



EFEK NUMBER OF ITEM DAN SAMPLE SIZE TERHADAP AKURASI ABILITY ESTIMATION PADA MODEL IRT TIGA PARAMETER LOGISTIK

Samritin
Universitas Muhammadiyah Buton
samritin75@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengetahui efek number of item dan sampel size terhadap akurasi ability estimation pada Model IRT 3PL. Penelitian ini merupakan penelitian simulasi yang menggunakan tiga variasi number of item yaitu 30, 40, dan 60 dengan tiga variasi sample size yaitu 500, 750, dan 1000. Data penelitian ini diperoleh dari hasil pembangkitan oleh program aplikasi komputer. Data yang diperoleh dianalisis secara deskriptif menggunakan correlation coefficient, mean square difference (MSD), dan bias. Hasil penelitian ini menunjukkan bahwa berdasarkan koefisien korelasi akurasi tertinggi diperoleh pada $N=1000$ dengan $n=60$ tetapi MSD terendah diperoleh pada $N=750$ dan $n=60$, bias terendah pada $N=500$ dan $n=60$. Berdasarkan hasil analisis MSD dan bias penelitian ini menyimpulkan bahwa sampel size memiliki efek yang tidak konsisten terhadap estimasi ability. Penelitian ini juga menemukan bahwa semakin banyak butir yang diujikan maka estimasi kemampuan peserta tes semakin akurat.

Kata kunci: *sample size, ability estimation.*

Abstract

This study aims to determine the effect of the number of items and sample size on the estimated accuracy of the 3PL IRT Model. This research is a simulation research that uses three variations of the number of items, namely 30, 40, and 60 with three variations of the sample size, namely 500, 750, and 1000. The research data is obtained from the results generated by a computer application program. The data obtained were analyzed descriptively using the correlation coefficient, mean square difference (MSD), and bias. The results of this study indicate that based on the correlation coefficient the highest accuracy was obtained at $N=1000$ with $n=60$ but the lowest MSD was obtained at $N=750$ and $n=60$, the lowest bias was at $N=500$ and $n=60$. Based on the results of the MSD analysis and the bias of this study it was concluded that sample size has an unadjusted effect on estimation ability. This study also found that the more items tested, the more accurate the estimation of the ability of the test takers.

Keywords: *sample size, ability estimation.*

Pendahuluan

Tes dapat dimaknai sebagai instrumen atau prosedur penentuan suatu atribut. Pengujian merupakan suatu pengukuran perilaku atau cara mengukur suatu perilaku guna memahami dan memprediksi suatu perilaku (Kaplan, 1993). Baik dalam psikologi maupun pendidikan, pengujian dilakukan dengan tujuan tertentu, seperti penempatan seseorang sesuai dengan bidangnya, sebagai bahan pertimbangan dalam pengambilan keputusan, atau memberikan informasi untuk mengukur seberapa baik hasil dari suatu pembelajaran (Febrayosi, 2012). Pengukuran di bidang pendidikan dikenal dua pendekatan yaitu teori tes klasik (classical test theory-CTT) dan teori tes modern (item response theory-IRT).

CTT merupakan pendekatan yang telah digunakan secara luas dan mendominasi pelaksanaan pengukuran di sekolah-sekolah. Namun, CTT memiliki banyak kelemahan antara lain ketergantungan parameter item terhadap kemampuan peserta tes sangat tinggi. Butir-butir tes akan tampak sederhana dan memiliki tingkat kesulitan yang rendah jika dijawab oleh peserta tes yang memiliki kemampuan tinggi. Butir-butir tes akan tampak menantang dan memiliki tingkat kesulitan yang tinggi jika dikerjakan oleh peserta tes yang memiliki kemampuan rendah. IRT hadir dengan berbagai kelebihan yang dimilikinya. Berbeda dengan CTT, IRT memiliki karakteristik independensi parameter butir yang diukur terhadap kemampuan peserta tes, dan sebaliknya. Parameter butir suatu butir tes tidak akan berubah meskipun dikerjakan oleh kelompok peserta tes memiliki kemampuan berbeda dengan yang sebelumnya dan kemampuan peserta tes tidak akan mengalami perubahan meskipun mengerjakan tes dengan tingkat kesulitan yang berbeda (Retnawati, 2014). Namun IRT tidak hanya memiliki kelebihan, tetapi juga memiliki kelemahan, utamanya yang berkaitan penggunaan konsep matematika yang lebih kompleks dibandingkan IRT dalam penentuan parameter.

Proses penentuan parameter dikenal sebagai estimasi dalam teori modern (IRT). Estimasi parameter disebut juga sebagai kalibrasi (Huriaty, 2015) merupakan suatu upaya menemukan nilai parameter yang mendekati keadaan sebenarnya. Menurut Diana dan Soehardjoepri (2016), estimasi adalah teknik yang menggunakan nilai sampel untuk memperkirakan nilai populasi. Estimasi parameter terdiri dari estimasi parameter kemampuan (ability) peserta tes dan estimasi parameter butir yang dilakukan berdasarkan data yang diperoleh melalui pengukuran. Pengukuran peserta tes yang menjawab banyak pertanyaan

digunakan untuk mengestimasi parameter kemampuan peserta tes, sedangkan pengukuran beberapa peserta tes yang menjawab pertanyaan yang sama digunakan untuk mengestimasi parameter untuk satu item tes.

