

KESTABILAN ESTIMASI PARAMETER KEMAMPUAN PADA MODEL LOGISTIK *ITEM RESPONSE THEORY* DITINJAU DARI PANJANG TES

Ilham Falani¹, Siti Ayu Kumala²

Program Studi Teknik Industri, Universitas Indraprasta PGRI Jakarta^{1,2}

E-mail: ilhamfalani@gmail.com¹

Abstrak

Penelitian ini bertujuan untuk menguji kestabilan estimasi parameter kemampuan pada Model Parameter Logistik *Item Response Theory* (IRT) berdasarkan panjang tes yang digunakan. Model Parameter Logistik yang digunakan dalam penelitian ini adalah Model 2 Parameter Logistik (2PL). Data yang digunakan adalah data simulasi yang dibangkitkan dengan menggunakan *software* Wingen, dengan ukuran sampel 1000 partisipan dengan variasi panjang tes yang digunakan adalah 20, 40, dan 60. Selanjutnya estimasi dilakukan untuk masing-masing panjang tes tersebut, estimasi tersebut dilakukan dengan menggunakan *software* Wingen dengan replikasi sebanyak 10 kali. Berdasarkan hasil analisis pada hasil estimasi yang telah dilakukan dapat diketahui bahwa semakin panjang tes yang digunakan, estimasi parameter kemampuan model 2PL semakin akurat.

Kata Kunci: Estimasi kemampuan, panjang tes, model logistik, IRT

Abstract

This study aims to test the stability of ability parameter estimation of Logistic Parameter Item Response Theory (IRT) model based on the length of test used. Logistic Parameter Model used in this research is 2 Logistic Parameter (2PL) Model. The data used are simulation data generated by using Wingen software, with sample size of 1000 participants with variation of test length used is 20, 40, and 60. The estimation is then made for each length of the test by using Wingen software with replications for 10 times. Based on the results of the analysis on the estimation results, it can be seen that the longer the test is used, the more accurate estimation of ability parameter of 2PL model is.

Keywords: Ability Estimation, Length of Test, Logistic Model, IRT

PENDAHULUAN

Pendekatan IRT berupaya mengatasi kelemahan-kelemahan pendekatan klasik seperti *item dependent*, *sample dependent*, *test oriented*, dan pemberlakuan *measurement error* yang sama untuk semua peserta tes [1].

Keunggulan IRT dikenal dengan parameter karakteristik item dan kemampuan yang *invariant* (invariance property), yaitu karakteristik item (atau tingkat kesulitan soal) yang tidak bergantung pada kelompok peserta tes yang berasal dari populasi yang sama [1]. Demikian pula sebaliknya, estimasi kemampuan peserta tidak tergantung pada karakteristik tes yang diberikan. Sehingga dapat dilakukan perbandingan antar individu penempuh tes serta perbandingan antar item tes.

Pada teori response butir, setiap butir atribut dapat dimintakan tanggapannya dari banyak subjek atau responden. Berdasarkan tanggapan tersebut, parameter kemampuan subyek atau responden serta parameter karakteristik butir atribut dapat diestimasi. Keakuratan dalam estimasi sangat diperlukan, hal ini bertujuan agar kemampuan dan karakteristik parameter item dapat memberikan gambaran yang tepat terkait tingkat kemampuan dan karakteristik sebuah item. Oleh karena itu, faktor-faktor yang mempengaruhi keakuratan estimasi perlu diperhatikan, salah satunya adalah panjang tes yang digunakan, perlu dipertimbangkan, dalam mendukung keakuratan dalam estimasi [3].

Penelitian ini bertujuan untuk menguji kestabilan estimasi parameter kemampuan pada Model Parameter Logistik *Item Response Theory* (IRT) berdasarkan panjang tes yang digunakan. Sehingga dapat diketahui keterkaitan panjang tes terhadap akurasi estimasi. Hal ini penting dilakukan dalam upaya untuk terus meningkatkan akurasi estimasi parameter kemampuan.

