

Stabilitas Estimasi Parameter Pada Regresi Logistik (Suatu Penerapan Pada Pengukuran)

Heri Retnawati

Pend. Matematika UNY (Riwati@yahoo.com)

Abstrak

Pada paper ini dibahas tentang efek panjangnya tes, distribusi kemampuan peserta tes, dan banyaknya peserta tes terhadap kestabilan parameter (tingkat kesulitan (a), daya pembeda (b), tebakan semu (c) dan kemampuan peserta tes (θ)) pada teori respons butir unidimensi model regresi logistik tiga parameter. Data dibangkitkan dengan program DGEN, dengan variable panjang tes (20 butir dan 50 butir), distribusi kemampuan ($N(-1,1)$, $N(0,1)$ dan $N(+1,1)$), dan banyaknya peserta tes (300, 750, dan 1200). Pola kecenderungan diamati berdasarkan *Mean Square of Error* (MSE) dari parameter, koefisien korelasi antara parameter sebenarnya dengan parameter hasil estimasi, dan nilai fungsi informasi estimasi. Hasil penelitian menunjukkan bahwa ada kecenderungan MSE paling rendah terjadi pada data yang dibangkitkan dengan parameter kemampuan berdistribusi normal baku dan korelasi antara parameter estimasi dan parameter sebenarnya tidak memiliki pola yang pasti, jika dilihat dari distribusinya. Pada parameter a dan c , ada kecenderungan semakin besar ukuran sampel, semakin besar keakuratan pengestimasiannya, namun untuk b dan θ tidak ada pola yang pasti. Melihat korelasinya, ada kecenderungan semakin semakin besar ukuran sample peserta tes, semakin dekat korelasi antara parameter hasil estimasi dengan parameter sebenarnya, namun hal ini tidak berlaku untuk parameter b . Berdasarkan rerata MSE dan rerata korelasi, ada kecenderungan semakin panjang suatu tes, akan semakin besar keakuratannya untuk mengestimasi parameter a , c dan θ . Namun sebaliknya, pada pengestimasi parameter tingkat kesulitan (b), semakin panjang tes akan semakin kurang akurat, karena semakin besar MSE-nya. Demikian pula berdasarkan korelasinya. Kesalahan pengukuran estimasi (SEE), yang tidak dipengaruhi oleh distribusi dan ukuran sampel, tetapi pada studi ini hanya dipengaruhi oleh panjang tes. Berdasarkan hasil analisis signifikansi dengan analisis varians, distribusi kemampuan, panjang tes dan interaksi panjang tes dengan ukuran sample yang berpengaruh pada stabilitas estimasi parameter b saja.

Latar Belakang

Permasalahan panjang tes, distribusi kemampuan peserta tes dan ukuran sample peserta tes sering dibahas oleh pengguna di dunia pengukuran, terlebih dalam pendidikan. Selain permasalahan di lapangan, beberapa peneliti juga mengakses ketiga variable ini dalam penelitiannya.

Hambleton & Cook (tanpa tahun) meneliti tentang ketegaran model respons butir dan efek panjang tes dan ukuran sample terhadap presisi estimasi kemampuan. Hasil penelitian ini menunjukkan bahwa panjang tes dan ukuran

sample keduanya merupakan factor penting yang mempengaruhi ketepatan kurva estimasi kesalahan pengukuran (*standard error estimate, SEE*).

Stone & Bo Zang (2003) meneliti tentang mengakses kecocokan model teori respons butir dengan membandingkan prosedur tradisional dan alternative dan Van Abswoude, dkk. (2004) yang melakukan studi comparative procedure penilaian dimensionalitas data tes di bawah model teori respons butir non parametrik, yang juga melibatkan panjang tes sebagai variable. Pada penelitian pendeteksian DIF dengan berbagai metode, Budiono (2005) memasukkan distribusi peserta tes sebagai salah satu variable yang perlu diteliti.

Tujuan Penulisan

Paper ini bertujuan untuk mengetahui :

1. efek panjang tes terhadap stabilitas estimasi parameter butir dan parameter kemampuan pada model logistic 3 parameter,
2. efek distribusi kemampuan terhadap stabilitas estimasi parameter butir dan parameter kemampuan pada model logistic 3 parameter,
3. ukuran sample/banyaknya peserta tes terhadap stabilitas estimasi parameter butir dan parameter kemampuan pada model logistic 3 parameter.

Metode

Model Pembangkitan Data

Data yang dibangkitkan diasumsikan uni dimensi model logistic 3 parameter yang memenuhi persamaan :

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

Keterangan :

$P_i(\theta)$: probabilitas peserta tes yang memiliki kemampuan θ dipilih
secara acak dapat menjawab butir I dengan benar

θ : tingkat kemampuan subjek

a_i : indeks daya beda dari butir ke-i

b_i : indeks kesukaran butir ke-i

c_i : indeks tebakan semu butir ke-i

e : bilangan natural yang nilainya mendekati 2,718

n : banyaknya item dalam tes

D : faktor penskalaan yang dibuat agar fungsi logistik mendekati
fungsi ogive normal yang harganya 1,7.

Fungsi informasi butir untuk model logistic tiga parameter ini dinyatakan oleh Birnbaum (Hambleton dan Swaminathan , 1985) dalam persamaan berikut.

$$I_i(\theta) = \frac{2,89a_i^2(1 - c_i)}{\left[(c_i + \exp(Da_i(\theta - b_i))) \right] \left[1 + \exp(-Da_i(\theta - b_i)) \right]^2} \dots\dots\dots (2)$$

keterangan :

$I_i(\theta)$: fungsi informasi butir i

θ : tingkat kemampuan subyek

a_i : parameter daya beda dari butir ke-i

- b_i : parameter indeks kesukaran butir ke- i
- c_i : parameter indeks peluang kebenaran jawaban tebakan semu (*pseudoguessing*) butir ke- i
- e : bilangan natural yang nilainya mendekati 2,718

Fungsi informasi tes atau sekumpulan butir tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut (Hambleton dan Swaminathan, 1985). Berhubungan dengan hal ini, fungsi informasi perangkat tes akan tinggi jika butir tes mempunyai fungsi informasi yang tinggi pula. Fungsi informasi perangkat tes secara matematis dapat dituliskan sebagai berikut.

$$I_i(\theta) = \sum_{i=1}^n I_i(\theta) \dots\dots\dots (3)$$

Fungsi informasi dengan *SEM* mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka *SEM* semakin kecil atau sebaliknya (Hambleton, Swaminathan dan Rogers, 1991). Jika nilai fungsi informasi dinyatakan dengan $I_i(\theta)$ dan nilai estimasi *SEM* dinyatakan dengan $SEM(\hat{\theta})$, maka hubungan keduanya, menurut Hambleton, Swaminathan, dan Rogers (1991) dinyatakan dengan

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots (4)$$

Data

Data digenerasikan dengan menggunakan program DGEN, yang masing-masing kasus terdiri dari 5 replikasi. Variabel dalam replikasi meliputi panjang tes (20 dan 50 butir), distribusi kemampuan peserta tes ($N(-1,1)$, $N(0,1)$, dan

$N(1,0)$) dan ukuran/banyaknya peserta tes (300, 600, dan 1200 orang). Parameter sebenarnya butir tes yang dibangkitkan adalah :

1. parameter b berdistribusi uniform antara $-2,0$ sampai dengan $2,0$.
2. parameter a berdistribusi uniform antara $0,0$ sampai dengan $2,0$.
3. parameter c berdistribusi uniform antara $0,0$ sampai dengan $0,25$.

