

Efek ukuran sampel pada parameter GRM yang didasarkan pada korelasi inter item

Hari Purnomo Susanto^{1,2,a*}, Heri Retnawati^{2,b}, Farida Agus Setiawati^{2,c}

¹ Sekolah Tinggi Keguruan dan Ilmu Pendidikan PGRI Pacitan. Jl. Cut Nya' Dien 4a Pacitan, 63515, Indonesia

² Universitas Negeri Yogyakarta. Jl Colombo No. 1, Yogyakarta, 55281, Indonesia

^a haripurnomosusanto@gmail.com; ^b heri_retnawati@uny.ac.id; ^c farida_as@uny.ac.id

* Corresponding Author.

Received: 14 June 2022; Revised: 24 July 2022; Accepted: 28 July 2022

Abstrak: Teori respon butir memiliki banyak kelebihan Ketika digunakan untuk menentukan karakteristik butir instrumen jika ukuran sampel yang digunakan besar. Sebaliknya, jika ukuran sampel kecil IRT memiliki kecenderungan kurang akurat dalam menentukan karakteristik suatu instrumen. Penelitian ini bertujuan untuk menentukan efek ukuran sampel pada GRM yang didasarkan pada data awal yang memiliki korelasi inter item yang sama secara statistik. Penelitian ini merupakan penelitian simulasi yang didasarkan pada data empiris. Data skunder dari 43 responden digunakan sebagai data awal. Akurasi efek dari ukuran sampel baru dihitung dengan menggunakan RMSE dan koelasi. Teknik membangkitkan data yang digunakan yaitu metode multivariat variabel random diskrit. perangkat lunak R. Hasil penelitian menunjukkan nilai RMSE berkisar $0,6 > 0,33$ dan korelasi sekitar $0,5 < 0,7$ tidak memenuhi standar akurasi.. Dapat disimpulkan bahwa ukuran sampel dengan matriks korelasi yang sama, tidak memiliki efek pada parameter GRM dan data baru yang dihasilkan tidak dapat digunakan untuk pemulihan parameter GRM

Kata Kunci: GRM, distribusi marginal, korelasi Inter item

The effect of sample size on GRM parameters based on the correlation between items

Abstract: Item response theory has many advantages in determining the instrument's item characteristics when the sample size used is large. On the other hand, if the sample size is small, the IRT tends to be less accurate in determining an instrument. This study aims to determine the effect of sample size on GRM based on the inter-item correlation matrix of initial data. This research was conducted on simulation research based on empirical data. Secondary data from 43 respondents was used as initial data. The effect accuracy of the new sample size was calculated using RMSE and correlation. The data generation method used was the multivariate method of discrete random variables. The results show that the RMSE value is around $0.6 > 0.33$, and the correlation of about $0.5 < 0.7$ did not meet the standard of accuracy. It could be concluded that the sample size with the same inter-item correlation matrix has no effect on the GRM parameters, and the new data generated couldn't be used for GRM parameter recovery.

Keywords: GRM, marginal distribution, inter-item correlation

How to Cite: Susanto, H., Retnawati, H., & Setiawati, F. (2022). Efek ukuran sampel pada parameter GRM yang didasarkan pada korelasi inter item. *Measurement In Educational Research (Meter)*, 2(1), 11-21. doi:<http://dx.doi.org/10.33292/meter.v2i1.171>



PENDAHULUAN

Teori respon butir memiliki banyak kelebihan ketika digunakan untuk menentukan parameter butir instrumen dan kemampuan, jika ukuran sampel yang digunakan besar. Sebaliknya, jika ukuran sampel kecil *Item Respon Theory* (IRT) memiliki kecenderungan kurang akurat dalam menentukan karakteristik suatu instrumen. Kalibrasi yang akurat akan diperoleh jika ukuran sampel yang digunakan besar (Hambleton, 1989; Kutscher et al., 2019). Sampel minimum yang dapat digunakan untuk model IRT dikotomus unidimensi yaitu 150 (Şahin & Anil, 2017), 200 sampel dapat digunakan, tetapi 1000 sampel akan memberikan hasil yang lebih akurat (Suwanto et al., 2019). Pada model polytomous *Generalized Partial Credit Model* (GPCM) 200

sampel dapat digunakan, tetapi 1000 sampel dengan 20 item akan memberikan hasil yang lebih akurat (Huang, 2016), dan 3000 sampel dengan 10 item akan memberikan hasil kalibrasi yang akurat pada *Partial Credit Model* (PCM) (Cho, 2014). Pada *Graded Response Model* (GRM) banyak sampel yang sesuai untuk digunakan yaitu 500 sampel dengan 240 item (Reise & Yu, 1990), 750 sampel (Bahry, 2012). Lebih extreme lagi dengan ukuran sample 1000 model *Multidimensional Graded Response Model* (MGRM) belum menunjukkan hasil yang akurat (Jiang et al., 2016), dan masih banyak lagi hasil penelitian yang sama yang tidak disebutkan pada artikel ini. Besarnya ukuran sampel yang harus digunakan menjadi kendala bagi para praktisi Pendidikan untuk dapat menggunakan IRT dalam melakukan kalibrasi terhadap instrumen yang dikembangkan untuk mengevaluasi pembelajaran yang dilakukan.

Beberapa peneliti mengatasi kelemahan dari model IRT agar dapat menggunakan sampel kecil dengan cara membangkitkan data untuk memulihkan estimasi parameter butir. Bahry (2012) melakukan penelitian dengan menggunakan data empiris sebanyak 100 sampel pada model GRM, selanjutnya parameter yang dihasilkan digunakan untuk membangkitkan data yang lebih besar menggunakan software MULTILOG dan diperoleh hasil yang akurat pada ukuran sampel 750. Barnes dan Wise (1991) menggunakan model satu parameter atau Rasch Model untuk kalibrasi model 3 PL dengan sampel kecil dengan software LOGIST 5. Penelitian lain yang juga memanfaatkan Rasch Model berpasangan dapat digunakan dengan *missing data* dan sampel berukuran kecil pada model 3 PL (Finch & French, 2019). Beberapa penelitian tersebut melakukan pemulihan parameter dengan kondisi awal menggunakan data empiris (Bahry, 2012) dan data bangkitan (Barnes & Wise, 1991) yang berukuran kecil. de la Torre dan Hong (2010)) menggunakan metode *Markov Chain Monte Carlo* (MCMC) untuk melakukan kalibrasi *Higher Order Iterm Respons Theory* (HO-IRT) dengan sampel kecil. Sebagian besar penelitian dalam membangkitkan data baik untuk pemulihan parameter ataupun untuk menentukan ukuran sampel yang sesuai sebagai syarat minimal menggunakan IRT kurang memperhatikan karakteristik data awal yang digunakan, yang diamati hanya hasil *output* prediksinya saja.

