

INFO ARTIKEL

Riwayat Artikel:

Diterima : 12 Januari 2020

Disetujui : 15 Februari 2020

PENDIDIKAN

ANALISIS SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN GEOGRAFI DENGAN MENGGUNAKAN PEMODELAN RASCH**Rusiyah¹, Sunarty S. Eraku², Supadmi³**¹⁻² Jurusan Ilmu dan Teknologi Kebumian, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo³ Madrasah Aliyah Negeri 1 Limboto, Provinsi Gorontalo(✉) rusiyah@ung.ac.id**ABSTRAK**

Penilaian merupakan bagian penting dalam pendidikan. Tes merupakan salah satu teknik penilaian yang biasa digunakan oleh pendidik atau guru. Instrumen penilaian hasil belajar berkualitas apabila mampu memberikan gambaran kompetensi peserta didik yang sebenarnya. Penelitian ini bertujuan untuk mengetahui tingkat kesulitan butir soal, tingkat kesesuaian butir soal, bias butir soal, tingkat abilitas siswa, tingkat kesesuaian individu dan kualitas instrumen. Teknik pengumpulan data menggunakan studi dokumentasi yaitu lembar jawaban soal ujian akhir semester mata pelajaran geografi semester ganjil tahun ajaran 2017/2018 kelas X di Madrasah Aliyah Negeri 1 Limboto. Berdasarkan analisis dengan Pemodelan Rasch diperoleh kesimpulan: (1) Soal dengan kategori sulit 48%, mudah 28%, sangat sulit 12%, sangat mudah juga 12%. Soal yang sangat sulit yaitu S13 dan soal yang sangat mudah yaitu S2, (2) Soal S13, S15, dan S6 tidak memenuhi kriteria nilai MNSQ dan *Pt Measure Corr*, tetapi nilai ZSTQ masih dapat memenuhi kriteria, (3) Butir soal S9 nilai probabilitas kurang dari 0,05 sehingga mengandung bias karena soal tersebut mudah dijawab oleh siswa laki-laki dibandingkan oleh siswa perempuan, (4) Abilitas siswa termasuk dalam kategori sedang 63%, rendah 18,5%, sangat rendah 14,8% dan sangat rendah sekali 3,7%. Siswa yang memiliki abilitas sangat rendah sekali 03L, (5) Siswa 13L tidak memenuhi ketiga kriteria nilai MNSQ, ZSTD dan *Pt Measure Corr*, sehingga memiliki respon yang tidak sesuai, (6) *Person reliability* termasuk dalam kategori bagus dan kualitas butir-butir soal reliabilitasnya cukup. Rata-rata *logit person* - 0,04 menunjukkan bahwa prestasi rata-rata siswa di bawah rata-rata tingkat kesukaran standar soal.

Kata Kunci: Kesulitan butir soal, abilitas siswa, kualitas instrumen, pemodelan Rasch**PENDAHULUAN**

Penilaian merupakan kegiatan yang penting dalam pembelajaran. Melaksanakan penilaian merupakan salah satu tugas pokok guru atau pendidik. Hasil pembelajaran dapat memberikan informasi keberhasilan program pembelajaran. Informasi keberhasilan pembelajaran sangat berguna bagi siswa, guru, orang tua dan sekolah. Mardapi (2012) berpendapat bahwa upaya meningkatkan kualitas pendidikan dapat ditempuh

melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaiannya. Oleh karena itu, seorang guru atau pendidik harus memiliki kompetensi melakukan penilaian untuk menghasilkan penilaian yang berkualitas.

Salah satu bentuk evaluasi yang biasa dilaksanakan di lembaga pendidikan yaitu evaluasi sumatif. Tes sumatif dilaksanakan pada akhir program pengajaran seperti ujian akhir semester. Evaluasi sumatif bertujuan untuk mengetahui

tingkat keberhasilan peserta didik dalam kurun waktu tertentu. Alat ukur yang dipakai untuk melakukan penilaian adalah instrumen. Jenis instrumen tes yang umum digunakan adalah tes pilihan ganda (*multiple choice*).

