

Analisis Item Soal Kimia SMA Menggunakan Rasch Model

Rina I Octaviana^{1*)}, Muhammad B Anggara², Rayhan Jamilah³,
Ayi Darmana⁴ dan Retno D Suyanti⁵

^{1,2,3,4,5} Universitas Negeri Medan, Sumatera Utara, Indonesia

*E-mail: rinaikaoctaviana@gmail.com

ARTICLE INFO

Article History:

Received May 2022

Revised form June 2022

Accepted June 2022

Published online June 2022

Abstract: Assessment is a technique for evaluating student learning outcomes and deciding whether or not the learning objectives have been achieved. This study aims to determine the validity, reliability, and difficulty level of the tested items. The data collection technique used a documentation technique in the form of an answer sheet for the odd semester final exam for the 2021/2022 academic year for class XI at SMAS Al-Amjad Medan with a total of 222 respondents, and the form of the question was 45 multiple-choice questions (dichotomies) with five categories of answers. The Rasch model is used to obtain fit items. This analysis was carried out with the help of Winsteps software version 3.7.3. Based on the results of the analysis of the results of the item analysis using the Rasch model, the following conclusions can be drawn: in terms of validity, the questions consist of 86,67% valid and 13,33% invalid; Cronbach's alpha value of 0,95 in the very good category, the Person Reliability value is 0,93, which is included in the excellent category; the Item Reliability value is 0,95 in the special category; the distribution of the difficulty level of the item items is by the rules for making questions because 8 items (17,78%) are in the very difficult category, 9 items are about (20%) are in the difficult category, 23 items are about (51,11%) are in the easy category and 5 items (11,11%) in the very easy category.

Keywords: difficulty level, rasch model, reliability, validity

Abstrak: Penilaian merupakan salah satu teknik untuk mengevaluasi hasil belajar siswa dan memutuskan tercapai atau tidaknya tujuan pembelajaran. Penelitian ini bertujuan untuk mengetahui validitas, reliabilitas, dan tingkat kesulitan dari item soal yang diujikan. Teknik pengumpulan data menggunakan teknik dokumentasi berupa lembar jawaban ujian akhir semester ganjil tahun ajaran 2021/2022 kelas XI di SMAS Al-Amjad Medan dengan jumlah 222 responden dan

soal yaitu 45 soal pilihan ganda (dikotomi) dengan lima kategori jawaban. Rasch model digunakan untuk mendapatkan butir soal yang *fit*. Analisa ini dilakukan dengan bantuan *software* Winsteps versi 3.7.3. Berdasarkan hasil analisis hasil analisis item soal dengan menggunakan analisis Rasch model maka dapat diambil kesimpulan sebagai berikut: ditinjau dari validitasnya, soal terdiri dari 86,67% sudah valid dan 13,33% tidak valid; nilai *cronbach alpha* sebesar 0,95 yang menempatkannya pada kategori sangat baik, nilai *Person Reliability* sebesar 0,93 yang menempatkan pada kategori bagus sekali, dan nilai *Item Reliability* sebesar 0,95 yang menempatkannya pada kategori istimewa; penyebaran tingkat kesulitan item soal sesuai dengan aturan pembuatan soal, karena 8 item soal (17,78%) dengan kategori sangat sulit, 9 item soal (20%) dengan kategori sulit, 23 item soal (51,11%) dengan kategori mudah, dan 5 item soal (11,11%) dengan kategori sangat mudah.

Kata Kunci: rasch model, reliabilitas, tingkat kesulitan, validitas

PENDAHULUAN

Kegiatan penilaian merupakan salah satu teknik untuk mengevaluasi hasil belajar siswa dan memutuskan tercapai atau tidaknya tujuan pembelajaran. Untuk mengetahui apakah suatu kegiatan telah dilaksanakan sesuai dengan tujuan yang telah ditetapkan, maka dilakukan evaluasi (Purniasari et al., 2021). Evaluasi hasil belajar adalah urutan tindakan yang digunakan oleh guru untuk memperoleh dan menganalisis informasi tentang proses dan hasil belajar siswa mereka selama proses belajar (Syadiah & Hamdu, 2020). Penilaian yang baik juga akan memberikan hasil yang positif dari segi kualitas pembelajaran. Akibatnya, penting bagi guru untuk melaksanakan tugas penilaian yang diperlukan ketika menentukan kemampuan akademik siswa mereka (Susdelina et al., 2018). Pendidik atau guru dituntut untuk tidak hanya memiliki kemampuan mengevaluasi proses pembelajaran, tetapi juga kemampuan mengembangkan alat atau instrumen evaluasi yang digunakan dalam proses pembelajaran dengan cara yang sesuai dengan jenis hasil belajar yang diinginkan (Rusiyah et al., 2020). Selain itu, metode, alat, atau instrumen penilaian perlu disesuaikan dengan metode pengukuran dan pengumpulan data, yang kemudian dapat digunakan sebagai indikasi tercapai atau tidaknya tujuan pembelajaran.