Pada penelitian ini mencoba melakukan pemulihan parameter GRM yang didasarkan pada korelasi inter item dengan menggunakan data empiris dan ukuran sampel kecil. Korelasi ini merupakan hubungan antar item yang mengindikasikan bahwa item-item tersebut dapat memprediksi variabel laten yang sama pada sampel dengan ukuran yang berbeda. Agar tidak terjadi pemahaman dengan multikolinieritas, maka korelasi inter item tidak boleh lebih besar dari 0.9 (Hair et al., 2010). Karakteristik korelasi inter-item data awal dengan ukuran sampel kecil digunakan untuk menentukan sampel dengan ukuran yang lebih besar dan memiliki korelasi inter item yang sama secara statistik, sehingga posisi sample baru tetap fit dengan setiap item dalam konstruk dan memiliki reliabilitas yang tidak jauh berbeda dengan data awal.

Metode pembangkitan data pada penelitian ini mengikuti prosedur pembangkitkan data multivariat variabel random diskrit dengan matriks korelasi yang diketahui dan distribusi marginal yang dikembangkan oleh Barbiero dan Ferrari (2015); Ferrari dan Barbiero (2012). Fungsi utama dari prosedur ini yaitu membangkitkan data dengan matrix korelasi dan distribusi marginal yang sudah diketahui. Pada penerapannya metode ini dapat menggunakan software R dengan package GenOrd. Mengacu pada procedure ini, peneliti memodifikasi prosedur dengan menentukan korelasi matriks inter item dari pengambilan data empiris dan dilanjutkan menentukan distribusi marginal yang sesuai dengan matriks tersebut dengan memodifikasi pada penggunaan package GenOrd. Mengacu dari beberapa informasi tersebut, maka tujuan dari penelitian ini yaitu menentukan efek dari sampel size terhadap parameter GRM yang didasarkan pada matriks korelasi inter item.

METODE

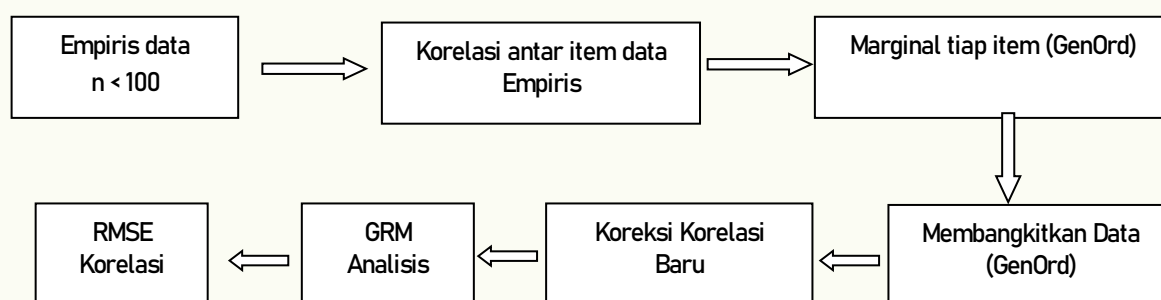
Model GRM merupakan salah satu jenis model IRT unidimensional, setiap item terdiri dari batas K di mana (K+1) adalah banyak opsi respons. Mereka mewakili probabilitas merespons dengan semua opsi di bawah batas vs. semua opsi di atas batas, merupakan fungsi-fungsi trait dalam θ . Fungsi-fungsi ini adalah fungsi logistik dua parameter dengan diskriminasi item konstan pada setiap itemnya. Fungsi-fungsi tersebut dapat dituliskan pada Persamaan 1 dan Persamaan 2.

$$P_{jk}(\theta) = P_{jk}^* - P_{j(k+1)}^*(\theta) \dots\dots\dots (1)$$

$$P_{jk}^*(\theta) = \frac{\exp(-Da_j(\theta - b_{jk}))}{1 + \exp(-Da_j(\theta - b_{jk}))} \dots\dots\dots (2)$$

Dengan $P_{j_0}^*(\theta) = 1$ dan $P_{j_{(m+1)}}^*(\theta) = 0$. a_j merupakan Indeks Daya Beda Butir. θ merupakan parameter kemampuan peserta tes. b_{jk} merupakan Indeks kesukaran kategori k butir ke j. $P_{jk}(\theta)$ Probabilitas peserta berkemampuan θ yang memperoleh skor kategori k pada butir j. $P_{jk}^*(\theta)$ merupakan Probabilitas peserta berkemampuan θ yang memperoleh skor kategori k atau lebih pada butir j. D merupakan faktor skala yang ekuivalen dengan nilai 1.7

Berdasarkan langkah-langkah dalam menggunakan package GenOrd (Barbiero & Ferrari, 2015), maka peneliti memodifikasi langkah-langkah penggunaan GenOrd untuk dapat membangkitkan data yang memiliki korelasi inter item yang hampir sama. Adapun prosedur dalam mengestimasi parameter butir dengan model GRM menggunakan ukuran sampel kecil dilakukan melalui langkah-langkah. (1) Mengambil data dengan $n < 100$ responden menggunakan instrumen dengan k item; (2) Menentukan korelasi antar item; (3) Menentukan marginal yang sesuai dengan matriks korelasi dengan menggunakan Package Genord; (4) Membangkitkan data menggunakan Package GenOrd; (5) Melakukan cek pada korelasi data hasil bangkitan; (6) Analisis GRM; dan (7) Mengevaluasi estimasi parameter dengan melihat *Root Mean Square Error* (RMSE), dan korelasi.