IRT memberikan syarat-syarat tertentu yang harus dipenuhi dalam estimasi parameter. Syarat-syarat yang harus dipenuhi oleh butir tes, peserta, atau gabungan keduanya dalam IRT, antara lain: 1) model IRT yang akan digunakan; 2) jumlah parameter yang akan diestimasi; 3) butir soal yang akan digunakan dalam tes harus memenuhi persyaratan unidimensi dan peserta tes harus memenuhi persyaratan dependensi lokal; dan 4) karakteristik peserta tes (Naga, 1992). Estimasi parameter dalam IRT tidak dapat dipisahkan dari model yang digunakan. Tiga komponen dasar yang saling terkait dalam estimasi parameter yaitu butir tes, peserta tes, dan jawaban peserta terhadap butir tes (Hikamudin, 2012). Ketiga komponen ini saling berkaitan sehingga dapat dinyatakan dalam bentuk fungsi matematika atau kurva karakteristik. Naga (1992) menjelaskan bahwa karakteristik peserta dinyatakan melalui parameter kemampuan (ability) peserta, karakteristik butir dinyatakan melalui tiga parameter butir, yaitu daya pembeda (a), tingkat kesulitan (b), dan tebakan atau *pseudo-guessing* (c), sedangkan karakteristik respon dinyatakan dalam bentuk probabilitas jawaban benar.

Kemampuan peserta tes (θ) dalam IRT model tiga parameter logistik (3PL) dinyatakan dalam bentuk probabilitas kemampuan. Menurut Hambleton, Swaminathan, dan Rogers, (1991), van der Liden dan Hambleton (1997) dan Hambleton dan Swaminathan (1985) probabilitas kemampuan peserta tes (θ) pada butir ke- i pada model 3PL diformulasikan sebagai berikut:

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

Estimasi parameter kemampuan peserta tes dalam IRT dapat dipengaruhi oleh faktor-faktor tertentu antara lain banyak peserta tes (*sample size*) dan banyak butir tes (*number of item*) (Hambleton dan Cook, 1983). Penelitian tentang faktor-faktor yang berpengaruh terhadap akurasi estimasi kemampuan sangat penting dilakukan guna menghasilkan estimasi yang tepat. Ketidakuratan estimasi akibat ketidaksesuaian syarat analisis dapat mengakibatkan kesalahan dalam pengambilan keputusan tentang kemampuan seorang peserta tes. Berdasarkan pertimbangan tersebut penelitian ini mengkaji tentang pengaruh ukuran sampel dan banyak butir tes terhadap akurasi estimasi kemampuan peserta tes. Penelitian difokuskan pada besarnya

koefisien korelasi *true ability* dengan *estimate ability*, *mean square difference* (MSD) antara *true ability* dengan *estimate ability* serta bias estimasinya.

Metode Penelitian

Penelitian ini merupakan penelitian deskriptif yang didesain menggunakan data simulasi (bangkitan). Penelitian ini menggunakan WinGen sebagai data generator. Generating data dilakukan berdasarkan karakteristik yang telah ditentukan. Karakteristik data bangkitan terdiri atas *number of item*, karakteristik butir tes, *sample size*, serta model parameter logistik. *Number of item* (n) dibangkitkan yaitu 30, 40 dan 60 butir. Menurut Mislevy & Bock (1990) tes ini termasuk dalam kategori tes panjang. Tes dibangkitkan menurut model IRT 3PL dan berdistribusi normal dengan parameter taraf kesukaran (b) pada selang $-2 \leq b \leq +2$ (Hambleton & Swaminathan, 1985), taraf daya beda (a) dari 0 sampai 2 (Hambleton & Swaminathan, 1985), dan *pseudo-guessing* dari 0 sampai 1 (Hullin, 1983). *Sample size* diset dalam tiga variasi yaitu 500, 750, dan 1000 dengan kemampuan peserta (θ) pada rentang $-4 \leq \theta \leq +4$. Untuk menghasilkan analisis yang akurat replikasi dilakukan 5 kali setiap kasus, sehingga banyaknya replikasi yang dilakukan sebanyak 45 replikasi. Analisis data dilakukan dengan menghitung korelasi rata-rata koefisien korelasi (ρ), *mean square difference* (MSD), dan bias.

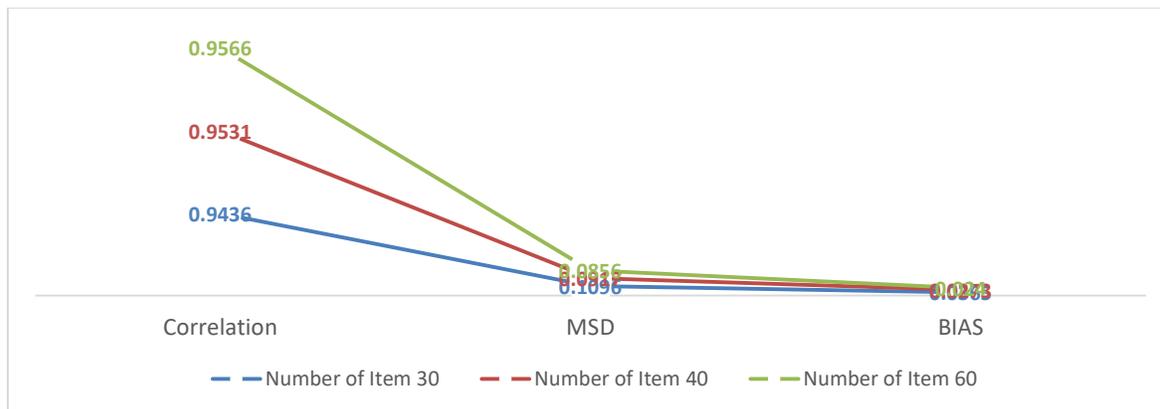
Hasil Penelitian dan Pembahasan

Data penelitian ini diperoleh dari hasil pembangkitan sesuai dengan karakteristik yang dibutuhkan. Data yang diperoleh merupakan data kuantitatif yang berupa koefisien korelasi antara *true ability* dengan *estimate ability*, *mean square difference* (MSD) antara *true parameter* dengan *estimate ability*, dan bias. Koefisien korelasi, MSD, dan bias dihitung rata-ratanya dari 5 kali replikasi. Data-data setiap kasus perubahan korelasi, MSD, dan bias tersebut dapat dilihat pada tabel dan grafik berikut.

