

METODE PENGUKURAN RASCH DALAM PEMBELAJARAN FISIKA

Godelfridus Hadung Lamanepa

godelfridus29lamanepa@gmail.com

Abstrak

Dalam usaha mengubah paradigma pendekatan penilaian dari *classical test theory* (CTT) menjadi *item respon theory* (IRT), maka penelitian ini berfokus pada estimasi kemampuan individu (Mahasiswa) menurut pengukuran rasch model. Komponen yang ditinjau yakni kemampuan mahasiswa dan tingkat kesukaran tiap aitem tes. Metode penelitian ini dilakukan dengan tes yang diberikan kepada mahasiswa pendidikan fisika semester V, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Katolik Widya Mandira. Proses analisis dilakukan dengan mengurutkan kemampuan mahasiswa dan parameter tingkat kesukaran dalam skala *logit*nya masing-masing. Hasil analisis menunjukkan bahwa rata-rata kemampuan mahasiswa dalam mengerjakan tes Optika memiliki *logit* lebih tinggi yakni (0,76 *logit*) dari rata-rata *logit* kesukaran aitem tes atau *mean item* (0,00 *logit*). Analisis ini mampu mengklasifikasikan kemampuan tiap-tiap individu dan kesulitan aitem-aitem tes dengan lebih tepat.

Kata Kunci: Rasch model, IRT, CTT, Pembelajaran Fisika.

Abstract

In an effort to change the paradigm of the assessment approach from classical test theory (CTT) to item response theory (IRT), this study focuses on estimating individual abilities (students) according to the Rasch model measurement. The components reviewed are the student's ability and the level of difficulty of each test item. This research method was carried out with tests given to physics education students in the fifth semester, Faculty of Teacher Training and Education, Widya Mandira Catholic University. The analysis process is carried out by sorting students' abilities and the level of difficulty parameters in their respective logit scales. The results of the analysis show that the average logit of students' ability in taking the Optics test is higher, namely (0.76 logit) than the average logit of the test item difficulty or the mean item (0.00 logit). This analysis is able to classify each individual's ability and difficulty of test items more precisely.

Keywords: rasch model, IRT, CTT, physics learning.

PENDAHULUAN

Prestasi atau hasil belajar siswa dari tahun ke tahun selalu dikaitkan dengan fakta bahwa mata pelajaran tertentu sulit, media belajar yang kurang dan metode pengajaran yang belum sesuai serta faktor lainnya. Namun, beberapa peneliti (Siri & Freddano, 2011), (Gugiu & Gugiu, 2013), (Caldwell & Pate, 2013) telah mengidentifikasi faktor lain yang mempengaruhi hasil belajar atau kinerja siswa tersebut. Menjadi perhatian atau penekananan diantara penelitian mereka adalah selain faktor-faktor yang disebutkan di atas, faktor lain yang mempengaruhi hasil belajar adalah sifat dari butir/aitem tes dan karakteristik peserta tes (individu), hasil tes pada suatu aitem dapat diprediksi atau dijelaskan oleh kemampuan individu dan karakteristik aitem.

Penelitian ini bertujuan mengestimasi hasil tes siswa melalui pendekatan IRT dengan 1 parameter (model rasch), yang diasumsikan bahwa abilitas individu tidak bergantung pada aitem tes yang diberikan begitu juga sebaliknya, sukar atau mudah suatu aitem tidak ditentukan oleh abilitas individu. Motivasi utama dari penelitian pada mahasiswa pendidikan fisika adalah mengidentifikasi kemampuan aktual mahasiswa sebelum pelaksanaan perkuliahan optika, pemetaan topik yang dianggap sukar, mengukur efektivitas berlangsungnya pembelajaran yang dilaksanakan, memperluas penggunaan model rasch dalam analisis kemampuan dan aitem serta menyiapkan penilaian formatif yang lebih baik selama pembelajaran karena estimasi kemampuan individu dan aitem tes dengan model rasch akan meningkatkan reliabilitas standar tes (Lee et al., 2008) selain itu model rasch dapat menghasilkan data kemampuan siswa yang

lebih akurat (Saidudin et al., 2010), karena penilaian yang akurat diyakini memberikan informasi lebih lengkap untuk proses perbaikan.

