</i> (a)	<i>Threshold</i> (b)
Koefisien Korelasi	GRM	0,9899	0,9416
	GPCM	0,9897	0,9416
RMSD	GRM	0,0978	0,3895
	GPCM	0,1022	0,3897
MAD	GRM	0,0764	0,1541
	GPCM	0,0798	0,1538
Bias	GRM	0,0034	0,1388
	GPCM	0,0042	0,1393
Varians	GRM	0,0061	0,0130
	GPCM	0,0063	0,0125

Tabel 1 menunjukkan nilai-nilai perbandingan koefisien korelasi, RMSD, MAD, varians, dan bias estimasi parameter daya beda (*slope-a*) dan parameter tingkat kesukaran (*threshold-b*) tes dikotomus bentuk *multiple choice* hasil kalibrasi menggunakan prosedur *Graded Response Model* (GRM) dan prosedur *Generalized Partial Credit Model* (GPCM). Pada tabel 1 terlihat bahwa koefisien korelasi *slope* hasil kalibrasi GRM lebih tinggi dibandingkan dengan hasil kalibrasi GPCM. Namun, selisihnya sangat kecil yaitu 0,0002 atau 0,02%. Sementara, koefisien korelasi *threshold* hasil kalibrasi GRM dan GPCM tidak memiliki perbedaan. Koefisien korelasi *slope* dan koefisien korelasi *threshold* masing-masing mendekati nilai korelasi maksimum.

Pada tabel 1 tampak bahwa RMSD estimasi parameter *slope* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Selisih RMSD hasil estimasi *slope* sebesar 0,044 atau mencapai 4,31%. RMSD estimasi parameter *threshold* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Selisih RMSD *threshold* 0,0002 atau 0,05%. Selisih RMSD estimasi *threshold* lebih kecil dibandingkan dengan selisih RMSD estimasi *slope*.

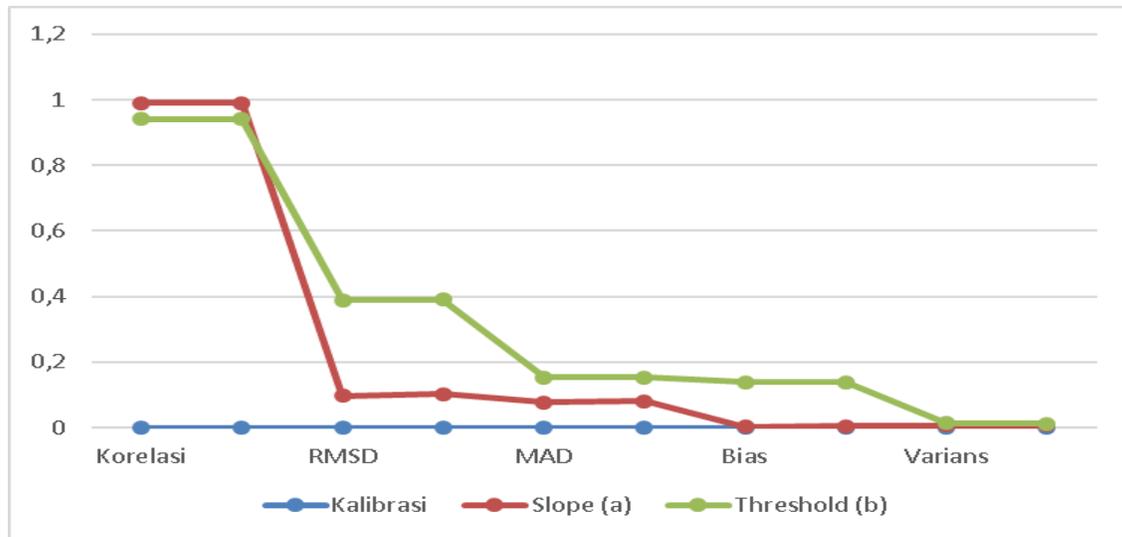
*Mean square difference* (MAD) estimasi parameter *slope* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM pada Tabel 1 terlihat lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Selisih MAD parameter *slope* sebesar 0,034 atau mencapai 4,26%. Selisih yang cukup besar. Namun, MAD estimasi parameter *threshold* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM lebih besar dibandingkan dengan hasil kalibrasi GPCM. Selisih MAD *threshold* 0,003 atau 0,20%. Meskipun nilai selisih MAD *threshold* sangat kecil tetapi temuan ini menunjukkan ketidakkonsistenan hasil estimasi parameter *slope* dengan parameter *threshold*.

