

Desain Instrumen Diagnostik Berbasis Model IRT Materi Sistem Persamaan Linier Perguruan Tinggi

Jesi Irwanto¹

¹ Institut Teknologi dan Bisnis Widya Gama Lumajang
Telp: (0331) 330224
E-mail: jesyrwanto@gmail.com

RIWAYAT ARTIKEL

Received: 2025-05-21

Revised : 2025-05-23

Accepted: 2025-05-30

KEYWORD

*Linear Equation System,
Diagnostic Instrument, Item
Response Theory (IRT), 3PL
Model, Student Understanding.*

KATA KUNCI

*Sistem Persamaan Linier,
Instrumen Diagnostik, Item
Response Theory (IRT), Model
3PL, Pemahaman Mahasiswa.*

ABSTRACT

Students' understanding of the Two-Variable Linear Equation System is still relatively low, so a diagnostic instrument is needed that is able to identify conceptual weaknesses accurately. This study aims to design a diagnostic instrument based on the Item Response Theory (IRT) model of three-parameter logistic (3PL) on the material. The research method used is a quantitative approach with the stages of designing question grids, expert validation, trials on 80 students, and data analysis using R Studio software. The results showed that 21 of the 25 questions met the validity criteria and unidimensionality assumptions. The instrument has high reliability (0.92) and is able to map students' abilities into three main categories. This fact shows that the IRT model is effective in developing diagnostic instruments and can help lecturers design data-based learning strategies to improve understanding of the two variables equations system concept in higher education. this instrument provides educators with a reliable tool for identifying learning gaps, enabling more targeted interventions and personalized support for students struggling with linear equation systems.

ABSTRAK

Pemahaman mahasiswa terhadap Sistem Persamaan Linier Dua Variabel (SPLDV) masih tergolong rendah, sehingga diperlukan instrumen diagnostik yang mampu mengidentifikasi kelemahan konsep secara tepat. Penelitian ini bertujuan untuk merancang instrumen diagnostik berbasis Item Response Theory (IRT) model tiga parameter logistik (3PL) pada materi SPLDV. Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan tahapan perancangan kisi-kisi soal, validasi ahli, uji coba terhadap 80 mahasiswa, dan analisis data menggunakan perangkat lunak R Studio. Hasil penelitian menunjukkan bahwa 21 dari 25 butir soal memenuhi kriteria validitas dan asumsi unidimensionalitas. Instrumen memiliki reliabilitas tinggi (0.92) serta mampu memetakan kemampuan mahasiswa dalam tiga kategori utama. Fakta ini menunjukkan bahwa model IRT efektif digunakan dalam pengembangan instrumen diagnostik dan dapat membantu dosen merancang strategi pembelajaran berbasis data untuk meningkatkan pemahaman konsep SPLDV di perguruan tinggi.

1. Pendahuluan

Matematika merupakan fondasi penting dalam berbagai disiplin ilmu, terutama di bidang ekonomi, teknik, dan ilmu alam. Salah satu topik dasar yang diajarkan di tingkat perguruan tinggi adalah Sistem Persamaan Linier Dua Variabel (SPLDV)(1). Topik ini tidak hanya menjadi batu loncatan untuk memahami sistem yang lebih kompleks, tetapi juga menjadi indikator awal terhadap kemampuan berpikir logis dan analitis mahasiswa. Namun, berbagai studi dan pengalaman empiris menunjukkan bahwa mahasiswa masih menghadapi kesulitan signifikan dalam memahami dan menyelesaikan SPLDV, baik secara konseptual maupun prosedural