Gambar 1. Procedure estimasi parameter butir dan item fit

Tabel 1. Data awal

No.	X1	X2	X3	X4	Y1	Y2	Y3	Y4	Y5	Y6	No	X1	X2	X3	X4	Y1	Y2	Y3	Y4	Y5	Y6
1.	3	3	3	3	3	3	2	2	1	3	23	4	3	4	3	3	3	3	3	3	3
2.	2	3	3	4	4	3	2	3	2	2	24	3	3	4	4	3	4	4	4	1	3
3.	3	3	3	3	2	3	2	3	2	4	25	3	2	3	3	2	3	3	3	3	2
4.	4	3	4	4	3	4	3	4	2	4	26	3	3	3	3	2	3	2	3	3	3
5.	4	3	4	4	3	4	4	4	2	3	27	3	3	3	3	2	3	2	3	3	3
6.	3	3	3	3	3	3	2	3	2	3	28	4	4	2	4	3	3	2	4	2	2
7.	3	4	4	4	2	4	4	3	1	3	29	4	4	4	4	1	4	1	1	3	2
8.	4	4	4	4	4	4	1	4	4	4	30	4	4	2	4	3	3	2	4	2	2
9.	3	3	3	3	2	3	3	3	2	3	31	2	2	2	3	3	3	3	3	2	2
10.	3	3	3	3	3	3	2	3	3	3	32	3	3	4	3	3	4	3	4	2	4
11.	2	3	3	3	3	3	2	3	3	3	33	3	4	4	4	3	4	3	4	1	3
12.	3	3	3	3	2	4	2	4	2	2	34	3	3	4	3	3	3	2	3	2	3
13.	3	2	4	4	2	4	3	3	1	4	35	4	3	4	4	4	4	2	3	3	4
14.	3	2	3	3	2	3	3	3	2	3	36	3	1	3	3	3	4	3	4	2	4
15.	2	1	3	3	3	2	3	3	2	2	37	4	3	4	4	3	4	2	4	3	4
16.	3	3	3	2	2	2	3	3	2	3	38	3	3	3	3	2	3	2	3	3	3
17.	2	2	3	3	2	3	3	3	2	2	39	3	2	3	3	2	3	2	3	3	3
18.	4	4	4	4	4	4	3	4	4	3	40	4	4	4	4	2	4	3	3	4	4
19.	3	3	3	4	3	3	3	4	3	2	41	4	4	3	4	3	4	4	4	4	4
20.	2	3	3	3	3	3	2	3	3	3	42	4	3	4	4	3	4	4	4	4	3
21.	3	3	3	3	2	3	3	3	2	3	43	4	3	4	4	3	4	3	4	3	3
22.	1	3	4	3	2	3	3	3	2	3											

Langkah pertama, peneliti menggunakan data skunder dengan sampel berukuran 53. Angket yang digunakan yaitu angket efikasi diri pada olahraga yang terdiri dari 10 item dengan setiap item terdiri dari empat respon. Data tersebut dapat dilihat pada Tabel 1.

Langkah kedua, korelasi inter item dihitung. Jika terdapat nilai korelasi yang lebih dari 0.9, sangat mungkin terjadi kolineritas atau multikolinieritas (Hair et al., 2010), maka keluarkan item yang berpotensi

menyebabkan kasus ini terjadi. Berdasarkan data pada Tabel 1, maka diperoleh matriks korelasi inter item pada Tabel 2.

Table 2. Korelasi inter item

	X1	X2	X3	X4	Y1	Y2	Y3	Y4	Y5	Y6
X1	1	0.48	0.31	0.56	0.19	0.54	0.01	0.31	0.39	0.33
X2	0.48	1	0.23	0.50	0.13	0.35	-0.14	0.12	0.24	0.09
X3	0.31	0.23	1	0.41	0.11	0.63	0.24	0.09	0.11	0.51
X4	0.56	0.50	0.41	1	0.38	0.70	0.16	0.38	0.16	0.13
Y1	0.19	0.13	0.11	0.38	1	0.19	-0.01	0.51	0.17	0.15
Y2	0.54	0.35	0.63	0.70	0.19	1	0.22	0.40	0.14	0.48
Y3	0.01	-0.14	0.24	0.16	-0.01	0.22	1	0.39	-0.18	0.09
Y4	0.31	0.12	0.09	0.38	0.51	0.40	0.39	1	0.10	0.22
Y5	0.39	0.24	0.11	0.16	0.17	0.14	-0.18	0.10	1	0.16
Y6	0.33	0.09	0.51	0.13	0.15	0.48	0.09	0.22	0.16	1

Langkah ketiga. Penentuan marginal dalam penelitian ini dilakukan dengan cara yang berbeda dengan penggunaan package GenOrd ini, yaitu dengan fungsi probabilitas uniform sehingga marginal yang diperoleh dilakukan secara otomatis dapat mengikuti bentuk korelasi inter item yang diberikan. Karena kuisiонер yang digunakan menggunakan 4 respon maka bentuk marginal berupa pasangan peluang terurut dengan tiga anggota. Misalnya marginal (m) untuk item pertama $m_1 = \{\rho_1, \rho_2, \rho_3\}$, dengan $\rho_1 \leq \rho_2 \leq \rho_3$ (Ferrari & Barbiero, 2012). Banyak item yang digunakan 10, maka terdapat 10 pasang marginal. Modifikasi penulisan marginal dalam program R pada penelitian dapat dilihat pada Persamaan 3. Sintak lengkap dapat dilihat pada artikel GenOrd (Barbiero & Ferrari, 2015).

$$\text{>marginal<-lapply(y, FUN=function(x){sort(runif(i-1))})} \dots\dots\dots (3)$$

Dengan / banyak respon dari skala likert, y banyak soal. Penentuan marginal pada penelitian ini tidak semudah yang dituliskan oleh peneliti pada awal pembahasan Langkah ketiga ini, diperlukan banyak waktu untuk dapat memperoleh marginal yang sesuai. Metode penentuan marginal seperti ini dilakukan peneliti karena keterbatasan peneliti dalam menentukannya secara matematis.

Langkah keempat. Himpunan marginal dan korelasi matrik inter item telah diperoleh. Selanjutnya dilakukan pembangkitan data dengan prosedur yang dikembangkan oleh Ferrari dan Barbiero (2012). Mereka juga menyatakan bahwa metode ini merupakan sebuah cara yang berbeda dalam menentukan distribusi sampling korelasi R untuk variabel diskrit/ordinal melalui simulasi Monte Carlo. Pengkondisian pada Langkah ini dilakukan sebanyak 7 x 30 x 1 (ukuran sampel baru x replikasi x kuisiонер).

Langkah kelima, korelasi inter item pada data hasil bangkitan dicek ekuivalensinya terhadap korelasi inter item awal. Sehingga pada proses ini akan terdapat 1 x 210 (matriks koelrasi awal x matriks korelasi data baru) proses perhitungan korelasi antara matriks korelasi awal dengan matriks koelasi hasil bangkitan.