Pada saat proses pembelajaran, salah satu metode yang digunakan guru untuk mengetahui kemampuan dan prestasi akademik siswa adalah dengan melaksanakan penilaian. Satu-satunya faktor yang dipertimbangkan guru ketika menentukan tingkat kemampuan siswa adalah jumlah pertanyaan yang siswa jawab dengan benar (Purniasari et al., 2021). Penilaian sumatif merupakan salah satu bentuk evaluasi yang sering dilakukan di lembaga pendidikan. Ujian akhir semester adalah contoh penilaian sumatif. Tujuan dari penilaian sumatif adalah untuk memastikan tingkat prestasi siswa setelah jangka waktu tertentu telah berlalu. Instrumen yang

dimaksud di sini adalah alat ukur yang digunakan untuk melakukan evaluasi. Tes pilihan ganda adalah jenis alat penilaian yang paling sering digunakan saat ini. Jika proses pembelajaran dan alat evaluasi berkualitas tinggi, hasil belajar yang baik dapat dicapai (Alfarisa & Purnama, 2019). Menurut Arikunto (2013), suatu tes dianggap sah sebagai alat untuk mengevaluasi sesuatu apabila memenuhi kriteria validitas, reliabilitas, objektivitas, kegunaan, dan keterjangkauan. Suatu instrumen yang kompeten akan memastikan bahwa temuan yang dihasilkannya valid dan dapat diandalkan untuk memberikan penilaian yang akurat terhadap kemampuan siswa (Darmana, Sutiani, Nasution, Ismanisa, et al., 2021). Pendekatan analisis teori klasik merupakan salah satu yang sering digunakan oleh pendidik dalam proses penilaian kualitas instrumen penilaian. Di sisi lain, analisis berdasarkan teori klasik memiliki banyak kelemahan. Oleh karena itu, salah satu opsi untuk mengatasi kelemahan pada teori klasik adalah analisis instrumen dengan menggunakan analisis berdasarkan Rasch model (Alfarisa & Purnama, 2019).

Penggunaan model analisis Rasch untuk menganalisis kualitas instrumen dapat dikatakan layak. Analisis yang dilakukan dengan menggunakan Rasch model tidak hanya sangat sederhana untuk dilakukan tetapi juga menghasilkan temuan analitis yang akurat. Selain itu, analisis item soal dengan menggunakan Rasch model memberikan sejumlah manfaat (Darmana, Sutiani, Nasution, Ismanisa, et al., 2021). Rasch model dapat mendeteksi jawaban yang salah dan skor data yang hilang, serta fakta bahwa abilitas tidak sepenuhnya bergantung pada jawaban yang benar (Eliza & Yusmaita, 2021). Untuk menentukan kualitas instrumen penilaian harus memberikan bukti validitas, reliabilitas (Sofia et al., 2020), tingkat kesukaran, daya pembeda butir soal (Son, 2019) dan distraktor (pengecoh) (Damanik & Irfandi, 2021) untuk melihat apakah instrumen tersebut dapat mengukur apa yang seharusnya diukur.

Penelitian yang telah dilakukan menunjukkan bahwa instrumen penilaian pemecahan masalah kolaboratif memenuhi kriteria valid dengan skor rata-rata 28 dari 32 dan estimasi reliabilitas 0,70 jika menggunakan kriteria reliabel. Sebanyak 117 siswa dari kemungkinan total 124 memberikan profil kemampuan pemecahan masalah kolaboratif mereka, dengan hasil yang menunjukkan kategori sedang (Luthfiyah & Susilaningih, 2021). Selanjutnya dalam penelitian analisis *expert judgement* butir soal dilakukan dengan menggunakan pendekatan model pengukuran Rasch yang menunjukkan lima panel ahli menilai butir soal berkualitas 23 dari 40 (57,5%), aspek kecocokan statistik, terlihat bahwa pakar Val.A dan Val.F paling konsisten (berdasarkan MNSQ dan ZSTD), dan reliabilitas butir soal menunjukkan taraf baik (0,97) (Darmana, Sutiani, Nasution, Sylvia, et al., 2021). Penelitian lain menunjukkan bahwa menggunakan pemodelan Rasch dapat sangat bermanfaat bagi pendidik dan peneliti penilaian pendidikan dengan memberikan prinsip-prinsip dasar yang benar dan model pemrosesan data yang tepat untuk analisis hasil tes, khususnya untuk skala ordinal, serta guru akan dapat menilai pertanyaan ujian yang diberikan kepada murid mereka dengan menyediakan data yang dapat dipercaya dan mendorong analisis pertanyaan atau upaya penelitian untuk menghasilkan hasil yang sesuai (Andrich & Pedler, 2019).