Varians estimasi parameter *slope* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Tabel 1 menunjukkan bahwa selisih varians estimasi *slope* tes *multiple choice* sebesar 0,002 atau 3,17%. Kondisi yang berbeda ditemukan pada varians estimasi parameter *threshold*. Varians estimasi parameter *threshold* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM lebih tinggi dibandingkan dengan hasil kalibrasi GPCM. Selisih varians estimasi *threshold* tes MC sebesar 0,005 atau mencapai 4%. Hal ini berarti bahwa ketepatan estimasi parameter *threshold* tidak konsisten dengan parameter *slope* bila kalibrasi dilakukan menggunakan GRM dan GPCM. Kecenderungan ketidak konsistenan tersebut dapat dilihat pada Gambar 1.

Bias estimasi parameter *slope* butir-butir tes dikotomus *multiple choice* hasil kalibrasi GRM yang ditunjukkan lebih kecil dibandingkan dengan hasil kalibrasi GPCM, sebagaimana yang ditunjukkan oleh Tabel 1. Selisih bias estimasi *slope* tes dikotomus menggunakan GRM dan GPCM sebesar 0,0008. Selisih bias tersebut cukup besar karena mencapai 19,05%. Bias estimasi parameter *threshold* butir-butir



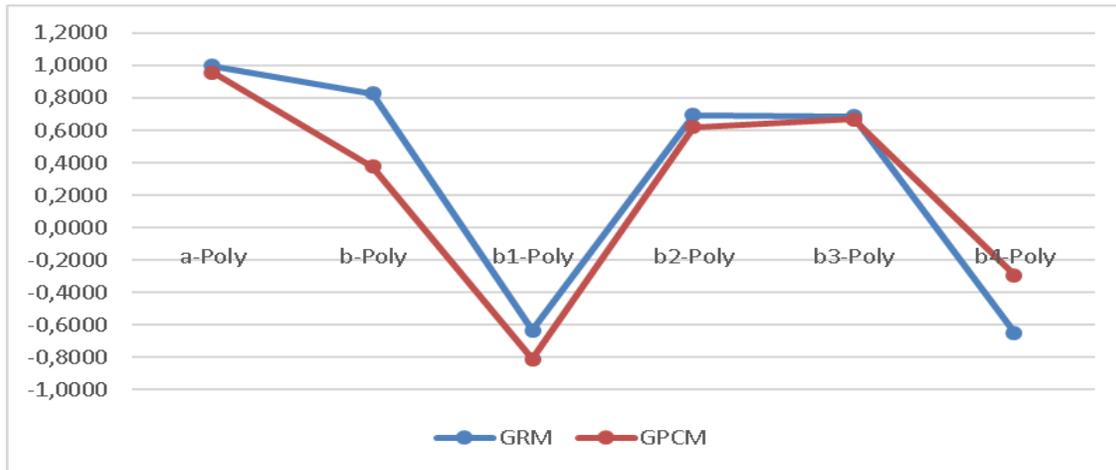
tes dikotomus *multiple choice* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Selisih bias estimasi *threshold* hasil kalibrasi kedua prosedur sebesar 0,005. Selisih bias estimasi ini sangat kecil karena hanya 0,36%. Secara grafik tampak berimpit sebagaimana yang ditunjukkan oleh Gambar 1.



Gambar. 1 Grafil Koef. Korelasi, RMSD, MDA, Varians, dan Bias Estimasi Tes Dikotomus