Langkah keenam, Proses yang terjadi pada estimasi parameter GRM sebanyak proses yang ada pada langkah empat. Dan **Langkah ketujuh,** estimasi parameter yang diperoleh kemudian digunakan untuk mendapatkan indeks pemulihan parameter menggunakan RMSE (Huang, 2016; Jiang et al., 2016; Kieftenbeld & Natesan, 2012). Selanjutnya untuk melihat keakuratan model digunakan korelasi (Şahin & Anıl, 2017; Swaminathan et al., 2003). Persamaan 4 menyatakan bahwa $\hat{\pi}_{Rl}$ adalah parameter ke l dari replikasi ke R dan π_i merupakan parameter ke-l dari kondisi awal.

$$RMSE = \sqrt{\left(\frac{\sum_1^n (\hat{\pi}_l - \pi_i)^2}{n}\right)} \dots\dots\dots (4)$$

HASIL DAN PEMBAHASAN

Menentukan Marginal

Proses dalam menentukan marginal dilakukan melalui beberapa percobaan dengan menjalankan sintak program yang telah dikondisikan. Dari beberapa marginal yang diperoleh, pada penelitian ini menggunakan marginal pada Tabel 3.

Tabel 3. Marginal vector

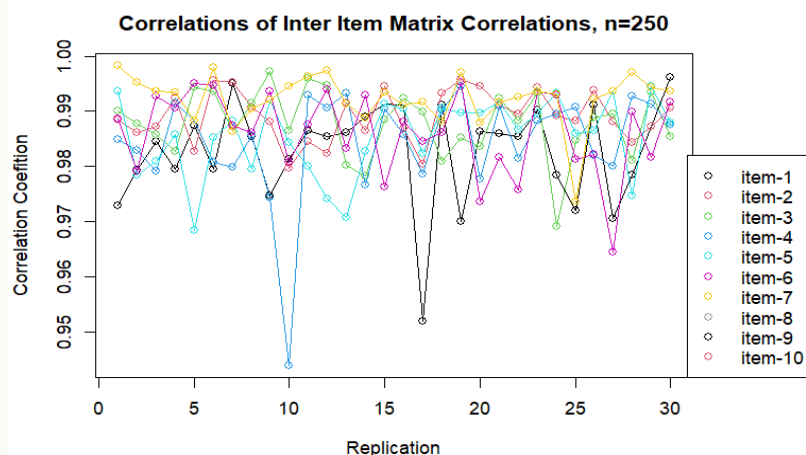
Marginal Terpilih
c(0.18 ,0.32 ,0.80),
c(0.35 ,0.45 ,0.75),
c(0.43 ,0.45 ,0.84),
c(0.28 ,0.55 ,0.67),
c(0.24 ,0.34 ,0.41),
c(0.35 ,0.56 ,0.76),
c(0.35 ,0.63 ,0.87),
c(0.30 ,0.50 ,0.76),
c(0.12 ,0.33 ,0.84),
c(0.59 ,0.69 ,0.88))

Membangkitkan data dan mengkonfirmasi korelasi antar matriks korelasi interitem

Berdasarkan matriks korelasi inter item data awal (Tabel 1) dan marginal (Tabel 2), dilakukan sampling ulang dengan ukuran sampel yang bervariasi yaitu 250, 500, 750, 1000, 1500, 2000 dan 3000 sample. Proses resampling ini dilakukan sebanyak 30 replikasi pada setiap ukuran sampel. Sehingga pada setiap ukuran sampel akan memiliki matriks korelasi inter item sebanyak 30, dan secara keseluruhan terdapat matriks korelasi inter item sebanyak 150. Setiap item pada setiap matriks korelasi inter item selanjutnya dikorelasikan dengan setiap item pada matriks korelasi awal (Tabel 2). Rerata koefisien korelasi dari matriks korelasi inter item antara data awal dan data bangkitan dapat dilihat pada Tabel 4.

Tabel 4. Rata-rata korelasi antara matriks korelasi interitem data awal dengan data bangkitan

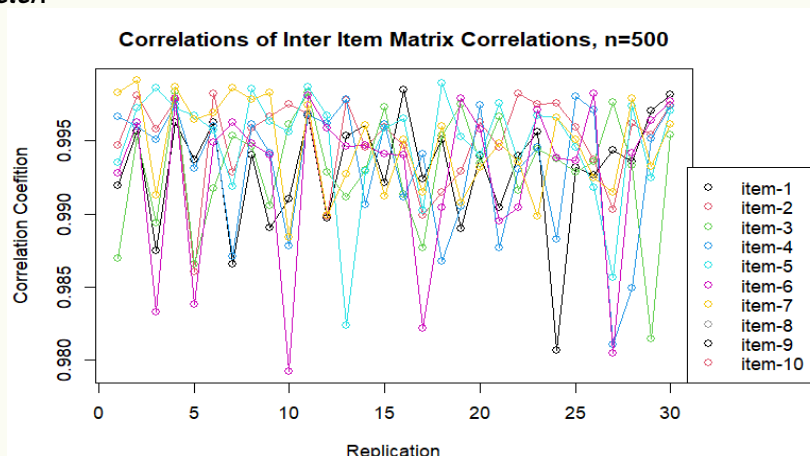
n	item 1	item 2	item 3	item 4	item 5	item 6	item 7	item 8	item 9	item 10
250	0.9828	0.9891	0.9877	0.9846	0.9854	0.9862	0.9923	0.9848	0.9886	0.9870
500	0.9930	0.9950	0.9931	0.9932	0.9949	0.9928	0.9946	0.9918	0.9945	0.9921
750	0.9955	0.9963	0.9964	0.9947	0.9956	0.9957	0.9967	0.9954	0.9964	0.9956
1000	0.9959	0.9968	0.9973	0.9962	0.9963	0.9965	0.9976	0.9954	0.9968	0.9964
1500	0.9976	0.9979	0.9978	0.9979	0.9976	0.9977	0.9982	0.9970	0.9981	0.9974
2000	0.9980	0.9986	0.9981	0.9982	0.9981	0.9977	0.9986	0.9975	0.9985	0.9976
3000	0.9988	0.9991	0.9989	0.9987	0.9990	0.9989	0.9991	0.9987	0.9991	0.9989