Hasil observasi menunjukkan bahwa soal-soal ujian akhir semester ganjil di SMAS Al-Amjad Medan adalah soal-soal yang disiapkan oleh guru kimia belum

pernah divalidasi oleh guru kimia lainnya, dan belum pernah diujicobakan kepada siswa sebelumnya, sehingga belum diketahui tingkat kualitas soal-soal tersebut. Menurut penjelasan tersebut, Rasch model merupakan salah satu jenis analisis penilaian yang disarankan untuk digunakan oleh guru dalam proses menganalisis kualitas item soal yang akan digunakan pada ujian akhir semester. Berdasarkan pertimbangan tersebut, tujuan dari penelitian ini adalah dengan menggunakan Rasch model untuk menyelidiki kualitas instrumen item soal yang digunakan untuk ujian akhir semester ganjil pada mata pelajaran kimia SMA kelas XI.

METODE PENELITIAN

Desain Penelitian

Penelitian ini merupakan penelitian deskriptif menggunakan analisis data kuantitatif dengan data primer instrumen soal ujian akhir semester ganjil kelas XI. Penelitian deskriptif kuantitatif dalam hal ini meliputi melihat, mengkaji, dan mendeskripsikan secara kuantitatif topik penelitian untuk menarik kesimpulan tentangnya berdasarkan peristiwa yang diamati (Putra, 2015).

Sasaran Penelitian

Penelitian ini dilaksanakan di SMAS Al-Amjad Medan tahun ajaran 2021/2022. Subjek penelitian ini adalah siswa kelas XI MIPA yang berjumlah 222 siswa.

Data Penelitian

Data dalam penelitian ini menggunakan data kuantitatif dan data primer yaitu data hasil ujian akhir semester ganjil yang terdiri dari 45 soal dengan jenis soal berupa pilihan ganda (dikotomi) dengan lima kategori jawaban. Total siswa dalam penelitian berjumlah 222 orang. Hasil jawaban dari setiap siswa dilakukan analisis untuk mendapatkan soal yang layak menggunakan Rasch model.

Instrumen Pengumpulan Data

Instrumen penelitian yang digunakan merupakan item soal tersusun atas 45 item soal berbentuk pilihan ganda (dikotomi) dengan lima kategori jawaban. Teknik pengumpulan data menggunakan teknik dokumentasi berupa lembar jawaban ujian akhir semester ganjil tahun ajaran 2021/2022 kelas XI di SMAS Al-Amjad Medan pada mata pelajaran kimia.

Analisis Data

Dalam pengolahan data, teknik analisis data penelitian meliputi analisis data kuantitatif dan deskriptif. Data hasil belajar siswa yang dikumpulkan dalam bentuk soal ujian akhir semester mata pelajaran kimia, lembar respon siswa, dan lembar kunci jawaban dengan menggunakan pendekatan dokumentasi. Hasil lembar dokumentasi memberikan analisis kuantitatif dari pertanyaan, dengan fokus khusus pada validitas, reliabilitas dan tingkat kesulitan. Dengan menggunakan *software* Winstep versi 3.7.3, kualitas soal dianalisis dengan program komputer (*software*) yang dibuat untuk mengolah data dari tes pilihan ganda.

Dari output program Winsteps, berbagai item parameter telah diambil yang sesuai dengan model Rasch. Informasi kesesuaian butir soal dengan kriteria yaitu

dengan melihat nilai item Outfit MNSQ, Outfit ZSTD dan Pt Measure Corr. Untuk memeriksa kesesuaian item soal adalah jika memenuhi beberapa kriteria yang dikemukakan oleh Boone (Susdelina et al., 2018) yaitu:

- a) Nilai outfit MNSQ (*Mean Square*) yang diterima adalah: $0,5 < \text{MNSQ} < 1,5$.
- b) Nilai outfit ZSTD (*Z-Standard*) yang diterima adalah: $-2,0 < \text{ZSTD} < +2,0$.
- c) Nilai Pt Measure Corr (*Point Measure Correlation*): $0,4 < \text{Pt Measure Corr} < 0,85$.