Menurut para ahli tes ini yang paling baik dalam mengukur berbagai macam tujuan pengajaran. Penskorannya mudah dan sampel materi yang diukur lebih luas (Suyanto dan Djihad, 2012). Hasil Pembelajaran yang baik dapat terwujud apabila proses pembelajaran dan instrumen penilaian berkualitas. Menurut Arikunto (2009) sebuah tes dikatakan baik sebagai alat pengukur harus memenuhi persyaratan, yaitu memiliki validitas, realibilitas, objektivitas, praktikalitas, dan ekonomis.

Instrumen penilaian hasil belajar yang berkualitas apabila mampu memberikan gambaran kompetensi siswa atau peserta didik yang sebenarnya. Untuk mengetahui kualitas instrumen penilaian dapat dilakukan dengan melakukan analisis soal. Selain itu analisis soal juga untuk mengetahui pemahaman dan kemampuan siswa untuk mater-materi yang telah diajarkan sehingga guru dapat memberi umpan balik dalam memperbaiki pembelajaran di kelas.

Analisis yang banyak digunakan dalam bidang pendidikan umumnya menggunakan pendekatan teori tes klasik atau *classical test theori*. Pendekatan klasik ini memiliki kelemahan sehingga diperlukan pendekatan yang berbeda. Pemodelan Rasch memiliki kelebihan memberikan informasi yang sangat lengkap dari abilitas yang dimiliki oleh peserta didik dan pada saat yang sama juga menentukan kualitas soal. Tujuan utamanya adalah menghasilkan suatu skala pengukuran dengan interval yang sama yang nantinya dapat memberikan informasi secara akurat tentang peserta tes maupun kualitas soal yang dikerjakan (Sumintono dan Widhiarso, 2015).

Menurut Sumintono dan Widhiarso (2014) Pemodelan Rasch juga memiliki banyak kelebihan yaitu dapat mengakomodasi pendekatan probabilitas dalam memandang atribut seluruh objek ukur, mengatasi masalah perbedaan metrik

antarbutir, mengatasi masalah keintervalan data, cukup tahan terhadap data hilang, dan telah memenuhipeukuran yang objektif. Oleh karena itu, pendidik diharapkan menggunakan model Rasch dalam menganalisis hasil ujian peserta didik agar dapat memberikan gambaran kemampuan peserta didik yang sebenarnya.

METODOLOGI PENELITIAN

Penelitian ini menggunakan metode penelitian deskriptif kuantitatif. Teknik Pengumpulan data menggunakan tehnik dokumentasi berupa lembar jawaban ujian akhir semester ganjil tahun ajaran 2017/2018 kelas X siswa Madrasah Aliyah Negeri 1 Limboto pada mata pelajaran geografi yang berjumlah 27 siswa. Setiap lembar jawaban siswa terdiri dari 25 soal pilihan ganda. Pengolahan data dalam pemodelan Rasch menggunakan *Software Ministep*.

HASIL DAN PEMBAHASAN

A. Analisis Butir Soal

1. Analisis Kesulitan Butir Soal

Tingkat kesulitan butir soal dapat diketahui berdasarkan output *table Item Measure*. *Item Measure* menunjukkan nilai logit untuk setiap item yang diurutkan dari yang tertinggi ke yang terendah. Nilai logit menunjukkan tingkat kesulitan butir soal. Semakin tinggi nilai logit maka tingkat kesulitan butir soal juga semakin tinggi. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik (Arifin, 2012).

Soal ke-13 atau S13 memiliki nilai logit paling tinggi yaitu 1,35. Hal ini menunjukkan bahwa butir soal S13 merupakan soal yang tersulit. Soal ke-20 atau S20 memiliki nilai logit paling rendah yaitu 3,27. Hal ini menunjukkan bahwa S20 merupakan butir soal paling mudah. Nilai logit berkorespondensi dengan jumlah siswa yang mampu menjawab dengan benar (total score). S13 hanya dapat dijawab dengan benar oleh 7 siswa (26%), sedangkan S20 dapat dijawab dengan benar oleh 25 siswa (93%).

