



**PENGARUH SAMPLE SIZE TERHADAP KESTABILAN ITEM PARAMETER
ESTIMATE PADA ITEM RESPONSE THEORY
MODEL TIGA PARAMETER LOGISTIK**

Samritin
Universitas Muhammadiyah Buton
samritin75@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengetahui kestabilan estimasi parameter butir tes pada teori tes modern (IRT) model tiga parameter logistik (3PL). Penelitian ini menggunakan metode simulasi. Data simulasi dibangkitkan menggunakan program WinGen for Windows. Analisis kestabilan estimasi dilakukan menggunakan korelasi antara true parameter dengan estimate parameter dan RMSE. Hasil penelitian menunjukkan bahwa pada sample size 500, 750, dan 1000 diperoleh koefisien korelasi (ρ) masing-masing sebesar 0.901, 0.988, dan 0.992 dengan RMSE 0.901, 0.351, dan 0.273 untuk parameter tingkat kesukaran (b), koefisien korelasi (ρ) sebesar 0.746, 0.807, dan 0.830 dengan RMSE 0.746, 0.304, dan 0.281 untuk parameter daya beda (a), koefisien korelasi (ρ) sebesar 0.642, 0.667, dan 0.679 dengan RMSE sebesar 0.642, 0.057, dan 0.056 untuk parameter pseudo-guessing. Berdasarkan hasil analisis data disimpulkan bahwa semakin sebesar sample size, estimasi parameter butir semakin stabil.

Kata kunci: sample size, parameter estimate.

Abstract

This study aims to determine the stability of the parameter estimates of the test items on the modern test theory (IRT) for the three-parameter logistics model (3PL). This study uses a simulation method. Simulation data was generated using the WinGen for Windows program. Estimation stability analysis was carried out using the correlation between true parameters with estimate parameters and RMSE. The results showed that for sample sizes 500, 750, and 1000, the correlation coefficient (ρ) was 0.901, 0.988, and 0.992 respectively with RMSE 0.901, 0.351, and 0.273 for the difficulty parameter (b), correlation coefficient (ρ) of 0.746, 0.807, and 0.830 with an RMSE of 0.746, 0.304, and 0.281 for the power of difference parameter (a), the correlation coefficient (ρ) of 0.642, 0.667, and 0.679 with an RMSE of 0.642, 0.057, and 0.056 for the pseudo-guessing parameter. Based on the results of data analysis it was concluded that the larger the sample size, the more stable the item parameter estimates.

Key words: sample size, parameter estimate.

Pendahuluan

Kualitas tes merupakan salah satu aspek penting dalam pengukuran hasil belajar. Penggunaan tes yang berkualitas memberikan informasi hasil tes yang benar. Sebaliknya penggunaan tes yang tidak berkualitas akan memberikan hasil tes yang tidak sesuai dengan kemampuan yang sebenarnya. Hal ini berdampak pada pengambilan keputusan berdasarkan hasil tes. Tes yang berkualitas dapat diketahui karakteristik atau parameter butir-butirnya. Tes yang berkualitas hanya dapat diperoleh melalui proses pengembangan sampai pada estimasi parameter butir tes. Proses estimasi ini disebut juga kalibrasi (Huriaty, 2015). Estimasi parameter butir tes hanya dapat dilakukan apabila data respons peserta tes telah diperoleh.

Dalam teori tes modern (*item response theory-IRT*), dikenal 3 model yang sangat populer yaitu model satu parameter logistik (1PL), model dua parameter logistik (2PL) dan model tiga parameter logistik (3PL) (Hambleton, Swaminathan, dan Rogers, 1991). Ketiga model ini cocok diterapkan pada model respon dikotomis. Model 1PL hanya memiliki parameter tingkat kesukaran. Model ini dikenal juga sebagai model Rasch. Model 2PL memiliki parameter tingkat kesukaran dan parameter daya beda butir. Model ini pertama kali dikembangkan oleh Lord tahun 1952 yang mengacu pada model ogive normal. Model 3PL memiliki tambahan parameter tebakan (*pseudo-guessing/chance*) yang membedakannya dengan model 2PL. Parameter *pseudo-guessing* lahir dari fenomena jawaban tebakan ketika peserta tes merespon soal yang tidak diketahui atau tidak diyakini kebenarannya.

