

Analisis Butir Soal Penilaian Akhir Semester MI Menggunakan Rasch Model

Zainul Anwar¹, Yuni Masrifatin², Nurani Rahmania³

¹Universitas Islam Negeri Sunan Kalijaga Yogyakarta,
Indonesia, ^{2,3}STAI Miftahul Ula Nganjuk Indonesia

Email: anwarzein205@gmail.com, yunimasrifatin@gmail.com, nuranirahmania@gmail.com

Abstrak

Penilaian merupakan satu dari sekian pendekatan yang digunakan untuk mengevaluasi hasil dari belajar siswa serta memutuskan ketercapaian tujuan dalam proses pembelajaran tersebut. Berangkat dari permasalahan tersebut, penelitian ini ingin mengetahui validitas, reliabilitas serta tingkat kesulitan butir soal menggunakan Rasch Model. Dalam pengumpulan, data peneliti menggunakan teknik dokumentasi yang meliputi lembar jawaban Penilaian Akhir Semester genap 2022/2023 kelas V di salah satu MI Kabupaten Nganjuk, dengan total 30 responden serta 15 tes pilihan ganda dengan 4 opsi jawaban. Pada penelitian ini analisa data menggunakan bantuan Winstep 3.7.3. Dengan hasil analisa dilihat dari validitas, 14 soal valid (93,3%) dan 1 soal tidak valid (6,3%); dengan nilai Cronbach Alpha sebesar 0,27 dengan interpretasi buruk, nilai Person Reliability 0,1 dengan interpretasi lemah serta nilai Item Reliability 0,83 dengan interpretasi bagus; tingkat kesulitan butir soal telah sesuai dengan kaidah pengembangan instrumen tes, dikarenakan terdapat penyebaran kesulitan butir soal tersebut meliputi 4 butir soal (26,66%) dengan kriteria sangat sulit, 2 butir soal (13,33%) dengan kriteria sulit, 4 butir soal (26,66%) mudah serta 5 butir soal (33,33%) dengan kriteria sangat mudah. Hasil penelitian ini dapat ditarik kesimpulan butir soal Penilaian Akhir Semester genap mata pelajaran matematika kelas V MI memiliki pola item yang fit. Sehingga, dapat diasumsikan butir soal Penilaian Akhir Semester genap mata pelajaran matematika kelas V MI dapat mengukur apa yang ingin diukur.

Kata Kunci: Validitas, Reliabilitas, Tingkat Kesulitan, Rasch Model

PENDAHULUAN

Penilaian merupakan satu dari bentuk evaluasi hasil belajar peserta didik serta mampu memutuskan tercapai atau tidak tujuan pembelajaran tersebut. Guna mengetahui suatu pembelajaran tersebut apakah telah dilakukan sesuai dengan tujuan pembelajaran, maka penting adanya evaluasi (Purniasari & Masykuri, 2021). Evaluasi hasil dari belajar merupakan rangkaian kegiatan yang diimplementasikan guru dalam rangka memperoleh serta menganalisis data mengenai proses pembelajaran yang telah dilaksanakan dalam waktu tertentu (Agustian, Mufidah, Chakra Setiawan, Suklani, & Chakra Setiawan, 2023). Evaluasi yang tepat tentunya akan menginterpretasikan hasil yang baik pula dalam konteks mutu pembelajaran. Maka menjadi kewajiban bagi seorang pendidik mampu melaksanakan penilaian yang diperlukan guna menilai kemampuan akademik pada siswa (Susdelina, Perdana, & Febrian, 2018). Tidak hanya

kemampuan mengevaluasi saja yang menjadi tuntutan bagi guru, namun kemampuan dalam pengembangan instrumen evaluasi dalam proses pembelajaran tersebut juga harus sesuai dengan jenis dari hasil belajar (Rusiyah, Eraku, & Supadmi, 2020). Selanjutnya, instrumen, alat serta metode penilaian juga perlu disesuaikan dengan pendekatan dalam pengumpulan serta pengukuran, yang nantinya dapat diartikan sebagai indikator ketercapaian tujuan pembelajaran. (Octaviana, Anggara, Jamilah, Darmana, & Suyanti, 2022).