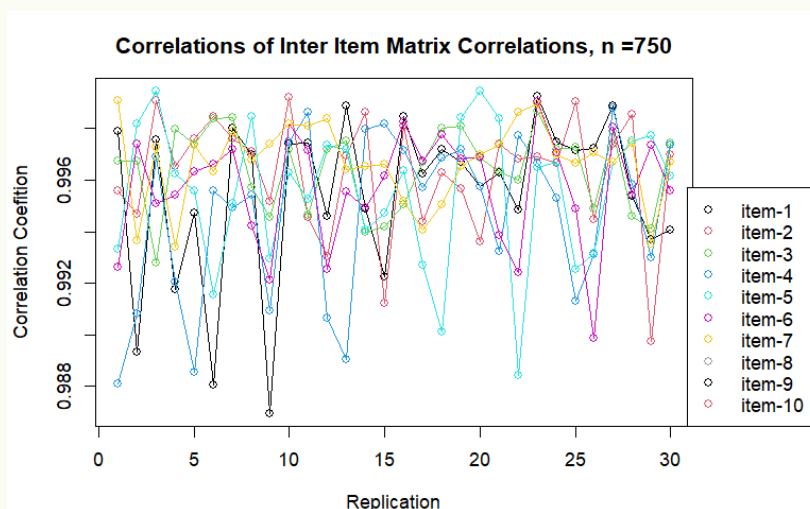
Gambar 2A. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=250

Tabel 4 menjelaskan bahwa rata-rata koefisien korelasi antara matriks korelasi inter item data awal dengan data bangkitan sangat tinggi. Hal ini mengindikasikan bahwa data bangkitan memiliki karakteristik korelasi inter item yang sama dengan data awal, sehingga dapat menggambarkan instrumen dalam mengukur laten yang sama. Visual data asli dari Tabel 4 dapat dilihat pada Gambar 2A-2G. Korelasi dari matriks inter item pada setiap itemnya untuk n=250, n=500, n=750, n=1000, n=1500, n= 2000, dan n=3000 menunjukkan bahwa koefisien terletak pada inerval 0.89 –0.999. Sesuai dengan tujuan penelitian yaitu recovery dilakukan dengan membangkitkan data yang didasarkan korelasi inter item. Seperti yang dituliskan pada Tabel 4. Setiap rata-rata korelasi item dari semua ukuran sampel dan replikasinya menunjukkan koefisiennya lebih

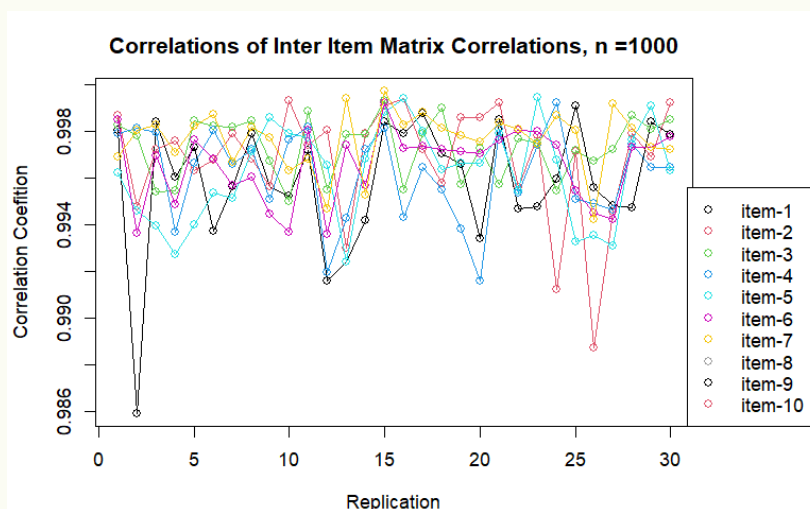
besar sama dengan $0.9 > 0.7$ (Şahin & Anil, 2017) dan >0.85 (Jiang et al., 2016), sehingga data yang dikorelasikan memiliki karakteristik yang sama dan data bangkitan dapat digunakan sebagai syarat dasar dalam *recovery parameter*.



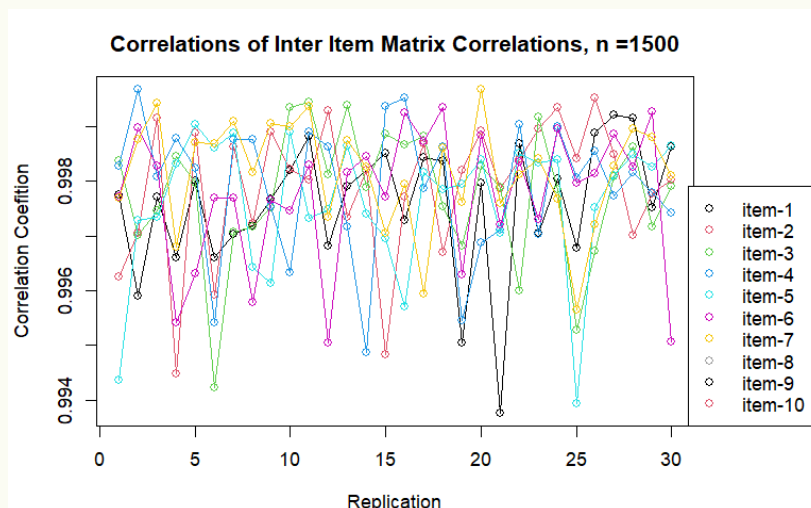
Gambar 2B. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=500



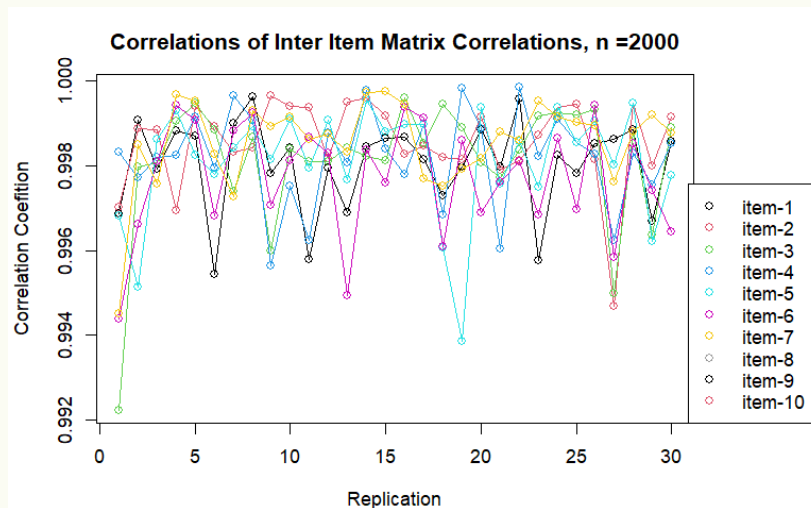
Gambar 2C. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=750