Sebuah tes dapat dipelajari dari perspektif yang berbeda dan aitem dalam tes tersebut dapat dievaluasi menurut teori berbeda. Teori yang dimaksud adalah teori tes klasik (CTT) dan teori respon aitem (IRT), kedua teori ini menjadi kerangka acuan dalam pengukuran pendidikan untuk mengembangkan, mengevaluasi dan mempelajari aitem tes. Kedua pendekatan ini didasarkan pada asumsi yang berbeda dan menggunakan pendekatan statistik yang berbeda pula (O. A. & E. R. I., 2016). Kedua teori ini tidak terbatas sampai pada mengembangkan, mengevaluasi, atau menentukan reliabilitas dan validitas tes tetapi juga meningkatkan kualitas item secara holistik, (Magno, 2009).

Dalam teori pengukuran klasik, terjadi fenomena dimana gambaran kemampuan orang dan parameter butir tidak dapat dipisahkan (*inseparability of person and item parameters*) yang membuat pengukuran menjadi subjektif, (Nurchahyo, 2017). Dalam arti lain, ketika tes yang diberikan terlalu rumit, sedikit saja individu yang berhasil mengerjakannya, namun ketika tes yang diberikan itu mudah, maka banyak individu akan berhasil. Oleh karena itu dibutuhkan sebuah teknik analisis yang membuat pengukuran menjadi objektif dengan cara memisahkan antara kemampuan individu dengan dengan karakteristik tesnya. Tes objektif yang dimaksud menurut adalah tes berdiri secara terpisah dengan individu yang dites sehingga memenuhi sifat independen (Sumintono, 2018) serta bersifat unidimensi (Planinic et al., 2019).

Analisis rasch model didasarkan pada model stokastik atau probabilitistik yang menggambarkan interaksi individu peserta tes dan aitem tes dengan memperhitungkan dua parameter yakni rasio antara kemampuan individu dan tingkat kesulitan aitem tes, (Mahmud et al., 2013) (Souza et al., 2017). Kemampuan individu tidak mengacu pada kemampuan intelektual umum seseorang tetapi mengacu hanya sejauh sifat laten (*laten trait*) yang diselidiki, misalnya pengetahuan tentang Optik, pengetahuan tentang mekanika dan sebagainya.

Pada model rasch, aitem (pertanyaan) dalam tes diurutkan secara hirarki menurut tingkat kesulitannya dan secara bersamaan individu diatur secara hirarki menurut kemampuannya, (Lange et al., 2004). Pemisahan tersebut menggunakan pendekatan probabilitistik dimana skor mentah individu pada tes diubah menjadi rasio jawaban benar dan salah, selanjutnya menjadi peluang logaritma dari jawaban benar dinyatakan dalam skala *logit* sebagai gambaran kemampuan individu. Untuk aitem dalam tes dengan cara yang sama, yakni memeriksa proporsi aitem yang dijawab salah, kemudian diubah menjadi rasio dan peluang logaritma dari aitem yang dijawab salah. *Logit* diplot dalam satu skala. Skor *logit* individu kemudian dapat digunakan sebagai perkiraan kemampuan individu, skor *logit* aitem digunakan sebagai perkiraan tingkat kesulitan aitem tersebut (Mahmud et al., 2013). Hubungan kedua parameter ini dinyatakan (Bond dan Fox, 2007) dalam (Sumintono, 2014) sebagai berikut:

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

Dimana: $P_{ni}(X_{ni} = 1 / \beta_n, \delta_i)$ adalah peluang individu n dalam aitem i untuk jawaban benar ($x=1$); nilai e (*base of natural logarithm*) atau konstanta euler= 2,7183 dan kemampuan individu β_n dan tingkat kesulitan aitem δ_i . Persamaan di atas dapat disederhanakan melalui fungsi logaritma dan menjadikannya sebagai berikut:

$$\ln(P_{ni}(X_{ni} = 1 / \beta_n, \delta_i)) = \beta_n - \delta_i$$

Sehingga peluang individu menjawab benar suatu aitem dapat dinyatakan: Peluang menjawab benar = (Kemampuan individu dikurangi tingkat kesulitan aitem).

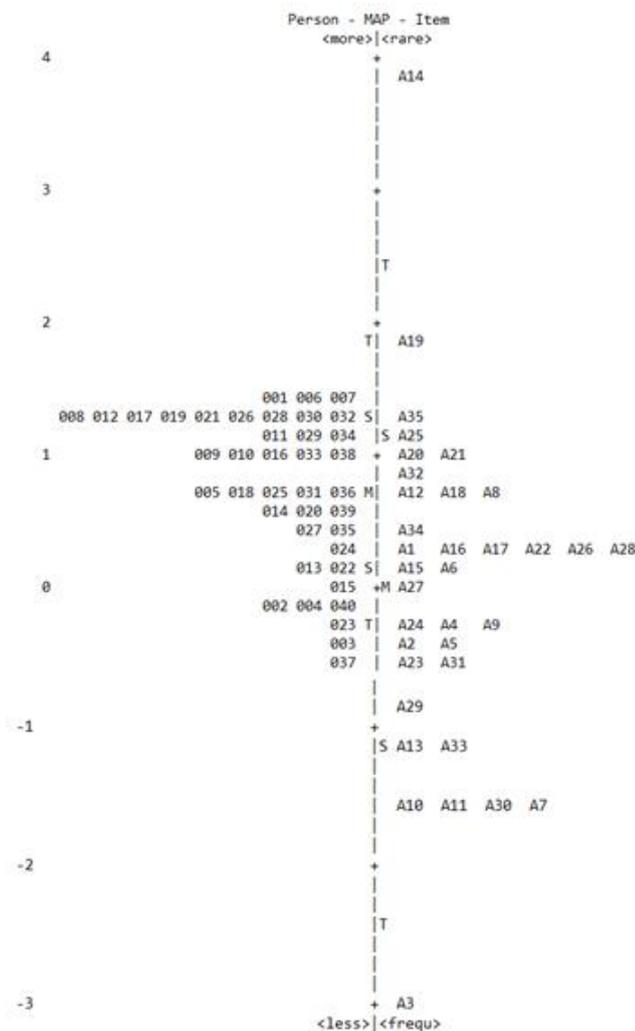
Untuk mengidentifikasi kemampuan individu dan aitem tes dalam model rasch digunakan alat ukur (matriks Guttman) yang sudah mempunyai metrik yang tetap dan interval yang sama dalam satuan *logit* (Arsad et al., 2013). Nilai probabilitas individu menjawab dengan benar suatu aitem tes antara 0 sampai 1, (Planinic et al., 2019). Individu dengan *logit* kemampuan yang lebih tinggi dari *logit* aitem mempunyai probabilitas yang lebih besar dari 0.5 ($P > 0.5$) untuk menjawab dengan benar aitem tersebut. Jika *logit* kemampuan individu sama dengan *logit* aitem maka probabilitas menjawab dengan benar item tersebut sama dengan 0.5 ($P = 0.5$).

