

PENGARUH PANJANG TES DAN UKURAN CONTOH PADA PENDUGAAN PARAMETER MODEL IRT

Siti Hasanah¹, Dewi Zulyani Pomalingo², Anang Kurnia³, Alona Dwinata⁴

Statistika dan Sains Data, IPB University^{1,2,3}

Pendidikan Matematika, Universitas Maritim Raja Ali Haji⁴

Email: stathanah@apps.ipb.ac.id

Abstrak

IRT (*item response theory*) dapat menjelaskan hubungan antara jawaban seseorang terhadap butir dengan variabel laten (*ability*). Jumlah butir atau soal latihan yang diberikan oleh seorang guru SD hingga SMA biasanya tidak begitu banyak. Begitu pun dengan jumlah siswa yang biasanya terbagi dalam kelas berbeda, berjumlah antara 30 hingga 35 siswa di tiap kelasnya. Penelitian ini dilakukan pada Juni 2023 dan bertujuan untuk mengetahui pengaruh ukuran contoh dan panjang tes terhadap pendugaan parameter model IRT 1 PL pada data berukuran kecil. Parameter yang dianalisis melalui model IRT 1 PL mencakup parameter butir (*item*) yaitu *difficulty* dan parameter *ability*. Data didapatkan dari bangkitan hasil simulasi. Simulasi pembangkitan data dengan beberapa skenario digunakan dalam mengkonstruksi model IRT 1 PL, yaitu berdasarkan panjang tes (10, 15, 20) dan ukuran contoh (32, 64). Hasil penelitian ditunjukkan dengan keamatan dugaan parameter θ (*ability*) dan parameter sebenarnya yang semakin tinggi seiring meningkatnya panjang tes. Saat Panjang tes dianggap tetap, keamatan dugaan parameter *difficulty* semakin tinggi dengan ukuran contoh yang meningkat pada skala kecil.

Kata Kunci : *Item Response Theory*, IRT 1 PL, Ukuran Contoh, Panjang Tes, Parameter Logistik

Abstract

IRT (*Item Response Theory*) can explain the relationship between a person's response to an item and a latent variable (*ability*). The number of items or exercise questions given by an elementary to high school teacher is usually not very large. The same goes for the number of students, who are usually divided into different classes, ranging from 30 to 35 students per class. This study was conducted in June 2023 and aims to determine the influence of sample size and test length on the estimation of parameters in the item response theory (IRT) 1 PL model with small-sized data. The parameters analyzed through the IRT 1 PL model include item parameters, namely *difficulty*, and *ability* parameters. The data were obtained from generation of simulation results. Data generation simulations with several scenarios were used to construct the IRT 1 PL model, based on test lengths (10, 15, 20) and sample sizes (32, 64). The research results are indicated by the closeness of the estimated parameter θ (*ability*) to the true parameter, which increases as the test length increases. When the test length is considered constant, the closeness of the estimated *difficulty* parameter increases with an increase in the sample size in small-scale scenarios.

Key Words : *Item Response Theory*, IRT 1 PL, Sample Size, Test Length, Logistic Parameter

PENDAHULUAN

Pengukuran adalah proses mengidentifikasi satu atribut yang dilakukan pada satu waktu [1]. Dua teori pengukuran butir yang masih terus berkembang sampai saat ini yaitu teori tes klasik atau *classical test theory* (CTT) dan teori tes modern atau *item response theory* (IRT). Teori pengukuran ini dimanfaatkan untuk merancang dan menganalisis alat ukur atau tes. IRT adalah paradigma psikometrik

untuk konstruksi, penilaian, dan analisis formulir serta item tes [2]. *Item response theory* memerlukan lebih banyak asumsi tetapi juga memberikan informasi yang lebih lengkap dibandingkan *classical test theory* [3]. IRT merupakan metode yang lebih terbaru sebagai usaha menangani kelemahan-kelemahan yang ada pada CTT [4]. Beberapa asumsi pada IRT adalah unidimensi, antar butir saling bebas, dan butir tidak mempengaruhi respon

seseorang berdasarkan karakteristik individu seperti latar belakang budaya atau jenis kelamin. Karakteristik kelompok peserta tes mengalami perubahan jika kelompok tersebut menempuh tes yang berbeda [5].