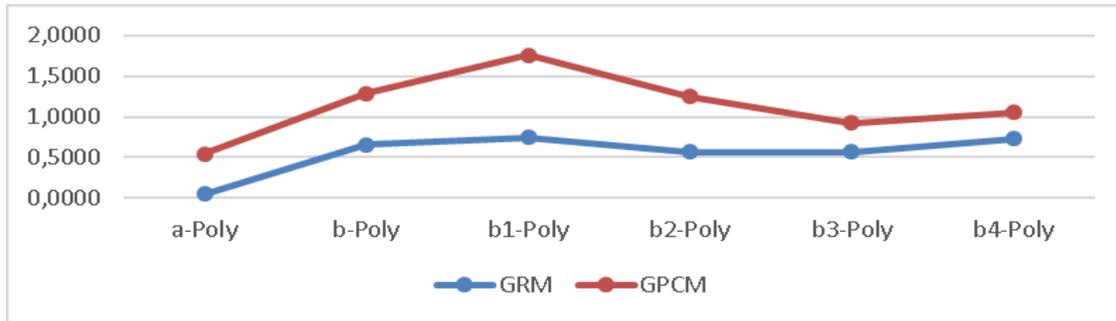
Tabel 2. Data Parsial Hasil Kalibrasi Parameter Tes Politomus

Aspek Analisis	Kalibrasi	Slope (a)		Threshold (b)			
		a-Poly	b-Poly	b <sub>1</sub> -Poly	b <sub>2</sub> -Poly	b <sub>3</sub> -Poly	b <sub>4</sub> -Poly
Koef. Korelasi	GRM	0,9968	0,8258	-0,6331	0,6928	0,6879	-0,6497
	GPCM	0,9562	0,3760	-0,8139	0,6196	0,6676	-0,2930
RMSD	GRM	0,0485	0,6575	0,7490	0,5674	0,5653	0,7258
	GPCM	0,5432	1,2855	1,7600	1,2456	0,9267	1,0498
MAD	GRM	0,0382	0,5713	0,7132	0,4594	0,4442	0,6683
	GPCM	0,5198	1,0703	1,6246	1,0178	0,6743	0,9645
Bias	GRM	0,0003	0,4311	0,5597	0,3213	0,3190	0,5244
	GPCM	0,2931	1,6290	3,0863	1,5233	0,8262	1,0803
Varians	GRM	0,0020	0,0012	0,0013	0,0007	0,0006	0,0023
	GPCM	0,0019	0,0234	0,0111	0,0281	0,0327	0,0218



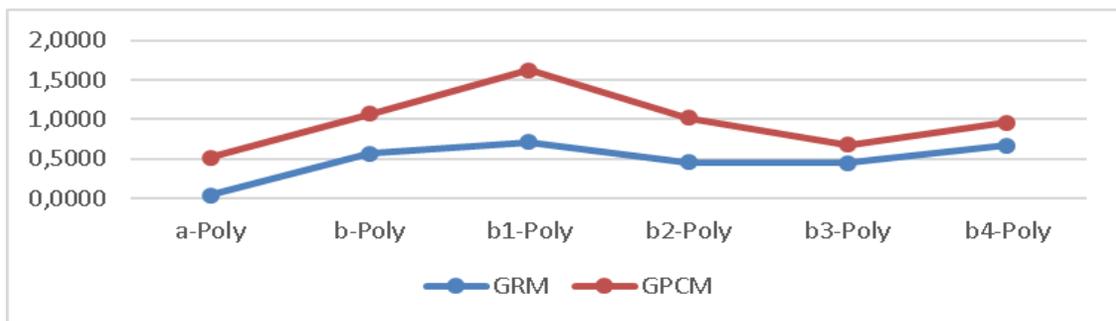
Gambar 2. Grafik Koefisien Korelasi Tes Politomus

Gambar 2 menunjukkan kecenderungan koefisien korelasi parameter *slope* dan koefisien korelasi parameter *threshold* hasil estimasi GRM dan GPCM. Koefisien korelasi parameter *slope* hasil kalibrasi GRM lebih tinggi dibandingkan dengan hasil kalibrasi GPCM. Selisih koefisien korelasi *slope* (a) sebesar 0,0406 atau mencapai 4,25%. Koefisien korelasi parameter *threshold* untuk semua jenjang skor pada gambar terlihat bahwa hasil kalibrasi GRM lebih tinggi daripada hasil kalibrasi GPCM. Selisih koefisien korelasi *threshold* skor 1 ( $b_1$ ) memiliki selisih korelasi sebesar 0,1808 atau 22,214%. Selisih korelasi *threshold*  $b_2$  memiliki sebesar 0,0732 atau 11,814%. Selisih koefisien korelasi *threshold*  $b_3$  sebesar 0,0203 atau 3,041%. Selisih koefisien korelasi *threshold*  $b_4$  sebesar 0,3567 atau 121,74%. Selisih koefisien korelasi *threshold*  $b$ -total sangat tinggi yaitu sebesar 0,4498 atau 119,63%. Nilai-nilai selisih koefisien korelasi, baik *slope* maupun *threshold* setiap jenjang skor menunjukkan bahwa parameter *slope* (a) dan *threshold* (bi) tes politomus bentuk GRM lebih tepat bila dikalibrasi menggunakan prosedur GRM.