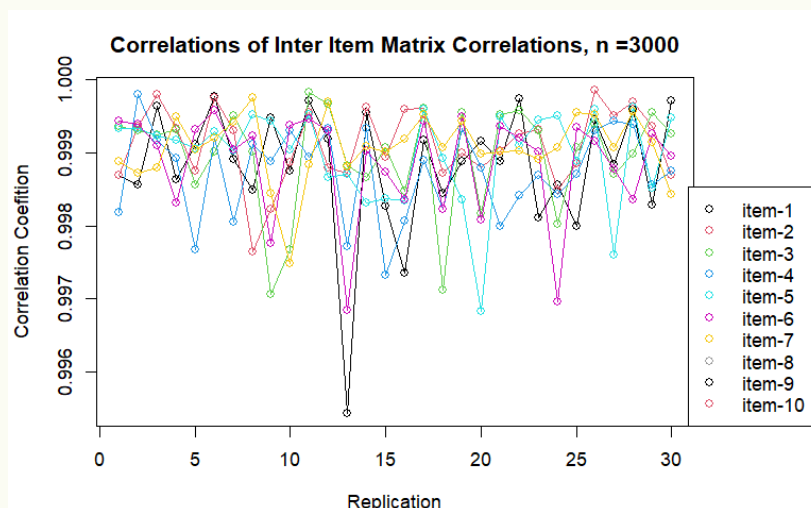
Gambar 2D. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=1000



Gambar 2E. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=1500



Gambar 2F. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=2000



Gambar 2G. Korelasi antara matriks korelasi inter item dari data awal dengan data bangkitan n=3000

Menentukan parameter butir model GRM

Data awal yang digunakan untuk menentukan parameter butir awal yang dijadikan sebagai acuan penentuan RMSE, dan korelasi parameter. Perhitungan parameter butir pada penelitian ini menggunakan

program R studio dengan Package MIRT. Adapun parameter butir yang dihasilkan dapat dilihat pada Tabel 5.

Tabel 5. Parameter butir model GRM menggunakan data awal

Parameter	item 1	item 2	item 3	item 4	item 5	item 6	item 7	item 8	item 9	item 10
a	2.25	1.35	1.61	4.21	0.87	6.23	0.3	1.56	0.67	0.9
b1	-2.39	-2.75	-2.18	-1.96	-4.67	-1.61	-10.4	-3	-3.36	-1.59
b2	-1.28	-1.47	0.22	0.09	-0.64	0.14	-0.96	-2.54	0.18	1.51
b3	0.54	1.3	-	-	2.96	-	6.89	0.41	3.2	-
Difficulty	-0.058	-0.105	-0.117	0.781	0.387	1.587	2.178	-0.698	0.775	0.272

a= parameter diskriminan, b1= *threshold* pertama parameter kesulitan, b2= *threshold* kedua parameter kesulitan, b3= *threshold* ketiga parameter kesulitan, difficulty= lokasi parameter kesulitan ditentukan dengan metode "mean"

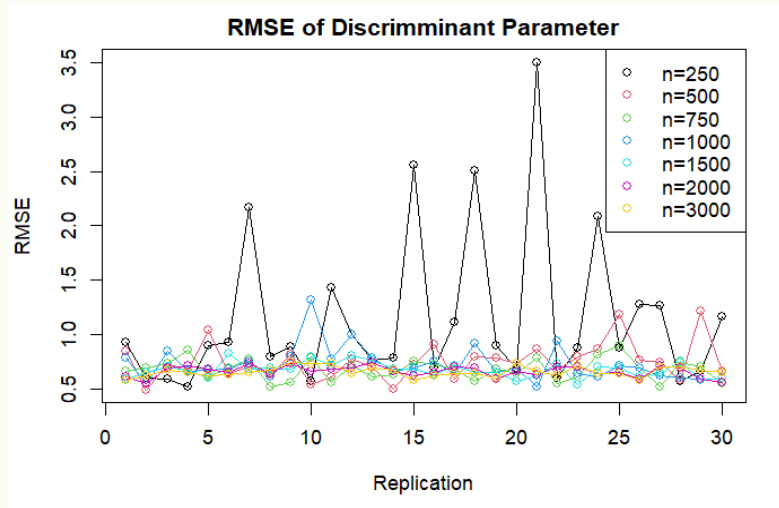
Pada Tabel 5 terlihat bahwa pada *threshold* ketiga untuk item 3, 4, 6, dan 10 tidak dapat ditentukan oleh *software* R, untuk itu pada penelitian ini *threshold* tidak digunakan dalam menentukan RMSE dan korelasi parameter. Untuk parameter kesulitan yang digunakan adalah lokasi parameter kesulitan yang ditentukan dengan metode "mean". Selain itu, parameter diskriminan juga digunakan dalam penentuan RMSE, dan korelasi parameter.

Tabel 5. Rerata RMSE dan korelasi dari diskriminan dan kesulitan

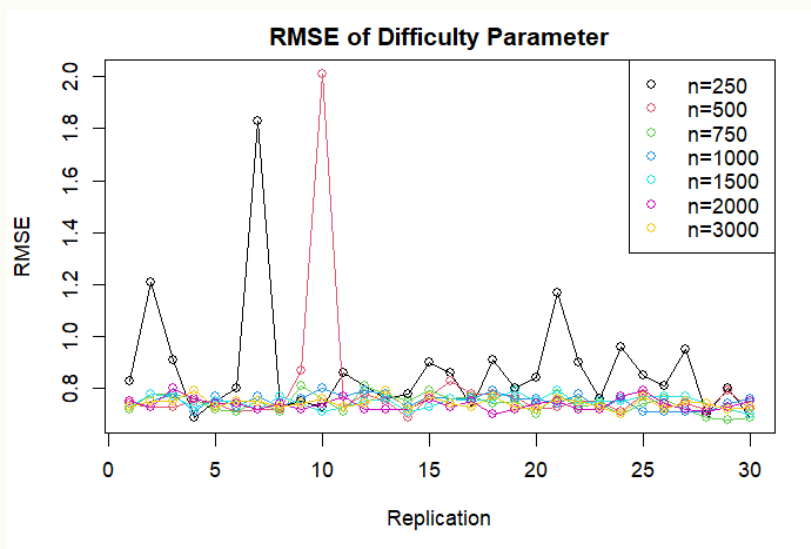
n	Diskriminan		Kesulitan	
	RMSE	Correlation	RMSE	Correlation
250	1.13926	0.93483	0.86919	0.55703
500	0.75173	0.95640	0.79214	0.58200
750	0.67493	0.96570	0.74299	0.54067
1000	0.72325	0.96033	0.75090	0.52950
1500	0.67272	0.96240	0.74783	0.52920
2000	0.65910	0.96543	0.73985	0.54560
3000	0.65509	0.96427	0.74307	0.53527

Pada Tabel 5 terlihat jelas bahwa pada parameter diskriminan dan lokasi kesulitan RMSE tidak menunjukkan perubahan yang signifikan dan Rerata Koefisien korelasi menunjukkan bahwa parameter data awal dengan data bangkitan memiliki hubungan yang tinggi, dan menunjukkan hubungan yang sedang pada parameter lokasi kesulitan. Untuk lebih jelasnya hasil perhitungan RMSE dan korelasi dapat dilihat pada Gambar 3 sampai Gambar 4.