Dalam proses pembelajaran, metode yang sering digunakan oleh guru guna mengetahui kompetensi serta prestasi peserta didik yakni melakukan penilaian (Kirom & Hasyim, 2021). Dalam hal ini, kuantitas soal yang dijawab peserta didik menjadi faktor pertimbangan guru dalam menentukan tingkat kompetensi peserta didik (Purniasari & Masykuri, 2021). Salah satu penilaian yang mayoritas digunakan oleh lembaga pendidikan adalah penilaian sumatif (Pohan, Ledy, Setiawan, Darmana, & Suyanti, 2022). Penilaian akhir semester merupakan contoh dari penilaian sumatif tersebut. Penilaian sumatif mempunyai tujuan mengindeks tingkat prestasi yang dimiliki siswa setelah pembelajaran dalam waktu tertentu yang telah lalu (Octaviana dkk., 2022). Dalam konteks ini, instrumen merupakan alat ukur yang dipakai dalam melaksanakan penilaian. Penilaian yang marak dan mayoritas digunakan pada saat ini adalah tes pilihan ganda.

Tes pilihan ganda merupakan salah satu bentuk tes yang memberikan alternatif jawaban bagi setiap peserta didik, kemudian peserta didik memilih salah satu jawaban yang dianggapnya paling benar (Wahab, Isa, & Laliyo, 2022). Tes pilihan ganda diujikan guna menghindari penilaian yang subjektif oleh guru kepada jawaban peserta didik, dikarenakan sudah tersedia kunci jawaban yang tidak dapat dirubah kembali jawabannya (Sakahuni & Ramadhanti, 2021). Penyusunan serta pengembangan tes pilihan ganda oleh guru yang digunakan untuk Penilaian Akhir Semester harus sesuai dengan pedoman-pedoman pembuatan soal yang baik, hal ini dimaksudkan tes tersebut benar adanya mampu mengukur tingkat prestasi peserta didik pada mata pelajaran tertentu (Wartoni & Benyamin, 2020). Pada tes pilihan ganda terdapat hal-hal yang perlu diperhatikan dalam proses pengerjaannya, antara lain (1) petunjuk untuk pengerjaan soal pilihan ganda jelas, (2) hanya 1 jawaban yang benar, (3) isi serta bentuk dari alternatif jawaban homogen, (4) menghindari tata Bahasa yang sama dengan buku yang telah diajarkan dan (5) membuat empat alternatif jawaban (Arvianto, 2016).

Mayoritas ahli evaluasi mengemukakan, penilaian semacam ini yang paling efisien guna mengukur berbagai tujuan dari pembelajaran (Priowuntato, 2020). Pemberian nilai yang sederhana dan ruang lingkup sampel yang luas (Rusiyah dkk., 2020). Hasil dari proses pembelajaran yang baik tentunya dapat terwujud jika proses pembelajaran tersebut dengan adanya instrumen penilaian yang juga berkualitas. Menurut (Arvianto, 2016), sebuah tes dapat dikategorikan baik sebagai alat ukur harus

memenuhi beberapa kriteria. Kriteria tersebut diantaranya, mempunyai validitas, reliabilitas, obyektivitas, praktikabilitas serta ekonomis.

Instrumen tes hasil belajar yang memiliki kualitas baik jika dapat menggambarkan kemampuan peserta didik dengan nyata (Rusiyah dkk., 2020). Dalam rangka mengetahui kualitas instrument tersebut, guru dapat melakukan kegiatan analisis sosial. Analisis sosial digunakan untuk mengetahui kompetensi serta pemahaman peserta didik terhadap materi yang telah diajarkan oleh guru (Chan, Looi, & Sumintono, 2021). Sehingga, sebagai guru mampu memberikan umpan balik dalam pembelajaran guna memperbaiki kualitas pembelajaran.

Dalam bidang pendidikan, analisis yang mayoritas digunakan adalah teori tes klasik (Risdiyanto, Syarkowi, & Jumiarni, 2021). Pendekatan ini dinilai mempunyai kekurangan sehingga diperlukan adanya perbedaan dalam analisis (Fernanda & Hidayah, 2020; Purnama & Alfarisa, 2020). Sehingga, muncul model pendekatan Rasch. Rasch model mempunyai keunggulan dalam memberikan informasi yang lebih lengkap dari abilitas peserta didik, serta pada saat bersamaan mampu menentukan kualitas soal. Hal senada diungkapkan oleh (Sumintono & Widhiarso, 2014), tujuan pokok *rasch model* adalah mampu menghasilkan skala pengukuran dengan interval yang sama sehingga mampu memberikan data yang akurat mengenai peserta tes dan kualitas soalnya. Selain itu, kelebihan dari *rasch model* juga mampu mempertimbangkan pendekatan probabilitas dalam melihat atribut semua objek yang diukur (Asmin Lukman, Hujairah Hi Muhammad, 2022; Rizki, 2022). Selanjutnya, dalam model rasch ini juga mampu masalah mengenai permasalahan perbedaan metrik antar butir, masalah interval data serta mampu menganalisis data yang hilang (Sumintono & Widhiarso, 2014).