Nilai *Person Reliability* dan nilai *Item Reliability* yang merupakan hasil uji reliabilitas item soal. Item yang valid atau dapat diterima harus memenuhi setidaknya dua dari tiga kriteria, diperbaiki jika memenuhi salah satunya, dan dibuang ketika tidak ada yang cocok dengan kriteria ini (Palimbong et al., 2019). Kualitas item soal dari aspek reliabilitas item soal menggunakan nilai *Cronbach Alpha* jika memenuhi kriteria yang ditunjukkan pada Tabel 1. Sedangkan reliabilitas item soal dengan Rasch model memiliki kriteria dalam menentukan nilai *Item Reliability* dan *Person Reliability* berdasarkan kriteria (Erfan et al., 2020) dapat dilihat pada Tabel 2 sebagai berikut.

Tabel 1. Kriteria reliabilitas menggunakan nilai *Cronbach Alpha*

Koefisien Korelasi	Korelasi	Interpretasi
$0,90 \leq r \leq 1,00$	Sangat Tinggi	Sangat Baik
$0,70 \leq r \leq 0,90$	Tinggi	Baik
$0,40 \leq r \leq 0,70$	Sedang	Cukup Baik
$0,20 \leq r \leq 0,40$	Rendah	Buruk
$r < 0,20$	Sangat Rendah	Sangat Buruk

Tabel 2. Kriteria reliabilitas pada pemodelan Rasch

Nilai Reliability (<i>Person/Item</i>)	Interpretasi
$> 0,94$	Istimewa
$0,91 - 0,94$	Bagus Sekali
$0,81 - 0,90$	Bagus
$0,67 - 0,80$	Cukup
$< 0,67$	Lemah

Tingkat kesulitan item (*item measure*) memberikan gambaran yang komprehensif dari nilai logit masing-masing item. Hasil *output wright map* merupakan suatu peta yang memberikan informasi mengenai persebaran kemampuan *person* dan tingkat kesulitan *item* (Fajri & Yusmaita, 2021) meliputi informasi item-item yang diurutkan dari yang memiliki nilai *logit measure* tertinggi hingga yang memiliki nilai *logit measure* terendah, menunjukkan bahwa soal diurutkan dari yang sangat sulit hingga yang sangat mudah.

HASIL DAN PEMBAHASAN

Uji Validitas

Pemodelan Rasch menghubungkan parameter butir dengan kemampuan peserta tes. Analisis model Rasch dianalisis dengan menggunakan program Winstep.

Terdapat beberapa asumsi yang harus dipenuhi sebelum dianalisis dengan model Rasch, yaitu asumsi unidimensionalitas dan independensi lokal. Uji unidimensionalitas dengan program *Winstep* dilihat pada *output item: dimensionality*. Hasil output analisis dapat dilihat pada Gambar 1.

		-- Empirical --	Modeled
Total raw variance in observations	=	70.4 100.0%	100.0%
Raw variance explained by measures	=	25.4 36.1%	35.4%
Raw variance explained by persons	=	12.7 18.1%	17.8%
Raw Variance explained by items	=	12.7 18.0%	17.7%
Raw unexplained variance (total)	=	45.0 63.9%	100.0%
Unexplned variance in 1st contrast	=	3.3 4.7%	7.4%
Unexplned variance in 2nd contrast	=	3.0 4.3%	6.7%
Unexplned variance in 3rd contrast	=	2.3 3.3%	5.2%
Unexplned variance in 4th contrast	=	2.2 3.2%	5.0%
Unexplned variance in 5th contrast	=	2.0 2.8%	4.4%

Gambar 1. Hasil asumsi unidimensionalitas dan independensi lokal

Gambar 1 menggambarkan temuan validitas konstruk pada nilai *Raw variance explained by measures* dijelaskan oleh pengukuran empiris, yang mencapai 36,1% sedangkan model Rasch diprediksi 35,4%. Dalam contoh ini, validasi konstruk empiris hampir sama dengan nilai yang diprediksi oleh Rasch model. Evaluasi validasi konstruk dinyatakan sangat baik jika *Raw variance explained by measures* > 40% (Hidayat et al., 2020). Penilaian validasi konstruk pada penelitian ini dinilai kurang baik. Sedangkan nilai *unexplained variance* yang diperoleh semuanya 15% yang tergolong kurang baik (Palimbong et al., 2019). Validasi konstruk tergolong kurang baik karena validasi konstruk tidak dilakukan oleh guru kimia dalam penelitian ini. Guru membuat pertanyaan sendiri dan tidak mencari validasi dari guru kimia lain. Validasi konstruk dapat dilakukan dengan menggunakan banyak validator untuk mendapatkan hasil yang lebih baik.