METODE PENELITIAN

Tahapan awal penelitian diawali dengan identifikasi aitem soal menurut taksonomi Bloom. Identifikasi ini diperlukan untuk mendapatkan variasi kesulitan aitem. Aitem yang diperlukan dalam penelitian mencakup aitem yang mengukur aspek pengetahuan (20%) aspek penerapan dan analisis (60%) serta aspek analisis dan sistesis (20%) dari total aitem tes. metode penelitian ini adalah deskriptif yakni mendeskripsikan kemampuan aktual mahasiswa dan tingkat kesulitan aitem-aitem tes. Subyek dalam penelitian yakni mahasiswa semester V Program Studi Pendidikan Fisika, Universitas Katolik Widya Mandira. Instrumen dalam penelitian ini adalah soal/aitem tes berjumlah 35 butir. Langkah selanjutnya, hasil tes diinput dalam dan dianalisis menurut pengukuran rasch model menggunakan software Winstep versi 3.73.

HASIL DAN PEMBAHASAN

Gambaran tentang abilitas atau kemampuan mahasiswa pada topik Optika menurut rasch model didasarkan pada kemampuan mereka dalam menjawab benar setiap aitem tes. Pada gambar 1, menunjukkan kemampuan individu (mahasiswa) dan tingkat kesulitan aitem diplot dalam skala yang sama yakni skala logit. Peta persebaran individu dan aitem (Person-item distribution map) menggambarkan bagaimana individu dengan kemampuan β sebagai ciri laten merespon aitem i yang memiliki tingkat kesulitan δ . Dari gambar 1, diperoleh informasi pencapaian atau kemampuan aktual individu sesuai tes yang diberikan.



Gambar 1. Peta Persebaran Kemampuan Individu dan Tingkat Kesulitan aitem

Gambar 1 di atas adalah peta Wrihgt yang menunjukkan persebaran kemampuan mahasiswa dan tingkat kesukaran aitem setiap aitem tes. Paling kiri adalah skala logit dari (-3 sampai +4). Skala logit merupakan skala interval dimana setiap logit unit memiliki nilai yang sama, selanjutnya kemampuan mahasiswa dan tingkat kesulitan aitem dinyatakan dalam skala logit. Semakin positif nilai logit dari mahasiswa maka makin besar pula peluang mahasiswa tersebut menjawab dengan benar setiap aitem tes (Arsad et al., 2013). Hal yang sama terjadi pada aitem tes, jika logit aitem tes lebih tinggi dari logit mahasiswa maka peluang menjawab dengan benar suatu aitem menjadi lebih kecil.

Nilai rata-rata logit aitem (*Mean item*) selalu ditetapkan dalam 0,0 *logit* (M+). Dari gambar 1, diperoleh informasi rata-rata *logit* mahasiswa yang mengikuti tes mata kuliah optika didapati 0,76 *logit* (M) yang berarti di atas rata-rata logit aitem, yang menunjukkan bahwa kemampuan rata-rata mahasiswa yang mengikuti tes di atas rata-rata tingkat kesulitan aitem. Mahasiswa dengan kode 001, 006 dan 007 dengan 1,47 *logit* dipastikan mempunyai peluang menjawab dengan benar semua aitem kecuali aitem A19 (1, 79 *logit*) dan A14 (3,82 *logit*) yang mempunyai kesulitan lebih tinggi dari kemampuan mahasiswa.

Mahasiswa dengan kode 037 memiliki logit paling rendah dari teman-temannya yang mengikuti tes. Gambar 1 menunjukkan bahwa siswa tersebut mempunyai peluang yang baik untuk mengerjakan dengan benar 8 (delapan) aitem tes yakni A3, A10, A11, A30, A7, A13, A33 dan A29, karena logit kemampuan siswa tersebut lebih tinggi dari aitem-aitem yang disebutkan di atas. Informasi tentang individu dan item lainnya dapat dibaca satu persatu seperti ditunjukkan pada gambar di atas. Dengan informasi yang lengkap seperti gambar 1, guru atau dosen mudah untuk mengelompokkan aitem dan individu sesuai tingkat pencapaiannya secara akurat dengan demikian dapat mengungkapkan tingkat kemampuan aktual individu meskipun dengan jumlah sedikit sampel (Arsad et al., 2013). Melalui pengukuran yang sama ini pula pengukuran dengan rasch model dapat dalam penelitian pendidikan fisika berguna untuk diagnostik kemampuan siswa, konstruksi soal serta evaluasi hasil belajar siswa, (Planinic et al., 2019), (Zamora-Araya et al., 2018). Informasi lain dari gambar 1 di atas adalah M-S-T (*mean, 1SD dan 2SD*) terlihat bahwa sebaran kesukaran aitem tes (bagian kanan) terlihat lebih lebar jika dibandingkan dengan kemampuan siswa. Dalam konteks ini, tingkat kesulitan aitem tes memiliki keragaman lebih besar, sehingga aitem-aitem tes bisa memberikan informasi yang diperlukan dalam konteks *assessment for learning*.