Pada penelitian sebelumnya yang dilakukan oleh (Luthfiyah & Susilaningsih, 2021), didapatkan penilaian dengan model pemecahan masalah secara kolaboratif telah memenuhi kategori valid, yang mempunyai rata-rata skor 28 dari 32 serta reliabilitas 0,7 menggunakan kriteriaa reliabilitas. 117 dari jumlah total 124 peserta didik memberikan kemampuan pemecahan masalah kolaboratif, sehingga dikesimpulan terakhir menunjukkan pada kategori sedang. Penelitian yang lain menunjukkan hasil validitas soal 86,67% valid dan 14,33% tidak valid, nilai *Cronbach Alpha* 0,95 dikategorikan sangat baik, nilai *Person Reliability* 0,93 dengan kategori baik sekali dan *Item Reliability* 0.95 dengan kategori istimewa. Sehingga didapat 8 soal sangat sulit, 9 soal sulit, 23 soal mudah serta 5 soal sangat mudah (Octaviana dkk., 2022).

Pada hasil observasi memberikan informasi bahwa soal Penilaian Akhir Semester genap di salah satu kelas V Madrasah Ibtidaiyah di Kabupaten Nganjuk merupakan soal yang dikembangkan oleh guru matematika yang belum divalidasi oleh guru lainnya serta belum diujikan pada peserta didik, sehingga masih belum diketahui kualitas soal tersebut. Berangkat dari permasalahan tersebut, penelitian ini bertujuan

menyelidiki kualitas instrument penilaian yang digunakan pada Penilaian Akhir Semester genap 2022/2023 pada mata pelajaran matematika Madrasah Ibtidaiyah.

METODE PENELITIAN

Pendekatan penelitian ini menggunakan diskriptif dengan analisis kuantitatif (Fadli, 2021). Data yang digunakan dalam penelitian ini adalah hasil Penilaian Akhir mata Pelajaran matematika semester genap 2023/2024 yang terdiri dari 15 soal tes pilihan ganda (*multiple choice*) dengan empat opsi pilihan. Jumlah total peserta didik dalam penelitian ini adalah 30 siswa. Sedangkan dalam teknik pengumpulan data pada penelitian ini menggunakan teknik dokumentasi yang berupa lembar jawaban Penilaian Akhir Semester genap 2023/2024 kelas V disalah satu Madrasah Ibtidaiyah Kabupaten Nganjuk. Adapun langkah-langkah analisis pada penelitian ini meliputi, mengumpulkan berbagai informasi berupa soal tes, lembar jawaban siswa dan kunci jawaban. Kemudian dianalisis guna mengetahui validitas, reliabilitas serta tingkat kesulitan soal. Dalam langkah ini peneliti menggunakan bantuan aplikasi Winstep versi 3.73. Output yang diambil dari aplikasi Winstep ini adalah informasi mengenai kualitas soal Penilaian Akhir Semester siswa meliputi abilitas responden serta kualitas butir soal, kesesuaian responden. Pada langkah akhir, peneliti menyimpulkan dari berbagai informasi dari analisis yang telah dilakukan pada penelitian ini.

HASIL DAN PEMBAHASAN

Hasil

Validitas

Sebelum menganalisis dengan *Rasch Model*, diperlukan beberapa prasyarat yang harus dipenuhi, yakni asumsi unidimensional dan independensi. Berikut hasilnya:

		-- Empirical --	Modeled
Total raw variance in observations	=	42.7 100.0%	100.0%
Raw variance explained by measures	=	12.7 29.7%	30.0%
Raw variance explained by persons	=	4.0 9.4%	9.5%
Raw Variance explained by items	=	8.7 20.4%	20.6%
Raw unexplained variance (total)	=	30.0 70.3% 100.0%	70.0%
Unexplnd variance in 1st contrast	=	6.0 14.0% 19.9%	
Unexplnd variance in 2nd contrast	=	4.6 10.8% 15.3%	
Unexplnd variance in 3rd contrast	=	3.3 7.7% 10.9%	

Gambar 1. Hasil Analisis Asumsi Unidimensional dan Independen

Pada gambar 1 menjelaskan hasil dari validitas kontruk sebesar 29,7% (lihat *Raw variance explained by measures*) sedangkan besaran prediksi pada *Rasch Model* sebesar 30,0%. Selanjutnya, untuk mengukur validitas membutuhkan hasil dari outfit MNSQ (*Mean Square*), ZSTD (*Z-standartd*) serta Pt Measure Corr (*Point Measure Correlation*). Pada analisis ini dihasilkan pada tabel 1.

