
ANALISIS SOAL PILIHAN GANDA DENGAN RASCH MODEL

¹Untung Kurniawan, ²Kusuma Dewi Kris Andriyani

¹Badan Pusat Statistik Kabupaten Klaten

²Badan Pusat Statistik Kota Surakarta

Email : untungk@bps.go.id

ABSTRAK

Proses belajar mengajar selalu melibatkan penilaian sebagai hal yang sangat penting dilakukan, tanpa hal tersebut sulit diketahui secara pasti apakah kemajuan belajardicapai atau tidak. Suatu bentuk ujian yang bisa menginformasikan prestasi antar siswa atau antar sekolah bahkan antar daerah memerlukan tes standar yang baik. Bentuk tes standar yang umum dipakai adalah jenis soal pilihan ganda dengan satu pilihan jawaban yang benar. Cara yang umum dilakukan adalah menjumlahkan skor jawaban yang benar untuk menggambarkan kemampuan siswa. Pada dasarnya penggunaan skor sebagai ukuran prestasi memiliki kelemahan yang tidak bisa mendukung umpan balik yang efektif. Makalah ini bertujuan memberikan informasi tingkat abilitas siswa dan tingkat kesulitan soal dengan Rasch model. Hasil Rasch model diperoleh siswa dengan abilitas tertinggi (05P) sampai siswa dengan abilitas terendah (07L) dan tingkat kesulitan soal dari yang paling sulit (S3) sampai ke yang paling mudah (S1).

Kata kunci : Abilitas Siswa, Rasch Model, Soal Pilihan Ganda.

PENDAHULUAN

Evaluasi adalah proses yang tidak terpisahkan dari kegiatan pembelajaran. Ujian adalah prosedur evaluasi yang dilakukan oleh seorang pengajar terhadap pengetahuan dan keterampilan siswa untuk mengetahui capaian kinerjanya. Bentuk soal yang paling populer digunakan adalah pilihan ganda. Hampir semua ujian umumnya menggunakan pendekatan skor untuk menjelaskan pencapaian prestasi siswa. Cara yang umum dilakukan adalah menjumlahkan skor jawaban yang benar. Jumlah skor yang benar dianggap sebagai kemampuan siswa.

Pada dasarnya penggunaan skor sebagai ukuran prestasi memiliki kelemahan, diantaranya adalah skor mentah pada dasarnya bukanlah hasil pengukuran, skor mentah memiliki makna kuantitatif yang lemah, skor

mentah tidak menunjukkan kemampuan seseorang terhadap tugas, dan skor mentah dan persentase jawaban yang benar tidak selalu bersifat linier [1].

Karena itu, pendekatan yang berbeda dengan penggunaan skor mentah sangat diperlukan dalam evaluasi pembelajaran. Pemodelan Rasch tidak hanya mengukur jumlah jawaban benar yang didapatkan siswa saja, namun saat yang sama juga menghitung probabilitas *odd ratio* untuk setiap butir soal yang dikerjakan.

Pemodelan Rasch akan membuat hubungan hierarki antara peserta ujian (*person*) dan butir soal yang digunakan. Karena dihasilkan skala interval yang sama dengan satuan *logit* yang sama untuk *person* dan butir soal, maka kedua hal ini dapat dibandingkan secara langsung yang menghasilkan informasi yang lebih lengkap tentang tes yang

dilakukan dan abilitas siswa yang mengerjakan tes.

Makalah ini bertujuan untuk memberikan informasi secara akurat abilitas yang dimiliki siswa dan pada saat yang sama juga menentukan kualitas soal yang diberikan melalui pendekatan Rasch model.

METODE PENELITIAN

Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder hasil ujian siswa-siswi di Bimbingan Belajar New Neutron Klaten. Sampel sebanyak 10 siswa dengan soal pilihan ganda.

Metode Analisis

Pemodelan Rasch pertama kali dibuat oleh Dr. Georg Rasch pada tahun 1950, ahli matematika dari Denmark. Rasch mengembangkan model matematika yang dapat mengukur hubungan probabilistik antara kemampuan seseorang dan tingkat kesulitan soal dengan menggunakan fungsi logaritma untuk menghasilkan pengukuran dengan interval yang sama [5]. Hasilnya adalah satuan baru yang disebut *logit* (*log odds* unit) yang menunjukkan abilitas siswa dan kesulitan soal. Nilai *logit* yang diperoleh dapat disimpulkan bahwa tingkat kesuksesan siswa dalam mengerjakan soal sangat tergantung dari tingkat abilitasnya dan tingkat kesulitan soal [4].

