

UKURAN SAMPEL DAN PENGARUHNYA TERHADAP BIAS ESTIMASI PARAMETER ITEM DALAM RASCH MODEL

Tina Deviana ^{*1}
Rahmawati ²

¹ Program Studi Psikologi, Universitas Paramadina¹

² Balai Pengelolaan Pengujian Pendidikan (BPPP) Kemdikbud

*e-mail: tina.deviana@paramadina.ac.id¹

Abstrak

Ukuran sampel yang tidak memenuhi syarat dalam penerapan Rasch model seringkali digunakan, padahal terdapat konsekuensi yang terjadi jika ukuran sampel bermasalah. Tujuan penelitian ini adalah untuk mengetahui pengaruh ukuran sampel terhadap bias estimasi parameter item dari Rasch model. Metode yang akan digunakan adalah simulasi monte carlo dengan kondisi ukuran sampel: 50, 100, 200 dan 500, serta replikasi sebanyak 1000 kali. Hasil penelitian ini menunjukkan bahwa ukuran sampel memberi pengaruh terhadap bias estimasi parameter dalam Rasch model (de Ayala, 2009; Linacre, 2018). Semakin kecil ukuran sampel, bias yang terjadi akan semakin besar. Saran dan implikasi untuk penelitian mendatang turut didiskusikan.

Kata kunci: Rasch model, ukuran sampel, bias estimasi, parameter item, simulasi Monte Carlo

Abstract

Inadequate sample sizes in the application of the Rasch model are often used, even though there are consequences that occur if the sample size is problematic. The purpose of this study was to determine the effect of sample size on the bias of item parameter estimates from the Rasch model. The method used is Monte Carlo simulation with sample size conditions: 50, 100, 200 and 500, and 1000 replications. The results of this study indicate that sample size affects the bias of parameter estimates in the Rasch model (de Ayala, 2009; Linacre, 2018). The smaller the sample size, the greater the bias that occurs. Suggestions and implications for future research are also discussed.

Keywords: Rasch model, sample size, estimation bias, item parameter, Monte Carlo simulation

PENDAHULUAN

Ketika Andrich (2011) mengemukakan bahwa terdapat tiga teori tes: (1) CTT, (2) IRT dan (3) Rasch measurement model, banyak pihak khususnya di Indonesia yang belum memahami hal tersebut. Umumnya Rasch model masih dianggap sebagai kasus khusus dari IRT yang sama dengan model IRT 1-PL. Padahal, terdapat perbedaan filosofis yang besar dan membatasi model pengukuran Rasch dengan model IRT (Linacre, 2018). Perbedaan filosofis ini terkadang tidak dipahami khususnya pada penelitian di Indonesia, diantaranya adalah kesalahan memandang model pengukuran Rasch hanya sebagai metode analisis data tanpa memahami batasan dengan IRT sebagai model statistik serta penggunaan ukuran sampel yang tidak representatif (Putra, Suryadi, & Hayati, 2018).

Ukuran sampel adalah faktor penting dalam penerapan latent trait theory termasuk Rasch model sebagai model pengukuran di dalamnya. Telah banyak studi yang membahas ukuran sampel minimum dalam penerapan latent trait theory seperti ukuran sampel minimal 200 (Drasgow, 1989), 265 (Muthen & Muthen, 2002) dan 500 (Wright, 1977). Namun, terdapat hal yang mengkhawatirkan ketika melihat penerapan Rasch model di Indonesia masih menggunakan ukuran sampel kecil seperti 50 dan besaran lainnya dibawah 100. Padahal, riset telah membuktikan bahwa penerapan IRT maupun Rasch model pada sampel kecil membutuhkan penanganan khusus (lihat, Setiadi, 1997), apabila tidak ditangani maka akan mengalami permasalahan berupa bias sehingga hasil estimasi sebenarnya tidak tepat dan mengalami kesalahan sistematis (Andrich, 2010), penganganan khusus inilah yang seakan diabaikan dalam riset-riset aplikasi Rasch model di Indonesia.