Tabel 1. Tabel *Outfit MNSQ, ZSTD & Pt. Measure Corr*

No. Soal	Jenis Outfit		Pt Measure
	MNSQ	ZSTD	Corr
1	0,52	-0,9	0.58
2	1,17	0,5	0.13
3	0,9	-0,4	0,42
4	1,15	1,0	0,15
5	0,39	-0,2	0,32
6	0,8	-0,4	0,34
7	1,12	0,5	0,16
8	1,93	3,6	-0,34
9	1,01	0,2	0,34
10	1,14	0,7	0,28
11	0,62	-0,6	0,57
12	0,78	-0,6	0,48
13	0,78	-0,8	0,54
14	0,47	-1,1	0,63
15	1,33	1,4	0,24

Reliabilitas

Nilai reliabilitas dapat diketahui dari hasil nilai *Cronbach Alpha*. Pada proses reliabilitas juga dihasilkan tingkat konsistensi dari responden berupa nilai *Person Reliability*. Kemudian, agar mampu melihat kualitas soal dalam setiap butir dapat dilihat dari hasil nilai *Item Reliability*. Dalam penelitian ini nilai dari *Cronbach Alpha*, *Person Reliability* serta *Item Reliability* dihasilkan pada Tabel 2.

 Tabel 2. Hasil Analisis *Cronbach Alpha, Person Reliability dan Item Reliability*

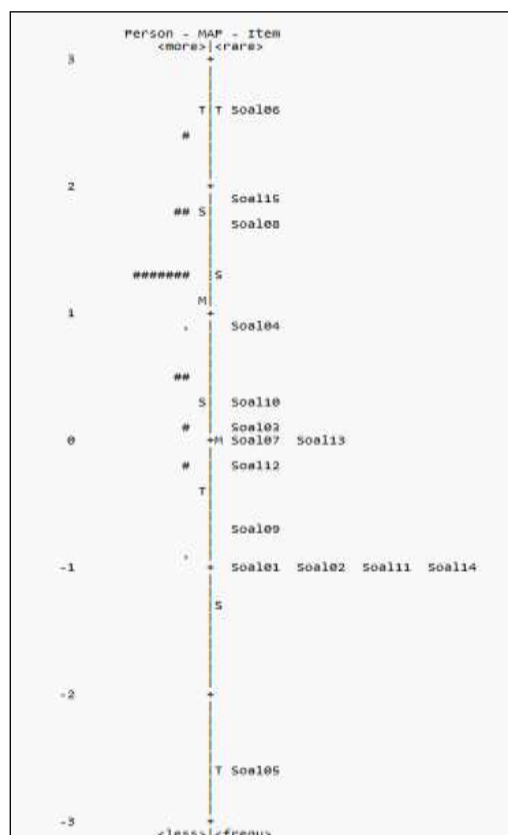
Analisa Reliabilitas	Nilai Reliabilitas
<i>Cronbach Alpha</i>	0.27
<i>Person Reliability</i>	0,1
<i>Item Reliability</i>	0,83

Tingkat Kesulitan Soal

Untuk dapat mengukur tingkat kesulitan soal pada *Rasch Model*, dapat diamati dari hasil analisa *Item Measure*. Sehingga, hasil analisis *Item Measure* pada penelitian ini dapat dilihat pada table 3. Sedangkan untuk melihat sebaran tingkat kesulitan soal dapat dilihat pada *Wright Map* pada Gambar 2.

Tabel 3. Hasil Analisis *Item Measure*

No. Soal	Nilai <i>Item Measure</i>
1	-0,99
2	-0,99
3	0,14
4	0,94
5	-2,57
6	2,61
7	-0,04
8	1,7
9	-0,71
10	0,32
11	-0,99
12	-0,24
13	-0,04
14	-0,99
15	1,86



Gambar 2. Wright Map

Pembahasan

Pada analisis yang menggunakan *Rasch Model* mampu mengakomodasi instrumen butir dengan kompetensi yang dimiliki oleh peserta didik. Hal ini dapat diamati pada gambar 1 yang mampu menjelaskan bahwa *raw variance explained by measures* (pengukuran empiris) sebesar 29,7% di bandingkan dengan *Rasch Model* sebesar 30% tidak jauh berbeda. Hal ini menandakan validasi pada penelitian ini dapat diprediksi oleh *Rasch Model*. Namun, menurut (Hidayat, Patras, Harijanto, & Hasanah, 2020) validasi dapat dinilai baik jika nilai *raw variance explained by measures* adalah >40%. Dengan hasil ini menunjukkan bahwa validitas pada soal PAS mata pelajaran matematika kelas V MI kurang baik. Hal ini disebabkan beberapa factor, salah satunya adalah soal yang dikembangkan oleh guru matematika dibuat secara individu dan tidak melakukan usaha validasi dari guru matematika yang lain. Dalam kasus ini, validasi harusnya mampu dilaksanakan dengan berbagai macam validator dengan tujuan mendapatkan hasil lebih sempurna (Octaviana dkk., 2022).