Pengujian akurasi estimasi parameter pada model 3PL telah dilakukan oleh Gao dan Chen (2005) yang menemukan bahwa estimasi parameter butir menggunakan $N=500$ dengan panjang tes 10 butir hasilnya tidak akurat dalam mengestimasi parameter butir a . Penelitian yang dilakukan oleh Sahin dan Ain (2017) menemukan akurasi estimasi parameter butir model 3PL pada sampel 750 dan 1000 dengan panjang tes 10 butir kedua penelitian ini menguji pengaruh sampel terhadap estimasi parameter butir menggunakan kriteria *root mean square difference* (RMSD). Jauh sebelum kedua penelitian tersebut, Wright (1977) mengemukakan bahwa untuk sample size yang semakin besar pada tes dengan panjang terbatas, estimasi parameter item tidak mendekati nilai sebenarnya. Oleh karena itu penelitian tentang keakuratan atau kestabilan estimasi parameter dengan sample size dan panjang tes yang bervariasi serta pendekatan yang berbeda perlu terus dilakukan. Berbeda dengan studi sebelumnya tersebut penelitian ini bertujuan pengaruh sample size terhadap kestabilan item

parameter estimate pada IRT model tiga parameter logistik menggunakan koefisien korelasi dan *root mean square error* (RMSE).

Metode Penelitian

Jenis Penelitian

Penelitian ini termasuk dalam penelitian simulasi yang bertujuan untuk mengkaji pengaruh *sample size*, terhadap kestabilan estimasi *item parameter*. *Item parameter estimate* dalam studi ini merupakan variabel dependen yang ingin diketahui kestabilannya apabila variabel independen bervariasi. Variabel independen dalam studi ini yaitu *sample size*. *Sample size* merupakan jumlah peserta tes dalam suatu administrasi tes yang menempuh sejumlah *number of item* (jumlah item dalam suatu tes). Penelitian ini menggunakan Model Item Response Theory (IRT) tiga parameter logistik (3PL).

Penelitian ini didesain menggunakan metode simulasi. Simulasi dalam penelitian ini dirancang dan dijalankan menggunakan Program WinGen for Windows dan Program Bilog-MG. Simulasi pengaruh *sample size* terhadap kestabilan *Item Parameter Estimate* dirancang menggunakan tiga kategori *sample size* yaitu N=500, N=750, dan N=1000, panjang tes n=30 (konstan). Desain simulasi ini disajikan pada Tabel 1.

Tabel 1. Desain Simulasi Pengaruh *Sample Size* terhadap Kestabilan *Item Parameter Estimate*

	V1	V2	V3
Sample size	500	750	1000
Number of Item	30	30	30
Model	3PLM	3PLM	3PLM
File Output WinGen	T1A2V1	T1A2V2	T1A2V3

Prosedur Penelitian

Penelitian simulasi ini dilakukan sesuai tahapan sebagai berikut.

1. Menentukan Desain Pembangkitan Data Simulasi
2. Membangkitkan Items Response Menggunakan Program WinGen for Windows untuk menghasilkan data true parameter. Data true parameter dan estimate parameter dibangkitkan berdasarkan model, *Sample Size*, dan *Number of Item* yang ditentukan. Pembangkitan data menggunakan faktor penskalaan $D=1.7$. Kriteria data yang dibangkitkan yaitu data berdistribusi normal dengan $mean = 0$, $deviation\ standard = 1$, kategori respons = 2 (benar = 1 dan salah = 0), item parameter berdistribusi *uniform*

dengan pada rentang $0,4 \leq a \leq 2$, $-3 \leq b \leq 3$, dan $0 \leq c \leq 0,25$. Replikasi data dilakukan 5 kali untuk setiap kasus