Georg Rasch mengembangkan satu model analisis dari teori respon butir (atau *Item Response Theory*, IRT) pada tahun 1960-an biasa disebut 1PL (satu parameter logistic). Model matematika ini kemudian dipopulerkan oleh Ben Wright. Dengan data mentah berupa data dikotomi (berbentuk benar dan salah) yang mengindikasikan kemampuan siswa, Rasch memformulasikan hal ini menjadi satu model yang menghubungkan antara siswa dan aitem [4].

Sebagai ilustrasi, seorang siswa yang mampu mengerjakan 80% soal dengan benar tentu mempunyai abilitas yang lebih baik dari siswa lain yang hanya bisa mengerjakan 65% soal. Data tersebut (persentase) menunjukkan bahwa data mentah yang diperoleh tidak lain adalah jenis data ordinal yang menunjukkan peringkat dan tidak linier. Oleh karena data ordinal tidak mempunyai interval yang sama, maka data tersebut perlu diubah menjadi data rasio untuk keperluan analisis statistik. Sehingga bila seseorang mendapat skor 80%, maka nilai *odds ratio*-nya adalah 80:20, yang tidak lain adalah data rasio yang lebih tepat untuk tujuan pengukuran. Melalui data rasio ini Rasch mengembangkan model pengukuran yang menentukan hubungan antara tingkat kemampuan siswa (*person ability*) dan tingkat kesulitan aitem (*item difficulty*) dengan menggunakan fungsi logaritma untuk menghasilkan pengukuran dengan interval yang sama. Hasilnya adalah satuan

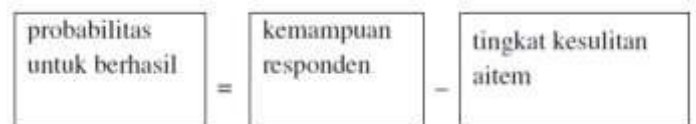
baru yang disebut logit (*log odds unit*) yang menunjukkan abilitas siswa dan kesulitan aitem; sehingga nantinya dari nilai logit yg didapat, disimpulkan bahwa tingkat kesuksesan siswa dalam 3 mengerjakan soal sangat tergantung dari tingkat abilitasnya dan tingkat kesulitan soal [5]. Untuk data yang berbentuk dikotomi, pemodelan Rasch menggabungkan suatu algoritma yang menyatakan hasil ekspektasi probabilistik dari aitem ‘i’ dan responden ‘n’, yang secara matematis dinyatakan sebagai:

$$P_{ni}(X_{ni} = 1/\beta_n, \delta_i) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$

dimana: $P_{ni}(X_{ni} = 1/\beta_n, \delta_i)$ adalah probabilitas dari responden n dalam aitem i untuk menghasilkan jawaban betul ($x = 1$); dengan kemampuan responden, β_n , dan tingkat kesulitan aitem δ_i . Persamaan di atas dapat lebih disederhanakan dengan memasukkan fungsi logaritma dan menjadikannya:

$$\log P_{ni} \left(X_{ni} = \frac{1}{\beta_n}, \delta_i \right) = \beta_n - \delta_i$$

Sehingga probabilitas akan satu keberhasilan dapat dituliskan sebagai:



Dalam lingkup ilmu sosial, maka mendapatkan data berupa angka sumber yang bias didapat biasanya dalam bentuk sikap (*attitude*) dan opini terhadap aitem pernyataan atau pertanyaan dalam satu instrumen yang diberikan. Instrumen tersebut dirancang dari variable yang sudah didefinisikan secara memuaskan, kemudian diidentifikasi konstruk-konstruk yang relevan; dari sana lah aitem-aitem dibuat dan dikembangkan untuk bisa mengukur variabel yang dimaksud. Pada saat yang sama pilihan jawaban yang

disediakan umumnya kemudian mengikuti pola penskoran yang dianut oleh teori test klasik (CTT). Dalam konteks model rasch, pola penskoran yang ‘menetap’ ini tidak lain adalah pengukuran yang hasilnya bergantung pada siapa yang diukur (*test dependent scoring*); sedangkan yang harus dilakukan dalam riset kuantitatif dalam ilmu sosial adalah pengukuran yang objektif (*objective measurement*).