Sedangkan pada uji reliabilitas diujikan dengan tujuan kualitas soal yang digunakan untuk PAS kelas V MI. Menurut (Erfan, Maulyda, Hidayati, Astria, & Ratu,

2020), memiliki kategori dalam menentukan nilai *Cronbach Alpha*, *Item Reliability* dan *Person Reliability* terhadap reliabilitas butir soal. Berikut kategori tersebut:

1. Kategori reliabilitas menggunakan *Cronbach Alpha*
 - a. Koefisiennya lebih besar atau sama dengan 0,90 dan lebih kecil atau sama dengan 1,00 memiliki korelasi sangat tinggi dan interpetasikan sangat baik
 - b. Koefisiennya lebih besar atau sama dengan 0,70 dan lebih kecil atau sama dengan 0,90 memiliki korelasi tinggi dan interpetasikan baik
 - c. Koefisien lebih besar atau sama dengan 0,40 dan lebih kecil atau sama dengan 0,70 memiliki korelasi sedang dan interpetasikan cukup baik
 - d. Koefisiennya lebih besar atau sama dengan 0,20 dan lebih kecil atau sama dengan 0,40 memiliki korelasi rendah dan interpetasikan buruk
 - e. Koefisiennya lebih kecil dari 0,20 memiliki korelasi sangat rendah dan interpetasikan sangat buruk
2. Kategori reliabilitas pada *Rasch Model*
 - a. Nilai *reliability* yang memiliki besaran $>0,94$ diinterpretasikan sebagai istimewa
 - b. Nilai *reliability* yang memiliki besaran $0,91 - 0,94$ diinterpretasikan sebagai baik sekali
 - c. Nilai *reliability* yang memiliki besaran $0,81 - 0,90$ diinterpretasikan sebagai baik
 - d. Nilai *reliability* yang memiliki besaran $0,67 - 0,80$ diinterpretasikan sebagai cukup
 - e. Nilai *reliability* yang memiliki besaran $<0,67$ diinterpretasikan sebagai sangat buruk

Menurut (Erfan dkk., 2020), nilai *reliability* yang dimaksud adalah nilai *Item Reliability* dan *Person Reliability* sama. Jika melihat hasil analisis pada table 2, pada reliabilitas *Cronbach Alpha* sebesar 0,27 memiliki korelasi yang rendah dan diinterpretasikan buruk. Kemudian, pada *Person Reliability* besaran reliabilitasnya adalah 0,1 hal ini menunjukkan bahwa *Person Reliability* lemah, sedangkan pada *Item Reliability* menunjukkan besaran sebesar 0,83 dapat diinterpretasikan dengan kategori bagus.

Kemudian mengenai kecocokan butir soal dengan asumsi menggunakan pendekatan *Rasch Model* menurut (Sumintono & Widhiarso, 2015) terdapat beberapa kriteria yang harus dipenuhi, antara lain:

- a. Nilai *Outfit* MNSQ sebesar 0,5 sampai 1,5
- b. Nilai *Outfit* ZSTD adalah sebesar -2 sampai 2
- c. Nilai *Pt Measure Corr* harus bernilai positif

Menilik dari ketiga kriteria diatas, menurut (Sumintono & Widhiarso, 2015) butir soal dapat dinilai fit jika salah satu kriteria diatas terpenuhi. Sehingga analisis pada 15 soal PAS matematika MI kelas V (lihat tabel 1) sebanyak 14 soal (93,3%) dinilai cocok terhadap butir atau model fit, hanya 1 item soal (6,7%) pada nomor 8 yang tidak

memenuhi kriteria. Sehingga pada kasus ini, khusus soal nomor 8 perlu diganti atau dihilangkan, hal ini disebabkan karena soal nomor 8 kurang dapat dipercaya.