3. Data-data true parameter dan estimate parameter selanjutnya dianalisis menggunakan Program Excel for Windows.

Teknik Analisis Data

Data yang diperoleh dari hasil bangkitan Program WinGen dianalisis menggunakan program Excel for Windows. Pengaruh sample size sebagai variabel independen terhadap item parameter estimate sebagai variabel dependen dianalisis dari koefisien korelasi antara true parameter dengan estimate parameter dan RMSE parameter estimate. Koefisien korelasi (ρ) dan RMSE tersebut selanjutnya disajikan dalam bentuk grafik yang menggambarkan kecenderungan pengaruh variabel terhadap variabel independen

Hasil Penelitian dan Pembahasan

Kestabilan item parameter dan ability estimate sangat penting dalam pengukuran kualitas tes dan hasil belajar atau kemampuan peserta didik. Pada kasus jumlah tes tertentu dan atau jumlah peserta tes yang dianalisis berpengaruh pada kesalahan informasi yang diperoleh yang berimplikasi pada kesalahan pengambilan keputusan baik tentang kualitas tes, butir-butir tes, maupun kemampuan peserta didik. Penelitian ini melakukan simulasi kestabilan estimasi parameter butir pada model IRT dengan *sample size* yang bervariasi. Analisis parameter butir dilakukan pada model 3 parameter logistic sehingga diperoleh informasi kestabilan estimasi parameter kesukaran butir (b), daya beda (a) dan factor tebakan (c). Data hasil penelitian ini dianalisis menggunakan koefisien korelasi (ρ) true parameter dengan estimate parameter, dan *root mean square error* (RMSE). Berikut disajikan ringkasan dalam bentuk tabel dan grafik untuk masing-masing studi.

Koefisien Korelasi

Koefisien korelasi *true parameter* dengan *estimate parameter* menggambarkan hubungan antara nilai parameter yang sebenarnya dengan hasil estimasi. Koefisien korelasi setiap parameter dapat dilihat pada Tabel 2.

Tabel 2. Koefisin Korelasi (ρ) *True Paramater* dan *Estimate Parameter*

Parameter	Variasi N		
	500	750	1000
b	0.901	0.988	0.992
a	0.746	0.807	0.830
c	0.642	0.667	0.679

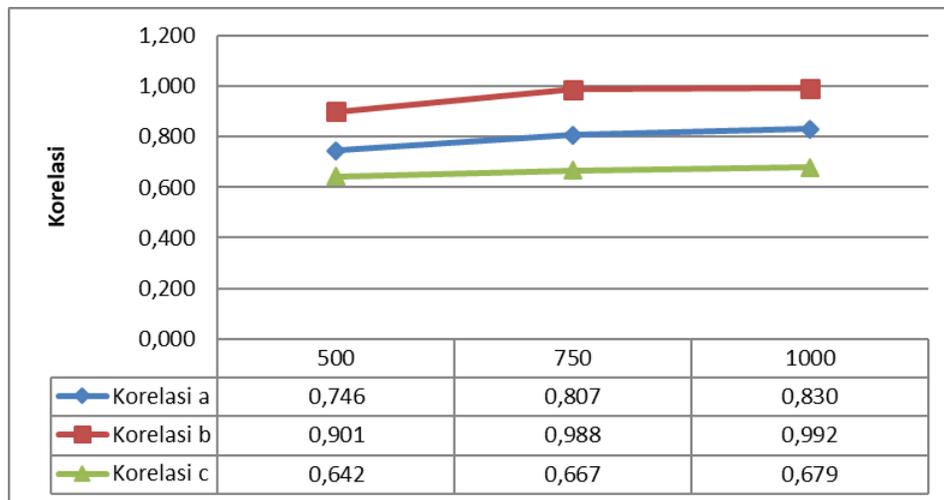
Tabel 2 menunjukkan rerata korelasi *true parameter* dengan *estimate parameter* pada *sample size* yang bervariasi yaitu N=500, N=750, dan N=1000. Tampak bahwa untuk parameter taraf kesukaran (b), korelasi *true parameter* dengan *estimate parameter* memiliki indeks sangat tinggi dengan indeks terendah 0,901 pada N=500, kemudian meningkat 0,087 poin pada N=750 dan meningkat 0,091 poin pada N=1000 atau meningkat 0,004 poin dari sampel N=750. Peningkatan korelasi tersebut mengindikasikan bahwa dalam estimasi parameter b, perubahan *sample size* mempengaruhi perubahan korelasi *true parameter* dengan *estimate parameter*. Berdasarkan hasil analisis tersebut, dapat dikatakan bahwa estimasi parameter b pada model 3PL akan diperoleh hasil yang semakin stabil dan semakin mendekati *true parameter* bila kalibrasi dilakukan menggunakan *sample size* yang besar.