Tingkat kesulitan soal dapat dikelompokkan berdasarkan nilai deviasi standar (SD) dan nilai rata-rata *logit*. Soal dengan nilai *logit* 0.0 *logit* +1SD termasuk dalam kelompok sulit 48%. Soal dengan nilai *logit* 0.0 *logit* -1SD termasuk dalam kelompok mudah 28%. Soal dengan nilai *logit* lebih besar 1SD termasuk dalam kelompok sangat sulit 12%. Soal dengan nilai *logit* lebih kecil -1SD termasuk dalam kelompok sangat mudah juga 12%.

Tabel 1. Tingkat Kesulitan Butir Soal

<i>Entry Number</i>	<i>Total Score</i>	<i>Measure</i>	<i>Item</i>
13	7	1.35	S13
11	8	1.13	S11
18	8	1.13	S18
9	9	.93	S9
2	10	.73	S2
6	10	.73	S6
23	10	.73	S23
24	10	.73	S24
3	11	.54	S3
4	12	.36	S4
7	12	.36	S7
15	12	.36	S15
22	12	.36	S22
25	12	.36	S25
17	13	.17	S17
10	15	-.22	S10
5	16	-.42	S5
16	16	-.42	S16
21	16	-.42	S21
1	17	-.62	S1
8	18	-.84	S8
12	18	-.84	S12
14	20	-1.32	S14
19	21	-1.60	S19
20	25	-1.27	S20
Mean	13.5	00.0	
S. D.	4.4	1.01	

Sumber : Analisis Data, 2018

Menurut Arifin (2012), tingkat kesukaran soal yang ideal untuk meningkatkan koefisien reliabilitas adalah soal yang menghasilkan sebaran skor berbentuk kurva normal. Untuk memperoleh prestasi belajar yang baik, proporsi antara tingkat kesukaran soal tersebar secara normal. Dengan demikian dapat dikatakan bahwa soal yang disusun oleh guru geografi proporsi tingkat kesulitan butir soal sebarannya belum normal.

2. Analisis Tingkat Kesesuaian Butir Soal

Untuk mengetahui tingkat kesesuaian butir soal (*item fit*) menggunakan kriteria yang dikemukakan oleh Boone et al., (dalam Sumintono dan Whidhiarso, 2015) yaitu:

- a) Nilai *outfit mean square* (MNSQ) yang diterima: $0.5 < MNSQ < 1.5$
- b) Nilai *Outfit Z-Standard* (ZSTD) yang diterima $-2,0 < ZSTD < +2,0$
- c) Nilai *Point Measure Correlation* (*Pt Mean Corr*) $0,4 < Pt Measure Corr < 0,85$.

Butir soal yang tidak memenuhi ketiga kriteria tersebut berarti butir soal tersebut tidak bagus sehingga harus diperbaiki. Nilai MNSQ, ZSTD dan *Pt Measure Corr* disajikan pada Tabel 2.

Tabel 2. Nilai MNSQ, ZSTD dan Pt Measure Corr

<i>Entri</i>	<i>Outfit</i>		<i>Measure Corr</i>	<i>Item</i>
	MNSQ	ZSTD		
13	1.95	1.4	-.02	S13
15	1.68	1.7	.15	S15
1	1.59	1.7	.48	S1
6	1.52	1.2	.32	S6
2	1.35	.9	.19	S2
9	1.27	.7	.19	S9
20	1.22	.6	.40	S20
21	1.17	.6	.35	S21
22	1.06	.3	.35	S22
16	.97	.0	.46	S16
23	.90	-.1	.41	S23
4	.96	.0	.44	S4
17	.93	-.1	.48	S17
14	.80	-.4	.55	S14
18	.94	.1	.38	S18
8	.92	-.1	.54	S8
11	.73	-.3	.43	S11
5	.87	-.4	.56	S5
10	.84	-.4	.56	S10
7	.75	-.6	.55	S7
24	.65	-.8	.56	S24
12	.65	-1.1	.66	S12
3	.64	-.9	.60	S3
25	.60	-1.1	.65	S25
19	.39	-1.5	.77	S19
MEAN	1.01	.0		
SD	.37	.8		