Data yang dibangkitkan sebanyak $5 \times 2 \times 3 \times 3 = 90$ set data.

Parameter b dibangkitkan dengan rentang $-2,0$ sampai dengan $+2,0$; parameter a dibangkitkan dengan rentang 0 sampai dengan $+2,0$ dan parameter c dibangkitkan pada rentang 0 sampai dengan $0,2$.

Estimasi Parameter

Parameter butir dan kemampuan peserta tes dari data yang dibangkitkan diestimasi dengan menggunakan program BILOG 3 (Mislevy & Bock, 1990), dengan menggunakan metode estimasi MML (*marginal maximum likelihood*). Ringkasan dari parameter butir hasil estimasi dalam table 1.

Mengevaluasi Pengestimasi Parameter

Untuk mengevaluasi pengestimasi parameter, digunakan kesalahan kuadrat rata-rata (mean square of error, MSE), seperti yang dilakukan oleh Cohen, Kane dan Kim (2001)). MSE dihitung pada setiap replikasi, r , dan untuk tiap parameter, baik b , a , c , dan θ . Misalkan e parameter estimasi dan t parameter sebenarnya (true), MSE dihitung dengan rumus :

$$MSE(e_r) = \frac{\sum_{i=1}^n (e_i - t_i)^2}{n} \dots\dots\dots (5)$$

Untuk tiap replikasi dan kemudian dihitung rerata dan variansnya untuk tiap kasus. Misalkan $Y_r = \text{MSE}(e_r)$, maka rerata untuk 5 replikasi adalah :

$$\bar{Y} = \frac{1}{5} \sum_{r=1}^5 Y_r \dots\dots\dots(6)$$

Dengan varians :

$$s_{Y_r}^2 = \frac{1}{5} \sum_{r=1}^5 (Y_r - \bar{Y})^2 \dots\dots\dots(7)$$

Selain MSE rerata korelasi antara parameter hasil estimasi dan parameter sebenarnya digunakan pula untuk evaluasi ini.

Untuk mengetahui pengaruh panjang tes, distribusi kemampuan peserta tes dan ukuran sample peserta tes digunakan analisis varians tiga jalur (Keppel, 1982). Mengetahui efek signifikansi dengan cara ini juga telah dilakukan oleh Bastari (1998).

Tabel 1. Ringkasan Parameter Butir Hasil Estimasi

Ukuran	Distribusi θ	Para- meter	20 butir		50 butir	
			Rerata	Stdev	Rerata	Stdev
300	N(-1,1)	b	1.46008	0.180342	1.604164	0.022648
		a	1.06784	0.031318	1.126828	0.027542
		c	0.17676	0.017958	0.1844	0.004286
		0	3.4E-05	1.000303	2.67E-07	1.000334
	N(0,1)	b	0.56904	0.071223	0.725664	0.042348
		a	1.10377	0.022961	1.151048	0.026566
		c	0.21695	0.005909	0.206428	0.006511
		0	-1.9E-06	1.000334	-1.6E-06	1.000332
	N(1,1)	b	-0.55724	0.127601	-0.20475	0.031965

		a	1.05014	0.090788	1.08934	0.029365
		c	0.23526	0.006867	0.225004	0.008278
		0	6.67E-07	1.000332	6.67E-07	1.000334
750	N(-1,1)	b	1.5167	0.060079	-0.20475	0.031965
		a	1.03271	0.051631	1.08934	0.029365
		c	0.17213	0.003862	0.225004	0.008278
		0	1.07E-07	1.000133	-7.7E-07	1.000133
	N(0,1)	b	0.48545	0.085212	0.826608	0.128959
		a	1.03315	0.029771	0.998312	0.097257
		c	0.20376	0.007548	0.19756	0.003803
		0	-4.3E-07	1.000134	5.33E-07	1.000134
	N(1,1)	b	-0.56568	0.048852	-0.16627	0.052839
		a	1.00289	0.045647	1.020292	0.038165
		c	0.22553	0.009199	0.216452	0.004335
		0	-9.6E-07	1.000133	1.12E-06	1.000133
1200	N(-1,1)	b	1.54238	0.084771	1.838779	0.21195
		a	0.99337	0.043725	0.98717	0.080073
		c	0.17001	0.007628	0.176772	0.012564
		0	-1.7E-08	1.000084	-4.8E-07	1.000083
	N(0,1)	b	0.50333	0.084963	0.777712	0.029981
		a	1.01884	0.036727	1.040512	0.028988
		c	0.19286	0.004835	0.186796	0.004641
		0	-2.4E-18	1.000084	1.17E-07	1.000083
	N(1,1)	b	-0.49803	0.027601	-0.12048	0.048765
		a	1.02883	0.054982	1.006168	0.018073
		c	0.22071	0.003173	0.212516	0.004185

		0	-2.3E-07	1.000084	2.48E-18	1.000083
--	--	---	----------	----------	----------	----------

Hasil

Hasil perhitungan rerata MSE dan rerata korelasi untuk tiap kasus disajikan pada tabel 2 dan tabel 3.