Kesulitan tersebut sering kali tidak terdeteksi secara akurat dengan metode penilaian konvensional seperti tes uraian atau pilihan ganda. Sebanyak 7 dari setiap 10 mahasiswa mengalami kesulitan. Oleh karena itu, dibutuhkan sebuah instrumen diagnostik yang tidak hanya mengukur skor akhir, tetapi juga mampu memetakan kemampuan kognitif mahasiswa secara lebih detail. Salah satu pendekatan yang dapat digunakan adalah Model Item Response Theory (IRT), khususnya model tiga parameter logistik (3PL), yang memungkinkan analisis mendalam terhadap karakteristik butir soal serta kemampuan responden IRT merupakan teori pengukuran modern yang menggeser paradigma klasik dalam penilaian pendidikan(2). Berbeda dengan pendekatan klasik yang bergantung pada total skor dan asumsi kesetaraan antar butir, IRT memberikan informasi spesifik tentang kemampuan individu berdasarkan interaksi mereka dengan karakteristik tiap item. Hal ini sangat relevan dalam pembelajaran matematika yang menuntut kemampuan berpikir bertahap dan sistematis. (3) Dengan menggunakan IRT, instrumen penilaian dapat didesain secara lebih presisi, valid, dan adaptif terhadap kebutuhan pembelajaran. Lebih jauh lagi dengan semakin berkembangnya teknologi pembelajaran dan integrasi data dalam pendidikan, kebutuhan akan asesmen yang berbasis model matematis menjadi semakin relevan. Instrumen diagnostik berbasis IRT tidak hanya memberikan umpan balik formatif yang bermakna, tetapi juga membantu dalam mendesain pembelajaran berbasis data. (4) Hal ini sangat penting dalam konteks pendidikan tinggi yang mengarah pada pembelajaran adaptif dan personalisasi kurikulum Penelitian ini bertujuan untuk: (1) merancang instrumen diagnostik berbasis IRT untuk materi SPLDV; (2) menguji validitas dan reliabilitas instrumen melalui model 3PL; (3)

memetakan kemampuan mahasiswa berdasarkan parameter butir soal. Hasil penelitian ini diharapkan dapat memberikan kontribusi teoretis dan praktis dalam bidang pendidikan matematika, terutama dalam pengembangan instrumen asesmen yang mampu memberikan gambaran detail tentang tingkat pemahaman mahasiswa(5)

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif (6) dengan desain pengembangan instrumen. Subjek penelitian adalah 150 mahasiswa semester awal dari 3 kelas di Institut Teknologi dan Bisnis Widya Gama Lumajang, yang dipilih menggunakan teknik stratified sampling berdasarkan IP Matematika Ekonomi. Tahapan penelitian terdiri dari: (1) penyusunan indikator berdasarkan teori David Tall teori konstruktifisme (2) penyusunan soal pilihan ganda berjumlah 6 item; (3) uji validasi isi oleh ahli; (4) uji coba lapangan; dan (5) analisis data menggunakan Model IRT 3PL. Analisis data dilakukan dengan bantuan perangkat lunak R Studio dan SPSS menggunakan paket ltm dan mirt. Asumsi unidimensionalitas diuji menggunakan analisis faktor konfirmatori (CFA). Validitas butir diuji dengan parameter diskriminasi (a), kesulitan (b), dan peluang menebak (c). Kriteria kelayakan soal mengacu pada literatur(7). Reliabilitas instrumen diukur melalui indeks informasi tes total. Hasil analisis digunakan untuk merevisi butir yang tidak memenuhi syarat dan menghasilkan perangkat soal final yang dapat digunakan sebagai instrumen diagnostik untuk mendeteksi pemahaman mahasiswa dalam materi SPLDV. Fokus utama adalah pemetaan kemampuan individu dan identifikasi miskonsepsi yang sering muncul.

3. Hasil dan Pembahasan

Dari hasil uji coba instrumen pada 150 mahasiswa, diperoleh data respons terhadap 6 butir soal yang kemudian dianalisis menggunakan Model IRT 3PL. Berdasarkan hasil analisis, sebanyak 139 dari 150 responden dinyatakan valid dan reliabel. Adapun data-nya ditunjukkan pada tabel 1.1 berikut.