Sumber: Analisis Data, 2018

Berdasarkan hasil analisis yang disajikan pada Tabel 2 menunjukkan semua soal (100%) memenuhi kriteria nilai ZSTD. Soal S13, S15, S1, S6 dan S19 tidak memenuhi nilai MNSQ (20%). Soal S13, S15, S6, S2, S9, S20, S21, S22 dan S18 tidak memenuhi nilai *PtMeasure Corr* (36%). S13, S15, dan S6 (12%) tidak memenuhi kriteria nilai MNSQ dan *Pt Measure Corr*, tetapi nilai ZSTD masih dapat memenuhi kriteria. Oleh karena itu dapat diperoleh kesimpulan soal S13, S15, S6 tidak perlu diperbaiki karena masih memenuhi satu kriteria.

3. Analisis Bias Butir Soal

Butir soal yang baik tidak boleh mengandung bias. Suatu soal dikatakan mengandung bias apabila butir soal tersebut menguntungkan kelompok tertentu dibandingkan kelompok lain. Untuk mengetahui apakah butir soal mengandung bias atau tidak dapat dilihat nilai probabilitasnya. Nilai probabilitas disajikan pada Tabel 3.

Tabel 3. Tabel Nilai Probabilitas

<i>Person Classes</i>	Probabilitas	<i>Item Number</i>	<i>Name</i>
2	. 2278	1	S1
2	. 3612	2	S2
2	. 7463	3	S3
2	. 8884	4	S4
2	. 8739	5	S5
2	. 4237	6	S6
2	. 3273	7	S7
2	. 3944	8	S8
2	. 0403	9	S9
2	. 7941	10	S10
2	. 9434	11	S11
2	. 3944	12	S12
2	. 7552	13	S13
1	1.0000	14	S14
2	. 0784	15	S15
2	. 4699	16	S16
2	. 5562	17	S17
2	. 9434	18	S18
2	. 7562	19	S19
1	1.0000	20	S20
2	. 8739	21	S21
2	. 0784	22	S22
2	. 4237	23	S23
2	. 4237	24	S24
2	. 4790	25	S25

Sumber: Analisis Data, 2018

Berdasarkan Tabel 3. menunjukkan deteksi butir soal yang bias terhadap jenis kelamin. Nilai probabilitas butir Soal S9 adalah 0,040. Nilai probabilitas S9 kurang dari 5% (0,05). Soal S9 menunjukkan bahwa terdapat bias jender. Grafik bias butir soal pada Gambar 1 menggambarkan bias jender tiap butir soal. Kurva yang berada dibagian atas menunjukkan butir soal yang sulit dijawab siswa, sedangkan kurva yang berada dibawah menunjukkan butir soal yang mudah dijawab oleh siswa.

Garis berwarna biru menunjukkan kemampuan menjawab siswa berjenis kelamin laki-laki dan garis berwarna merah menunjukkan kemampuan menjawab siswa berjenis kelamin perempuan. Berdasarkan grafik dapat disimpulkan bahwa butir soal S9 mengandung bias karena butir soal S9 mudah dijawab oleh siswa laki-laki dibandingkan oleh siswa perempuan. Oleh karena itu dapat disimpulkan bahwa butir S9 perlu diperbaiki agar tidak merugikan kelompok jender tertentu.