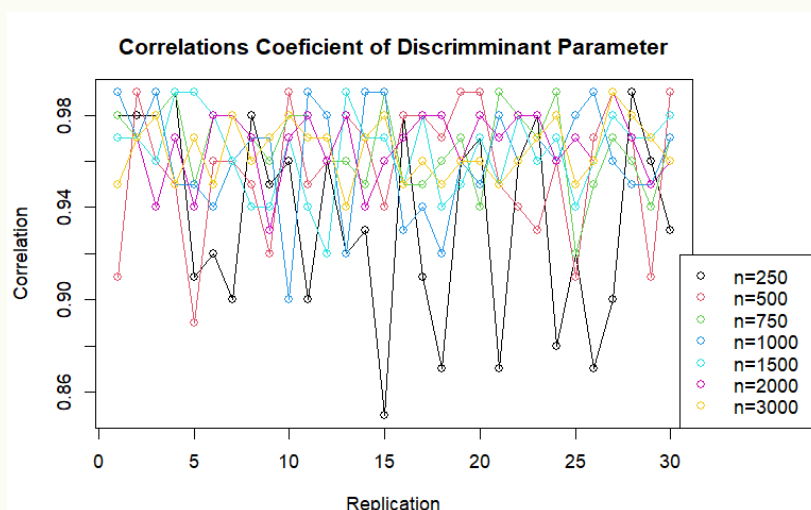
Gambar 3. menunjukkan bahwa pada parameter diskriminan dan kesulitan untuk n = 250 dan 500, memiliki nilai RMSE yang besar dan tidak stabil pada setiap replikasinya. Sebaliknya untuk n > 500 nilai RMSE pada setiap replikasi tidak menunjukkan perbedaan yang signifikan pada interval (0,5,1) untuk diskriminan dan (0, 1) untuk parameter kesulitan.



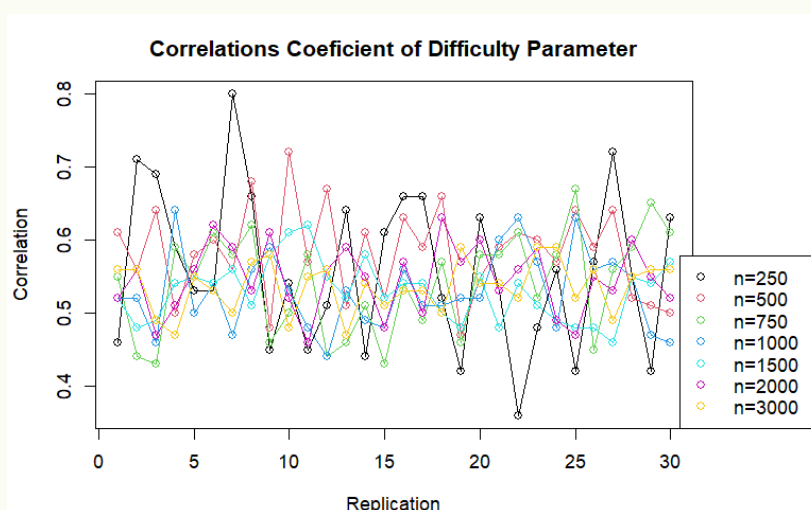
Gambar 3A. RMSE dari Parameter diskriminan



Gambar 3B. RMSE dari Parameter Kesulitan



Gambar 4A. Korelasi dari Parameter diskriminan



Gambar 4B. Korelasi dari Parameter Kesulitan

Selanjutnya korelasi setiap parameter antara data awal dan data bangkitan dapat dilihat pada Gambar 4. Hasil perhitungan korelasi untuk setiap n dengan 30 replikasi pada parameter diskriminan, menjelaskn

bahwa setiap nilai korelasi yang diperoleh ada pada interval (0.9,1), kecuali untuk $n = 250$. Sedangkan pada parameter kesulitan koefisien korelasi stabil pada interval (0.4, 0.7), kecuali untuk $n = 250$.

Karakteristik matriks yang sama ini, akan mempertahankan nilai reliabilitas dari data baru yang dihasilkan dari resampling atau data bangkitan, dan inilah kelebihan dari package GenOrd. Selain itu karakteristik ini, akan mempertahankan data yang sesuai dengan konstruk dari item yang digunakan. tetapi cocok dengan konstruk belum tentu data akan fit untuk bisa menjelaskan konstruk yang akan diukur. Untuk itu dalam proses membangkitkan data dengan tujuan kalibrasi, sebisa mungkin unsur validitas konstruk dan reliabilitas dapat dipenuhi, seperti pada saran dari penelitian yang dilakukan oleh Şahin dan Anıl (2017).

Tabel 3 merupakan hasil dari Langkah ketiga dari prosedur yang digunakan oleh peneliti. Nilai-nilai marginal pada Tabel 3 merupakan salah satu dari banyak nilai marginal yang dapat dihasilkan oleh package GenOrd. Selanjutnya pada Tabel 4 merupakan hasil dari langkah keempat. Hasil pada Tabel 4 merupakan salah satu bentuk korelasi inter-item yang dihasilkan. Hasil Random yang dapat dihasilkan oleh langkah keempat, untuk setiap hasil random hasil langkah ketiga menjadi kendala untuk peneliti. Sehingga peneliti hanya menggunakan salah satu hasil dari langkah ketiga yaitu pada Tabel 3 dan menggunakan salah satu korelasi inter-item yang dapat dihasilkan yaitu pada Tabel 4. Pembatasan banyaknyanya kemungkinan yang dapat terjadi ini menjadi kelemahan dari simulasi yang digunakan.