Menurut Boone et al., (2014), nilai *outfit means-square* (MNSQ), *outfit z-standard* (ZSTD), dan *point measure correlation* (Pt Measure Corr) adalah kriteria yang digunakan untuk melihat tingkat kesesuaian item (*item fit*). Hasil uji validitas dapat dilihat dari hasil outfit MNSQ, ZSTD, dan Pt Measure Corr pada Tabel 3.

Tabel 3. Hasil Outfit MNSQ, ZSTD, dan Pt Measure Corr

Item	Outfit		Pt Measure Corr.	Keterangan
	MNSQ	ZSTD		
1	0,78	-1,5	0,56	Fit
2	0,74	-1,9	0,59	Fit
3	0,32	-5,1	0,86	Fit
4	0,71	-1,7	0,54	Fit
5	1,50	2,4	0,44	Fit
6	0,88	-0,9	0,67	Fit
7	1,61	2,4	0,44	Misfit
8	0,88	-1,0	0,62	Fit
9	0,61	-3,5	0,76	Fit
10	1,41	1,4	0,50	Fit
11	2,06	5,5	0,14	Misfit

Item	Outfit		Pt Measure Corr.	Keterangan
	MNSQ	ZSTD		
12	0,64	-3,0	0,67	Fit
13	0,73	-2,2	0,63	Fit
14	0,86	-1,0	0,56	Fit
15	0,73	-2,3	0,65	Fit
16	1,06	0,4	0,48	Fit
17	2,73	4,4	0,12	Misfit
18	0,81	-1,5	0,61	Fit
19	0,75	-1,6	0,53	Fit
20	0,68	-2,6	0,66	Fit
21	0,86	-1,1	0,62	Fit
22	1,96	5,9	0,36	Misfit
23	1,44	2,4	0,45	Fit
24	2,56	4,0	0,17	Misfit
25	0,98	-0,1	0,61	Fit
26	0,74	-2,2	0,69	Fit
27	1,13	1,0	0,59	Fit
28	1,36	2,3	0,43	Fit
29	0,68	-2,3	0,60	Fit
30	0,84	-0,9	0,48	Fit
31	0,73	-2,2	0,65	Fit
32	0,63	-3,3	0,75	Fit
33	0,87	-0,8	0,65	Fit
34	0,81	-1,4	0,58	Fit
35	0,85	-1,2	0,64	Fit
36	0,91	-0,7	0,56	Fit
37	0,80	-1,6	0,60	Fit
38	2,25	3,8	0,33	Misfit
39	1,03	0,2	0,56	Fit
40	0,71	-2,3	0,63	Fit
41	0,92	-0,6	0,64	Fit
42	0,84	-1,3	0,64	Fit
43	1,35	2,4	0,44	Fit
44	1,21	1,6	0,42	Fit
45	0,83	-1,3	0,59	Fit

Berdasarkan Tabel 3 jika dilihat dari ketiga kriteria, terdapat 6 item soal (13,33%) yang *misfit*, yaitu item soal nomor 7, 11, 17, 22, 24, dan 38 yang tidak memenuhi syarat outfit MNSQ, ZSTD dan Pt Measure Corr. Adanya *outliers* dalam analisis menyebabkan hasil analisis kesesuaian item soal terganggu sehingga menjadi kurang dapat dipercaya, oleh karena itu item soal nomor 7, 11, 17, 22, 24, dan 38 perlu dihilangkan. Sedangkan untuk 39 item soal (86,67%) lain yang hanya memenuhi satu atau dua kriteria saja dinyatakan *fit* dengan model (Azizah & Wahyuningsih, 2020) maka item soal hanya perlu diperbaiki (Purniasari et al., 2021).