IRT adalah sebuah model probabilitas yang dapat menjelaskan hubungan antara respon seseorang terhadap sebuah butir dengan variabel laten (kemampuan/*ability* atau sifat/*trait*) yang diukur oleh tes tersebut [6]. Pada IRT dapat digunakan model karakteristik butir di antaranya model IRT 1 Parameter Logistik (IRT 1PL). Parameter pada IRT model 1 PL adalah tingkat kesukaran butir (*difficulty*) dan kemampuan peserta (*ability*). Model khusus dari IRT 1 PL saat parameter daya beda bernilai 1 adalah model *rasch*. Terdapat banyak model IRT yang dikembangkan untuk keadaan khusus dan tidak berlaku umum [7]. Penelitian kali ini berfokus pada model IRT 1 PL non *rasch*.

Dalam konteks analisis IRT (*Item Response Theory*), respon dari setiap butir soal diambil dari jawaban yang diberikan oleh responden atau peserta tes. Jawaban ini biasanya berupa pilihan ganda, benar/salah, atau jenis respon lainnya yang telah ditentukan sebelumnya dalam desain soal. Setiap jawaban yang diberikan oleh peserta akan direkam dan dianalisis untuk menentukan parameter butir (seperti tingkat kesulitan) dan parameter individu (seperti kemampuan peserta) menggunakan model IRT yang sesuai. Dengan memperhitungkan pola jawaban dari sejumlah peserta dan pertanyaan tertentu, model IRT dapat memberikan estimasi yang akurat tentang kemampuan individu dan karakteristik butir soal, serta hubungan antara keduanya.

Dengan demikian, informasi yang diperoleh dari respon peserta sangat penting dalam proses analisis IRT, karena

akan menjadi dasar bagi pembuatan estimasi parameter-parameter yang diperlukan dalam evaluasi dan pembuatan keputusan pendidikan.

Model IRT 1 PL [8] :

$$P_i = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} ; i = 1, 2, 3, \dots, n$$

dimana, P_i adalah peluang peserta tes yang memiliki kemampuan (θ) dipilih secara acak dapat menjawab butir i dengan benar, θ adalah tingkat kemampuan peserta, b_i adalah indeks kesulitan butir ke- i , e adalah bilangan natural yang nilainya mendekati 2,718, dan n adalah banyaknya butir dalam tes. Pada model IRT 1 PL, parameter yang dilibatkan hanya parameter *difficulty* (b_i). Secara teoretis, nilai b_i terletak pada rentang $-\infty$ dan $+\infty$. Suatu butir dikatakan baik jika nilai b_i berkisar antara -2 dan $+2$ [8]. Jika nilai b_i mendekati -2 , maka indeks kesulitan butir sangat rendah atau dengan kata lain tipe soal sangat mudah, sedangkan jika nilai b_i mendekati $+2$ maka indeks kesulitan butir sangat tinggi. Pengaruh Panjang tes pada data simulasi berukuran besar telah diteliti oleh [3] dan [9]. Penelitian oleh menggunakan model IRT 2 PL dengan banyaknya contoh 200, 400, dan 1000 [3]. Hasil dari penelitian adalah kestabilan dugaan parameter *difficulty* dipengaruhi oleh *sample size*, dan kestabilan parameter *ability* dipengaruhi oleh ukuran *test length* [3]. Pendugaan parameter perlu dilakukan untuk memperoleh parameter populasi yang tidak diketahui dengan menggunakan data sampelnya [10]. Semakin besar *sample size* dan ukuran *test length*, maka pendugaan parameter semakin stabil. Penelitian lainnya juga menggunakan model IRT 2 PL dengan banyaknya contoh 1000 dan ukuran *test length* 20, 40, dan 60 [9]. Hasil dari penelitian penelitian tersebut adalah semakin panjang tes yang digunakan, estimasi parameter kemampuan model 2PL semakin akurat. Penelitian ini memiliki tujuan yang sangat relevan dengan kondisi