Oleh karena itu, penelitian ini akan membahas pengaruh ukuran sampel terhadap bias estimasi parameter dalam Rasch model. Adapun bias estimasi merupakan ukuran dari kesalahan

sistematis dalam perhitungan yang dilakukan oleh software ketika ukuran sampel divariasikan dari yang representatif hingga yang tidak representatif. Penelitian ini bertujuan untuk memberikan gambaran terhadap apa yang terjadi ketika penerapan Rasch model tidak diikuti dengan tersedianya ukuran sampel yang representatif.

METODE

1. Kondisi

Dalam penelitian ini, akan diuji berbagai kondisi berupa ukuran sampel (50, 100, 200, 500) dalam pengaruhnya terhadap bias hasil estimasi parameter item dalam Rasch model. Kondisi ukuran sampel 50 dan 100 merupakan gambaran dari ukuran sampel yang tidak representatif (Setiadi, 1997), ukuran sampel 200 merupakan kriteria minimum penerapan Rasch model (Drasgow, 1989) sedangkan ukuran sampel 500 merupakan ukuran sampel yang sangat optimal (Wright, 1977). Metode yang digunakan adalah simulasi monte carlo dengan program Mplus. Jumlah item tidak memiliki variasi kondisi yaitu seluruh kondisi sampel akan menghasilkan estimasi parameter item sebanyak 20 item. Masing-masing kondisi ukuran sampel akan direplikasi sebanyak 1000 kali. Dalam penelitian simulasi, dibutuhkan distribusi prior untuk membangkitkan data. Adapun dalam penelitian ini, distribusi prior yang digunakan adalah distribusi normal (0,1) untuk kemampuan responden serta uniform (2,2) untuk distribusi tingkat kesukaran item yang dibangkitkan dengan program WinGen.

Data akan dibangkitkan sesuai dengan model Rasch dikotomus dimana seluruh daya pembeda untuk seluruh item sebesar 1. Model populasi yang dianalisis juga telah memenuhi asumsi-asumsi dari Rasch model yaitu unidimensionalitas, local independence dan monotonicity. Hal ini dilakukan agar ukuran sampel dapat diuji pengaruhnya terhadap bias, dimana kemungkinan bias terjadi akibat pelanggaran asumsi telah dihindari.

2. Dichotomous Rasch model

Karena dalam penelitian ini data yang dibangkitkan berbentuk dikotomi, maka model yang digunakan adalah dichotomous rasch model. Model ini sama dengan bentuk paling sederhana dari model logistik yaitu:

$$p(x) = \frac{e^z}{1 + e^z}$$

Dimana $p(x)$ adalah probabilitas untuk mendapatkan skor 1 saat prediktor memperhitungkan nilai dari x , sedangkan e adalah konstan yang nilainya sebesar 2.7183 dan z adalah kombinasi linier dari variabel prediktor dan konstan. Dengan melakukan spesifikasi pada z dengan benar, kita akan dapat memodelkan prediksi untuk '*response behavior*' pada item j (de Ayala, 2009).

Dalam menspesifikasikan z , kita dapat kembali kepada ide mengenai perbedaan jarak antara lokasi penempuh tes dan lokasi item (yaitu, $(\theta - \delta_j)$) yang merupakan hal penting yang mempengaruhi probabilitas dari respon seseorang tersebut. Apabila kita menentukan bahwa $z = (\theta - \delta_j)$ maka akan dihasilkan model dimana kita dapat memprediksi probabilitas menjawab benar (skor 1) sebagai fungsi dari lokasi item dan penempuh tes. Dengan melakukan substitusi z kedalam model logistik, maka didapat (de Ayala, 2009):

$$P(x_j = 1 | \theta, \delta_j) = \frac{e^{(\theta - \delta_j)}}{1 + e^{(\theta - \delta_j)}}$$

Dimana $P(x_j = 1 | \theta, \delta_j)$ merupakan probabilitas untuk menjawab benar (skor 1), θ adalah lokasi penempuh tes dan δ_j adalah lokasi item ke- j . Model ini dinamakan *Rasch model* (Rasch, 1960, dalam de Ayala, 2009). Persamaan di atas menyatakan bahwa probabilitas untuk dapat menjawab benar pada sebuah item ke- j (mendapatkan skor 1) merupakan fungsi dari perbedaan jarak antara lokasi orang θ dan lokasi item δ .