Akurasi estimasi parameter diskriminan dan kesulitan yang hanya didasarkan pada Tabel 3 dan Tabel 4 tidak dapat ditentukan oleh RMSE. Nilai RMSE tidak semakin kecil Ketika ukuran sampel semakin besar. Hasil ini menunjukkan bahwa semakin besar replikasi dan banyak sampel yang digunakan tidak menunjukkan perubahan nilai dari parameter diskriminan. selanjutnya, setiap nilai rata-rata RMSE dari setiap ukuran sampel > 0.33 (Şahin & Anıl, 2017; Yen, 1987). hasil ini menunjukkan bahwa ukuran sampel 250, 500, 750, 1000, 1500, 2000 dan 3000 dengan karakteristik korelasi inter item yang secara statistik sama, tidak dapat mewakili sampel awal dalam melakukan recovery parameter pada GRM.

Korelasi antara parameter diskriminan data awal dan data bangkitan memiliki koefisien korelasi lebih besar sama dengan $0.9 > 0.7$ (Şahin & Anıl, 2017). Akurasi dengan koefisien korelasi menunjukkan estimasi parameter diskriminan tidak jauh berbeda dengan parameter diskriminan yang menjadi acuan. Ukuran sampel yang dibangkitkan dapat digunakan untuk recovery parameter diskriminan. sebaliknya pada parameter kesulitan nilai seluruh korelasi berkisar $0.5 < 0.7$, sehingga berdasarkan akurasi korelasi, ukuran sampel yang didasarkan pada matriks korelasi inter item tidak dapat digunakan untuk melakukan recovery parameter kesulitan.

Hasil dalam penelitian ini tidak sesuai dengan hasil penelitian (Reise & Yu, 1990) bahwa dengan sampel 500 sudah dapat digunakan untuk melakukan kalibrasi GRM dengan akurat walaupun tanpa dibatasi dengan korelasi antar item. Sarstedt dan Mooi (2019) menyatakan bahwa GRM dapat akurat hanya dengan menggunakan sampel 150.

SIMPULAN

Ukuran sampel yang lebih besar dengan korelasi inter item yang sama secara statistik dengan data sampel awal menunjukkan efek yang berbeda dengan karakteristik parameter data awal. Hal ini berakibat bahwa ukuran sampel yang lebih besar dengan korelasi inter item yang disyaratkan tidak dapat digunakan untuk melakukan *recovery parameter*.

Hasil yang diperoleh mungkin kurang sesuai espektasi seperti pada penelitian-penelitian yang sudah ada. Tidak berhenti disini, pada penelitian selanjutnya akan dilakukan penelitian dengan pembatasan kondisi yang lebih ketat. (1) memperbanyak simulasi yang dapat dihasilkan oleh langkah ketiga dan keempat pada prosedur yang peneliti gunakan; (2) Data bangkitan yang diperoleh diharuskan memenuhi kondisi yang validitas konstruk dengan menggunakan uji CFA.

Kontribusi artikel terhadap bidang studi terkait

Pada penelitian simulasi ini, pemilihan parameter-parameter pada GRM dengan metode yang digunakan belum mampu mengatasi penggunaan ukuran sampel kecil pada analisis IRT. Tetapi, setidaknya studi ini telah mencoba untuk berkontribusi dalam mengatasi kelemahan IRT. Masih banyak kemungkinan yang belum dicoba terkait metode yang digunakan, khususnya pada objek yang digunakan dalam penentuan akurasi menggunakan RMSE.

DAFTAR PUSTAKA

- Bahry, M. L. (2012). *Polytomous item response theory parameter recovery: An investigation of non-normal distributions and small sample size* [University of Alberta]. <https://doi.org/10.7939/R36H8X>
- Barbiero, A., & Ferrari, P. A. (2015). *GenOrd: Simulation of discrete random variables with given correlation matrix and marginal distributions*. <https://rdrr.io/cran/GenOrd/>
- Barnes, L. L. B., & Wise, S. L. (1991). The utility of a modified one-parameter IRT model with small samples. *Applied Measurement in Education, 4*(2), 143–157. https://doi.org/10.1207/s15324818ame0402_4
- Cho, Y. (2014). *The mixture distribution polytomous rasch model used to account for response styles on rating scales: A simulation study of parameter recovery and classification accuracy*. [University of Maryland]. <http://hdl.handle.net/1903/14511>
- de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size a higher-order IRT model approach. *Applied Psychological Measurement, 34*(4), 267–285. <https://doi.org/10.1177/0146621608329501>
- Ferrari, P. A., & Barbiero, A. (2012). Simulating ordinal data. *Multivariate Behavioral Research, 47*(4), 566–589. <https://doi.org/10.1080/00273171.2012.692630>
- Finch, H., & French, B. F. (2019). A comparison of estimation techniques for IRT models with small samples. *Applied Measurement in Education, 32*(2), 77–96. <https://doi.org/10.1080/08957347.2019.1577243>
- Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Multivariate data analysis: A Global Perspective* (7th ed.). Prentice Hall.
- Hambleton, R. K. (1989). *Principles and selected applications of item response theory* (R. L. Linn (ed.); 3rd ed.). Macmillan.
- Huang, H.-Y. (2016). Mixture random-effect IRT models for controlling extreme response style on rating scales. *Frontiers in Psychology, 7*. <https://doi.org/10.3389/fpsyg.2016.01706>
- Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology, 7*. <https://doi.org/10.3389/fpsyg.2016.00109>
- Kieftenbeld, V., & Natesan, P. (2012). Recovery of graded response model parameters. *Applied Psychological Measurement, 36*(5), 399–419. <https://doi.org/10.1177/0146621612446170>
- Kutscher, T., Eid, M., & Crayen, C. (2019). Sample size requirements for applying mixed polytomous item response models: Results of a Monte Carlo simulation study. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.02494>
- Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27*(2). <https://doi.org/10.1111/j.1745-3984.1990.tb00738.x>
- Şahin, A., & Anil, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice*. <https://doi.org/10.12738/estp.2017.1.0270>
- Sarstedt, M., & Mooi, E. (2014). A concise guide to market research. In *A Concise Guide to Market Research*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-53965-7>
- Suwarto, S., Widoyoko, E. P., & Setiawan, B. (2019). The effects of sample size and logistic models on item parameter estimation. *Proceedings of the 2nd International Conference on Education, 323–330*. <https://doi.org/10.4108/eai.28-9-2019.2291082>
- Swaminathan, H., Hambleton, R. K., Sireci, S. G., Xing, D., & Rizavi, S. M. (2003). Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied Psychological Measurement, 27*(1), 27–51. <https://doi.org/10.1177/0146621602239475>
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika, 52*(2), 275–291. <https://doi.org/10.1007/BF02294241>