Konsep pengukuran yang objektif dalam ilmu-ilmu sosial harus mempunyai lima kriteria [6], yaitu:

1. Memberikan ukuran yang linear dengan interval yang sama;
2. Melakukan proses estimasi yang tepat;
3. Menemukan aitem yang tidak tepat (*misfits*) atau tidak umum (*outliers*);
4. Mengatasi data yang hilang;
5. Menghasilkan pengukuran yang *replicable* (independen dari parameter yang diteliti)

Dari kelima syarat tadi, sejauh ini hanya rasch model lah yang bisa memenuhi kelima syarat tersebut. Dengan kata lain kualitas pengukuran dalam ilmu sosial yang dilakukan dengan rasch model akan mempunyai kualitas yang sama seperti halnya pengukuran yang dilakukan dalam bidang fisika. Bila dilihat lebih lanjut, skala logit (*log odds unit*) yang dihasilkan dalam model rasch adalah skala dengan interval yang sama dan bersifat linear yang berasal dari data ratio (*odds ratio*) dan bukannya data mentah skor yang didapat (1). Oleh karena itu proses estimasi abilitas seseorang ataupun tingkat kesulitan soal akan mempunyai nilai estimasi yang lebih tepat dan bias saling dibandingkan karena mempunyai satuan yang sama (logit) (2). Berhubung algoritma yang digunakan akan melakukan pengurutan secara terstruktur antara responden dari abilitas tinggi ke rendah, yang secara bersamaan juga mengurutkan soal dari yang mudah ke yang sulit, maka adanya

ketidaktepatan/konsistensi jawaban dari responden (*misfit*) ataupun pola yang diluar kebiasaan (*outlier*) akan mudah dideteksi; demikian juga untuk pola respon yang diterima satu soal tertentu (3). Pengurutan abilitas responden dan kesulitan soal secara terstruktur juga membuat model rasch dapat melakukan prediksi bila terdapat data yang hilang (4). Skala logic yang dihasilkan akan memunculkan nilai yang tergantung dari pola respon yang diberikan, bukannya pada skor awal yang ditentukan, sehingga rasch model akan selalu menghasilkan pengukuran yang independen [7].

METODE

Penelitian diawali dengan membangkitkan data simulasi dengan ukuran sampel 1000 partisipan. Sedangkan model yang digunakan untuk membangkitkan data adalah model 2PL. Selanjutnya data yang telah dibangkitkan ini akan menjadi sebagai acuan terhadap estimasi yang akan dilakukan. Perhitungan korelasi dari data acuan dan data hasil estimasi dengan 10 replikasi dilakukan dengan bantuan Wingen. Analisis hasil estimasi dilakukan dengan membandingkan nilai median korelasi, RMSE, dan MAD yang dihasilkan.

HASIL DAN PEMBAHASAN

Berdasarkan replikasi yang telah dilakukan diperoleh *output* sebagai berikut Resume *output* Wingen untuk

Tabel. 1 Proficiency Parameter Report (Panjang tes = 20)

Replikasi	Nilai		
	Corr	RMSE	MAD
1	0,956	989230,300	803799
2	0,957	999298,000	812857
3	0,956	989233,200	804409
4	0,959	997694,000	811507
5	0,958	997464,600	818678
6	0,958	997811,600	814261
7	0,960	992959,200	809814
8	0,954	992855,900	808687
9	0,955	994951,700	811475

10	0,958	994812,100	809237
Median	0,957		

Sumber: Pengolahan Data

Tabel. 2 Proficiency Parameter Report (Panjang tes = 40)