B. Analisis Abilitas Siswa

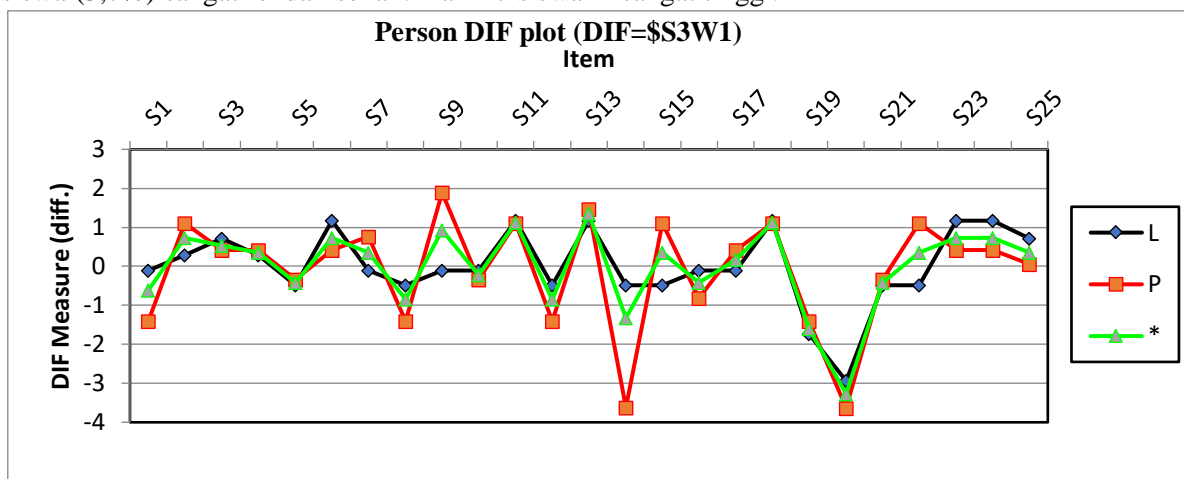
1. Analisis Tingkat Abilitas Siswa

Gambaran tingkat abilitas siswa diurutkan dari siswa yang memiliki abilitas tertinggi sampai abilitas terendah (Tabel 4.). Terdapat 6 (enam) siswa yang memiliki nilai *logit* yang sama yaitu 06L, 07L, 15P, 19P, 21P dan 24P (22%) dengan nilai *logit* 1.15. Hal ini menunjukkan bahwa ke 6 siswa tersebut memiliki kemampuan yang sama dalam menjawab soal.

Nilai *logit* menunjukkan kemampuan siswa dalam menyelesaikan soal. Nilai *logit* berkorespondensi dengan jumlah jawaban yang benar. Enam siswa tersebut mampu menjawab 18 jawaban yang benar. Siswa yang memiliki abilitas terendah adalah 03L dengan nilai *logit* -3.77 dan hanya 1 jawaban yang benar dari 25 soal yang ada.

Tingkat abilitas siswa dapat dikelompokkan berdasarkan nilai rata-rata *logit person* dan nilai deviasi standar (Tabel 5). Nilai rata-rata *logit* -0.40 dan nilai deviasi standar 1.2. Dari hasil pengelompokan tersebut didapatkan 17 siswa (63 %) siswa memiliki abilitas sedang, 5 siswa (18,5%)

abilitas rendah, 4 siswa (14,8%) sangat rendah dan tidak ada yang memiliki abilitas tinggi maupun 1 siswa (3,7%) sangat rendah sekali. Dari 27 siswa sangat tinggi.