Uji Reliabilitas

Penentuan reliabilitas dilakukan untuk melihat konsistensi instrumen dalam pengukuran jika digunakan secara berulang. Nilai reliabilitas secara keseluruhan dapat dilihat dari nilai *Cronbach Alpha*. Untuk mengetahui tingkat konsistensi

responden dapat dilihat dari hasil *Person Reliability*. Sedangkan untuk melihat kualitas per item dalam instrumen ditentukan dari nilai *Item Reliability*. Tabel 4 menampilkan hasil penelitian reliabilitas model Rasch terhadap soal ujian akhir semester ganjil mata pelajaran kimia.

Tabel 4. Hasil analisis reliabilitas

Analisis Reliabilitas	Reliabilitas	Kategori
<i>Cronbach Alpha</i>	0,95	Sangat Baik
<i>Person Reliability</i>	0,93	Bagus Sekali
<i>Item Reliability</i>	0,95	Istimewa

Berdasarkan Tabel 4 diperoleh nilai *cronbach alpha* sebesar 0,95 yang menempatkannya pada kategori sangat baik, nilai *Person Reliability* sebesar 0,93 yang menempatkan pada kategori bagus sekali, dan nilai *Item Reliability* sebesar 0,95 yang menempatkannya pada kategori istimewa. Oleh karena itu, dapat dikatakan bahwa reliabilitas komponen instrumen tersebut tinggi.

Tingkat Kesulitan Item Soal

Tingkat kesulitan item soal menentukan berapa banyak responden yang cenderung menjawab item dengan benar. Pada pemodelan Rasch, untuk melihat seberapa sulit suatu item berdasarkan hasil tabel *Item Measure*. Nilai logit untuk setiap item ditampilkan di *Item Measure*, yang diurutkan dari tertinggi ke terendah. Nilai logit adalah ukuran kesulitan item. Semakin sulit item soal, semakin tinggi nilai logitnya. Suatu soal dapat dikatakan baik jika tingkat kesulitannya proporsional (seimbang) (Rusiyah et al., 2020). Distribusi tingkat kesulitan item soal menjadi empat kategori sebagai berikut (Erfan et al., 2020). Hasil distribusi tingkat kesulitan butir soal dapat dilihat pada Tabel 5.

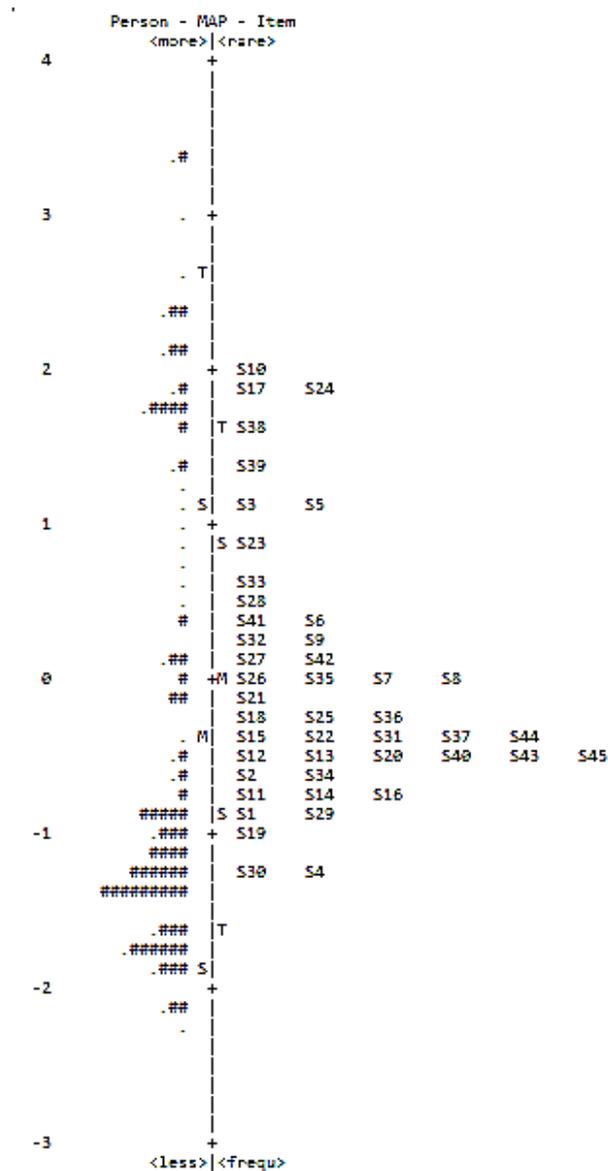
Tabel 5. Hasil distribusi tingkat kesulitan soal