Untuk data yang berbentuk dikotomi, pemodelan Rasch menggabungkan suatu algoritma yang menyatakan hasil ekspektasi probabilistik dari item 'i' dan responden 'n', yang secara matematis dinyatakan sebagai [2]:

$$P_{ni}(x_{ni} = 1 | b_n, d_i) = \frac{e^{(b_n - d_i)}}{1 + e^{(b_n - d_i)}} \quad (1)$$

dimana: $P_{ni}(x_{ni} = 1 | b_n, d_i)$ adalah probabilitas dari responden n dalam butir i untuk menghasilkan jawaban benar ($x=1$); dengan kemampuan responden,

b_n , dan tingkat kesulitan butir d_i . Dengan memasukkan fungsi logaritma pada 1 menjadi:

$$\log(P_{ni}(x_{ni} = 1 | b_n, d_i)) = b_n - d_i \quad (2).$$

Keberhasilan dapat dituliskan sebagai: Probabilitas untuk berhasil = kemampuan responden - tingkat kesulitan butir.

Kriteria yang digunakan untuk memeriksa kesesuaian butir soal dan person yang tidak sesuai (*outliers* atau *misfit*) [3] adalah:

- Nilai *Outfit means-square* (MNSQ) yang diterima: $0,5 < MNSQ < 1,5$.
- Nilai *Outfit Z-standard* (ZSTD) yang diterima: $-2,0 < ZSTD < +2$
- Nilai *Point Measure Correlation*: $0,4 < Pt Measure Corr < 0,85$

Langkah-langkah dalam melakukan analisis dengan Rasch model secara garis besar adalah sebagai berikut:

- Menyiapkan data mentah skor per orang dan data skor per butir soal dalam bentuk Microsoft Excel.
- Transformasi file Microsoft Excel ke dalam bentuk file Formatted Text (*space delimited*).
- Membuat probabilitas *Odd-Ratio* untuk skor per orang dan skor per butir soal.
- Merubah probabilitas *Odd-Ratio* untuk skor per orang dan skor per butir soal ke dalam fungsi logaritma.
- Membuat mistar *logit* Rasch dari fungsi logaritma *Odd-Ratio*.
- Melakukan interpretasi hasil.

Pengolahan Rasch model menggunakan *Software* Winsteps versi 3.73.

HASIL PENELITIAN

Salah satu keistimewaan pemodelan Rasch adalah menghasilkan skala yang kualitasnya menyamai pengukuran pada dimensi fisik dalam fisika, seperti mengukur panjang dengan mistar sentimeter ataupun mengukur berat

dengan neraca kilogram, dalam hal ini hasil yang didapat bisa dibandingkan karena mempunyai satuan yang sama, bersifat linier, dan mempunyai interval yang sama.

Analisi Butir Soal

Untuk mengetahui data tentang tingkat kesulitan soal dapat dilihat dari output *Item Measure*. Pada tabel 1 terlihat beberapa kolom yang memberikan informasi tentang tiap butir soal. Tabel ini diurutkan sesuai dengan tingkat kesulitannya, yaitu berdasarkan nilai *measure* yang tidak lain adalah nilai *logit* dari tiap butir soal. Nilai *logit* yang tertinggi menunjukkan tingkat kesulitan soal yang tinggi. Hal ini berkorespondensi dengan kolom total score, yaitu yang menyatakan berapa jumlah jawaban yang benar. Misalnya, untuk soal ke-3 nilai *logit*-nya adalah 3.91 *logit* dan hanya dua orang saja yang menjawab dengan benar. Soal ke-2, ke-7 dan ke-9 mempunyai nilai *logit* yang sama sebesar -0.15, dan mempunyai jawaban benar sama sebanyak 6 orang.

Tabel 1. Tingkat Kesulitan Butir Soal

Entry Number	Total Score	Measure	Item
3	2	2.03	S3
10	3	1.37	S10
8	4	0.82	S8
5	5	0.33	S5
2	6	-0.15	S2
7	6	-0.15	S7
9	6	-0.15	S9
4	7	-0.67	S4
6	8	-1.27	S6
1	9	-2.15	S1
Mean	5.6	0.00	
S.D.	2.1	1.16	

Tingkat kesulitan soal dapat dilihat pada kolom item, soal dengan tingkat kesulitan paling tinggi adalah soal S3 sampai soal paling mudah yaitu soal S1. Dalam hal tingkat kesulitan, jika soal S3 (+2.03*logit*) dibandingkan soal S10 (+1.37*logit*), bisa dikatakan bahwa soal

S3 tingkat kesulitannya hampir satu setengah kali S2.