yang umum terjadi di lapangan, terutama dalam konteks tes pada siswa sekolah dasar dan menengah. Pada kenyataannya, jumlah butir soal dan jumlah peserta tes pada level pendidikan tersebut biasanya berukuran kecil. Oleh karena itu, penelitian ini bertujuan untuk menggali pengaruh panjang tes dan ukuran contoh terhadap dugaan parameter butir (*difficulty*) dan parameter θ atau kemampuan individu (*ability*) pada data dengan ukuran yang kecil.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang berharga bagi pembuat kurikulum, guru, dan para pengambil keputusan pada bidang pendidikan. Dengan memahami pengaruh panjang tes dan ukuran contoh terhadap estimasi parameter dalam model IRT, mereka dapat membuat keputusan yang lebih baik dalam merancang soal-soal latihan yang efektif dan relevan untuk siswa. Dengan kata lain, temuan dari penelitian ini dapat menjadi masukan yang berharga dalam proses pengembangan kurikulum dan penyusunan materi pelajaran di sekolah.

Diharapkan bahwa dengan pemahaman yang lebih baik tentang bagaimana ukuran contoh dan panjang tes memengaruhi estimasi parameter dalam model IRT, maka pengajaran dan pembelajaran di sekolah dapat menjadi lebih efektif dan efisien. Hal ini akan membantu siswa dalam memperoleh pemahaman yang lebih baik tentang materi pelajaran serta meningkatkan kemampuan mereka dalam menghadapi tes dan evaluasi. Sehingga, penelitian ini memiliki potensi untuk memberikan dampak positif yang signifikan dalam meningkatkan kualitas pendidikan di tingkat dasar dan menengah.

METODE

Data yang digunakan dalam analisis ini adalah hasil dari simulasi yang dilakukan

menggunakan perangkat lunak Wingen 3.1.15. Simulasi pembangkitan data dilakukan dengan mempertimbangkan beberapa kriteria utama. Pertama, kemampuan peserta (θ , *ability*) diambil dari distribusi normal dengan mean 0 dan standar deviasi 1 ($\theta \sim N(0,1)$). Kedua, parameter butir yang menggambarkan tingkat kesulitan (b) diambil dari distribusi seragam dalam rentang $[-2, 2]$ ($b \sim U[-2, 2]$). Data dibangkitkan dalam variasi ukuran contoh, yaitu sebanyak 32 dan 64, serta dalam variasi panjang tes, yakni 10, 15, dan 20 butir soal. Model yang digunakan untuk menganalisis data ini adalah model IRT 1 PL. Selanjutnya, pengaruh dari ukuran contoh dan panjang tes terhadap estimasi parameter dianalisis melalui dua metrik utama: nilai korelasi dan *Root Mean Square Error* (RMSE). Korelasi digunakan untuk menilai seberapa erat hubungan antara parameter yang diestimasi dengan nilai sebenarnya, sementara RMSE digunakan untuk mengevaluasi seberapa besar kesalahan estimasi yang terjadi dalam mengukur parameter tersebut.

Dengan menggunakan data hasil simulasi ini, analisis akan mengungkap bagaimana variasi ukuran contoh dan panjang tes mempengaruhi kualitas estimasi parameter dalam konteks model IRT 1 PL. Ini akan memberikan wawasan dalam pemahaman tentang bagaimana pemodelan IRT dapat digunakan untuk mengukur dan menganalisis kemampuan peserta dalam tes dengan berbagai skenario ukuran contoh dan panjang tes. Nilai RMSE yang semakin kecil menunjukkan estimasi yang diujicobakan semakin akurat [11].