Sample size Tetap dan Number of Item Bervariasi

Tabel 1. Koefisien Korelasi, MSD, dan Bias pada N=500

<i>Number of item</i>	Koefisien Korelasi	MSD	Bias
30	0,9436	0,1096	0,0365
40	0,9531	0,0917	0,0273
60	0,9566	0,0856	0,0240

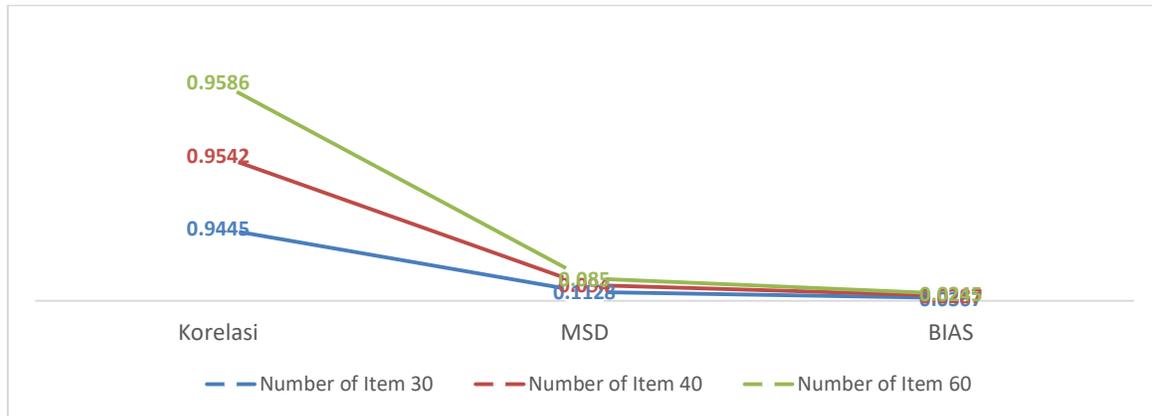


Gambar 1. Grafik Perbandingan Koefisien Korelasi, MSD, dan Bias pada N = 500

Pada tabel 1 terlihat adanya perbedaan koefisien korelasi (ρ) apabila *number of item* diubah. $\rho_{\theta_T:\theta_F}$ terendah diperoleh pada $n=30$. Penambahan 10 butir tes menghasilkan peningkatan korelasi sebesar 0,0053 poin, penambahan 30 butir tes menghasilkan peningkatan korelasi 0,0130 atau meningkat 0,0035 dari penambahan 10 butir tes. MSD juga berubah apabila *number of item* (n) berubah. MSD terendah diperoleh pada $n=30$. Berbeda dengan koefisien korelasi, penambahan n mengakibatkan MSD berkurang. Penambahan 10 butir tes dari $n=30$ mengakibatkan MSD berkurang 0,0179 poin, kemudian MSD berkurang lagi sebesar 0,0061 poin setelah penambahan 20 butir tes. Perubahan *number of item* juga mempengaruhi bias estimasi. Bias pada $n=30$ sebesar 0,0365. Penambahan 10 butir tes menghasilkan pengurangan bias sebesar 0,0022 poin. Bias berkurang 0,0125 poin setelah penambahan tes sebanyak 30 butir. Kecenderungan perubahan-perubahan tersebut dapat dilihat pada Gambar 1. Hal ini mengindikasikan bahwa penambahan *number of item* pada sample size $N=500$ menghasilkan peningkatan koefisien korelasi *true ability* dengan *estimate ability* sementara MSD dan bias mengalami penurunan. Temuan ini menunjukkan bahwa semakin banyak butir tes, maka akurasi estimasi *ability* semakin tinggi.

Tabel 2. Koefisien Korelasi, MSD, dan Bias pada $N=750$

<i>Number of item</i>	Koefisien Korelasi	MSD	BIAS
30	0,9445	0,1128	0,0367
40	0,9542	0,0940	0,0287
60	0,9586	0,0850	0,0245



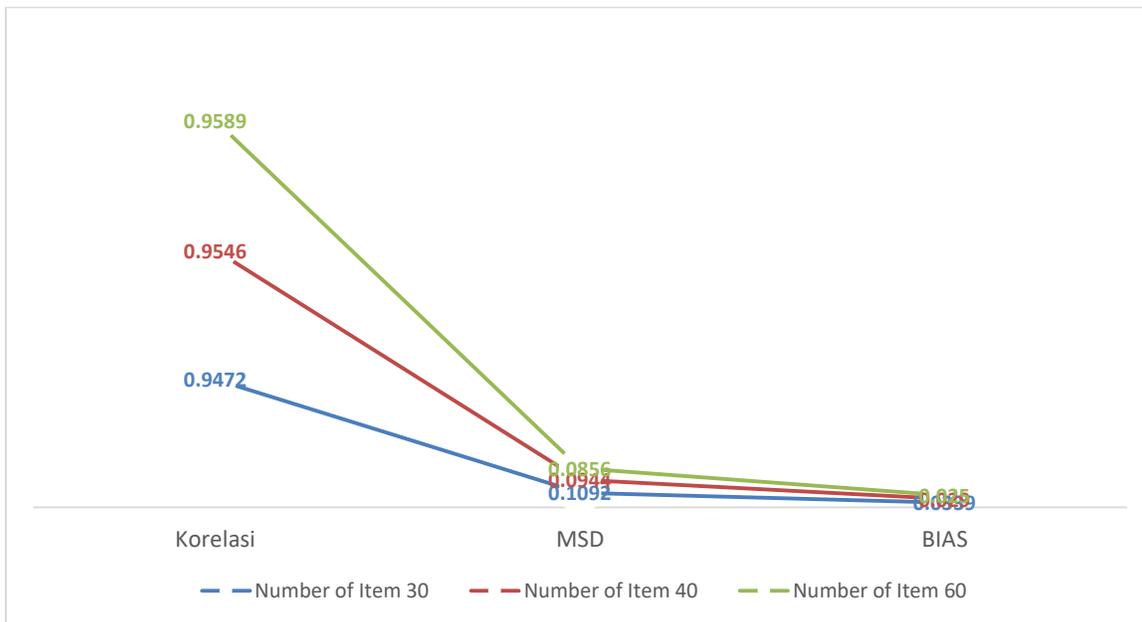
Gambar 2. Grafik Perbandingan Koefisien Korelasi, MSD, dan Bias pada N = 750