Perbandingan ini sedikit berbeda jika melihatnya dari skor yang didapat, untuk soal S3 jumlah jawaban yang benar ada 2 orang, sedangkan soal S10 ada 3 orang. Selisih yang kecil ini tidak bisa disimpulkan secara langsung karena pemodelan Rasch menganggap bahwa data skor 1 tidak mempunyai jarak yang sama. Algoritma perhitungan melalui probabilitas *odd ratio* dan transformasi *logit*-lah yang bisa mengatakan secara pasti tingkat kesulitan soal dalam interval yang sama.

Pengelompokan tingkat kesulitan soal dilakukan dengan menggunakan nilai deviasi standar (SD) dikombinasikan dengan nilai rata-rata *logit*. Misalkan 0.0 *logit* +1SD adalah satu kelompok soal sulit (S5 dan S8), lebih besar dari +1SD adalah soal yang sangat sukar (S3 dan S10), 0.0 *logit* -1SD adalah soal yang mudah (S2, S4, S7 dan S9), dan lebih kecil dari -1SD soal yang sangat mudah (S1 dan S6).

Tabel 2 menunjukkan kualitas kesesuaian butir dengan model (*item fit*). *Item fit* menjelaskan apakah butir soal berfungsi normal melakukan pengukuran atau tidak. Jika ada soal yang tidak fit, hal ini mengindikasikan bahwa terjadi miskonsepsi pada siswa terhadap butir soal. Informasi ini sangat berguna bagi guru untuk memperbaiki kualitas pengajarannya sehingga miskonsepsi bisa dihindari pada pelajaran berikutnya. Jika butir soal pada ketiga kriteria tersebut tidak terpenuhi, dapat dipastikan dipastikan bahwa butir soalnya kurang bagus sehingga perlu diperbaiki ataupun diganti. Hal ini akan menjamin di kemudian hari bahwa tingkat pemahaman siswa memang diuji melalui butir-butir soal yang sesuai dan berkualitas.

Tabel 2. Tingkat Kesesuaian Butir Soal

Entry	OUTFIT		PT-MEASURE	Item
	MNSQ	ZSTD	CORR.	
1	2.34	1.5	0.24	S1
2	1.52	1.3	0.27	S2
7	1.52	1.2	0.33	S7
3	1.46	0.9	0.35	S3
10	0.66	0.2	0.66	S10
8	0.86	-0.2	0.52	S8
15	0.51	-0.4	0.71	S5
6	0.59	-1.1	0.65	S6
4	0.52	-0.8	0.72	S4
9	0.52	-0.8	0.72	S9

Pada tabel 2 bahwa butir soal S1, S2, dan S7 cenderung tidak fit. Jika dilihat dari ketiga kriteria, butir soal S1, S2 dan S7 tidak memenuhi syarat *Outfit Mnsq* dan *Point Measure Correlation*, namun untuk kriteria *OutfitZ-standard* nilainya masih dalam batas wajar yang diperbolehkan. Karena itu butir soal S1, S2, dan S7 tidak perlu diubah. Untuk butir-butir soal yang lain hanya tidak memenuhi pada satu kriteria saja, sehingga kesimpulan akhirnya tidak ada soal yang perlu diubah.

Suatu pengukuran yang valid, salah satu ukurannya adalah instrumen dan butir-butir soal yang digunakan tidak mengandung bias. Suatu butir soal disebut bias jika didapati bahwa salah satu individu dengan karakteristik tertentu lebih diuntungkan dibandingkan individu dengan karakteristik lain. Misalnya, butir soal yang lebih mudah dijawab oleh siswa laki-laki dibandingkan perempuan, yang hal ini menunjukkan adanya butir soal yang bias gender. Suatu butir soal dikatakan mengandung bias jika nilai probabilitas butirnya di bawah 5%.

Tabel 3. Deteksi Bias Butir Soal dengan *Differential Item Functioning*

Item Number	Prob.	Name
1	0.8562	S1
2	0.2692	S2
3	0.5300	S3
4	1.0000	S4
5	1.0000	S5
6	1.0000	S6
7	0.2692	S7
8	0.9236	S8
9	1.0000	S9
10	1.0000	S10

Tabel 3 menunjukkan informasi deteksi adanya butir soal yang bias terhadap jenis kelamin. Semua butir soal mempunyai nilai probabilitas diatas 5%, hal ini menunjukkan bahwa seluruh butir soal tidak ada bias gender.

Analisis Abilitas Siswa

Dalam konteks pembelajaran, selain melakukan analisis butir soal, maka yang perlu juga penting dilakukan adalah analisis abilitas siswa dalam mengerjakan soal. Analisis abilitas siswa akan banyak membantu pengajar untuk bisa lebih efektif membantu proses pembelajaran. Melalui abilitas ini bisa dipetakan kemampuan siswa berdasarkan tingkat abilitasnya.