Tabel 2. Rerata MSE hasil estimasi

Ukuran	Distribusi θ	Parameter	20 butir		50 butir	
			Rerata	Stdev	Rerata	Stdev
300	N(-1,1)	b	2.438971	0.705635	2.281603	0.117386
		a	0.350489	0.479044	0.362733	0.415052
		c	0.01366	0.004422	0.016555	0.001248
		θ	1.322742	1.037555	1.237852	0.810819
	N(0,1)	b	0.793995	0.32515	0.919151	0.133166
		a	0.337473	0.486554	0.31429	0.437827
		c	0.023746	0.003041	0.02158	0.001736
		θ	0.184273	0.269212	0.091066	0.141649
	N(1,1)	b	1.024007	0.283923	1.091665	0.160674
		a	0.3268	0.493709	0.286104	0.453628
		c	0.027387	0.00328	0.025465	0.002353
		θ	1.075453	0.83704	0.968911	0.537215
750	N(-1,1)	b	1.024007	0.283923	2.615459	0.049558
		a	0.3268	0.493709	0.296926	0.447481
		c	0.027387	0.00328	0.014873	0.000131
		θ	1.273622	1.016642	1.193668	0.769786
	N(0,1)	b	0.753921	0.336192	1.618606	0.409706

		a	0.298066	0.507711	0.483782	0.610965
		c	0.019966	0.00292	0.021032	0.001641
		θ	0.179405	0.258197	0.306459	0.575433
	N(1,1)	b	0.983997	0.132099	1.402552	0.234269
		a	0.283126	0.515813	0.24812	0.474855
		c	0.02486	0.002644	0.025232	0.002056
		θ	1.131931	0.893325	1.014846	0.566962
	N(-1,1)	b	2.688341	0.505645	3.344067	0.644053
		a	0.308006	0.50217	0.211058	0.306134
		c	0.012883	0.002127	0.016442	0.002451
		θ	1.290783	1.032301	1.211663	0.816245
1200	N(0,1)	b	0.847166	0.384258	1.467922	0.20987
		a	0.27957	0.517851	0.245343	0.47643
		c	0.017919	0.002415	0.0199	0.001272
		θ	1.11557	0.879779	0.085718	0.135564
	N(1,1)	b	1.284437	0.30954	1.751527	0.250979
		a	0.266001	0.525545	0.239198	0.479744
		c	0.024089	0.001405	0.024175	0.00136
		θ	1.115107	0.885467	1.000226	0.557793

Tabel 2. Rerata Korelasi antara parameter hasil estimasi dengan parameter sebenarnya

Ukuran	Distribusi θ	Parameter	20 butir		50 butir	
			Rerata	Stdev	Rerata	Stdev
300	N(-1,1)	b	0.886659	0.057487	0.789262	0.049123
		a	0.785308	0.054097	0.951716	0.031079

		c	0.58733	0.068976	0.380843	0.061311
		θ	0.898048	0.004713	0.939552	0.001991
	N(0,1)	b	0.858324	0.055627	0.769279	0.03377
		a	0.825268	0.056354	0.876177	0.024907
		c	0.41122	0.077883	0.433281	0.04489
		θ	0.185429	0.020035	0.957176	0.003837
	N(1,1)	b	0.866672	0.05551	0.743276	0.031373
		a	0.856157	0.064594	0.896815	0.015931
		c	0.281671	0.05709	0.327435	0.040847
		θ	0.910617	0.003657	0.962748	0.003877
750	N(-1,1)	b	0.889524	0.060652	0.817937	0.006344
		a	0.876002	0.028185	0.841222	0.007903
		c	0.62792	0.035509	0.611073	0.002238
		θ	0.896274	0.002903	0.936566	0.000728
	N(0,1)	b	0.864683	0.059157	0.598966	0.141436
		a	0.901807	0.044218	0.775476	0.38242
		c	0.493713	0.058832	0.48881	0.021612
		θ	0.910178	0.001791	0.953656	0.00999
	N(1,1)	b	0.858432	0.039529	0.663398	0.048922
		a	0.931106	0.019201	0.952411	0.01364
		c	0.382665	0.0512	0.348717	0.053391
		θ	0.900527	0.006784	0.959531	0.001744
1200	N(-1,1)	b	0.889524	0.060652	0.741572	0.035564
		a	0.876002	0.028185	0.88689	0.021763
		c	0.62792	0.035509	0.60322	0.068411
		θ	0.894259	0.230741	0.93423	0.003313

	N(0,1)	b	0.847971	0.065936	0.658322	0.045666
		a	0.934383	0.024894	0.956693	0.016069
		c	0.515057	0.029416	0.500238	0.017667
		θ	0.902166	0.005393	0.957553	0.001251
	N(1,1)	b	0.789262	0.049123	0.596509	0.046723
		a	0.951716	0.031079	0.968576	0.003142
		c	0.380843	0.061311	0.425155	0.044725
		θ	0.902399	0.005122	0.960437	0.002499

Berdasarkan hasil perhitungan MSE, dapat dibuat diagram tiap kasus dengan melihat distribusi kemampuan peserta dari data yang dibangkitkan sebagai variable (Gambar 1). Berdasarkan gambar ini, ada kecenderungan MSE paling rendah terjadi pada data yang dibangkitkan dengan parameter kemampuan berdistribusi normal baku atau N(0,1). MSE tertinggi terjadi jika distribusi kemampuan N(-1,1), diikuti oleh MSE pada distribusi kemampuan N(1,1). Mencermati lebih lanjut tabel Ringkasan Parameter Butir Hasil Estimasi (table 1), dapat diperoleh bahwa meskipun data dibangkitkan dengan distribusi kemampuan yang berbeda-beda, namun pada setiap kasus, distribusi kemampuan hasil estimasi berdistribusi normal baku atau N(0,1). Hasil ini menunjukkan ada pengaruh distribusi parameter kemampuan pada saat membangkitkan data terhadap stabilitas estimasi parameter kemampuan hasil estimasi, dan yang paling stabil jika parameter kemampuan berdistribusi normal baku.

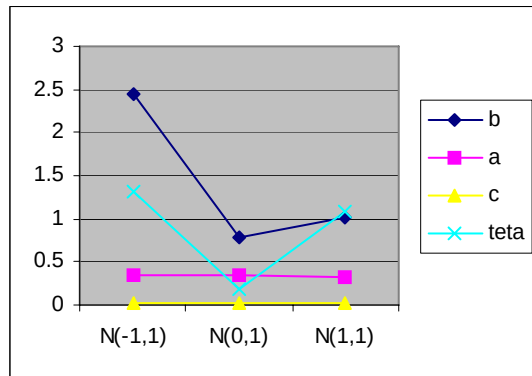
Hasil yang dideskripsikan pada MSE, kurang didukung jika indikator stabilitas parameter dilihat dari korelasi antara parameter estimasi dengan parameter sebenarnya. Korelasi antara parameter estimasi dan parameter sebenarnya tidak memiliki pola yang pasti, jika dilihat dari distribusinya, N(-1,1), N(0,1) dan N(1,1), seperti yang dideskripsikan pada gambar 2.

Diagram MSE untuk tiap kasus dengan variabel ukuran sampel peserta disajikan pada gambar 3. Berdasarkan gambar ini, dapat dicermati bahwa pada parameter a dan c, ada kecenderungan semakin besar ukuran sampel,

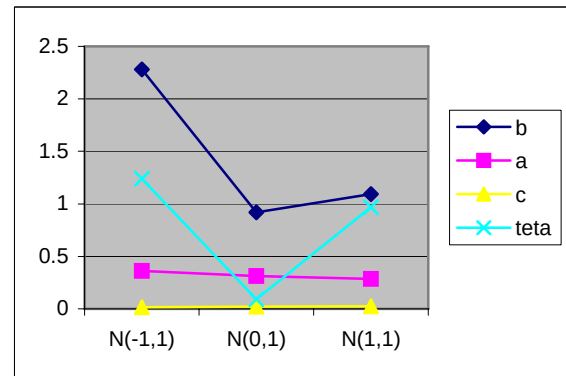
semakin besar keakuratan pengestimasiannya, dengan indikasi semakin menurunnya MSE. Namun untuk parameter b dan θ tidak ada pola yang pasti.