		B01	B02	B03	B04	B05	B06	TOTAL
B01	Pearson Correlation	1	.712**	.733**	.183	.341**	-.150	.762**
	Sig. (2-tailed)		<.001	<.001	.031	<.001	.078	<.001
	N	139	139	139	139	139	139	139
B02	Pearson Correlation	.712**	1	.598**	.158	.298**	-.125	.714**
	Sig. (2-tailed)	<.001		<.001	.087	<.001	.143	<.001
	N	139	139	139	139	139	139	139
B03	Pearson Correlation	.733**	.598**	1	.129	.373**	-.070	.750**
	Sig. (2-tailed)	<.001	<.001		.129	<.001	.410	<.001
	N	139	139	139	139	139	139	139
B04	Pearson Correlation	.183	.158	.129	1	.222**	.249**	.496**
	Sig. (2-tailed)	.031	.067	.129		.009	.003	<.001
	N	139	139	139	139	139	139	139
B05	Pearson Correlation	.341**	.298**	.373**	.222**	1	.283**	.683**
	Sig. (2-tailed)	<.001	<.001	<.001	.009		<.001	<.001
	N	139	139	139	139	139	139	139
B06	Pearson Correlation	-.150	-.125	-.070	.249**	.283**	1	.316**
	Sig. (2-tailed)	.078	.143	.410	.003	<.001		<.001
	N	139	139	139	139	139	139	139
TOTAL	Pearson Correlation	.762**	.714**	.750**	.496**	.683**	.316**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	
	N	139	139	139	139	139	139	139

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Hasil uji validitas menggunakan korelasi Pearson antara masing-masing butir soal dan total skor menunjukkan bahwa seluruh item (B01–B06) memiliki nilai signifikansi di bawah 0,05, yang berarti valid secara statistik. Item B01, B02, B03, dan B05 menunjukkan korelasi yang kuat dengan skor total ($r = 0,762; 0,714; 0,750; \text{ dan } 0,683$) dengan tingkat signifikansi 0,000, mengindikasikan bahwa butir-butir tersebut sangat representatif terhadap konstruk yang diukur. Item B04 memiliki korelasi sedang ($r = 0,496$) namun tetap signifikan, sehingga dianggap valid. Sementara itu, item B06 memiliki korelasi yang paling rendah ($r = 0,318$), meskipun signifikan secara statistik. Korelasi yang lemah pada B06 menunjukkan bahwa item tersebut kurang kuat dalam merepresentasikan keseluruhan konstruk, sehingga perlu ditinjau kembali dari sisi redaksi, tingkat kesulitan, atau kesesuaian konteks. Secara keseluruhan, keenam butir soal layak digunakan dalam instrumen diagnostik dengan catatan revisi minor pada B06. Berdasarkan fakta tersebut dapat disimpulkan bahwa semua item soal yang diberikan memiliki signifikansi < 0.05 yang dapat dinyatakan valid. Butir soal 1,2,3 dan 5 menunjukkan validitas yang kuat. Sedangkan Butir soal 4 memiliki validitas dengan korelasi sedang.

Sementara untuk uji reliabilitas ditunjukkan pada tabel 1.2 sebagai berikut:

Reliability Statistics	
Cronbach's Alpha	N of Items
,683	6

Dari hasil uji reliabilitas 6 butir soal tersebut, menunjukkan nilai 0,683 yang mana nilai ini membuktikan bahwa 6 butir soal tersebut sudah reliable. Fakta -fakta tersebut menunjukkan bahwa butir soal yang disusun berdasarkan indikator teori David Maar yang kemudian di kembangkan menjadi instrumen soal yang diberikan kepada mahasiswa menunjukkan hasil uji yang valid dan reliabel.

Sementara itu hasil uji IRT menggunakan 3 parameter logistik dari penelitian ini dijabarkan sebagai berikut. Hasil uji kecocokan model butir soal pada penelitian ini menggunakan aplikasi R studio dengan Uji fit models di tunjukan pada tabel berikut:

Likelihood Ratio Table

	AIC	BIC	log.Lik	LRT	df	p.value
Out1	1010.95	1028.51	-			
Out2	921.65	956.77	499.47	101.3	6	<0.001
			-			
			448.82			

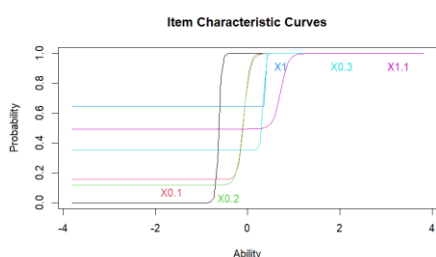
Out2	921.65	956.77	-	19.5	6	0.003
Out3	914.15	966.84	448.82			
			-			
			439.07			
Out1	1010.95	1028.51	-	120.8	12	<0.001
Out3	914.15	966.84	499.47			
			-			
			439.07			