Tabel 4. Tingkat Abilitas Individu

Entry Number	Total Score	Measure	Person
5	9	2.66	05P
6	8	1.72	06P
3	6	0.52	03P
4	6	0.52	04L
9	6	0.52	09P
1	5	0.01	01L
8	5	0.01	08P
2	4	-0.51	02L
10	4	-0.51	10L
7	3	-1.06	07L
Mean	5.6	0.39	
S.D.	1.7	1.05	

Tabel 4 pada kolom person dapat dilihat abilitas siswa, yang diurutkan dari abilitas tinggi (05P) menuju ke abilitas paling rendah (07L). Nilai *logit* yang tinggi menunjukkan tingkat kemampuan menyelesaikan soal yang tinggi. Hal ini

berkorespondensi dengan kolom total score, yaitu yang menyatakan berapa jumlah jawaban yang benar. Misalnya untuk 05P, nilai *logit*-nya adalah 2.66 *logit*, siswa ini dapat menyelesaikan soal dengan benar sebanyak 9 dari 10 soal yang diberikan. Siswa 03P, 04L, dan 09P mempunyai nilai *logit* yang sama yaitu 0.52 *logit*, ketiganya hanya dapat menjawab benar 6 dari 10 soal yang diberikan.

Kolom *measure* merupakan nilai *logit* dari masing-masing siswa dapat digunakan untuk perbandingan kemampuan siswa. Dilihat dari abilitasnya maka terlihat bahwa siswa 05P mempunyai kemampuan kurang lebih satu setengah kali 06P. Nilai *logit* yang sama pada siswa 03P, 04L, dan 09P menunjukkan abilitas yang sama.

Dengan diketahuinya nilai deviasi standar, maka informasi ini bisa digunakan untuk pengelompokan abilitas siswa. Misalnya 0.39 *logit* + 1SD adalah kelompok sedang (03P, 04L, dan 09P), 0.39 *logit* + 2SD adalah kelompok tinggi (06P), 0.39 *logit* + 3SD adalah kelompok sangat tinggi (05P), 0.39 *logit* - 1SD adalah kelompok rendah (01L, 08P, 02L, dan 10L), dan 0.39 *logit* - 2SD adalah kelompok sangat rendah (07L).

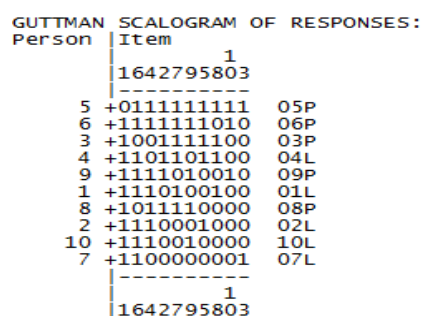
Tabel 5. Tingkat Kesesuaian Individu

OUTFIT		PT-MEASURE	Person
MNSQ	ZSTD	CORR.	
9.90	3.3	-0.62	05P
2.55	1.9	0.26	07L
1.37	0.8	0.26	03P
0.93	0.0	0.45	04L
0.80	-0.3	0.53	09P
0.83	-0.3	0.56	08P
0.75	-0.5	0.59	01L
0.64	-0.7	0.66	02L
0.48	-0.4	0.61	06P
0.53	-1.1	0.74	10L

Selain bisa memetakan kemampuan siswa sesuai kepentingan pengelompokan prestasi, pemodelan Rasch juga bisa

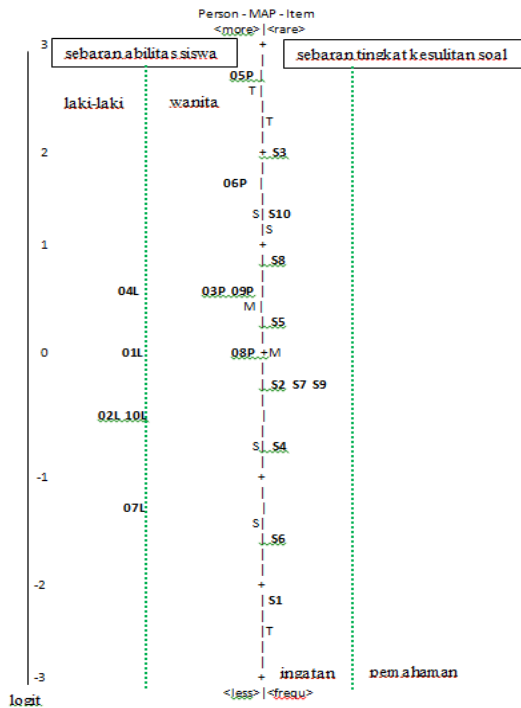
mendeteksi individu yang mempunyai pola respon yang tidak sesuai. Artinya adanya ketidaksesuaian jawaban yang diberikan berdasarkan abilitasnya dibandingkan model ideal. Hal ini bisa digunakan oleh pengajar untuk mengetahui konsistensi berpikir siswa maupun dapat digunakan untuk mengetahui jika terdapat kecurangan yang dilakukan. Tabel 5 diurutkan menurut tingkat ketidaksesuaian individu dengan model, terdapat satu orang siswa yang pola responnya dinilai tidak fit, yaitu siswa 05P. Nilai *outfit means-square*, *outfit z-std*, dan nilai *point-measure correlation* di luar batas yang diperbolehkan.