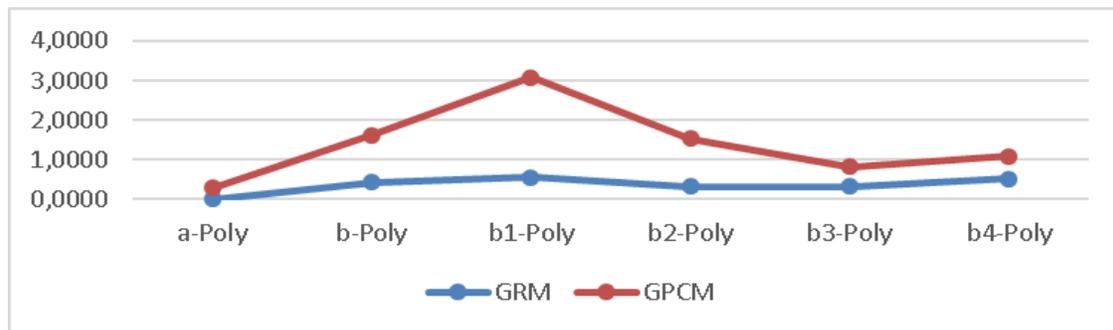
Gambar 3. Grafik RMSD Tes Politomus

Gambar 3 menunjukkan kecenderungan RMSD parameter *slope* dan RMSD parameter *threshold* hasil estimasi GRM dan GPCM tes politomus 4 jenjang skor. Berdasarkan tabel 1, RMSD parameter *slope* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Selisih RMSD *slope* (a) sebesar 49,47 atau mencapai 91,071%. RMSD parameter *threshold* untuk semua jenjang skor pada gambar terlihat bahwa hasil kalibrasi GRM lebih rendah daripada hasil kalibrasi GPCM. Selisih RMSD *threshold*  $b_1$  sebesar 1,011 atau 57,443%. Selisih RMSD *threshold*  $b_2$  sebesar 0,678 atau 57,45%. Selisih RMSD *threshold*  $b_3$  sebesar 0,3614 atau 38,999%. Selisih RMSD *threshold*  $b_4$  sebesar 0,324 atau 30,863%. Selisih RMSD *threshold* b-total cukup tinggi yaitu sebesar 0,628 atau 48,53%. Nilai-nilai selisih koefisien korelasi, baik *slope* maupun *threshold* setiap jenjang skor menunjukkan bahwa parameter *slope* (a) dan *threshold* (bi) tes politomus bentuk GRM lebih tepat bila dikalibrasi menggunakan prosedur GRM.



Gambar 4. Grafik MAD Tes Politomus

Gambar 4 menunjukkan kecenderungan MAD parameter *slope* dan MAD parameter *threshold* hasil estimasi GRM dan GPCM tes politomus 4 jenjang skor. MAD parameter *slope* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Berdasarkan Tabel 2 diperoleh selisih MAD *slope* (a) sebesar 48,16 atau mencapai 92,651%. MAD parameter *threshold* untuk semua jenjang skor pada gambar terlihat bahwa hasil kalibrasi GRM lebih rendah daripada hasil kalibrasi GPCM. Selisih MAD *threshold*  $b_1$  sebesar 0,9114 atau 56,1%. Selisih MAD *threshold*  $b_2$  sebesar 0,5584 atau 54,863%. Selisih MAD *threshold*  $b_3$  sebesar 0,2301 atau 34,124%. Selisih MAD *threshold*  $b_4$  sebesar 0,2962 atau 30,71%. Selisih MAD *threshold* b-total cukup tinggi yaitu sebesar 0,499 atau 46,622%. Nilai-nilai selisih koefisien korelasi, baik *slope* maupun *threshold* setiap jenjang skor menunjukkan bahwa parameter *slope* (a) dan *threshold* (bi) tes politomus bentuk GRM lebih tepat bila dikalibrasi menggunakan prosedur GRM.