Tabel 2 menampilkan koefisien korelasi, MSD, dan bias estimasi *ability* (θ) pada sampel size N=750. Pada tabel tersebut terlihat perbedaan koefisien korelasi (ρ) apabila *number of item* diubah. $\rho_{\theta_T; \theta_F}$ terendah diperoleh pada n=30. Penambahan 10 butir tes menghasilkan peningkatan korelasi sebesar 0,0097 poin, penambahan 30 butir tes menghasilkan peningkatan korelasi 0,0141 atau meningkat 0,0044 dari penambahan 10 butir tes. MSD juga berubah apabila *number of item* (n) berubah. MSD terendah diperoleh pada n=30. Berbeda dengan koefisien korelasi, penambahan n mengakibatkan MSD berkurang. Penambahan 10 butir tes dari n=30 mengakibatkan MSD berkurang 0,0188 poin, kemudian MSD berkurang lagi sebesar 0,0061 poin setelah penambahan 20 butir tes.

Pada tabel 2 terlihat bahwa perubahan *number of item* juga mempengaruhi bias estimasi. Bias pada n=30 sebesar 0,0365. Penambahan 10 butir tes menghasilkan pengurangan bias sebesar 0,0009 poin. Bias berkurang 0,0278 poin setelah penambahan tes sebanyak 30 butir. Kecenderungan perubahan-perubahan tersebut dapat dilihat pada Gambar 2. Hal ini mengindikasikan bahwa penambahan *number of item* pada sample size N=750 menghasilkan peningkatan koefisien korelasi *true ability* dengan *estimate ability* sementara MSD dan bias mengalami penurunan. Temuan ini menunjukkan bahwa semakin banyak butir tes, maka akurasi estimasi *ability* semakin tinggi.

Tabel 3. Koefisien Korelasi, MSD, dan Bias pada *Sample Size* (N) = 1000

<i>Number of Item</i>	Koefisien Korelasi	MSD	BIAS
30	0,9472	0,1092	0,0339
40	0,9546	0,0944	0,0290
60	0,9589	0,0856	0,0250



Gambar 3. Grafik Perbandingan Koefisien Korelasi, MSD, dan Bias pada N = 1000

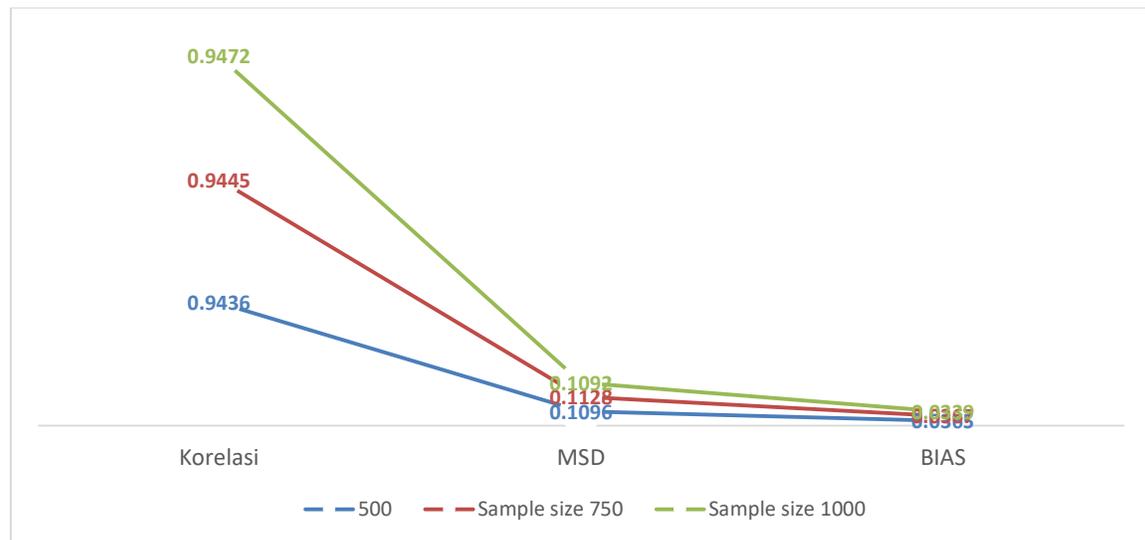
Koefisien korelasi, MSD, dan bias estimasi *ability* (θ) pada sampel size N=1000 disajikan dalam tabel 3. Pada Tabel 3 terlihat perbedaan koefisien korelasi (ρ) apabila *number of item* diubah. $\rho_{\theta_T:\theta_R}$ terendah diperoleh pada n=30. Penambahan 10 butir tes menghasilkan peningkatan korelasi sebesar 0,0074 poin, penambahan 30 butir tes menghasilkan peningkatan korelasi 0,0117 atau meningkat 0,0043 dari penambahan 10 butir tes. MSD juga berubah apabila *number of item* (n) berubah. MSD tertinggi diperoleh pada n=30. Berbeda dengan koefisien korelasi, penambahan n mengakibatkan MSD berkurang. Penambahan 10 butir tes dari n=30 mengakibatkan MSD berkurang 0,0148 poin, kemudian MSD berkurang lagi sebesar 0,0088 poin setelah penambahan 20 butir tes.