Tabel 1 menunjukkan ringkasan informasi tentang pola respon mahasiswa secara keseluruhan, kualitas instrumen yang digunakan juga informasi tentang interaksi antara mahasiswa dan aitem tes.

Tabel 1. Ringkasan statistik

SUMMARY OF 40 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	22.6	35.0	.76	.40	1.00	.0	.99	.0
S.D.	3.6	.0	.55	.02	.18	1.1	.40	1.0
MAX.	27.0	35.0	1.47	.44	1.52	2.8	2.48	2.8
MIN.	14.0	35.0	-.51	.38	.76	-1.7	.64	-1.3
REAL RMSE	.42	TRUE SD	.35	SEPARATION	.85	Person	RELIABILITY	.42
MODEL RMSE	.40	TRUE SD	.37	SEPARATION	.92	Person	RELIABILITY	.46
S.E. OF Person MEAN = .09								
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .50								
SUMMARY OF 35 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	25.9	40.0	.00	.41	.99	.0	.99	.0
S.D.	7.8	.0	1.18	.14	.11	.8	.22	1.0
MAX.	39.0	40.0	3.82	1.02	1.17	1.4	1.41	1.8
MIN.	2.0	40.0	-3.06	.33	.73	-2.3	.53	-2.4
REAL RMSE	.44	TRUE SD	1.10	SEPARATION	2.52	Item	RELIABILITY	.86
MODEL RMSE	.43	TRUE SD	1.10	SEPARATION	2.57	Item	RELIABILITY	.87
S.E. OF Item MEAN = .20								

Dari nilai *person reliability* 0,42 menunjukkan bahwa konsistensi jawaban mahasiswa lemah karena konsistensi baik, jika *person reliability* lebih besar dari 0,67 (Rashid & Abdullah, 2008) dan kualitas butir-butir dalam instrumen aspek reliabilitasnya 0,86 yang berarti baik. Hasil analisis yang ditampilkan pada gambar 1, menggambarkan kemampuan aktual mahasiswa pada topik Optika yang diukur, sedangkan pada Tabel 1 menunjukkan bahwa secara umum instrumen tes Optika mempunyai kualitas yang baik atau sesuai yang diharapkan.

KESIMPULAN

Hasil analisis dalam penelitian ini menunjukkan bahwa kemampuan aktual mahasiswa dalam kaitan pengetahuan Optika paling tinggi adalah 1,47 *logit* dan paling rendah -0,51 *logit* dengan rata-rata logit seluruh mahasiswa 0,76 *logit* lebih tinggi dari rata-rata tingkat kesulitan seluruh aitem tes. Meskipun demikian, ditemukan bahwa terdapat beberapa mahasiswa (gambar 1) memiliki logit dibawah rata-rata tingkat kesulitan siswa. Penggunaan pendekatan IRT dalam penelitian ini memberikan akses informasi lebih lengkap tentang kemampuan aktual mahasiswa serta alat ukur yang dipakai sehingga perancangan metode perkuliahan dapat disesuaikan. Tidak hanya itu tetapi penggunaan IRT dalam pengembangan alat ukur pembelajaran fisika perlu ditingkatkan.