Estimasi parameter daya pembeda butir (a) yang ditampilkan dalam Tabel 2 terlihat perubahan meningkat dari setiap penambahan *sample size*. Pada *sample size* N=500 korelasi *estimate parameter* dengan *true paramater* sebesar 0,746, kemudian meningkat sebesar 0,057 poin pada N=750, dan meningkat lagi sebesar 0,084 poin pada *sample size* N=1000 atau meningkat 0,023 poin dari kasus N=500. Peningkatan korelasi tersebut mengindikasikan bahwa dalam estimasi parameter daya beda (a), perubahan *sample size* mempengaruhi perubahan korelasi *true parameter* dengan *estimate parameter*. Hal ini menunjukkan bahwa estimasi parameter a pada model 3 PL memberikan hasil yang stabil dan mendekati *true parameter* bila kalibrasi dilakukan pada *sample size* yang semakin besar.

Pada tabel 2 terlihat bahwa koefisien korelasi *true parameter* dengan *estimate paramater pseudo-guessing* (c) memiliki indeks masing-masing di bawah 0,7. Meskipun tergolong sedang, rata-rata koefisien korelasi *true parameter* dengan *estimate parameter* terlihat meningkat pada setiap kasus. Pada *sample size* N=500 diperoleh koefisien korelasi

true parameter dengan *estimate parameter* sebesar 0,642, kemudian meningkat sebesar 0,025 pada kasus *sample size* $N=750$ dan meningkat lagi sebesar 0,037 poin pada kasus *sample size* $N=1000$ atau meningkat 0,012 poin dari kasus *sample size* $N=500$. Peningkatan koefisien korelasi tersebut mengindikasikan bahwa dalam estimasi parameter *pseudoguessing* (c), perubahan *sample size* mempengaruhi perubahan korelasi *true parameter* dengan *estimate parameter*. Hal ini menunjukkan bahwa estimasi parameter *pseudo-guessing* (c) pada model 3 PL memberikan hasil yang stabil dan mendekati *true parameter* bila kalibrasi dilakukan pada *sample size* yang semakin besar.

Hasil analisis yang terhadap korelasi *true parameter* dengan *estimate parameter* baik parameter b, a, maupun c dipengaruhi oleh besarnya sampel yang digunakan. Berdasarkan hasil analisis terhadap korelasi dan kecenderungan perubahan koefisien korelasi, diketahui bahwa perubahan koefisien korelasi terbesar terjadi pada perubahan *sample size* dari $N=500$ menjadi $N=750$. Sementara perubahan koefisien korelasi dari *sample size* $N=750$ menjadi $N=1000$ menghasilkan perubahan koefisien korelasi yang lebih kecil. Untuk lebih jelasnya, kecenderungan perubahan koefisien korelasi pada setiap perubahan *sample size* dapat dilihat pada Gambar 1.



Gambar 1. Grafik Korelasi Item Parameter Estimate

Root Mean Square of Error

Kestabilan estimasi parameter perlu dianalisis dari aspek penyimpangan yang terjadi. Hal ini penting dilakukan untuk menghindari besarnya penyimpangan (kesalahan) hasil estimasi

dengan parameter murninya. Data penyimpangan hasil estimasi dalam penelitian ini disajikan dalam bentuk *Root Mean Square Error* (lihat Tabel 3).