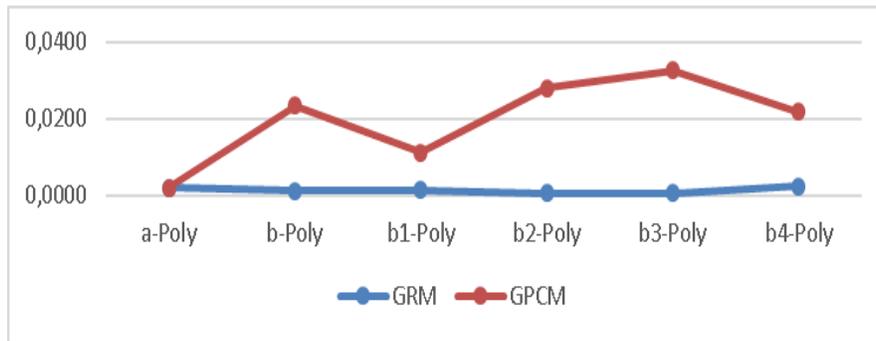


Gambar 5. Grafik Bias Estimasi Tes Politomus

Gambar 5 menunjukkan kecenderungan bias estimasi parameter *slope* dan bias estimasi parameter *threshold* hasil kalibrasi GRM dan GPCM tes politomus 4 jenjang skor. Bias parameter *slope* hasil kalibrasi GRM lebih kecil dibandingkan dengan hasil kalibrasi GPCM. Berdasarkan Tabel 2 diperoleh selisih bias *slope* (a) sebesar 0,2928 atau mencapai 99,898%. Bias parameter *threshold* untuk semua jenjang skor pada gambar terlihat bahwa hasil kalibrasi GRM lebih rendah daripada hasil kalibrasi GPCM. Selisih Bias *threshold*  $b_1$  sebesar 2,5266 atau 81,863%. Selisih Bias *threshold*  $b_2$  sebesar 1,202 atau 78,908%. Selisih Bias *threshold*  $b_3$  sebesar 0,5072 atau 61,3895%.



Selisih Bias *threshold*  $b_4$  sebesar 0,5559 atau 51,4579%. Selisih Bias *threshold*  $b$ -total sangat tinggi yaitu sebesar 1,1979 atau 73,5359%. Nilai-nilai selisih koefisien korelasi, baik *slope* maupun *threshold* setiap jenjang skor menunjukkan bahwa parameter *slope* (a) dan *threshold* ( $b_i$ ) tes politomus bentuk GRM lebih tepat bila dikalibrasi menggunakan prosedur GRM.



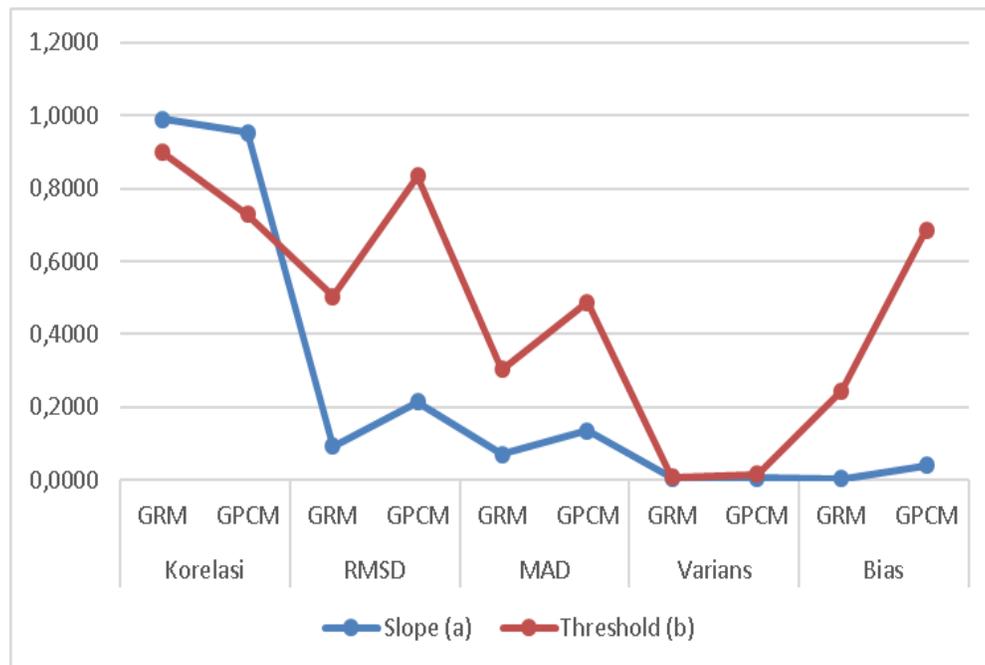
Gambar 6. Grafik Varians Estimasi Tes Politomus