Tingkat Kesulitan	Nilai Measure	Item Soal	Jumlah Item Soal	Persentase
Sangat Sulit	$Measure\ logit > 0,82$	10, 17, 24, 38, 39, 3, 5, 23	8	17,78%
Sulit	$0,00 \leq Measure\ logit \leq 0,82$	33, 28, 41, 6, 32, 9, 27, 42, 35	9	20%
Mudah	$-0,82 \leq Measure\ logit \leq 0,00$	36, 7, 8, 21, 18, 25, 36, 15, 22, 31, 37, 44, 12, 13, 20, 40, 43, 45, 2, 34, 11, 14, 16	23	51,11%
Sangat Mudah	$Measure\ logit < -0,82$	1, 29, 19, 30, 4	5	11,11%

Rasch model digunakan untuk memeriksa item, dan setiap item dalam Rasch model memiliki nilai *Measure logit* yang menunjukkan tingkat kesulitannya. Tabel 5 menunjukkan bahwa item soal sebanyak 8 item soal (17,78%) dengan kategori sangat sulit, 9 item soal (20%) dengan kategori sulit, 23 item soal (51,11%) dengan

kategori mudah, dan 5 item soal (11,11%) dengan kategori sangat mudah.

Kategori tingkat kesulitan item ditentukan dari jumlah siswa menjawab benar item soal. Nilai logit dapat pula dilihat dengan menggunakan peta persebaran hubungan *person-item* pada *Wright map*. Peta pada tabel digambarkan dari penggabungan nilai logit pada *item* dan *person measure*. Di bagian kiri *Wright map* menunjukkan nilai logit *person* dan bagian kanan menunjukkan nilai logit *item*.



Gambar 2. Distribusi kemampuan siswa dan kesulitan item soal (*person-item map*)

Dapat dilihat pada Gambar 2 bahwa tidak terdapat kesenjangan yang signifikan pada tingkat kemampuan yang dimiliki siswa dalam menjawab pertanyaan. Hal ini ditunjukkan oleh fakta bahwa posisi mereka di peta cukup dekat dengan siswa lain. Di bagian kanan *Wright map* memberikan penjelasan tentang nilai logit *item*. Angka ini menunjukkan bahwa item soal S10 merupakan soal dengan tingkat kesulitan sangat sulit. Oleh karena itu, kemungkinan siswa menjawab pertanyaan ini dengan

benar cukup rendah (Kurniawan & Andriyani, 2018). Sedangkan item soal S4 merupakan item soal dengan tingkat kesulitan yang sangat mudah. Hal tersebut dibuktikan juga jumlah siswa yang menjawab benar pada item soal S4, yaitu 137 dari total 222 siswa menjawab benar item soal nomor 4. Bila item soal memiliki nilai logit yang sama, hal ini menunjukkan bahwa tingkat kesulitan soal-soal yang berkaitan dengan objek-objek tersebut tidak berbeda satu sama lain.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis item soal dengan menggunakan analisis Rasch model maka dapat diambil kesimpulan sebagai berikut: ditinjau dari validitasnya, soal terdiri dari 86,67% sudah valid dan 13,33% tidak valid; nilai *cronbach alpha* sebesar 0,95 yang menempatkannya pada kategori sangat baik, nilai *Person Reliability* sebesar 0,93 yang menempatkan pada kategori bagus sekali, dan nilai *Item Reliability* sebesar 0,95 yang menempatkannya pada kategori istimewa; penyebaran tingkat kesulitan item soal sesuai dengan aturan pembuatan soal, karena 8 item soal (17,78%) dengan kategori sangat sulit, 9 item soal (20%) dengan kategori sulit, 23 item soal (51,11%) dengan kategori mudah, dan 5 item soal (11,11%) dengan kategori sangat mudah. Dari hasil penelitian di atas, dapat disimpulkan bahwa sebagian besar soal ujian akhir semester ganjil pada mata pelajaran kimia SMA kelas XI memiliki pola item yang *fit*, oleh karena itu diasumsikan bahwa keseluruhan instrumen soal ujian akhir semester ganjil mata pelajaran kimia mampu mengukur apa yang ingin diukur.

Saran dalam penelitian ini yaitu guru kimia SMA dapat memanfaatkan soal-soal ujian akhir semester ganjil yang telah akurat dan tervalidasi sebagai bank soal untuk ujian siswanya. Selanjutnya, sangat penting bahwa soal-soal yang dibuat guru kimia harus divalidasi dan diuji coba terlebih dahulu pada siswa sebelum digunakan.