Gambar 1. Diagram MSE tiap kasus dengan variabel distribusi kemampuan peserta sebenarnya

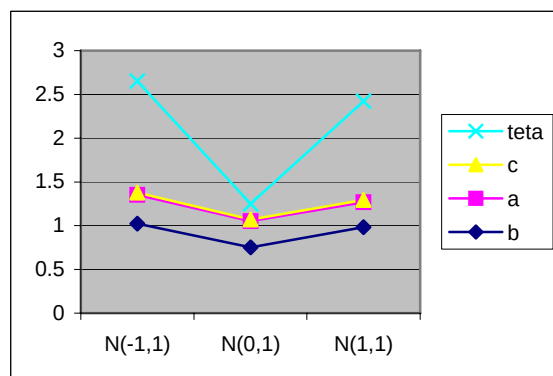
MSE (20 butir 300 peserta)



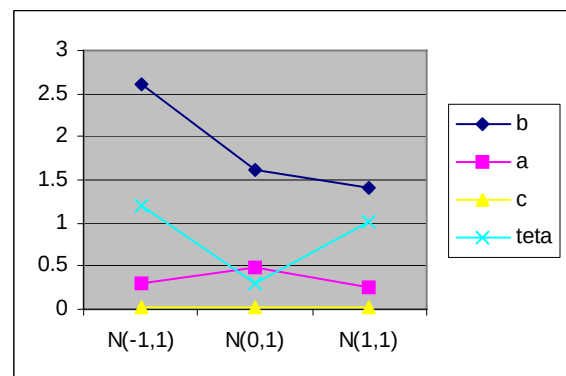
MSE (50 butir 300 peserta)



MSE (20 butir 750 peserta)

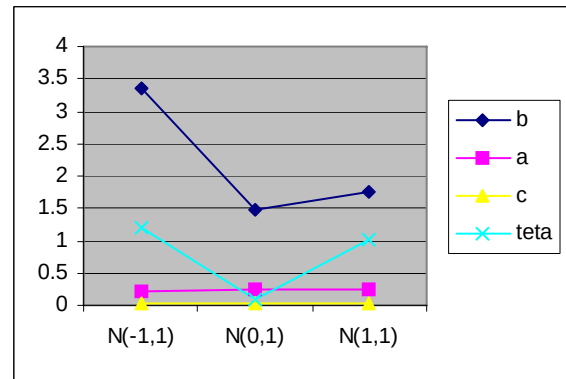
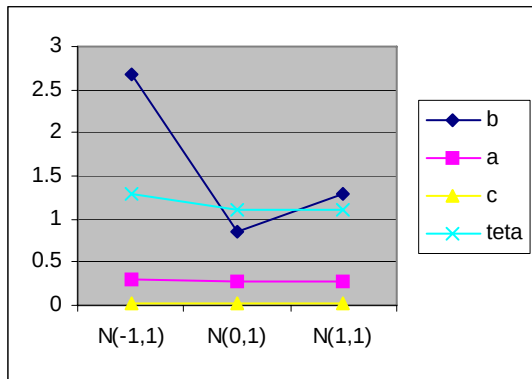


MSE (50 butir 750 peserta)



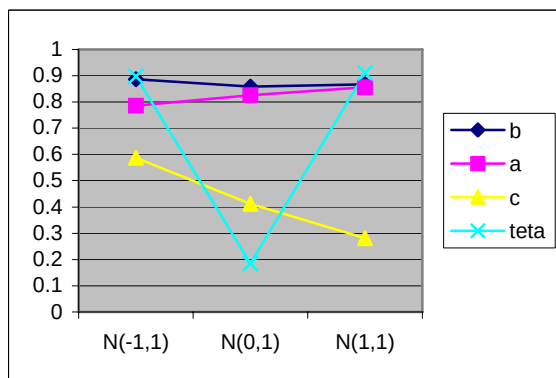
MSE (20 butir 1200 peserta)

MSE (50 butir 1200 peserta)

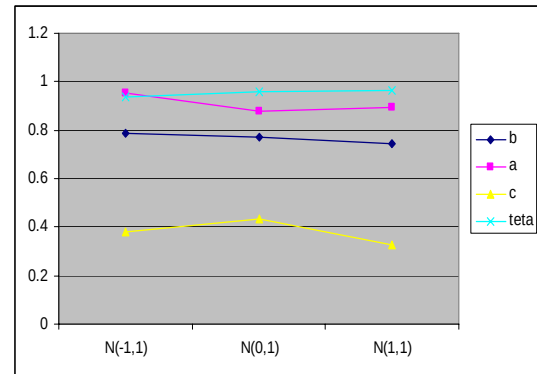


Gambar 2. Diagram korelasi untuk tiap kasus dengan variabel distribusi kemampuan peserta sebenarnya (true)

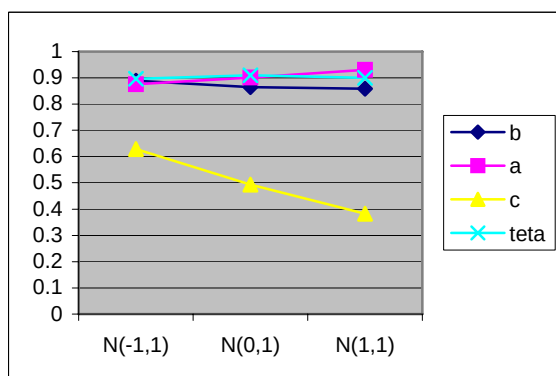
Korelasi (20 butir 300 peserta)



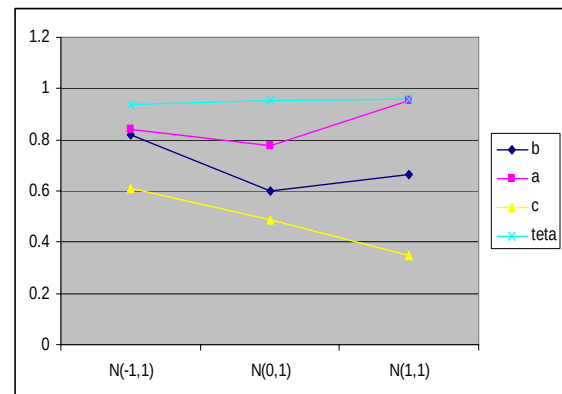
Korelasi (50 butir 300 peserta)