Sama halnya dengan koefisien korelasi, perubahan *number of item* juga mempengaruhi bias estimasi pada N=1000. Bias pada n=30 sebesar 0,0339. Penambahan 10 butir tes menghasilkan pengurangan bias sebesar 0,0049 poin. Bias berkurang 0,0089 poin setelah penambahan tes sebanyak 30 butir. Kecenderungan perubahan-perubahan tersebut dapat dilihat pada Gambar 3. Hal ini mengindikasikan bahwa penambahan *number of item* pada sample size N=1000 menghasilkan peningkatan koefisien korelasi *true ability* dengan *estimate ability* sementara MSD dan bias mengalami penurunan. Temuan ini menunjukkan bahwa semakin banyak butir tes, maka akurasi estimasi *ability* semakin tinggi.

Number of Item Tetap Sample Size Berubah

Tabel 4. Koefisien Korelasi, MSD, dan Bias pada *number of item* (n) = 30

Sample Size	Koefisien Korelasi	MSD	BIAS
500	0,9436	0,1096	0,0365
750	0,9445	0,1128	0,0367
1000	0,9472	0,1092	0,0339



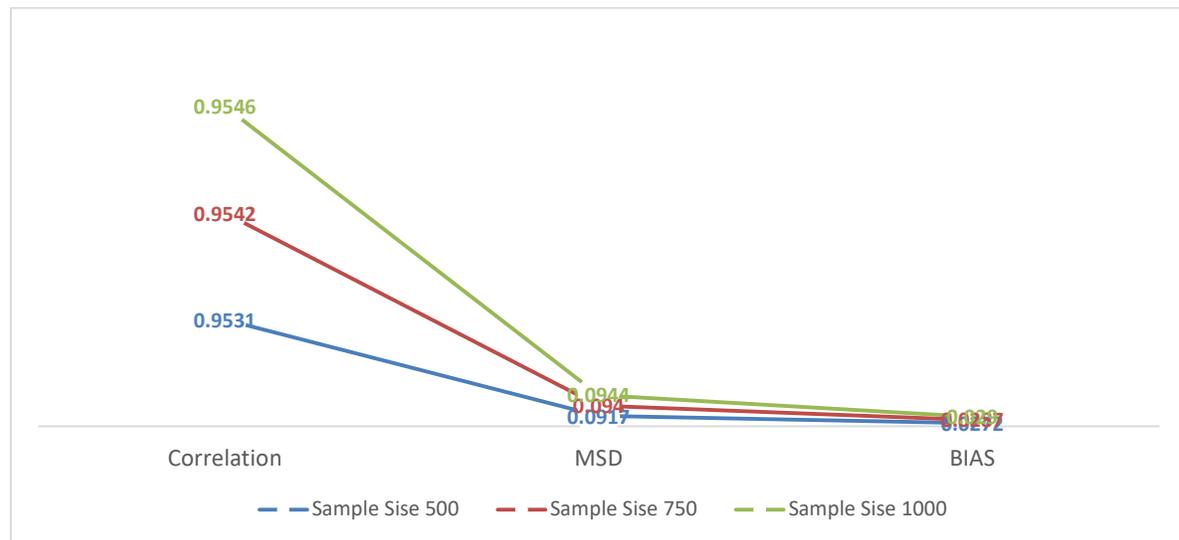
Gambar 4. Grafik Perbandingan Koefisien Korelasi, MSD, dan Bias Ability pada n = 30

Tabel 4 menampilkan korelasi, MSD, dan bias estimasi *ability* (θ) pada *number of item* (n)=30 . Pada Tabel 4 terlihat perbedaan koefisien korelasi (ρ) apabila *sample size* diubah. Korelasi terendah diperoleh pada N=500. Penambahan 250 sample size menghasilkan peningkatan korelasi sebesar 0,0009 poin, penambahan 750 sample size menghasilkan peningkatan korelasi 0,0036 atau meningkat 0,0027 dari penambahan 250 sample size. MSD juga berubah apabila *sample size* berubah. MSD terendah diperoleh pada N=500. Penambahan 250 sample size dari N=500 mengakibatkan MSD meningkat 0,0032 poin, kemudian MSD berkurang lagi sebesar 0,0036 poin setelah penambahan 250 sample size. Hal ini berarti penambahan sample size menghasilkan MSD yang tidak konsisten. Perubahan *number of item* juga mempengaruhi bias estimasi. Bias pada N=500 sebesar 0,0365. Penambahan 250 *sample size* menghasilkan peningkatan bias sebesar 0,0002 poin. Bias berkurang 0,0026 poin setelah penambahan sample size sebanyak 750. Kecenderungan perubahan-perubahan tersebut dapat

dilihat pada Gambar 4. Hal ini mengindikasikan bahwa penambahan *sample size* pada *number of item* (n)=30 menghasilkan peningkatan koefisien korelasi *true ability* dengan *estimate ability* sementara MSD dan bias mengalami perubahan yang tidak konsisten. Temuan ini menunjukkan bahwa akurasi estimasi *ability* diperoleh pada *sample size* sebanyak 1000 jika *number of item* yang ditempuh sebanyak 30 butir.

Tabel 5. Koefisien Korelasi, MSD, dan Bias pada *number of item* (n) = 40

<i>Sample Size</i>	Koefisien Korelasi	MSD	Bias
500	0,9531	0,0917	0,0272
750	0,9542	0,0940	0,0287
1000	0,9546	0,0944	0,0290