Uji kecocokan model dilakukan untuk menentukan model IRT yang paling sesuai dengan data. Perbandingan dilakukan antara tiga model, yakni Out1, Out2, dan Out3 yang masing-masing merepresentasikan model 1PL, 2PL, dan 3PL. Hasil uji menunjukkan bahwa model Out2 memiliki nilai AIC (921.65) dan BIC (956.77) yang jauh lebih rendah dibandingkan Out1 (AIC = 1010.95, BIC = 1028.51). Nilai log-likelihood juga meningkat dari -499.47 pada Out1 menjadi -448.82 pada Out2. Uji Likelihood Ratio menunjukkan nilai LRT sebesar 101.3 dengan derajat bebas 6 dan nilai signifikansi < 0.001 , yang berarti model Out2 secara signifikan lebih baik daripada Out1. selanjutnya, perbandingan antara Out2 dan Out3 menunjukkan bahwa model Out3 lebih unggul dengan AIC 914.15 dan log-likelihood -439.07, disertai nilai LRT sebesar 19.5 ($df = 6; p = 0.003$). Meskipun perbedaannya tidak sebesar sebelumnya, peningkatan ini tetap signifikan secara statistik. Terakhir, model Out3 dibandingkan dengan Out1 menunjukkan hasil yang sangat signifikan (LRT = 120.8; $df = 12; p < 0.001$), mengukuhkan bahwa model Out3 adalah yang paling sesuai. Berdasarkan seluruh indikator AIC, BIC, log-likelihood, dan LRT model Out3 (3PL) terbukti paling cocok dengan data. Dengan mempertimbangkan kemampuan model ini menangkap tiga parameter penting (tingkat kesulitan, diskriminasi, dan kemungkinan menebak), model 3PL dipilih sebagai dasar analisis lebih lanjut dalam pengembangan instrumen diagnostik. Sehingga dari penjabaran hasil uji kecocokan tersebut menunjukkan bahwa butir-butir soal tersebut memenuhi IRT yang mengandung 3 parameter logistik.

Selain itu hasil uji butir soal yang diberikan kepada mahasiswa dengan model IRT 3PL di tunjukan pada gambar 1.3 berikut

	Gussng	Dffclt	Dscrmm
X0	0.0000656248	-0.62401579	32.97454
X0.1	0.1562896032	-0.07869836	17.09544
X0.2	0.1170951598	-0.08872806	14.22420
X1	0.6449221982	0.37147204	232.11938
X0.3	0.3524042079	0.33373998	39.20027
X1.1	0.4945829892	0.71262693	12.15906

Hasil estimasi parameter model IRT 3PL untuk sejumlah butir soal menunjukkan beragam karakteristik dari segi tebakan (*guessing*), kesulitan (*difficulty*), dan daya pembeda (*discrimination*). Pada parameter *guessing*, item X0 memiliki nilai sangat rendah (0.00006), menunjukkan bahwa hampir tidak ada peluang siswa menjawab benar secara kebetulan suatu indikasi kualitas soal yang baik. Item X0.1 dan X0.2 memiliki nilai *guessing* yang masih dalam batas wajar (sekitar 0.11–0.15), sedangkan item X1, X0.3, dan X1.1 menunjukkan nilai *guessing* yang terlalu tinggi (0.35 hingga 0.64), mengindikasikan bahwa soal tersebut berpotensi ditebak, sehingga perlu peninjauan ulang. Pada aspek kesulitan, sebagian besar item berada dalam rentang normal, dari yang agak mudah (X0 dengan -0.624) hingga cukup sulit (X1.1 dengan 0.712), menunjukkan bahwa soal memiliki tingkat kesulitan yang variatif dan seimbang. Namun, pada parameter diskriminasi, hampir semua item menunjukkan nilai yang tidak wajar, dengan kisaran ekstrem dari 12 hingga 232. Nilai yang terlalu tinggi ini mengindikasikan potensi ketidaksesuaian model, kemungkinan *overfitting*, atau masalah pada data, seperti jumlah responden yang terbatas atau butir soal yang terlalu ekstrem. Nilai diskriminasi ideal berada antara 0.35 hingga 2.0; oleh karena itu, temuan ini menunjukkan bahwa model atau data perlu ditelaah kembali. Secara keseluruhan, hasil ini mengindikasikan bahwa meskipun sebagian butir memiliki tingkat kesulitan dan tebakan yang wajar, aspek diskriminasi perlu perhatian khusus agar validitas dan reliabilitas instrumen tetap terjaga. Perlu dilakukan perbaikan pada konstruksi item atau re-estimasi model dengan pendekatan statistik yang lebih stabil. Berdasarkan penjabaran tersebut instrumen menunjukkan variasi tingkat kesulitan dan peluang tebakan yang wajar, namun nilai diskriminasi yang terlalu tinggi mengindikasikan perlunya revisi soal atau re-estimasi model agar instrumen lebih valid dan andal dalam mengukur pemahaman sistem persamaan linier. Visualisasi grafiknya ditunjukkan pada gambar 1.1 berikut