Informasi pola respon yang tidak biasa ini bisa diketahui lebih jauh dengan melihat pada skalogram.



Gambar 1. Skalogram guttman pola respon siswa

Gambar 1 menunjukkan pola respon secara sistematis antara siswa (diurutkan dari yang abilitas tinggi ke yang rendah secara vertical, atas ke bawah) dan butir soal (diurutkan dari yang mudah ke yang susah secara horizontal, kiri ke kanan). Abilitas setiap siswa dapat dianalisis lebih lanjut, siswa 05P termasuk kategori siswa yang tidak cermat, dimana soal pertama termudah tidak bisa mengerjakan dengan benar, padahal soal yang lebih sulit berikutnya bisa dikerjakan dengan benar semua. Siswa 07L hanya bisa mengerjakan dua soal termudah pertama dan menebak soal yang paling sulit dengan benar.



Gambar 2. Peta Wright person item

Gambar 2 peta Wright person-item sebelah kiri yang menggambarkan abilitas siswa terlihat ada satu siswa yang mempunyai abilitas tertinggi (05P). Nilai *logit* siswa ini di luar batas dua deviasi standar (T) yang menunjukkan kecerdasan yang tinggi yang berbeda (*outlier*). Peta Wright sebelah kanan menjelaskan nilai *logit* butir soal. Soal S3 adalah soal dengan tingkat kesukaran tertinggi, yang bermakna probabilitas siswa mengerjakan soal ini dengan benar kecil sekali. Adapun soal S1 merupakan soal yang paling mudah.

KESIMPULAN

Analisis soal pilihan ganda dengan Rasch model dapat menghasilkan suatu skala pengukuran dengan interval yang sama. Karena skor mentah tidak memiliki sifat keintervalan, maka skor ini tidak bisa digunakan secara langsung untuk memberikan penafsiran kemampuan siswa. Rasch model secara bersama-sama menggunakan data skor berdasarkan per orang (*person*) maupun data skor per

butir soal (*item*). Kedua skor ini menjadi basis untuk mengestimasi skor murni (*true score*) yang menunjukkan tingkat kemampuan individu serta tingkat kesulitan butir.

Hasil Rasch model diperoleh siswa dengan abilitas tertinggi (05P) sampai siswa dengan abilitas terendah (07L) dan tingkat kesulitan soal dari yang paling sulit (S3) sampai ke yang paling mudah (S1). Pengelompokan abilitas siswa menjadi 5 kelompok, yaitu siswa dengan abilitas sangat rendah, rendah, sedang, tinggi dan sangat tinggi. Sedangkan pengelompokan tingkat kesulitan soal terbagi menjadi 4 kelompok yaitu, tingkat kesulitan soal sangat mudah, mudah, sulit, dan sangat sulit.

UCAPAN TERIMA KASIH

Terima kasih kepada New Neutron Klaten untuk data penelitian.

DAFTAR PUSTAKA

[1] Alagumalai, S., Curtis, D.D. dan Hungi, N. 2005. *Applied Rasch Measurement: book of exemplars. papers in honour of John P. Keeves.* Dordrecht: Springer.

[2] Bond, T.G., dan Fox, C. 2015. *Applying the Rasch Model. Fundamental measurement in the Human Sciences.* 3rd edition. New York: Routledge.

[3] Boone, W. J., Staver, J.R. and Yale, M.S. 2014. *Rasch Analysis in the Human Sciences.* Dordrecht: Springer.

[4] Englehard, G. 2013. *Invariant Measurement, using rasch models in the social, behavioral and health sciences.* New York: Routledge.

[5] Sumintono, B dan Widhiarso, W. 2015. *Aplikasi Pemodelan Rasch pada Assesment Pendidikan.* Cimahi: Trim Komunikata Publishing House.