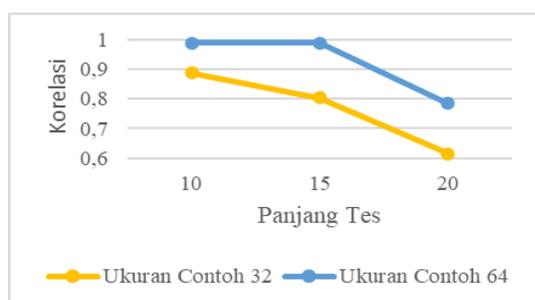
HASIL DAN PEMBAHASAN

Hasil pemodelan IRT 1 PL untuk data dengan ukuran contoh (N) 32 dan 64 dapat dilihat pengaruhnya terhadap parameter butir melalui nilai korelasi dan RMSE pada Tabel 1.

Tabel 1. Pengaruh Ukuran Contoh pada Dugaan Parameter *Difficulty*

No	N	Korelasi			RMSE		
		n=10	n=15	n=20	n=10	n=15	n=20
1.	32	0,89	0,80	0,61	5,79	22,33	85,17
2.	64	0,99	0,99	0,78	42,56	1,33	65,86

Semakin meningkatnya ukuran contoh, terlihat bahwa keeratan antara parameter dugaan *difficulty* dengan parameter sebenarnya semakin tinggi, seperti yang terungkap dalam Tabel 1. Meskipun demikian, perlu dicatat bahwa nilai RMSE pada contoh-contoh berukuran kecil ini menunjukkan tingkat konsistensi yang kurang. Secara khusus, pada panjang tes $n=10$, terdapat sebuah anomali dimana nilai RMSE untuk ukuran contoh $N=64$ justru lebih besar. Meskipun demikian, jika kita mengambil rata-rata dari ketiga panjang tes tersebut, ukuran contoh 64 ternyata memiliki nilai rata-rata RMSE yang lebih kecil. Fenomena ini menggambarkan kompleksitas dalam interpretasi data ketika mengkaji hubungan antara ukuran contoh dan keakuratan estimasi parameter. Selain itu, ada hal menarik lain yang terungkap, yaitu pengaruh panjang tes terhadap pendugaan parameter kesulitan butir. Menurut Tabel 1, semakin panjang suatu tes, nilai korelasi antara parameter kesulitan butir dengan panjang tes semakin menurun. Untuk pemahaman yang lebih mendalam, perbedaan ini dapat dilihat secara visual melalui Gambar 1.



Gambar 1. Grafik Pengaruh Panjang Tes terhadap *Difficulty*

Dengan demikian, kesimpulan yang dapat diambil adalah bahwa baik ukuran contoh maupun panjang tes merupakan faktor penting yang perlu dipertimbangkan secara cermat dalam menganalisis dan menginterpretasi data tes.

Pengaruh panjang tes pada kemampuan peserta (θ) dapat dilihat melalui Tabel 2.

Tabel 2. Pengaruh Panjang Tes pada Dugaan Parameter *Ability*

No	Ukuran Contoh	Panjang Tes	Korelasi	RMSE
1.	32	10	0,6068	5,5857
2.	32	15	0,8642	3,1833
3.	32	20	0,8686	3,1160
4.	64	10	0,7926	5,0274
5.	64	15	0,8376	4,3492
6.	64	20	0,8454	4,3980

Berdasarkan analisis yang tercantum dalam Tabel 2, dapat diamati bahwa terdapat korelasi dalam hubungan antara parameter kemampuan (*ability*) dengan panjang tes. Dengan meningkatnya panjang tes, terlihat bahwa korelasi parameter *ability* cenderung meningkat secara linier. Fenomena ini menunjukkan bahwa semakin panjang tes, semakin akurat pula kemampuan individu dalam diukur. Tak hanya itu, pada contoh dengan ukuran 64, terdapat pola yang serupa yang mengindikasikan bahwa hubungan antara parameter dugaan dan nilai sebenarnya semakin kuat seiring dengan peningkatan panjang tes. Hal ini menandakan bahwa pengukuran terhadap kemampuan individu menjadi lebih tepat seiring dengan peningkatan panjang tes. Selain itu, perlu dicatat bahwa nilai *Root Mean Square Error* (RMSE) dari parameter θ terhadap parameter sebenarnya juga menunjukkan kecenderungan yang menarik. Dengan meningkatnya panjang tes, terlihat bahwa RMSE tersebut mengalami penurunan. Hal ini menunjukkan bahwa semakin panjang tes, semakin kecil pula kesalahan estimasi antara parameter yang diukur dengan nilai

sebenarnya. Oleh karena itu, hasil dari analisis ini menyoroti pentingnya panjang tes dalam meningkatkan akurasi dalam mengukur kemampuan individu.