Gambar 1. Grafik Bias Butir Soal

Tabel 4. Tingkat Abilitas Siswa

Entry Number	Total Score	Meansur	Person
5	18	1.15	06L
7	18	1.15	07L
15	18	1.15	15P
19	18	1.15	19P
21	18	1.15	21P
24	18	1.15	24P
23	17	.93	23P
27	17	.93	27P
10	16	.73	10P
17	16	.73	17P
26	16	.73	26P
9	15	.54	09L
12	15	.54	12L
16	15	.54	16P
25	15	.54	25P
22	14	.35	22P
13	13	.16	13L
2	9	-.63	02L
11	9	-.63	11L
18	8	-.84	18P
1	7	-1.07	01L
14	7	-1.07	12L
5	6	-1.33	05L
8	6	-1.33	08L
20	5	-1.61	20P
4	3	-2.36	04L
3	1	-3.77	03L
EAN	12.5	-.04	
S.D.	5.3	1.25	

Sumber: Analisis Data, 2018

Tabel 5. Persentase Abilitas Siswa

Person	Persen	Kategori
06L,07L,15P,19P,21P,24P,23P,27P,10P,17P,26P,09L,12L,16P,25P,22P,13L.	63%	Sedang
02L,11L,18P,01L,12L.	18.5%	Rendah
05L,08L,20P	14.8%	Sangat Rendah
04L.		
03L	3.7%	Sangat rendah sekali

Sumber: Analisis Data, 2018

2. Analisis Kesesuaian Individu

Analisis kesesuaian individu bertujuan untuk mengetahui individu yang responnya tidak sesuai. Siswa yang memiliki respon tidak sesuai apabila jawaban yang diberikan tidak sesuai dengan model ideal berdasarkan abilitas siswa.

Untuk mengetahui tingkat kesesuaian individu (*item fit*) menggunakan kriteria yang dikemukakan oleh Boone et al. , 2014 (dalam Sumintono dan Widhiarso,2015) yaitu:

- Nilai *outfit mean square* (MNSQ) yang diterima: $0,5 < MNSQ < 1,5$
- Nilai *Outfit Z-Standard* (ZSTD) yang diterima $-2,0 < ZSTD < +2,0$

c. Nilai *Point Measure Correlation (Pt Mean Corr)*
 $0,4 < Pt \text{ Measure Corr} < 0,85$.

Tabel 6. Tingkat Kesesuaian Individu

Entri	Outfit		Pt-Measure Corr	Person
	MNSQ	ZSTD		
13	1.55	4.3	-.23	13L
4	1.03	1.0	.26	04L
12	1.29	1.2	.07	12L
3	1.29	.5	.13	03L
20	.98	.7	.40	20P
18	1.07	.7	.34	18P
11	1.03	.5	.37	11L
2	1.11	.4	.33	02L
14	1.04	.4	.38	14L
27	1.05	.0	.29	27P
21	1.01	.3	.27	21P
1	1.02	.2	.40	01L
24	1.01	.0	.30	24P
5	.97	-.3	.47	05L
22	.96	-.3	.42	22P
26	.94	-.3	.41	26P
7	.92	-.3	.39	07L
10	.91	-.3	.42	10L
15	.91	-.4	.41	15P
23	.91	-.4	.42	23P
17	.90	-.4	.44	17P
9	.89	-.5	.46	09L
19	.88	-.4	.42	19P
8	.88	-.4	.54	08L
6	.85	-.6	.46	06L
16	.79	-.9	.55	16P
25	.79	-.9	.55	25P
ME	1.00	.0		
AN				
S.D.	.16	1.0		

Sumber: Analisis Data, 2018

Berdasarkan nilai MNSQ, ZSTD dan *Pt Measure Corr* maka 3,7% atau 1 siswa tidak memenuhi ketiga kriteria. Siswa tersebut adalah 13L Siswa yang tidak memenuhi nilai *Pt Measure Corr* sebanyak 11,1% atau 3 siswa yaitu 04L, 12L, 03L. Untuk melihat gambaran lebih rinci mengenai siswa yang memiliki kecenderungan respon tidak fit dapat dilihat pada *Guttman Scalogram Of Responses*:



Gambar 2. Hasil *Guttman Scalogram of Responses*

Hasil *output Guttman Scalogram of responses* menunjukkan urutan pola respon siswa mulai siswa yang memiliki abilitas tertinggi ke siswa yang abilitas terendah (secara vertikal dari atas ke bawah). Urutan soal diurutkan mulai soal yang termudah sampai yang paling sukar secara horisontal dari kiri ke kanan. Angka 0 menunjukkan bahwa siswa jawabannya salah dan angka 1 menunjukkan siswa menjawab soal dengan benar.