Gambar 6 menunjukkan kecenderungan varians parameter *slope* dan varians parameter *threshold* hasil estimasi GRM dan GPCM tes politomus 4 jenjang skor. Varians parameter *slope* hasil kalibrasi GRM lebih tinggi dibandingkan dengan hasil kalibrasi GPCM. Berdasarkan Tabel 2 diperoleh selisih varians *slope* (a) sebesar 0,0001 atau 5,263%. Varians parameter *threshold* untuk semua jenjang skor pada gambar terlihat bahwa hasil kalibrasi GRM lebih rendah daripada hasil kalibrasi GPCM. Selisih varians *threshold*  $b_1$  sebesar 0,0098 atau 88,2883%. Selisih varians *threshold*  $b_2$  sebesar 0,0274 atau 97,5089%. Selisih varians *threshold*  $b_3$  sebesar 0,0321 atau 98,1651%. Selisih varians *threshold*  $b_4$  sebesar 0,0195 atau 89,4495%. Selisih varians *threshold*  $b_{total}$  cukup tinggi yaitu sebesar 0,0222 atau 94,8718%. Nilai-nilai selisih varians, baik *slope* maupun *threshold* setiap jenjang skor menunjukkan bahwa parameter *slope* (a) dan *threshold* ( $b_i$ ) tes politomus bentuk GRM lebih tepat bila dikalibrasi menggunakan prosedur GRM.

Setelah dianalisis secara parsial, data hasil kalibrasi selanjutnya dianalisis secara serentak tanpa memisahkan item-item tes dikotomus dan politomus. Data hasil estimasi koefisien korelasi, RMSD, MAD, varians, dan bias estimasi parameter *slope* (a) dan *threshold* (b) tes campuran dikotomus dan politomus disajikan pada Tabel 3.

Tabel 3. Data Hasil Kalibrasi Tes Campuran Dikotomus dan Politomus GRM

Aspek Analisis	Kalibrasi	Parameter	
		<i>Slope</i> (a)	<i>Threshold</i> (b)
Koefisien Korelasi	GRM	0,9903	0,8994
	GPCM	0,9531	0,7296
RMSD	GRM	0,0931	0,5038
	GPCM	0,2145	0,8351
MAD	GRM	0,0716	0,3040
	GPCM	0,1348	0,4875
Varians	GRM	0,0056	0,0084
	GPCM	0,0057	0,0168
Bias	GRM	0,0030	0,2454
	GPCM	0,0403	0,6870



Gambar 7. Grafik Hasil Kalibrasi Tes Campuran Dikotomus dan Politomus GRM



Dari Gambar 7 di atas tampak kecenderungan koefisien korelasi *slope* (a) dan *threshold* (b) tes campuran MC model 2PL dan uraian model GRM yang dikalibrasi menggunakan prosedur GPCM dan hasil kalibrasi menggunakan prosedur GRM. Koefisien korelasi hasil kalibrasi *slope* (a) menggunakan GPCM tampak lebih rendah dibandingkan dengan hasil kalibrasi menggunakan prosedur GRM. Berdasarkan Tabel 3, selisih koefisien korelasi kedua hasil kalibrasi sebesar 0,0372. Nilai selisih ini cukup besar yang mengindikasikan bahwa estimasi parameter daya beda (*slope*) tes campuran MC model 2PL dan uraian model GRM lebih tepat jika dikalibrasi menggunakan prosedur GRM. Tabel 3 juga memperlihatkan bahwa koefisien korelasi *threshold* (b) yang dikalibrasi menggunakan prosedur GRM lebih tinggi dibandingkan dengan hasil kalibrasi menggunakan GPCM. Koefisien korelasi hasil kalibrasi GPCM terpaut 0,1698. Hal ini juga mengindikasikan bahwa estimasi parameter *threshold* (b) dari tes campuran MC model 2PL dan uraian model GRM lebih tepat jika dikalibrasi menggunakan prosedur GRM.