3. Kriteria bias estimasi parameter

Menurut kamus American Psychological Association (2015) bias estimasi terjadi apabila nilai yang diperoleh dari data sampel yang secara konsisten underestimate atau overestimate dari nilai sebenarnya dalam populasi berjumlah besar yang diteliti. Dengan kata lain, sebuah estimator dikatakan bias ketika rata-rata nilainya berbeda dari nilai parameter yang dimaksudkan untuk mewakili nilai yang sebenarnya. Hal ini disebut juga sebagai bias statistik. Marsh, Hau, & Greyson (2005) menjelaskan bila nilai yang diharapkan dari statistik yang dilakukan pada sampel bervariasi secara sistematis dengan jumlah N, maka statistik yang dihasilkan akan menghasilkan bias dari parameter populasi sebenarnya.

Terdapat beberapa cara untuk melihat bias pada estimasi parameter dan *standard error* dari parameter tersebut, seperti *relative difference* atau *standardized difference* (Collins, Schafer & Kam, 2001; Hoogland & Boomsma, 1998, dalam Putra, Umar, Hayat, & Utomo, 2017). Umumnya, bias merupakan ukuran dari *systematic error*, SE merupakan ukuran dari *random error* dan MSE merupakan ukuran dari *total error* (Feinberg & Rubright, 2016). Dalam penelitian ini, bias estimasi parameter item dari Rasch model dihitung dengan rumus bias absolut yaitu (Pornprasertmanit, Lee, & Preacher, 2014):

$$\text{Absolute Difference } (\theta) = \bar{\theta}_{est} - \theta_{true}$$

Dimana $\bar{\theta}_{est}$ merupakan rata-rata dari estimasi parameter pada tiap kondisi dan θ_{true} adalah nilai populasi dari parameter tersebut. Jika perbedaan antara hasil estimasi dan true value secara keseluruhan melebihi 0.10, maka telah terjadi bias pada parameter item (Muthen, Kaplan, & Hollis, 1987). Kriteria itulah yang digunakan dalam penelitian ini.

HASIL DAN PEMBAHASAN

Setelah simulasi dilakukan berdasarkan kondisi-kondisi yang telah dipaparkan sebelumnya, didapat hasil estimasi parameter item sebagai berikut:

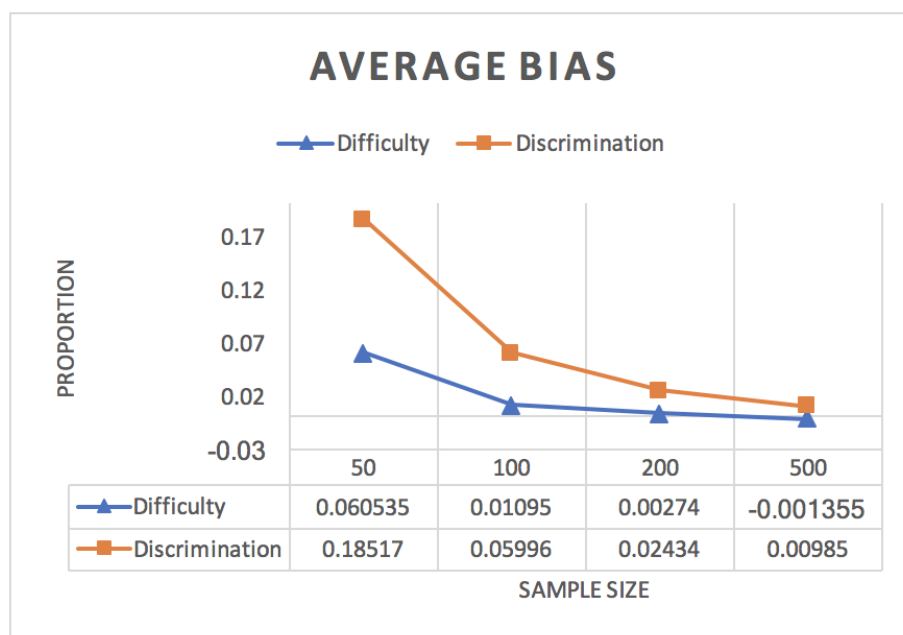
Tabel 1 Estimasi parameter item

Item	True Value		N = 50		N = 100		N = 200		N = 500	
	a	b	a	b	a	b	a	b	a	b
1	1.000	1.621	1.2640	1.9027	1.0811	1.7054	1.0272	1.6554	1.0017	1.6274
2	1.000	-0.755	1.1518	-0.8392	1.0652	-0.7877	1.0143	-0.7716	1.0139	-0.7653
3	1.000	-1.361	1.1983	-1.5463	1.0447	-1.423	1.0076	-1.384	1.0031	-1.3749
4	1.000	0.608	1.1537	0.6845	1.0410	0.6258	1.0152	0.6125	1.0085	0.6111
5	1.000	0.774	1.1288	0.8427	1.0615	0.8085	1.0292	0.7876	1.0101	0.7738
6	1.000	0.095	1.1409	0.1088	1.0715	0.1088	1.0272	0.1014	1.0128	0.0951
7	1.000	1.586	1.2429	1.8328	1.0717	1.6538	1.0400	1.6141	1.0173	1.5913
8	1.000	0.006	1.1359	0.0131	1.0617	0.0042	1.0289	0.0112	1.0044	0.0054
9	1.000	0.366	1.1489	0.4123	1.0691	0.3864	1.0320	0.3691	1.0196	0.3721
10	1.000	-0.393	1.1764	-0.4214	1.0539	-0.4081	1.0261	-0.4058	1.0069	-0.4014
11	1.000	1.702	1.2565	2.0149	1.0694	1.7835	1.0305	1.7365	1.0166	1.714
12	1.000	-1.164	1.2236	-1.3514	1.0758	-1.2372	1.0258	-1.1985	1.0115	-1.1783
13	1.000	-1.158	1.1538	-1.2931	1.0481	-1.2100	1.0156	-1.1873	1.0045	-1.1723
14	1.000	1.336	1.2142	1.5118	1.0702	1.3882	1.0439	1.3627	1.0169	1.3417
15	1.000	1.073	1.121	1.1815	1.0496	1.1156	1.0226	1.0844	1.0064	1.0706
16	1.000	-0.515	1.1092	-0.5551	1.0476	-0.5341	1.0166	-0.5245	1.0124	-0.5192
17	1.000	-0.118	1.1242	-0.147	1.0507	-0.1398	1.0184	-0.1315	1.0103	-0.1209
18	1.000	-0.597	1.1569	-0.678	1.0643	-0.6280	1.0289	-0.6145	1.0043	-0.6061
19	1.000	0.736	1.1973	0.8278	1.0482	0.7623	1.0165	0.7422	1.0073	0.7364
20	1.000	1.982	1.4051	2.5333	1.0539	2.0684	1.0203	2.0194	1.0085	1.9964

Keterangan: a = *average item discrimination*, b = *average item difficulty*

Seperti dapat dilihat pada Tabel 1, meskipun true value dari discrimination dan tingkat kesukaran masing-masing item telah diketahui sebelumnya, ketika ukuran sampel dirubah, akan dihasilkan estimasi parameter dengan angka yang berbeda. Padahal, jika tidak terjadi 'kesalahan' akibat perbedaan ukuran sampel, angka yang dihasilkan harusnya sama. Sekilas dapat terlihat bahwa terdapat pola yang konsisten dimana semakin kecil ukuran sampel, perbedaan dengan true

value semakin besar. Perbedaan akan menurun seiring ukuran sampel yang semakin besar. Pada grafik di bawah ini akan dipaparkan bias dari masing-masing ukuran sampel (lihat Gambar 1).