Korelasi (20 butir 750 peserta)

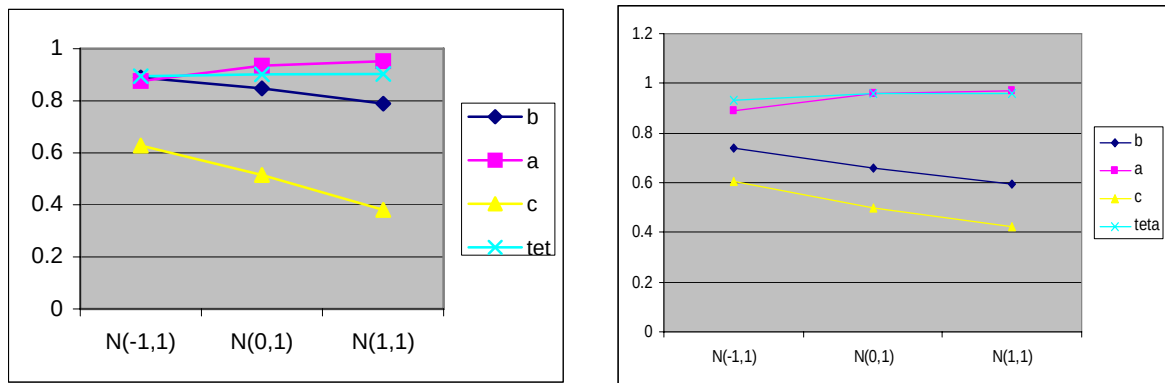


Korelasi (50 butir 750 peserta)



Korelasi (20 butir 1200 peserta)

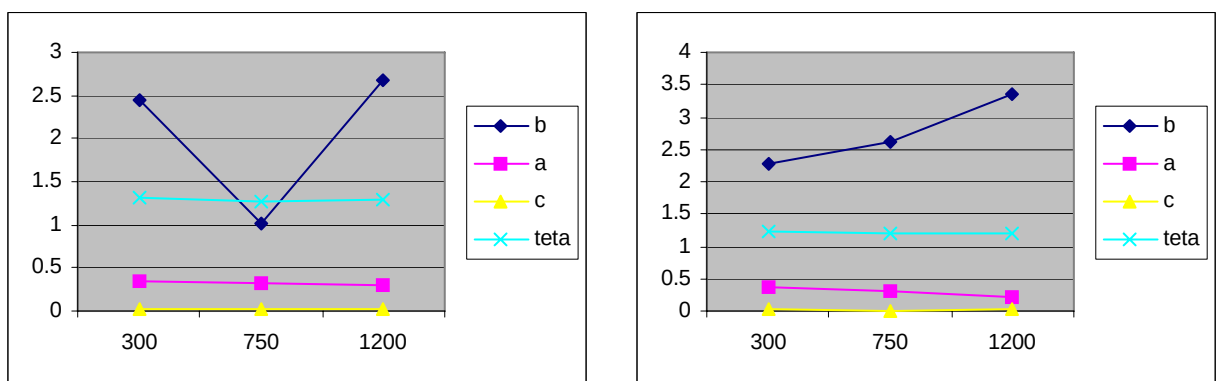
Korelasi (50 butir 1200 peserta)



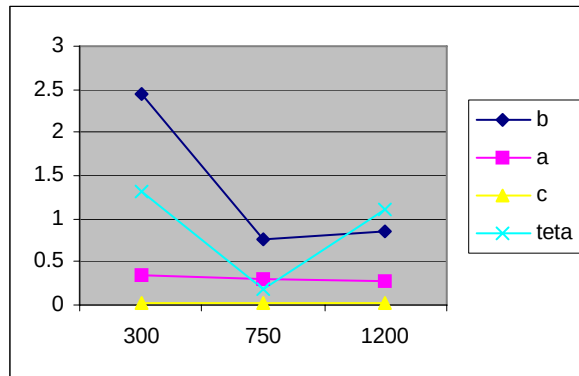
Gambar 3. Diagram MSE tiap kasus dengan variabel ukuran sampel peserta tes

MSE (N(-1,1) 20 butir)

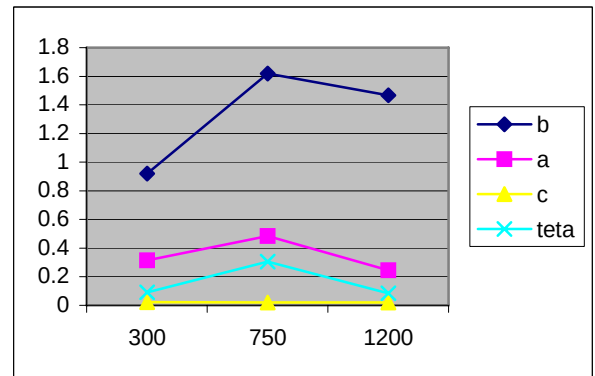
MSE (N(-1,1) 50 butir)



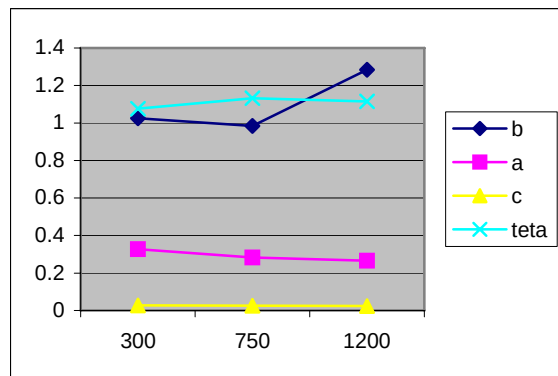
MSE (N(0,1) 20 butir)



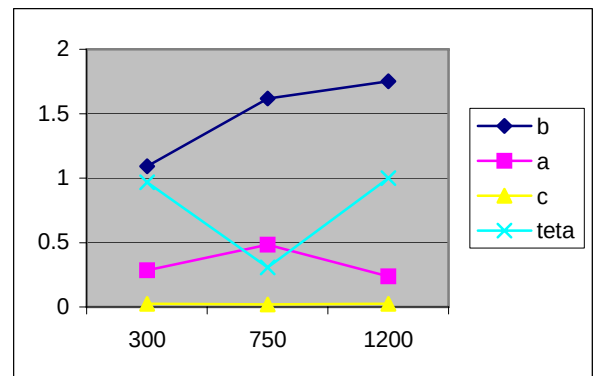
MSE (N(0,1) 50 butir)



MSE (N(1,1) 20 butir)