Tabel 3. *RMSE Item Parameter Estimate*

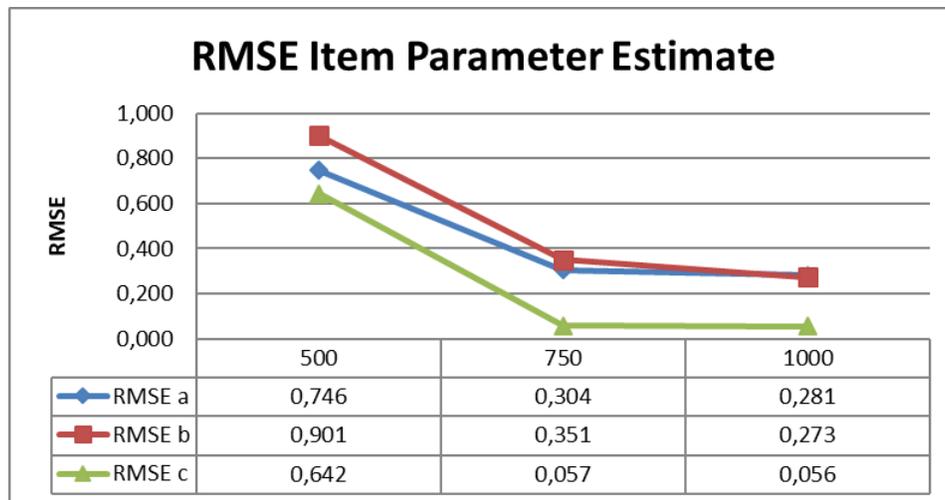
Parameter	Variasi N		
	500	750	1000
b	0.901	0.351	0.273
a	0.746	0.304	0.281
c	0.642	0.057	0.056

Tabel 3 memperlihatkan RMSE dari estimasi parameter pada *sample size* N=500, N=750, dan N=1000. RMSE estimasi parameter b terlihat sangat tinggi mendekati nilai 1 pada *sample size* rendah yaitu N=500. Namun RMSE turun sebesar 0,55 poin pada *sample size* N=750 dan RMSE turun lagi sebesar 0,098 poin pada *sample size* N=1000. Perubahan nilai tersebut mengindikasikan bahwa dalam estimasi parameter tingkat kesukaran (b), perubahan *sample size* mempengaruhi perubahan kesalahan estimasi. RMSE estimasi parameter tingkat kesukaran tertinggi terjadi pada *sample size* kecil, dan RMSE rendah terjadi pada *sample size* besar. Hal ini berarti bahwa semakin besar *sample size* yang digunakan dalam estimasi parameter b pada model 3 PL, hasil estimasi semakin mendekati *true parameter*.

Estimasi parameter daya pembeda (a) mengandung kesalahan sebagaimana dalam estimasi parameter tingkat kesukaran (b). Tabel 3 memperlihatkan bahwa RMSE estimasi parameter a yang cukup tinggi pada *sample size* N=500 yaitu sebesar 0,746 meskipun lebih rendah 0,155 poin dari RMSE parameter tingkat kesukaran (b). RMSE turun sebesar 0,406 poin pada *sample size* N=750 kemudian turun lagi pada sebesar 0,465 pada kasus *sample size* N=1000 atau turun sebesar 0,023 dari *sample size* N=750. Perubahan RMSE setiap kasus mengindikasikan bahwa dalam estimasi parameter daya beda (a), perubahan *sample size* mempengaruhi perubahan kesalahan estimasi. RMSE estimasi parameter daya beda (a) tertinggi terjadi pada kasus rendah dan RMSE rendah pada *sample size* besar. Hal ini berarti bahwa semakin besar *sample size* yang digunakan dalam estimasi parameter daya beda pada model 3 PL, hasil estimasi semakin mendekati *true parameter*.