Setelah melaksanakan analisis model fit atau kecocokan model, tahap yang selanjutnya adalah penilaian pada karakteristik butir soal PAS mata pelajaran matematika kelas V MI. Dalam pendekatan *Rasch Model*, karakteristik butir soal adalah tahap yang melihat tingkatan kesulitan pada butir soal tersebut (Rahayuni dkk., 2020). Pada *Rasch Model*, kesulitan butir soal ini dapat dilihat pada hasil analisis *Item Measure* pada tabel 3. Nilai logit pada setiap butir soal ditampilkan pada *Item Measure*. Nilai logit merupakan ukuran dalam kesulitan butir soal, semakin tinggi nilai logit, maka semakin tinggi pula tingkat kesulitan butir soal (Octaviana dkk., 2022). Sebuah butir soal yang baik merupakan apabila tingkat kesulitan dalam butir soal tersebut adalah seimbang (Rusiyah dkk., 2020). Namun, perlu diketahui tingkat kesulitan butir soal dikategorikan menjadi 4 kategori, sebagai berikut (Erfan dkk., 2020):

- Sangat sulit jika $Item Measure > 0,82$
- Sulit jika $0,00 \leq Item Measure \leq 0,82$
- Mudah jika $0,82 \leq Item Measure \leq 0,00$
- Sangat mudah jika $Item Measure < -0,82$

Sehingga pada analisis tingkat kesulitan butir soal pada kasus ini apabila di urutkan dari nilai logit yang paling tinggi berdasarkan kategori diatas adlah sebagai berikut:

- Pada kategori soal sangat sulit terdapat 4 butir soal (26,66%) yakni pada nomor soal 4,6,8, dan 15.
- Pada kategori soal sulit terdapat 2 butir soal (13,33%) yaitu pada nomor soal 3 dan 10.
- Pada kategori soal mudah terdapat 4 butir soal (26,66%) yakni pada nomor 7,9,12 dan 13.
- Pada kategori sangat mudah terdapat 5 butir soal (33,33%) yaitu pada nomor soal 1,2,5, 11 dan 14.

Namun jika diurutkan per butir soal berdasarkan tingkat kesulitan dari yang paling sulit sampai paling mudah yaitu soal 6, soal 15, soal 8, soal 4, soal 3, soal 10, soal 7, soal 13, soal 12, soal 9, soal 1, soal 2, soal 11, soal 14 dan yang paling mudah soal nomor 5. Apabila melihat hasil tingkat kesulitan diatas dapat disimpulkan soal pada PAS mata pelajaran matematika kelas V MI baik, dikarenakan tingkat kesulitan pada butir soal relatif proporsional.

Kategori tingkat kesulitan butir soal didapatkan dari jumlah jawaban peserta didik yang menjawab benar pada butir soal tersebut (Tyas, Hamdu, & Pranata, 2020). Tingkat kesulitan butir soal juga dapat dilihat pada gambar 2. *Wright Map*. Pada *Wright Map* juga menjelaskan persebaran hubungan antara *person* dengan *item*, dimana bagian kiri menjelaskan nilai logit pada *person* sedangkan pada bagian kanan menjelaskan sebaran nilai logit pada *item*. Pada Gambar 2 bagian kiri menjelaskan semakin kebawah posisinya menunjukkan kompetensi yang dimiliki peserta didik semakin rendah pula.

Pada kasus ini nilai abilitas paling rendah dari peserta didik pada 0. Hal ini memberi fakta, bahwa tidak ada peserta didik yang kompetensinya jauh tertinggal. Dapat dilihat juga pada *Wright Map* tidak terdapat kesenjangan jarak antar *person* yang nyata dibagian kiri. Sedangkan pada Gambar 2. *Wright Map* bagian kanan menjelaskan tingkat kesulitan antar butir soal, semakin kebawah maka tingkat kesulitan pada butir soal tersebut semakin rendah (Eliza & Yusmaita, 2021). Pada kasus ini dapat dilihat butir soal dengan tingkat kesulitan paling tinggi yang berada pada bagian paling atas yakni soal nomor 6. Butir soal dengan tingkat kesulitan paling mudah berada dibagian paling bawah yakni soal nomor 5.

SIMPULAN

Menilik hasil analisis butir soal dengan menggunakan *Rasch Model* dengan bantuan aplikasi Winstep dapat ditarik garis besarn apabila dilihat dari validitas, soal yang terdiri dari 14 soal valid (93,3%) dan 1 soal tidak valid (6,3%); dengan nilai *Cronbach Alpha* sebesar 0,27 dengan interpretasi buruk, nilai *Person Reliability* 0,1 dengan interpretasi lemah serta nilai Item Reliaability 0,83 dengan interpretasi bagus; tingkat kesulitan butir soal telah sesuai dengan kaidah pengembangan instrumen tes, dikarenakan terdapat penyebaran kesulitan butir soal tersebut meliputi 4 butir soal (26,66%) dengan kreiteria sangat sulit, 2 butir soal (13,33%) dengan kriteria sulit, 4 butir soal (26,66%) mudah serta 5 butir soal (33,33%) dengan kriteria sangat mudah. Hasil penelitian ini dapat ditarik kesimpulan secara garis besar butir soal Penilaian Akhir Semester genap mata pelajaran matematika kelas V MI memiliki pola item yang *fit*. Sehingga, dapat diasumsikan secara keseluruhan butir soal Penilaian Akhir Semester genap mata pelajaran matematika kelas V MI dapat mengukur apa yang ingin diukur.