Replikasi	Nilai		
	Corr	RMSE	MAD
1	0,963	1020433,000	836027
2	0,965	1025030,000	839237
3	0,966	1021808,000	835146
4	0,968	1026910,000	839324
5	0,967	1022412,000	837704
6	0,965	1014363,000	832916
7	0,966	1026150,000	840119
8	0,966	1021074,000	832022
9	0,965	1023383,000	838543
10	0,966	1029914,000	843698
Median	0,966		

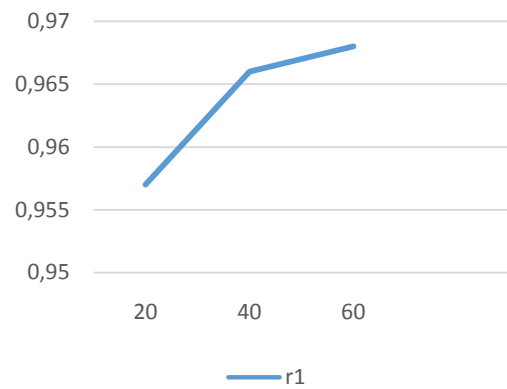
Sumber: Pengolahan Data

Tabel. 3 Proficiency Parameter Report (Panjang tes = 60)

Replikasi	Nilai		
	Corr	RMSE	MAD
1	0,964	1028685,000	849444
2	0,967	1047580,000	867426
3	0,968	1038442,000	858187
4	0,969	1043348,000	865056
5	0,967	1030578,000	851300
6	0,968	1047568,000	866217
7	0,968	1044721,000	859208
8	0,966	1047125,000	862938
9	0,968	1036830,000	853371
10	0,968	1029106,000	851525
Median	0,968		

Sumber: Pengolahan Data

Berdasarkan hasil olahan data yang telah dilakukan menggunakan Wingen dapat dilihat bahwa nilai median corr (Korelasi) pada Tabel 1, 2, dan 3 berturut-turut mengalami kenaikan berbanding lurus dengan jumlah item. Kenaikan tersebut dapat dilihat pada grafik di bawah ini:



Sumber: Pengolahan Data

Gambar 1. Grafik Peningkatan Median Korelasi untuk n=20, 40, 60

Berdasarkan grafik di atas dapat disimpulkan bahwa semakin panjang tes, maka akan semakin stabil estimasi θ (kemampuan) pada model 2PL. Hal tersebut dapat dilihat dari koefisien korelasi dari 10 replikasi yang dilakukan terus meningkat seiring bertambah panjangnya tes.

SIMPULAN

Semakin panjang tes, maka akan semakin stabil estimasi θ (kemampuan) pada model Logistik *Item Response*. Hal tersebut dapat dilihat dari koefisien korelasi dari 10 replikasi yang dilakukan cenderung terus meningkat seiring bertambah panjangnya tes.

DAFTAR PUSTAKA

[1] Hambleton, R.K, Swaminathan H, & Roger, H.J. *MMSS Fundamental Of Item Response Theory (Volume 2)*. California: Sage Publications.1991.

[2] Dali, S.Naga. *Psikometrika pada Pengukuran Model Rasch*. Jakarta: Nagrani Citrayasa. 2017.

[3] Cohen, A.S. & Kane, M.T. The Precision of simulation study results. *Applied Psychological Measurement Journal*. Vol. 25 No. 2. pp. 136- 145. 2007

- [4] Nathan A.Thompson. *Ability Estimation with Item Response Theory*. Minessota: Assesment System Corporation. USA: Scientific Software International, Inc. 2009.
- [5] Olsen, L. W. Essays on Georg Rasch and his contributions to statistics. Unpublished PhD thesis at Institute Of Economics University of Copenhagen. 2013.
- [6] Michael Edward. An Introduction to Item Response Theory Using the Need for Cognition Scale. *Social and Personality Psychology Compass* 3/4 : 507–529. 2009.
- [7] Sumintono, Bambang. Model Rasch dalam Penelitian Kuantitatif. Makalah dipresentasikan dalam kuliah umum di Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, 21 November 2014