Analisis menunjukkan bahwa siswa 13L tidak cermat dalam mengerjakan soal karena siswa tersebut dapat mengerjakan dengan benar soal yang paling sukar tetapi tidak dapat mengerjakan soal yang paling mudah. Siswa 04L memiliki kemampuan yang sangat rendah karena hanya mampu menjawab benar sebanyak 3 soal. Siswa 04L hanya mampu mengerjakan soal termudah ke 1, tidak mampu mengerjakan soal termudah ke dua sampai termudah ke 13, tetapi mampu menjawab

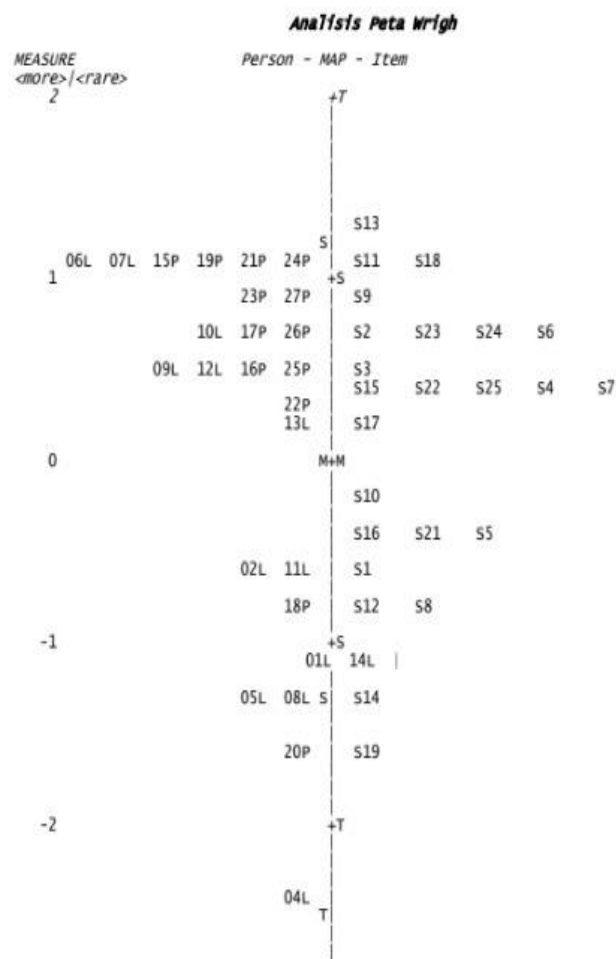
soal termudah ke 14 dan termudah ke 19 dengan benar. Hal ini menunjukkan adanya indikasi tebakan. Siswa 12L merupakan Siswa yang tidak cermat karena tidak mampu mengerjakan soal yang lebih mudah(termudah ke 3, 4, 5) tetapi mampu mengerjakan soal yang memiliki tingkat kesulitan yang lebih tinggi (soal S13). Siswa 03L memiliki kecerdasan sangat rendah sekali karena hanya mampu mengerjakan dengan benar satu soal yaitu soal ke 6 dan tidak mampu mengerjakan soal ke 1,2,3,4 dan 5 yang lebih mudah. Berdasarkan uraian tersebut dapat disimpulkan bahwa siswa S13L memiliki respon yang tidak sesuai. Berdasarkan analisis ini guru dapat melakukan penyelidikan lebih lanjut.