REFERENSI

- Agustian, I., Mufidah, N., Chakra Setiawan, H., Suklani, S., & Chakra Setiawan, H. (2023). Manajemen Evaluasi Kinerja Guru di Pondok Pesantren Al-Iman Putra Ponorogo. *COMSERVA Indonesian Jurnal of Community Services and Development*, 2(09), 1873–1882. <https://doi.org/10.59141/comserva.v2i09.567>
- Arvianto, F. (2016). Analisis Kualitas dan Keterbacaan Soal Ujian Nasional Bahasa Indonesia. *Utile: Jurnal Kependidikan*, 2(2), 184–203. <https://doi.org/10.37150/jut.v2i2.287>
- Asmin Lukman, Hujairah Hi Muhammad. (2022). *Analisis Tingkat Kesulitan Butir dan Kemampuan Matematika Siswa Berdasarkan Hasil Ujian Sekolah*. <https://doi.org/10.5281/ZENODO.7421825>
- Chan, S.-W., Looi, C.-K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: A Rasch model measurement analysis. *Journal of Computers in Education*, 8(2), 213–236. <https://doi.org/10.1007/s40692-020-00177-2>

- Eliza, W., & Yusmaita, E. (2021). Pengembangan Butir Soal Literasi Kimia pada Materi Sistem Koloid Kelas XI IPA SMA/MA. *JURNAL EKSAKTA PENDIDIKAN (JEP)*, 5(2), 197–204. <https://doi.org/10.24036/jep/vol5-iss2/621>
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). ANALISIS KUALITAS SOAL KEMAMPUAN MEMBEDAKAN RANGKAIAN SERI DAN PARALEL MELALUI TEORI TES KLASIK DAN MODEL RASCH. *Indonesian Journal Of Educational Research and Review*, 3(1), 11–19. <https://doi.org/10.23887/ijerr.v3i1.24080>
- Fadli, M. R. (2021). *Memahami desain metode penelitian kualitatif*. 21(1).
- Fernanda, J. W., & Hidayah, N. (2020). Analisis Kualitas Soal Ujian Statistika Menggunakan Classical Test Theory dan Rasch Model. *Square : Journal of Mathematics and Mathematics Education*, 2(1), 49–60. <https://doi.org/10.21580/square.2020.2.1.5363>
- Hidayat, R., Patras, Y. E., Harijanto, S., & Hasanah, L. (2020). Analisis Instrumen dan Prioritas Tindakan untuk Kepuasan Kerja Guru di Indonesia Berdasarkan Pemodelan Rasch. *Kelola: Jurnal Manajemen Pendidikan*, 7(2), 110–130. <https://doi.org/10.24246/j.jk.2020.v7.i2.p110-130>
- Kirom, A., & Hasyim, M. (2021). Analisis Abilitas Siswa pada Tes Penilaian Akhir Semester Mata Pelajaran PADB dengan Menggunakan Rasch Model di SD Ma'arif NU Pandaan Pasuruan. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, 13(2), 591–600. <https://doi.org/10.37680/qalamuna.v13i2.1017>
- Luthfiah, A., & Susilaningsih, E. (2021). Pengembangan Instrumen Penilaian Collaborative Problem Solving Pada Materi Larutan Elektrolit Nonelektrolit Dan Analisisnya Menggunakan Rasch Model. *Chemistry in Education*, 10(1), 1–7. <https://doi.org/10.15294/chemined.v10i1.39385>
- Octaviana, R. I., Anggara, M. B., Jamilah, R., Darmana, A., & Suyanti, R. D. (2022). Analisis Item Soal Kimia SMA Menggunakan Rasch Model. *Orbital: Jurnal Pendidikan Kimia*, 6(1), 26–37. <https://doi.org/10.19109/ojpk.v6i1.12248>
- Pohan, L. A., Ledy, K., Setiawan, V. R., Darmana, A., & Suyanti, R. D. (2022). PENGGUNAAN MODEL RASCH UNTUK MENDETEKSI KUALITAS SOAL KIMIA SMA PADA UJIAN AKHIR SEMESTER (UAS). *EKSAKTA : Jurnal Penelitian dan Pembelajaran MIPA*, 7(2), 280–289. <https://doi.org/10.31604/eksakta.v7i2.280-289>
- Prijowuntato, S. W. (2020). *Evaluasi Pembelajaran*. Sanata Dharma University Press.
- Purnama, D. N., & Alfarisa, F. (2020). KARAKTERISTIK BUTIR SOAL TRY OUT TEORI KEJURUAN AKUNTANSI SMK BERDASARKAN TEORI TES KLASIK DAN TEORI RESPON BUTIR. *Jurnal Pendidikan Akuntansi Indonesia*, 18(1), 36–46. <https://doi.org/10.21831/jpai.v18i1.31457>
- Purniasari, L., & Masykuri, M. (2021). ANALISIS BUTIR SOAL UJIAN SEKOLAH MATA PELAJARAN KIMIA SMA N 1 KUTOWINANGUN TAHUN PELAJARAN 2019/2020 MENGGUNAKAN MODEL ITEMAN DAN RASCH. *Jurnal Pendidikan Kimia*, 10(2).
- Rahayuni, G., Pd, M., Dwiyanti, A. N., Pd, M., Khotimah, K., & Ibtizam, R. N. (2020). *Analisis Butir Soal Ulangan Harian siswa Kelas IV Sekolah Dasar*.
- Risdianto, E., Syarkowi, A., & Jumiarni, D. (2021). Analisis Data Respon Mahasiswa Terhadap Sistem Pembelajaran Berbasis MOOCs pada Matakuliah Ilmu