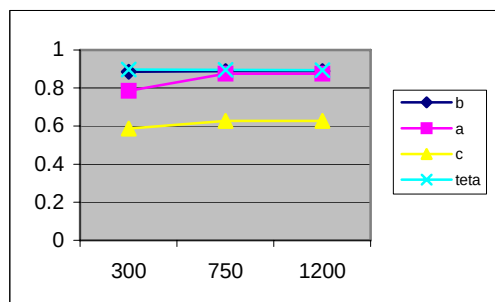
MSE (N(1,1) 50 butir)



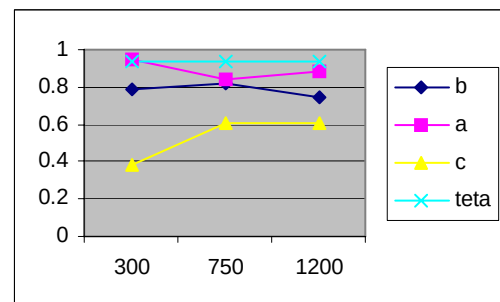
Pada gambar 4 dengan melihat ukuran sample peserta tes sebagai variable, untuk parameter daya pembeda (a), tebakan semu (c) dan parameter kemampuan (θ), ada kecenderungan semakin semakin besar ukuran sample peserta tes, semakin dekat korelasi antara parameter hasil estimasi dengan parameter sebenarnya. Namun hal ini tidak berlaku untuk parameter tingkat kesulitan, yang tidak memiliki pola yang pasti.

Gambar 4. Diagram Korelasi tiap kasus dengan variabel ukuran sampel peserta tes

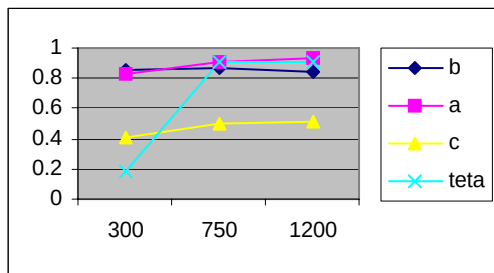
Korelasi (N(-1,1) 20 butir)



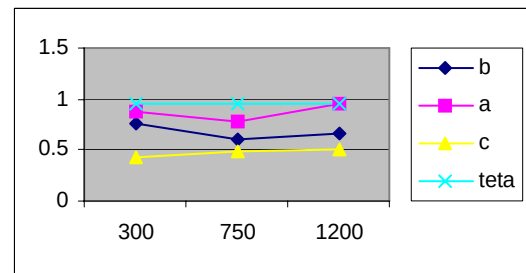
Korelas (N(-1,1) 50 butir)



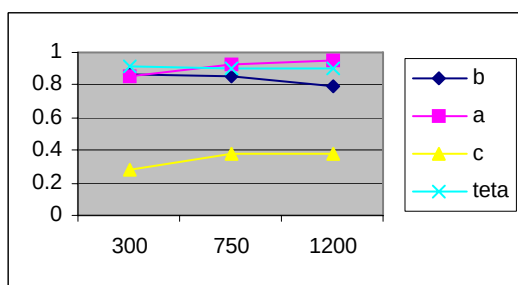
Korelasi (N(0,1) 20 butir)



Korelasi (N(0,1) 50 butir)



Korelasi (N(1,1) 20 butir)



Korelasi (N(1,1) 50 butir)

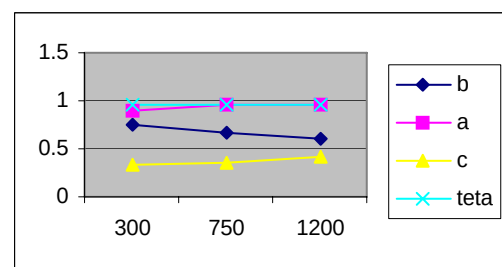


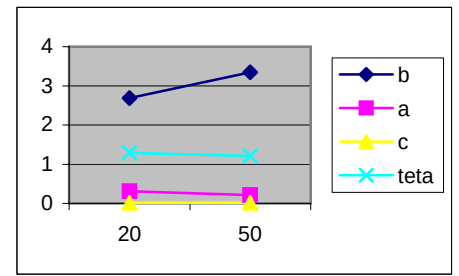
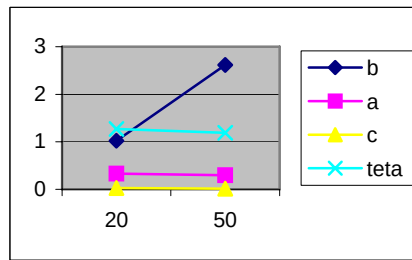
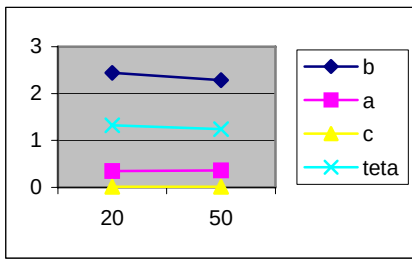
Diagram MSE dengan variable panjang tes disajikan pada gambar 5. Berdasarkan gambar ini, dapat dilihat bahwa ada kecenderungan semakin panjang suatu tes, akan semakin besar keakuratannya untuk mengestimasi parameter a, c dan θ . Namun sebaliknya, pada pengestimasi parameter tingkat kesulitan (b), semakin panjang tes akan semakin kurang akurat, karena semakin besar MSE-nya.

Gambar 5. Diagram MSE tiap kasus dengan variabel panjang tes

MSE N(-1,1) 300 butir

MSE N(-1,1) 750 butir

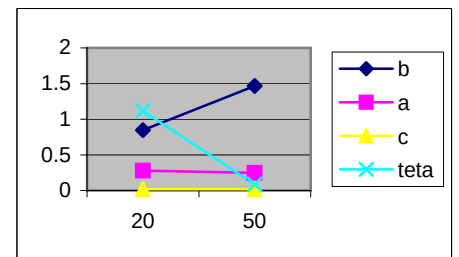
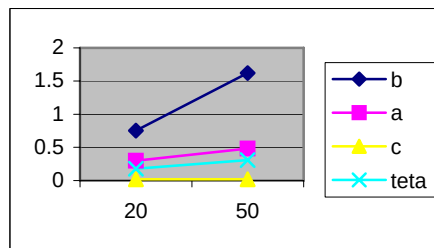
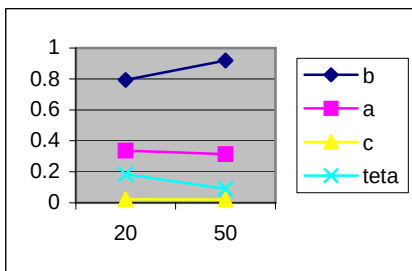
MSE N(-1,1) 1200 butir