Perbedaan hasil estimasi berdasarkan *root mean square difference* (RMSD) *slope* (daya beda) dan *threshold* (tingkat kesukaran) butir tes hasil kalibrasi menggunakan GRM dan GPCM berdasarkan RMSD-nya tampak pada tabel Tabel 3 dan Gambar 7. RMSD *slope* hasil kalibrasi menggunakan GRM sebesar 0,0931. RMSD ini lebih rendah 0,1214 dibandingkan dengan RMSD hasil kalibrasi *slope* menggunakan GPCM. Persentase ketepatan kalibrasi parameter *slope* menggunakan GRM 130,397% lebih tepat dibandingkan dikalibrasi menggunakan GPCM. Hal ini menunjukkan bahwa kalibrasi menggunakan model GRM terhadap parameter *slope* tes campuran 2PL dan uraian model GRM lebih tepat dibandingkan dengan hasil kalibrasi menggunakan GPCM. Hal yang serupa juga terlihat pada RMSD parameter *threshold*. RMSD *threshold* hasil kalibrasi menggunakan prosedur GPCM mencapai 0.8351. Nilai ini melampaui RMSD *threshold* hasil kalibrasi menggunakan prosedur GRM sebesar 0,3313. Persentase pelampauan ini mencapai 65,76%. Hal ini menunjukkan bahwa

kalibrasi parameter *threshold* tes campuran *multiple choice* model 2PL dan uraian model GRM lebih tepat menggunakan prosedur GRM dibandingkan dengan menggunakan prosedur GPCM.

Perbedaan hasil estimasi berdasarkan *mean absolute difference* (MAD) *slope* dan *threshold* butir tes hasil kalibrasi menggunakan GRM dan GPCM tampak jelas pada Tabel 3 dan Gambar 7. RMSD *slope* hasil kalibrasi menggunakan GRM sebesar 0,0716. RMSD ini lebih rendah 0,0632 dibandingkan dengan MAD hasil kalibrasi *slope* menggunakan GPCM. Persentase ketepatan kalibrasi parameter *slope* menggunakan GRM 88,268% lebih tepat dibandingkan dikalibrasi menggunakan GPCM. Hal ini menunjukkan bahwa kalibrasi menggunakan model GRM terhadap parameter *slope* tes campuran 2PL dan uraian model GRM lebih tepat dibandingkan dengan hasil kalibrasi menggunakan GPCM. MAD parameter *threshold* hasil kalibrasi menggunakan GRM juga tampak lebih rendah. MAD *threshold* hasil kalibrasi menggunakan prosedur GRM sebesar 0.3040. Nilai ini lebih tinggi 0,1835 dibandingkan MAD *threshold* hasil kalibrasi menggunakan prosedur GPCM. Persentase pelampauan ini mencapai 60,36%. Hal ini menunjukkan bahwa kalibrasi parameter *threshold* tes campuran *multiple choice* model 2PL dan uraian model GRM lebih tepat menggunakan prosedur GRM dibandingkan dengan menggunakan prosedur GPCM.

Berdasarkan varians estimasi yang ditampilkan Tabel 3 dan Gambar 7 di atas terlihat bahwa *varians true slope* dengan *estimate slope* tes campuran hasil kalibrasi GRM berbeda dengan hasil kalibrasi menggunakan GPCM. Varians *slope* hasil kalibrasi menggunakan GRM sebesar 0.0056. Varians hasil kalibrasi GRM tersebut lebih kecil dibandingkan dengan varians hasil kalibrasi *slope* menggunakan GPCM. Selisih keduanya sebesar 0,0001. Persentase ketepatan kalibrasi parameter *slope* menggunakan GRM sebesar 17,86% lebih tepat dibandingkan bila dikalibrasi menggunakan GPCM. Hal ini menunjukkan bahwa kalibrasi menggunakan model GRM terhadap parameter *slope* tes campuran 2PL dan uraian model GRM lebih tepat dibandingkan dengan hasil kalibrasi menggunakan GPCM. Varians perbedaan parameter *threshold* hasil kalibrasi menggunakan GRM juga tampak lebih rendah. Varians *threshold* hasil kalibrasi menggunakan prosedur GRM sebesar 0.0084,