- Lingkungan Menggunakan Rasch Model. *Jurnal Inovasi Dan Teknologi Pembelajaran*, 8(1), 47–57. <https://doi.org/10.17977/um031v8i12021p047>
- Rizki, M. (2022). *Pengembangan Butir Soal Literasi Kimia Pada Materi Ikatan Kimia Menggunakan Model Rasch* (Skripsi, Universitas Negeri Padang). Universitas Negeri Padang. Diambil dari <http://repository.unp.ac.id/43175/>
- Rusiyah, R., Eraku, S. S., & Supadmi, S. (2020). ANALISIS SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN GEOGRAFI DENGAN MENGGUNAKAN PEMODELAN RASCH. *JURNAL SWARNABHUMI : Jurnal Geografi Dan Pembelajaran Geografi*, 5(1), 11–19. <https://doi.org/10.31851/swarnabhumi.v5i1.4136>
- Sakahuni, S., & Ramadhanti, A. (2021). Perbandingan Hasil Belajar Menggunakan Tes Pilihan Ganda ditinjau dari Kemampuan Berfikir Kritis Siswa. *Integrated Science Education Journal*, 2(3), 89–93. <https://doi.org/10.37251/isej.v2i3.174>
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Cimahi, Indonesia: Trim Komunikata Publishing House. Diambil dari https://www.academia.edu/9478441/Aplikasi_Model_Rasch_untuk_Penelitian_Ilmu-Ilmu_Sosial_edisi_revisi_
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi, Indonesia: Trim Komunikata. Diambil dari <https://eprints.um.edu.my/14228/>
- Susdelina, S., Perdana, S. A., & Febrian, F. (2018). Analisis Perbandingan Teori Tes Klasik dan Rasch Model dalam Evaluasi Instrumen Pengukuran Pemahaman Konsep. *Analisis Perbandingan Teori Tes Klasik dan Rasch Model dalam Evaluasi Instrumen Pengukuran Pemahaman Konsep*.
- Tyas, E. H., Hamdu, G., & Pranata, O. H. (2020). Analisis Soal Pilihan Ganda dengan Menggunakan Pemodelan RASCH untuk Mengukur Kemampuan Siswa dalam Mengurutkan Bilangan Pecahan di Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 7(2), 1–12. <https://doi.org/10.17509/pedadidaktika.v7i2.24773>
- Wahab, M., Isa, I., & Laliyo, L. A. R. (2022). ANALISIS MISKONSEPSI LARUTAN PENYANGGA DENGAN TES PILIHAN GANDA EMPAT TINGKAT PADA SISWA. *Quantum: Jurnal Inovasi Pendidikan Sains*, 13(1), 118. <https://doi.org/10.20527/quantum.v13i1.12263>
- Wartoni, W., & Benyamin, P. I. (2020). STRATEGI PENGEMBANGAN TES OBJEKTIF (PILIHAN GANDA). *Diegesis : Jurnal Teologi*, 5(1). <https://doi.org/10.46933/DGS.vol5i1%p>