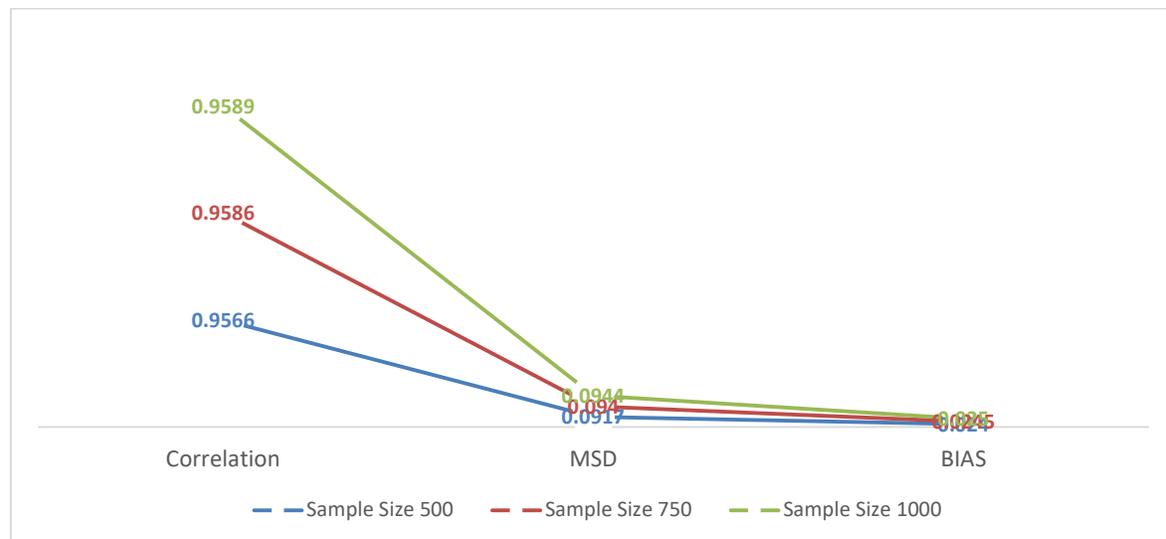
Gambar 5. Grafik Perbandingan Koefisien Korelasi, MSD, dan Bias pada $n = 40$

Tabel 4 menampilkan korelasi, MSD, dan bias estimasi *ability* (θ) pada *number of item* (n)=40. Pada Tabel 4 terlihat perbedaan koefisien korelasi (ρ) apabila *sample size* diubah. Korelasi terendah diperoleh pada $N=500$. Penambahan 250 *sample size* menghasilkan peningkatan korelasi sebesar 0,0011 poin, penambahan 750 *sample size* menghasilkan peningkatan korelasi 0,0015 atau meningkat 0,0004 dari penambahan 250 *sample size*. MSD juga berubah apabila *sample size* berubah. MSD terendah diperoleh pada $N=500$. Penambahan 250 *sample size* dari $N=500$ mengakibatkan MSD meningkat 0,0023 poin, kemudian MSD meningkat lagi sebesar 0,0004 poin setelah penambahan 250 *sample size*. Hal ini berarti

penambahan *sample size* menghasilkan peningkatan MSD. Perubahan *sample size* juga mempengaruhi bias estimasi. Bias pada N=500 sebesar 0,0272. Penambahan 250 *sample size* menghasilkan peningkatan bias sebesar 0,0015 poin. Bias meningkat lagi 0,0003 poin setelah penambahan *sample size* sebanyak 250. Kecenderungan perubahan-perubahan tersebut dapat dilihat pada Gambar 5. Hal ini mengindikasikan bahwa penambahan *sample size* pada *number of item* (n)=40 menghasilkan peningkatan koefisien korelasi *true ability*, MSD, bias. Temuan ini menunjukkan bahwa jika *number of item* yang ditempuh sebanyak 40 butir dengan penambahan *sample size* yang lebih besar mengakibatkan estimasi *ability* semakin tidak akurat karena menghasilkan MSD dan bias yang meningkat meskipun koefisien korelasi ditemukan meningkat.

Tabel 4. Koefisien Korelasi, MSD, dan Bias pada *number of item* (n) = 60

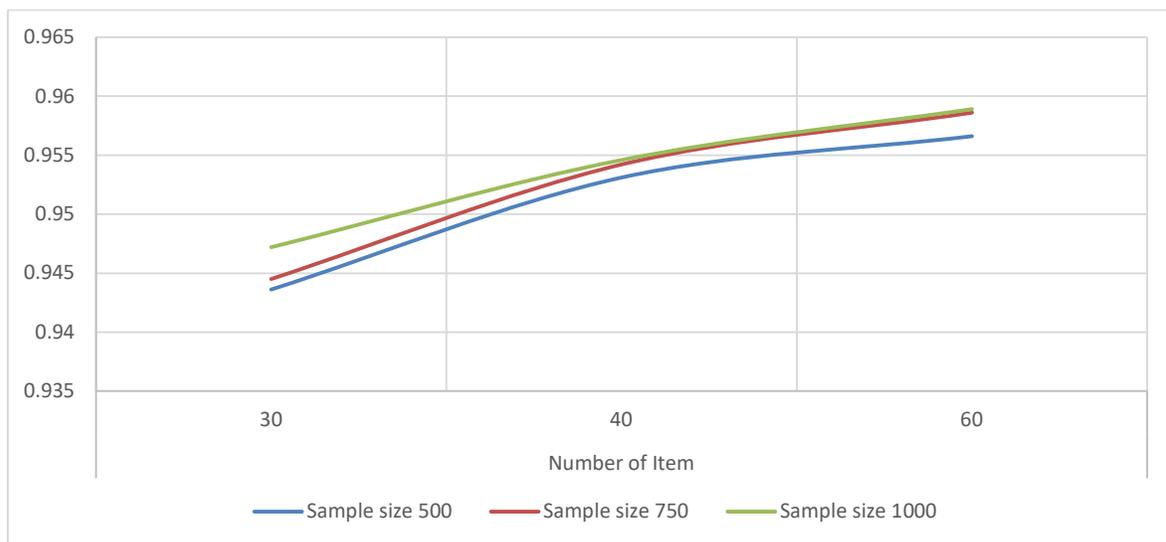
<i>Sample Size</i>	Koefisien Korelasi	MSD	Bias
500	0,9566	0,0917	0,0240
750	0,9586	0,0940	0,0245
1000	0,9589	0,0944	0,0250