SIMPULAN

Ukuran contoh berpengaruh pada parameter kesulitan butir meskipun *sample* berukuran kecil. Peningkatan ukuran contoh pada skala kecil juga meningkatkan keakuratan pendugaan parameter kemampuan pada model IRT 1PL. Panjang tes berpengaruh terhadap parameter kemampuan peserta tes dan kesulitan butir. Peningkatan panjang tes pada skala kecil linier dengan peningkatan korelasi *ability*, namun berbanding terbalik dengan korelasi kesulitan butir.

DAFTAR PUSTAKA

- [1] B. Sumintono, "Model Rasch untuk Penelitian Sosial Kuantitatif," in *Kuliah umum di Jurusan Statistika*, Surabaya, 2014. [Online]. Available: https://core.ac.uk/display/162013734?utm_source=pdf&utm_medium=banner&utm_campaign=pdf-decoration-v1
- [2] N. A. Thompson, "Ability Estimation with Item Response Theory," *Assesment Syst. Corp.*, pp. 1–11, 2009, [Online]. Available: www.assess.com
- [3] R. Budiarti, "Pengaruh Sample Size (N) dan Test Length (n) terhadap Item Parameter Estimate dan Examinee Parameter Estimate, suatu Studi Simulas," *J. Math. Its Appl.*, vol. 12, no. 1, pp. 25–36, Jul. 2013, doi: 10.29244/jmap.12.1.25-36.
- [4] M. Tarigan and F. Fadillah, "Analisis Item Response Theory Raven's Coloured Progressive Matrices pada Sampel Anak Usia Dini," *PSIKODIMENSIA*, vol. 20, no. 2, pp. 158–169, Dec. 2021, doi: 10.24167/psidim.v20i2.3101.
- [5] S. DS, "Pengaruh Panjang Tes dan Ukuran Sampel terhadap Kekekaran Estimasi Parameter pada Teori Respon Butir (Item Response Theory)," *Cakrawala Pendidik.*, vol. 3, 2006, doi: 10.21831/cp.v1i3.8593.
- [6] F. Fajrianti, W. Hendriani, dan B. G. Septarini, "Pengembangan Tes Berpikir Kritis dengan dengan Pendekatan Item Response Theory," *J. Penelit. dan Eval. Pendidik.*, vol. 20, no. 1, pp. 45–55, Jun. 2016, doi: 10.21831/pep.v20i1.6304.
- [7] M. C. Edwards, "An Introduction to Item Response Theory Using the Need for Cognition Scale," *Soc. Personal. Psychol. Compass*, vol. 3, no. 4, pp. 507–529, 2009, doi: 10.1111/j.1751-9004.2009.00194.x.
- [8] Hambleton, Swaminathan, and Rogres, *Item Response Theory*. Boston MA: Kluwer Inc, 1985.
- [9] I. Falani dan S. A. Kumala, "Kestabilan Estimasi Parameter Kemampuan pada Model Logistik Item Response Theory Ditinjau dari Panjang Tes," *SAP (Susunan Artikel Pendidikan)*, vol. 2, no. 2, Dec. 2017, doi: 10.30998/sap.v2i2.2028.
- [10] S. A. Purba, "Estimasi Parameter Data Berdistribusi Normal Menggunakan Maksimum Likelihood Berdasarkan Newton Raphson," *J. Sains Dasar*, vol. 9, no. 1, pp. 16–18, Feb. 2021, doi: 10.21831/jsd.v9i1.38564.
- [11] Rahayu, W., "Metode Estimasi Parameter dan Meter Equating pada Ukuran Sampel Kecil Berdasarkan Item Respons Theory," *Prosiding Semirata2015*, pp. 315–324.