DAFTAR PUSTAKA

- Alfarisa, F., & Purnama, D. N. (2019). Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Ekonomi SMA Menggunakan Rasch Model. *Jurnal Pendidikan Ekonomi*, 11(2), 366–374.
- Andrich, D., & Pedler, P. (2019). A law of ordinal random error: The Rasch measurement model and random error distributions of ordinal assessments. *Measurement: Journal of the International Measurement Confederation*, 131, 771–781. <https://doi.org/10.1016/j.measurement.2018.08.062>
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: PT Bumi Aksara.
- Azizah, & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen the Use of Rasch Model for Analyzing Test. *JUPITEK Jurnal Pendidikan Matematika*, 3(1), 45–50. <https://doi.org/https://doi.org/10.30598/jupitekvol3iss1ppx45-50>
- Boone, W., J., Staver, J., R., & Yale, M., S. (2014). *Rasch Analysis in The Human Science*. Dordrecht: Springer.
- Damanik, F. C., & Irfandi. (2021). Pengembangan Instrumen Tes Berbasis Higher Order Thinking Skill (HOTS) Bentuk Pilihan Ganda Berdasarkan Taksonomi Bloom Revisi Pada Materi Gelombang Bunyi. *Jurnal Inovasi*

Pembelajaran Fisika, 10(1), 14–20.

- Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, & Nurhaswinda. (2021). Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments. *Jurnal Pendidikan Sains Indonesia*, 9(3), 329–345. <https://doi.org/10.24815/jpsi.v9i3.19618>
- Darmana, A., Sutiani, A., Nasution, H. A., Sylvia, N. S. A., Aminah, N., & Utami, T. (2021). Analysis of Multi Rater with Facets on Instruments HOTS of Solution Chemistry based on Tawheed. *Journal of Physics: Conference Series*, 1819(1), 1–6. <https://doi.org/10.1088/1742-6596/1819/1/012038>
- 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.
- Fajri, N. M. A. K., & Yusmaita, E. (2021). Analisis Literasi Kimia Peserta Didik di SMAN 1 Batam pada Topik Hukum-Hukum Dasar Kimia dengan Model Rasch. *Jurnal Eksakta Pendidikan (Jep)*, 5(1), 102–109. <https://doi.org/10.24036/jep/vol5-iss1/576>
- 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.
- Kurniawan, U., & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda Dengan Rasch Model. *Statistika*, 6(1), 34–39.
- Luthfiyah, A., & Susilaningih, E. (2021). Pengembangan Instrumen Collaborative Problem Solving pada Materi Larutan Elektrolit Nonelektrolit Menggunakan Analisis Rasch. *Journal of Chemistry In Education*, 10(2252), 1–7.
- Palimbong, J., Mujasam, M., & Allo, A. Y. T. (2019). Item Analysis Using Rasch Model in Semester Final Exam Evaluation Study Subject in Physics Class X TKJ SMK Negeri 2 Manokwari. *Kasuari: Physics Education Journal (KPEJ)*, 1(1), 43–51. <https://doi.org/10.37891/kpej.v1i1.40>
- Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2022 Menggunakan Model Iteman dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205–214. <https://doi.org/https://doi.org/10.20961/jpkim.v10i2.48244>
- Putra, E. A. (2015). Anak Berkesulitan Belajar di Sekolah Dasar Se-Kelurahan Kalumbuk Padang. *Jurnal Ilmiah Pendidikan Khusus*, 1(3), 71–76. <http://103.216.87.80/index.php/jupekhu/article/viewFile/6065/4707>
- Rusiyah, 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. <https://doi.org/10.31851/swarnabhumi.v5i1.4136>

- Sofia, Permanasari, A., Sholihin, H., & Supriyanti, F. M. T. (2020). Konten Dan Konstruksi Asesmen Keterampilan Berpikir Evaluatif Pada Tema Karbohidrat Dalam Kimia Pangan. *Orbital: Jurnal Pendidikan Kimia*, 4(1), 88–98.
- Son, A. L. (2019). Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Gema Wiralodra*, 10(1), 41–52. <https://doi.org/10.31943/gemawiralodra.v10i1.8>
- Susdelina, Perdana, S. A., & Febrian. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, 6(1), 41–48. <https://doi.org/10.31629/kiprah.v6i1.574>
- Syadiah, A. N., & Hamdu, G. (2020). Analisis Rasch Untuk Soal Tes Berpikir Kritis Pada Pembelajaran STEM di Sekolah Dasar. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, 10(2), 138–148. <https://doi.org/10.25273/pe.v10i2.6524>