Gambar 6. Grafik Perbandingan Koefisien Korelasi, MSD, dan Bias pada n = 60

Tabel 4 menampilkan korelasi, MSD, dan bias estimasi *ability* (θ) pada *number of item* (n)=60 . Pada Tabel 4 terlihat perbedaan koefisien korelasi (ρ) apabila *sample size* diubah. Korelasi terendah diperoleh pada N=500. Penambahan 250 *sample size* menghasilkan peningkatan korelasi sebesar 0,0020 poin, penambahan 750 *sample size* menghasilkan

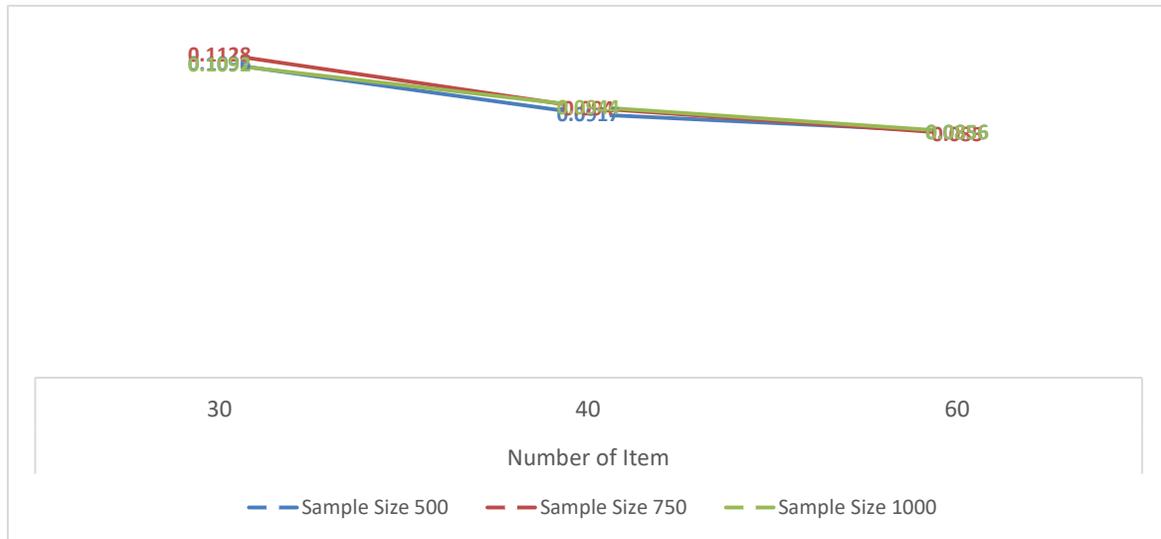
peningkatan korelasi 0,0023 atau meningkat 0,0003 dari penambahan 250 *sample size*. MSD juga berubah apabila *sample size* berubah. MSD terendah diperoleh pada N=500. Penambahan 250 *sample size* dari N=500 mengakibatkan MSD meningkat 0,0023 poin, kemudian MSD meningkat lagi sebesar 0,0004 poin setelah penambahan 250 *sample size*. Hal ini berarti penambahan *sample size* menghasilkan peningkatan MSD. Perubahan *sample size* juga mempengaruhi bias estimasi. Bias pada N=500 sebesar 0,0245. Penambahan 250 *sample size* menghasilkan peningkatan bias sebesar 0,0005 poin. Bias meningkat lagi 0,0005 poin setelah penambahan *sample size* sebanyak 250. Kecenderungan perubahan-perubahan tersebut dapat dilihat pada Gambar 6. Hal ini mengindikasikan bahwa penambahan *sample size* pada *number of item* (n)=60 menghasilkan peningkatan koefisien korelasi *true ability*, MSD, bias. Temuan ini menunjukkan bahwa jika *number of item* yang ditempuh sebanyak 60 butir dengan penambahan *sample size* yang lebih besar mengakibatkan estimasi *ability* semakin tidak akurat karena menghasilkan MSD dan bias yang meningkat meskipun koefisien korelasi ditemukan meningkat.



Gambar 7. Grafik Perbandingan Koefisien Korelasi

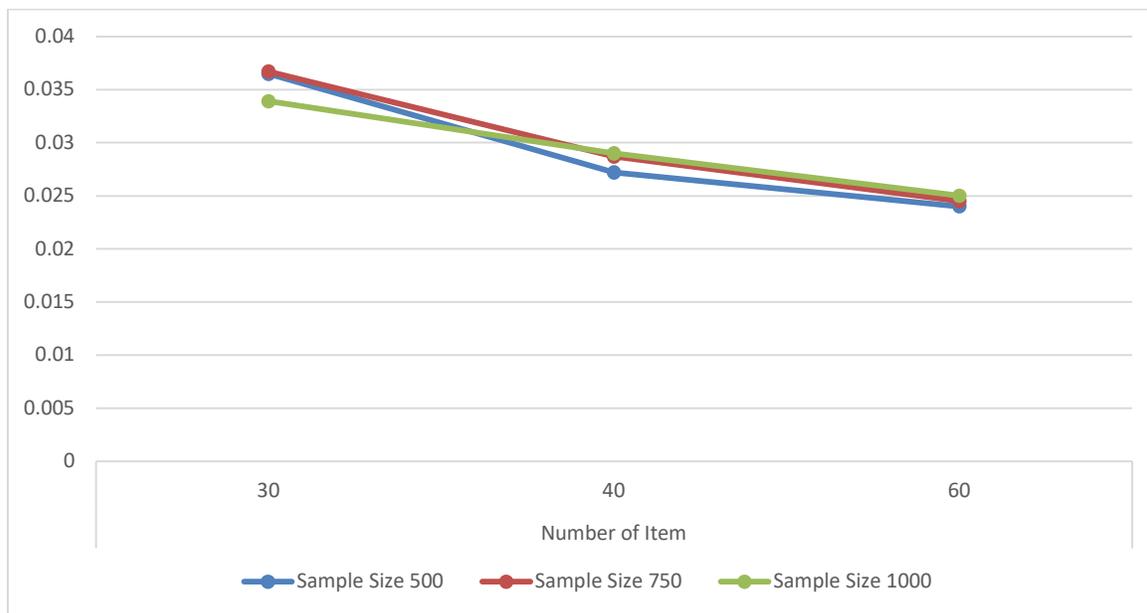
Berdasarkan data-data yang telah disajikan dalam Tabel 2 - Tabel 5, dapat dilihat bahwa koefisien korelasi setiap kasus berada pada rentang $0,9436 \leq \rho \leq 0,9589$. Koefisien terendah diperoleh pada tes yang memiliki *number of item* (n) = 30 dengan *sample size* (N) = 500. Koefisien tertinggi diperoleh pada tes yang memiliki *number of item* (n) = 60 dengan *sample size* (N) = 1000. Dalam Gambar 7 tampak bahwa koefisien korelasi *ability* (θ) pada *sample*