sedangkan varians hasil kalibrasi GPCM sebesar 0.0168. Selisih keduanya sebesar 0.0084 atau hasil kalibrasi menggunakan GPCM menghasilkan varians dua kali lebih tinggi dibandingkan hasil kalibrasi menggunakan GRM. Hal ini menunjukkan bahwa kalibrasi parameter *threshold* tes campuran *multiple choice* model 2PL dan uraian model GRM lebih tepat menggunakan prosedur GRM dibandingkan dengan menggunakan prosedur GPCM.

Estimasi parameter menggunakan prosedur GRM dan GPCM menghasilkan bias estimasi yang berbeda. Hal ini dapat dilihat pada Tabel 3 dan Gambar 4. Bias estimasi *slope* hasil kalibrasi menggunakan GRM sebesar 0.0030. Bias estimasi hasil kalibrasi GRM tersebut lebih kecil dibandingkan dengan varians hasil kalibrasi *slope* menggunakan GPCM. Selisih keduanya sebesar 0,0373. Persentase selisih bias hasil kalibrasi parameter *slope* menggunakan GRM sebesar 1243%. Hal ini menunjukkan bahwa kalibrasi menggunakan model GRM terhadap parameter *slope* tes campuran 2PL dan uraian model GRM lebih tepat dibandingkan dengan hasil kalibrasi menggunakan GPCM. Bias estimasi parameter *threshold* hasil kalibrasi menggunakan GRM juga tampak lebih rendah. Bias *threshold* hasil kalibrasi menggunakan prosedur GRM sebesar 0,2454, sedangkan varians hasil kalibrasi GPCM sebesar 0,6870. Selisih keduanya sebesar 0,4461 atau hasil kalibrasi menggunakan GPCM menghasilkan varians hampir dua kali lebih tinggi dibandingkan hasil kalibrasi menggunakan GRM. Hal ini menunjukkan bahwa kalibrasi parameter *threshold* tes campuran *multiple choice* model 2PL dan uraian model GRM lebih tepat menggunakan prosedur GRM dibandingkan dengan menggunakan prosedur GPCM.

#### D. KESIMPULAN

Berdasarkan pembahasan hasil penelitian di atas dapat disimpulkan bahwa (1) estimasi parameter *slope* dan *threshold* tes dikotomus model 2 paramater logistik dapat dilakukan melalui kalibrasi GRM atau GPCM, (2) estimasi parameter *slope* dan

*threshold* tes politomus model GRM lebih tepat bila di kalibrasi menggunakan GRM daripada menggunakan GPCM, (3) estimasi parameter slop dan *threshold* tes campuran dikotomus 2 parameter logistik model (2PLM) dan tes politomus *graded response model* (GRM) lebih tepat bila dikalibrasi menggunakan prosedur GRM daripada prosedur *generalized partial credit model* (GPCM)

#### DAFTAR PUSTAKA

- Awopeju, O. A. & Afolabi, E. R. I. (2016) Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics. *European Scientific Journal*, 12(28), 263-284.
- De Ayala, R.J., 1993. *An Introduction to Polytomus Item Response Theory Models*. Measurement and Evaluation in Counseling and Development, 25, 172-189.
- DeMars C. (2010). *Item Response Theory*, New York: Oxford University Press, Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- Kim, T.S. & Bolt, D.M. (2007). *Estimating Item Response Theory Models Using Markov Chain Monte Carlo Methods*. Educational Measurement: Issues and Practice. Pp. 38-41
- Mislevy, R.J. & Bock, R.D. (1990). *BILOG 3: Item analysis & test scoring with binary logistic models*. Moorseville: Scientific Software, Inc.
- Retnawati, H. (2014). *Teori respon butir dan penerapannya*. Yogyakarta: Nuha Medika.
- Van der Liden, W.J. & Hambleton, R.K. (1997). *Item response theory: brief history, common models and extention*. Dalam Van der Liden, W.J. dan Hambleton, R.K. (Eds). *Handbook of Item Response Theory*. New York: Springer.