DAFTAR PUSTAKA

- Arsad, N., Kamal, N., Ayob, A., Sarbani, N., Tsuey, C. S., Misran, N., & Husain, H. (2013). Rasch model analysis on the effectiveness of early evaluation questions as a benchmark for new students ability. *International Education Studies*, 6(6), 185–190. <https://doi.org/10.5539/ies.v6n6p185>
- Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *American Journal of Pharmaceutical Education*, 77(4). <https://doi.org/10.5688/ajpe77471>
- Gugiu, M. R., & Gugiu, P. C. (2013). Utilizing Item Analysis to Improve the Evaluation of Student Performance. *Journal of Political Science Education*, 9(3), 345–361. <https://doi.org/10.1080/15512169.2013.796248>
- Lange, R., Greiff, W. R., Moran, J., Ferro, L., & acl. (2004). A probabilistic Rasch analysis of question answering evaluations. *Hlt-Naacl 2004: Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*, 65–72.
- Lee, Y. J., Palazzo, D. J., Warnakulasooriya, R., & Pritchard, D. E. (2008). Measuring student

- learning with item response theory. *Physical Review Special Topics - Physics Education Research*, 4(1). <https://doi.org/10.1103/PhysRevSTPER.4.010102>
- Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11.
- Mahmud, Z., Ghani, N. A. M., & Rahim, R. A. (2013). Assessing Students' Learning Ability in a Postgraduate Statistical Course: A Rasch Analysis. *Procedia - Social and Behavioral Sciences*, 89, 890–894. <https://doi.org/10.1016/j.sbspro.2013.08.951>
- Nurchahyo, F. A. (2017). Aplikasi IRT dalam Analisis Aitem Tes Kognitif. *Buletin Psikologi*, 24(2), 64–75. <https://doi.org/10.22146/buletinpsikologi.25218>
- O. A., A., & E. R. I., A. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal*, ESJ, 12(28), 263. <https://doi.org/10.19044/esj.2016.v12n28p263>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Rashid, R. A., & Abdullah, R. (2008). Application of Rasch-based ESPEGS Model in Measuring Generic Skills of Engineering Students: A New Paradigm. *Advances in Engineering Education*, 5(8), 591–602.
- Saidfudin, M., Azrilah, A. A., Rodzo'An, N. A., Omar, M. Z., Zaharim, A., & Basri, H. (2010). Easier learning outcomes analysis using rasch model in engineering education research. *International Conference on Engineering Education and International Conference on Education and Educational Technologies - Proceedings, March 2014*, 442–447.
- Siri, A., & Freddano, M. (2011). The use of item analysis for the improvement of objective examinations. *Procedia - Social and Behavioral Sciences*, 29, 188–197. <https://doi.org/10.1016/j.sbspro.2011.11.224>
- Souza, M. A. P., Coster, W. J., Mancini, M. C., Dutra, F. C. M. S., Kramer, J., & Sampaio, R. F. (2017). Rasch analysis of the participation scale (P-scale): usefulness of the P-scale to a rehabilitation services network. *BMC Public Health*, 17(1), 934. <https://doi.org/10.1186/s12889-017-4945-9>
- Sumintono, B. (2014). Model Rasch untuk Penelitian Sosial Kuantitatif. *Makalah Kuliah Umum Di Jurusan Statistika, ITS Surabaya, 21 November 2014, November 201*, 1–9. <http://deceng3.wordpress.com>
- Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*. <https://doi.org/10.2991/icei-17.2018.11>
- Zamora-Araya, J. A., Smith-Castro, V., Montero-Rojas, E., & Moreira-Mora, T. E. (2018). Advantages of the Rasch Model for Analysis and Interpretation of Attitudes: the Case of the Benevolent Sexism Subscale. *Revista Evaluar*, 18(3). <https://doi.org/10.35670/1667-4545.v18.n3.22201>