3. Analisis Instrumen

Berdasarkan *output table summary statistick* dapat diperoleh informasi tentang *person reliability*, *item reliability*, nilai *alpha cronbach* dan *person measure*. *Person reliability* 0,82 termasuk dalam kategori bagus. Hal ini menunjukkan *person reliability* siswa bagus. *Item reliability* pada ujian sebesar 0,77 menunjukkan bahwa kualitas butir-butir soal aspek realibilitasnya cukup. *Person measure* atau rata-rata *logit person* -0,04 menunjukkan rata-rata nilai seluruh siswa dalam mengerjakan butir-butir soal. Nilai rata-rata *logit person* lebih rendah dari 0,0 menunjukkan bahwa prestasi rata-rata siswa di bawah rata-rata tingkat kesukaran standar soal. Nilai *alpha Cronbach* menunjukkan besarnya interaksi antara person dan butir-butir soal secara keseluruhan. Nilai *alpha Cronbach* 0,83 dalam kategori bagus sekali.

Peta Wright pada Gambar 2 menggambarkan sebaran abilitas siswa sekaligus sebaran tingkat kesulitan butir soal. Bagian kanan menggambarkan abilitas siswa Semakin ke bawah posisi siswa pada Peta Wright menunjukkan abilitas siswa rendah. Terdapat 1 siswa yang memiliki nilai *logit* lebih kecil -3 *logit* yang menunjukkan siswa tersebut memiliki kemampuan sangat rendah sekali (03L). Peta Wright sebelah kanan menunjukkan sebaran tingkat kesulitan soal. Soal yang memiliki tingkat kesulitan paling tinggi berada dibagian atas sedangkan soal yang berada dibagian paling bawah kesulitan paling rendah. Rata-rata *logit person* -0,04

menunjukkan bahwa prestasi rata-rata siswa di bawah rata-rata tingkat kesukaran standar soal. Berdasarkan analisis ini dapat memberikan informasi bagi guru dalam memperbaiki proses pembelajaran dan perbaikan kualitas instrumen selanjutnya.



Gambar 2. Peta Wright

SIMPULAN

1. Tingkat kesulitan butir soal dapat dikelompokkan berdasarkan nilai deviasi standar dan nilai rata-rata *logit*. Kelompok butir soalsulit 48%, mudah 28%, sangat sulit 12%, sangat mudah juga 12%.
2. Soal S13, S15, dan S6 tidak memenuhi kriteria nilai MNSQ dan *Pt Measure Corr*, tetapi nilai ZSTQ masih dapat memenuhi kriteria sehingga tidak perlu diperbaiki karena masih memenuhi satu kriteria
3. Butir soal S9 nilai probabilitas kurang dari 0,05 sehingga mengandung bias karena butir soal S9

mudah dijawab oleh siswa laki-laki dibandingkan oleh siswa perempuan. Oleh karena itu, butir S9 perlu diperbaiki supaya tidak merugikan kelompok jender tertentu.

4. Abilitas siswa termasuk dalam kategori sedang 63%, rendah 18,5%, sangat rendah 14,8% dan sangat rendah sekali 3,7%. Siswa yang memiliki abilitas sangat rendah sekali 03L.
5. Siswa 13L tidak memenuhi ketiga kriteria nilaiMNSQ, ZSTD dan *Pt Measure Corr* sehingga memiliki respon yang tidak sesuai.
6. *Person reliability* siswa bagus, kualitas butir-butir soal dalam instrument aspek realibilitasnya cukup. Nilai *alpha Cronbach* 0,83 dalam kategori bagus sekali. Rata-rata *logit person*-0,04 menunjukkan bahwa prestasi rata-rata siswa di bawah rata-rata tingkat kesukaran standar soal.

DAFTAR PUSTAKA

- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam. Kementerian Agama.
- Arikunto, S. (2009). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Mardapi, D. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta: Nuha Litera.
- Sumintono, B & Widhiarso, W. (2014). *Aplikasi Model Rasch Untuk Ilmu-Ilmu Sosial*. Cimahi: Trim Komunikata.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata.
- Suyanto, dan Djihad, A. (2012). *Bagaimana Menjadi Calon Guru dan Guru Profesional*. Yogyakarta: Multi Presindo.