size (N) = 750 dan N=1000 dan tes yang memiliki *number of item* (n) = 40 dan (n) = 60 grafiknya hampir berimpit. Grafik ini menunjukkan bahwa estimasi parameter menggunakan kedua *sample size* dan kedua *number of item* memiliki akurasi yang relatif sama. Namun demikian, estimasi *ability* menggunakan *sampel size* 1000 lebih akurat dibandingkan dengan *sampel size* 750. Koefisien korelasi ini menggambarkan bahwa peningkatan *number of item* dan *sample size* menghasilkan peningkatan keselarasan akurasi *ability estimation*.



Gambar 8. Grafik Perbandingan MSD

Berdasarkan data MSD yang disajikan pada Tabel 2 – tabel 4, dapat dilihat bahwa MSD terendah diperoleh pada *sample size* 750 dengan *number of item* (n) = 60, dan MSD tertinggi juga diperoleh pada *sampel size* (N) = 750 dengan *number of item* (n) = 30. Namun sesilih MSD pada N=750 dengan *sample size* lainnya untuk *number of item* (n) = 60 sangat kecil yaitu hanya 0,0006. Gambar 8 menggambarkan tentang hal ini. Grafik MSD untuk N=750 dan N=1000 sangat kecil jaraknya pada n = 60. Hal ini menunjukkan bahwa pada *number of item* (n) = 60, ketiga *sample size* menghasilkan akurasi yang relatif sama dalam mengestimasi *ability*. Sementara MSD pada *number of item* (n) = 40 berada di antara n = 30 dan n = 60. Hanya saja, MSD pada *number of item* (n) = 40 MSD memiliki trend meningkat. Dengan demikian dapat dikatakan bahwa penambahan *sample size* pada n=40 menghasilkan berkurangnya tingkat akurasi estimasi *ability*.



Gambar 9. Grafik Perbandingan Bias Estimasi

Bias estimasi setiap kalibrasi diharapkan serendah mungkin. Bias rendah menggambarkan bahwa estimasi menghasilkan akurasi sangat tinggi. Berdasarkan analisis data pada Tabel 2-tabel 5 diperoleh bahwa bias estimasi *ability* sangat rendah pada setiap kasus. Bias estimasi berada pada rentang 0,0240 sampai 0,0367. Bias tertinggi diperoleh pada hasil estimasi menggunakan sample size (N) = 750 dengan *number of item* (n) = 30 butir. Bias estimasi terendah diperoleh pada hasil estimasi *ability* menggunakan *sample size* (N) = 500 dengan *number of item* (n) = 60. Bias estimasi *ability* yang diperoleh pada setiap peningkatan sample size untuk *number of item* (n) = 40 dan 60, ditemukan meningkat secara konsisten. Sementara bias estimasi pada *nuber ofi item* (n) = 30 ditemukan tidak konsisten pada setiap peningkatan *sample size*. Hal ini jelas terlihat pada Gambar 9 di atas.

Simpulan

Berdasarkan hasil analisis data dapat disimpulkan bahwa (1) pengaruh *sample size* terhadap akurasi *ability estimation* ditemukan tidak konsisten karena MSD dan bias estimasi tidak konsisten, namun selisih inskonsistensi, baik MSD dan bias, relatif sangat kecil, dan (2) semakin besar *number of item* maka *ability estimation* semakin akurat. Perlu dilakukan penelitian lanjutan dengan melibatkan banyak variasi *sample size* dan *number of item* termasuk *sample size* dan *number of item* kecil.

DAFTAR PUSTAKA

- Kaplan, R.M., & Saccuzo, D.P. (1993). *Psychological Testing: Principles, Applications, and Issues*. Third edition. California: Brooks/Cole Publishing.
- Mislevy, R.J. & Bock, R.D. (1990). *BILOG 3: Item analysis & test scoring with binary logistic models*. Mooreseville: Scientific Software, Inc.
- Febrayosi, P. (2012). *Estimasi True Score pada Second Order Unidimensional Data: sebuah studi simulasi monte carlo tentang dampak panjang tes, tingkat kesukaran dan daya pembeda item*. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, Vol I, No 4, Oktober 2012
- Naga, D.S. (1992). *Pengantar teori sekor pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Hambleton, R. K. & Cook, L.L. (1983). *Robustness of item response models and effects of test length and sample size on the precision of ability estimates*. Dalam Weiss, David J. (ed.) *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. New York: Academic Press, Inc., p.31-49.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- Hikamudin, E. (2012). *Estimasi Kemampuan Siswa dalam Ujian Nasional Menggunakan Metode Bayes*. *Jurnal Penelitian Kebijakan Pendidikan*. Vol.10, Nomor 2.
- Hullin, C.L. et al. (1983). *Item Response Theoty: Aplication to Psichologycal Measurement*. Homewood, IL: Dow Jones-Irwin.
- 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.