MSE $N(0,1)$ 300 butir

MSE $N(0,1)$ 750 butir

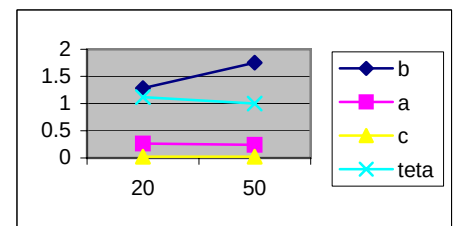
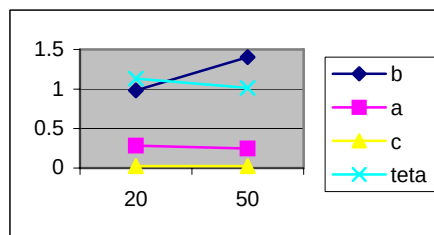
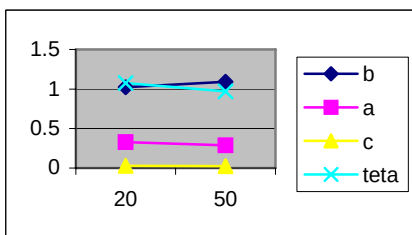
MSE $N(0,1)$ 1200 butir



MSE $N(1,1)$ 300 butir

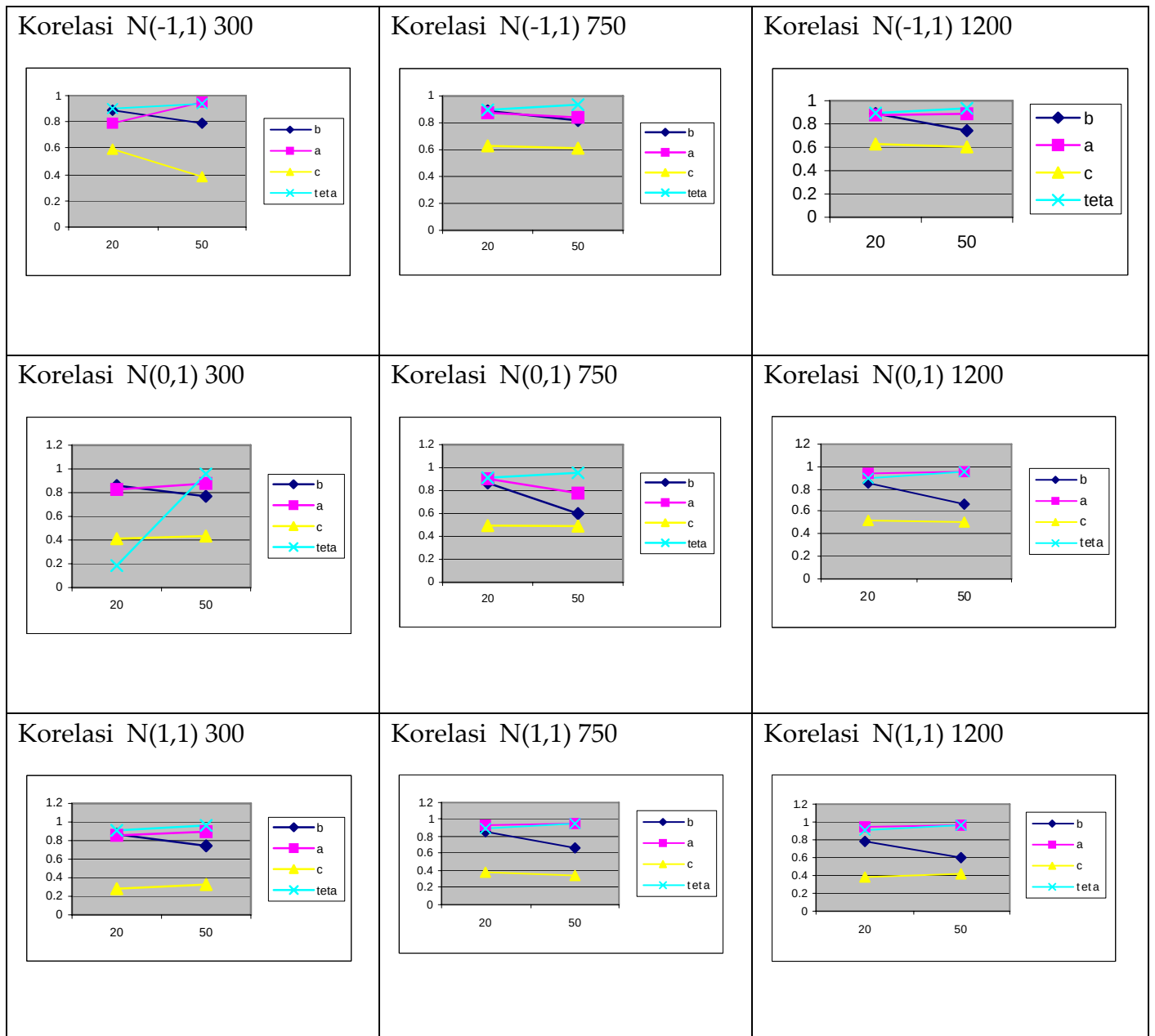
MSE $N(1,1)$ 750 butir

MSE $N(1,1)$ 1200 butir



Hasil perhitungan MSE untuk variable panjang tes didukung oleh hasil pada korelasi parameter sebenarnya dengan parameter hasil estimasi. Pada parameter a, c dan θ , ada kecenderungan semakin panjang suatu tes, pada kasus ini dari panjang 20 butir ke 50 butir, semakin dekat korelasi antara parameter hasil estimasi dengan parameter sebenarnya. Namun, pada parameter b, semakin besar panjang tes, korelasi antara parameter butir dengan parameter sebenarnya justru menurun. Hubungan ini disajikan pada gambar 6.

Gambar 6. Diagram korelasi tiap kasus dengan variabel panjang tes



Dari parameter-parameter hasil estimasi untuk tiap-tiap kasus, dapat diestimasi nilai fungsi informasi yang selanjutnya dirata-rata setiap 5 replikasi dari tiap kasus yang disajikan pada table 4.