Hasil analisis rata-rata jawaban benar menunjukkan bahwa sebagian besar butir soal berada pada tingkat kesulitan sedang hingga mudah. Item X1 memiliki nilai mean tertinggi (0,754), menunjukkan bahwa soal tersebut paling mudah di antara yang lain. Sebaliknya, item X0.3 memiliki nilai mean terendah (0,551), menandakan soal tersebut relatif paling sulit. Sementara itu, item lainnya berada pada kisaran 0,58 hingga 0,63. Secara keseluruhan, distribusi tingkat kesulitan soal cukup merata, meskipun dominan pada kategori mudah. Hal ini menunjukkan bahwa instrumen sudah cukup sesuai, namun perlu ditambahkan soal dengan tingkat kesulitan lebih tinggi untuk keseimbangan.

4. Kesimpulan

Instrumen diagnostik yang dikembangkan untuk mengukur pemahaman mahasiswa terhadap Sistem Persamaan Linier telah memenuhi sebagian besar kriteria validitas, dengan korelasi antar butir yang signifikan dan model IRT 3PL yang lebih sesuai dibanding model lainnya. Meskipun sebagian besar soal memiliki tingkat kesulitan dan peluang tebakan yang wajar, beberapa item menunjukkan nilai diskriminasi yang sangat tinggi, mengindikasikan potensi *overfitting*. Rata-rata jawaban benar menunjukkan soal berada pada tingkat kesulitan sedang hingga mudah. Perbaikan pada soal dengan nilai diskriminasi ekstrem dan re-estimasi model diperlukan untuk meningkatkan validitas dan keandalan instrumen.

5. Referensi

- Rohendi, D., Ramadhan, M. O., Rahim, S. S. A., & Zulnaldi, H. (2025). Enhancing student's interactivity and responses in learning geometry by using augmented reality. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(1), em2559–em2559.
- De Ayala, R. J. (2022). Appendix A. Maximum likelihood estimation of person locations. In *The theory and practice of item response theory* (2nd ed., pp. 77). Appendices A–G.
- History, A. (2024). *Jurnal Kependidikan*, 10(3), 847–858.
- Marsigit, M., Retnawati, H., Apino, E., Santoso, R. H., Arlinwibowo, J., Santoso, A., et al. (2020). Constructing mathematical concepts through external representations utilizing technology: An implementation in IRT course. *TEM Journal*, 9(1), 317–326.

- Hidajat, F. A., Haeruman, L. D., Wiraningsih, E. D., & Pambudi, D. S. (2023). The effect of digital technology learning based on guided discovery and self-regulated learning strategy on mathematical creativity. *International Journal of Information and Education Technology*, 13(3), 535–543.
- Sa'adah, N., Faizah, S., Sa'dijah, C., Khabibah, S., & Kurniati, D. (2023). Students' mathematical thinking process in algebraic verification based on crystalline concept. *Mathematics Teaching-Research Journal*, 15(1), 90–107.
- Reeve, B. B. (2020). Item response theory (IRT). In *Polling America: An encyclopedia of public opinion* (2nd ed., Vols. 1–2, pp. 343–345).



© 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution Share Alike (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).