Gambar 1. Average bias pada masing-masing kondisi sampel

Seperti dapat dilihat pada gambar di atas, dengan kriteria average bias > 0.10 , didapat hasil bahwa ukuran sampel sebesar 50 tidaklah cukup untuk dilakukan penerapan Rasch model. Hal ini terlihat dari besaran bias sebesar $0.18 > 0.10$ pada daya pembeda. Secara keseluruhan, semakin meningkatnya ukuran sampel, bias akan semakin menurun. Hal ini berlaku untuk parameter tingkat kesukaran dan juga daya pembeda. Ketika pada penelitian mendatang ukuran sampel yang dibutuhkan tidaklah mencukupi (misal, $N = 50$), maka sebelum dilakukan pelaporan hasil analisis Rasch, perlu dilakukan koreksi terhadap bias terlebih dahulu agar estimasi parameter yang dihasilkan dapat dipercaya.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dipaparkan, dapat disimpulkan bahwa terdapat pengaruh dari ukuran sampel terhadap bias estimasi parameter Rasch model. Bias akan terjadi lebih besar pada sampel yang semakin kecil. Rasa-rasanya, kriteria ukuran sampel minimum sebesar 200 (Drasgow, 1989) bukanlah hal yang dengan mudah diabaikan. Para peneliti sebaiknya mengacu pada kriteria minimum tersebut untuk mendapatkan hasil estimasi yang tidak mengandung kesalahan sistematis. Penelitian ini juga masih memiliki keterbatasan, dimana bias menjadi satu-satunya kriteria yang digunakan meskipun terdapat kriteria lain yang dapat digunakan begitupun faktor-faktor lain yang perlu diuji. Penelitian mendatang diharapkan dapat mengatasi keterbatasan ini.

DAFTAR PUSTAKA

- American Psychological Association. (2015). *APA dictionary of psychology* (2nd ed.). Washington, DC: American Psychological Association
- Andrich, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous rasch model. *Psychometrika*, 75(2), 292-308
- Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research*, 11(5), 571-585
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-

- parameter model. *Applied Psychological Measurement*, 13, 77–90.
- Feinberg, R. A. & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49
- Linacre, J. M. (2018). *Winsteps® Rasch measurement computer program user's guide*. Beaverton, OR: Winsteps.com
- Marsh, H. W., Hau, K-T & Grayson, D. (2005). Goodness of Fit Evaluation in Structural Equation Modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary Psychometrics. A Festschrift for Roderick P. McDonald* (pp. 275-340). Mahwah, NJ: Erlbaum.
- Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52, 431–462.
- Muthen, L. K. & Muthen, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599–620
- Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: some consequences for model fit and standardized parameter estimates. *Multivariate Behavioral Research*, 49, 518-543
- Putra, M. D. K., Suryadi, B., & Hayat, B. (2018). *What we can learn from Indonesian scholars about the Rasch model*. The paper is presented at The 2nd International Conference on Educational Research and Evaluation (ICERE), Jakarta, Indonesia, November 23-24
- Putra, M. D. K., Umar, J., Hayat, B., & Utomo, A. P. (2017). Pengaruh ukuran sampel dan intraclass correlation coefficients (ICC) terhadap bias estimasi parameter multilevel latent variable modeling: studi dengan simulasi monte carlo, *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(1), 34-50
- Setiadi, H. (1997). *Small sample IRT item parameter estimates*. Doctoral dissertation, University of Massachusetts Amherst
- Wright, B. D. (1977). Misunderstanding the Rasch model. *Journal of Educational Measurement*, 14(3), 219-225