Tabel 4. Nilai fungsi informasi (FI) dan kesalahan standar estimasi (*Standard Error Estimate, SEE*) hasil estimasi dari sekumpulan butir dari tiap kasus

Ukuran Sampel	Distribusi 0	20 butir		50 butir	
		FI	SEE	FI	SEE
300	N(-1,1)	5.6872	0.419325	19.338	0.227402
	N(0,1)	5.9608	0.409588	18.348	0.233456
	N(1,1)	6.0274	0.407319	17.486	0.239141
750	N(-1,1)	5.6992	0.418883	16.764	0.244237
	N(0,1)	5.802	0.415156	16.95	0.242893
	N(1,1)	4.9924	0.447554	15.648	0.252796
1200	N(-1,1)	5.4098	0.429942	16.422	0.246767
	N(0,1)	5.4122	0.429846	18.072	0.235232
	N(1,1)	5.6348	0.42127	15.92	0.250627

Berdasarkan tabel 4, diperoleh bahwa pada pola respons peserta tes dengan panjang tes 20 butir, rerata nilai fungsi informasi berkisar antara 4,9924 sampai dengan 6,0274. Besar nilai fungsi informasi ini tidak memiliki kecenderungan, baik dilihat dari distribusi kemampuan peserta maupun ukuran sample peserta. Pada kasus panjang tes 50 butir, rerata nilai fungsi informasi informasi berkisar antara 15.648 sampai dengan 19.338, dan tidak ada pola kecenderungan besarnya nilai dilihat dari distribusi kemampuan peserta maupun ukuran peserta tes. Berdasarkan hasil ini dapat disimpulkan bahwa ukuran sample peserta tes dan distribusi kemampuan tidak mempengaruhi besarnya nilai fungsi informasi dan pada tes yang lebih panjang, nilai fungsi informasi akan lebih besar. Demikian pula halnya dengan kesalahan pengukuran estimasi (SEE), yang tidak dipengaruhi oleh distribusi dan ukuran sampel, tetapi pada studi ini hanya dipengaruhi oleh panjang tes. Semakin panjang suatu tes, akan semakin kecil SEE-nya, atau akan semakin akurat estimasi parameter-parameternya.

Selanjutnya dilakukan analisis varians 3 jalur, untuk mengetahui signifikansi efek panjang tes, distribusi kemampuan peserta tes, dan ukuran sampel pada hasil perhitungan rerata MSE dan rerata korelasi. Pada table 5 disajikan hasil analisis varians yang signifikan.

Table 5. Hasil analisis varians yang signifikan

Sumber	Derajat kebebasan	Signifikansi
MSE Parameter b		
Distribusi kemampuan peserta tes	2	0,030
Panjang tes*ukuran sampel	2	0,082
Korelasi Parameter b		
Panjang tes	1	0,034

Berdasarkan hasil analisis ini, distribusi kemampuan, panjang tes dan secara bersama-sama interaksi panjang tes dengan ukuran sample yang berpengaruh pada stabilitas estimasi parameter, itupun hanya pada parameter tingkat kesulitan (b). Pada parameter yang lain (a, c, dan θ), panjang tes, distribusi kemampuan dan ukuran sample peserta tes tidak berpengaruh pada stabilitas parameter daya pembeda, tebakan semu, dan kemampuan peserta tes.

Kesimpulan dan Rekomendasi

Sesuai dengan tujuannya, studi ini dimaksudkan untuk mengetahui : (1) efek panjang tes terhadap stabilitas estimasi parameter butir dan parameter kemampuan pada model logistic 3 parameter, (2) efek distribusi kemampuan terhadap stabilitas estimasi parameter butir dan parameter kemampuan pada model logistic 3 parameter, dan (3) ukuran sample/banyaknya peserta tes terhadap stabilitas estimasi parameter butir dan parameter kemampuan pada model logistic 3 parameter. Hasil penelitian menunjukkan bahwa ada kecenderungan MSE paling rendah terjadi pada data yang dibangkitkan dengan parameter kemampuan berdistribusi normal baku dan korelasi antara parameter estimasi dan parameter sebenarnya tidak memiliki pola yang pasti, jika dilihat dari distribusinya. Pada parameter a dan c , ada kecenderungan semakin besar ukuran sampel, semakin besar keakuratan pengestimasiannya, , namun untuk b dan θ tidak ada pola yang pasti. Melihat korelasinya, ada kecenderungan semakin semakin besar ukuran sample peserta tes, semakin dekat korelasi antara parameter hasil estimasi dengan parameter sebenarnya, namun hal ini tidak berlaku untuk parameter b . Berdasarkan rerata MSE dan rerata korelasi , ada kecenderungan semakin panjang suatu tes, akan semakin besar keakuratannya untuk mengestimasi parameter a , c dan θ . Namun sebaliknya, pada pengestimasian parameter tingkat kesulitan (b), semakin panjang tes akan semakin kurang akurat, karena semakin besar MSE-nya. Demikian pula berdasarkan korelasinya. Kesalahan pengukuran estimasi (SEE), yang tidak dipengaruhi oleh distribusi dan ukuran sampel, tetapi pada studi ini hanya dipengaruhi oleh panjang tes. Berdasarkan hasil analisis signifikansi dengan analisis varians, distribusi kemampuan, panjang tes dan interaksi panjang tes dengan ukuran sample yang berpengaruh pada stabilitas estimasi parameter b saja.

Pada studi ini, hanya dibahas panjang tes 20 butir dan 50 butir saja, yang mewakili tes pendek tes panjang. Namun perlu dikaji lebih mendalam jika panjang tesnya kurang dari 20 butir atau lebih dari 50 butir. Distribusi parameter kemampuan peserta tes yang dibahas di studi ini hanya yang berdistribusi normal saja, padahal pada realitasnya masih ada distribusi-distribusi yang lainnya, misalnya distribusi miring. Hal ini perlu dikaji lebih

lanjut., termasuk juga variable ukuran sample peserta tes. Terlepas dari keterbatasan penelitian ini, yang masing-masing hanya dilakukan 5 replikasi tiap kasus, perlu dilakukan studi sejenis dengan replikasi yang lebih banyak, sehingga memadai untuk penarikan kesimpulan.

DAFTAR KEPUSTAKAAN

Bastari (1998). An Investigation of linear and non linear Estimates for Multidimensional Graded Response Model. *Paper*. Tidak dipublikasikan.

Cohen, A.S. & Kane, M.T. (2001). The Precision of simulation study results. *Applied Psychological Measurement Journal*. Vol. 25 No. 2. pp. 136-145

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.

Hambleton, R.K. & Cook, L.L. (tth). Robusness of item response models and effects of tes length and sample size on the precision of ability estimates. *New Horizons in Testing Journal*.

Keppel, G. (1982). *Design and analysis*. London : Prentice-Hall International Inc.

Mislevy, R.J. & Bock, R.D. (1990). *BILOG 3 : Item analysis & test scoring with binary logistic models*. Moorseville : Scientific Software Inc.

Stone, C.A. & Bo Zhang (2003). Assessing goodness of fit of item response theory models : a comparison of traditional and alternative. *Journal of Educational Measurement*. Winter, vol 40. N0. 4. pp. 331-352.

Swaminathan H, dkk. (2003). Small sample estimation in dichotomous item response models : effect of priors based on judgemental information on accuracy of item parameter estimates. *Journal of Educational Measurement*. Winter, vol 27. N0. 1. pp. 27-51.

Van Abswoude, A.A.H., dkk. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Journal of Educational Measurement*., vol 28. N0. 1. pp. 3-24.