Kehadiran kesalahan estimasi juga terjadi dalam estimasi parameter *pseudo-guessing* (c). Hal ini tampak jelas pada Tabel 3. Pada tabel tersebut terlihat bahwa RMSE hasil estimasi parameter *pseudo-guessing* pada N=500 diperoleh nilai 0,642. Nilai tersebut turun sebesar 0,545 pada *sample size* N=750, kemudian turun lagi 0,001 poin pada N=1000 atau turun 0,546 poin dari *sample size* N=500. Perubahan RMSE pada setiap *sample size* mengindikasikan bahwa perubahan *sample size* mempengaruhi nilai kesalahan dalam estimasi parameter *pseudo-guessing* (c). Perubahan kecil RMSE dari *sample size* N=750 menjadi N=1000 mengindikasikan bahwa pengaruh *sample size* pada N=750 dan N=1000 relatif sama atau tidak berbeda. Sementara pada perubahan *sample size* dari N=500 menjadi N=750, terjadi perubahan RMSE yang sangat besar.

Kesalahan estimasi ketiga parameter model 3 PL di atas menunjukkan bahwa bahwa semakin besar *sample size* RMSE cenderung turun, namun antara N=750 dan N=1000 relatif stabil. Jika *sample size* semakin diperbesar, kemungkinan stabilnya semakin besar. Dapat dikatakan bahwa semakin besar *sample size* RMSE semakin turun, sehingga estimasi terhadap item parameter semakin stabil. Kecenderungan penurunan RMSE pada setiap *sample size* dapat dilihat pada Gambar 2.



Gambar 2. Grafik RMSE Item Parameter Estimate

Gambar 1 dan Gambar 2 menunjukkan bahwa koefisien korelasi berbanding terbalik dengan RMSE untuk masing-masing item parameter estimate. Berdasarkan hasil analisis di atas, ditemukan bahwa setiap peningkatan *sample size* memberikan hasil estimasi parameter butir yang lebih stabil. Estimasi parameter yang sangat stabil ditemukan pada penggunaan

sample size sebanyak $N=750$ atau $N=1000$. Temuan hasil penelitian ini kurang konsisten dengan hasil penelitian Sukirno (2006) yang menyimpulkan bahwa penambahan *sample size* hanya berpengaruh pada kestabilan estimasi parameter tingkat kesulitan dan daya beda tetapi tidak berpengaruh terhadap estimasi parameter tingkat kesulitan namun. Ketidakkonsistenan ini dapat disebabkan oleh perbedaan ukuran *sample size* yang digunakan dalam kedua penelitian.

Simpulan

Berdasarkan hasil analisis data dapat disimpulkan bahwa semakin besar *sample size* maka estimasi parameter butir semakin stabil. Kestabilan estimasi parameter ditemukan pada *sample size* 1000. Namun kestabilan estimasi pada *sample size* 750 tidak jauh berbeda dengan kestabilan estimasi pada *sample size* 1000.

Keterbatasan penelitian

Penelitian ini hanya menggunakan 3 (tiga) ukuran sampel yaitu 500, 750, dan 1000 pada model 3PL. Penelitian ini juga hanya menguji efek pada tes yang memiliki panjang 30 butir. Oleh karena itu perlu dilakukan penelitian lebih lanjut dengan menggunakan variasi ukuran sampel yang lebih banyak, panjang tes yang bervariasi, dan model IRT lainnya.

DAFTAR PUSTAKA

- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: CBS College Publishing.
- DeMars, C. (2001). Group Differences Based On IRT Scores: Does the Model Matter?. *Educational and Psychological Measurement, Vol. 61 No. 1 p.60-70*.
- Gao, F., dan Chen, L. (2005). Bayesian or non Bayesian: A comparison study of item parameter estimation in the three parameter logistic model. *Applied Measurement in Education, 18(4)*, 351-380.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. California: Sage Publications.
- Huriaty, D. (2015). Metode kalibrasi dan desain tes berdasarkan teori respon butir. *Math Didactic: Jurnal Pendidikan Matematika, Vol. 1, No. 3, p.191-199*
- Sahin, A., & Anil, D. (2017). The effect of test length and sample size on item parameter in item response theory. *Educational Sciences: Theory & Practice. 17(1)*, p. 321-335. 10.12738/estp.2017.1.0270

Sukirno, D.S. (2006). Pengaruh panjang tes dan ukuran sampel terhadap kekekaran estimasi parameter pada teori respon butir (item response theory). *Jurnal Cakrawala Pendidikan*, XXV(6), 431-452.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97-116.