

Penggunaan Model Rasch untuk Mengembangkan Instrumen Penilaian Keterampilan Proses Sains pada Materi Gaya Antarmolekul

Ayu Wandyka, Faizah Qurrata Aini*

Departemen Kimia, FMIPA, Universitas Negeri Padang, Kota Padang, Indonesia

*Corresponding Author: faizah_qurrata@fmipa.unp.ac.id

Dikirim: 12-12-2024; Direvisi: 21-12-2024; Diterima: 25-12-2024

Abstrak: Instrumen penilaian yang digunakan guru lebih memfokuskan aspek pengetahuan daripada keterampilan proses sains sehingga belum sesuai dengan capaian pembelajaran kimia pada kurikulum merdeka. Tujuan penelitian ini yaitu untuk mengembangkan instrumen penilaian keterampilan proses sains pada materi gaya antarmolekul. Jenis penelitian ini adalah *Research and Development* (R&D) dengan menggunakan model Rasch. Subjek penelitian ini yaitu tiga orang dosen kimia FMIPA UNP, dua orang guru kimia dan peserta didik SMA Negeri 1 Padang Sago. Instrumen penilaian ini terdiri dari 24 butir soal yang disesuaikan dengan jumlah tujuan pembelajaran materi gaya antarmolekul dan indikator keterampilan proses sains yang akan diukur yaitu (1) mengamati, (2) mempertanyakan dan memprediksi, (3) memilih metode yang sesuai, (4) memproses, menganalisis data dan informasi, (5) mengevaluasi dan refleksi dan (6) mengomunikasikan hasil. Data mentah dari hasil validitas logis oleh ahli/pakar kimia dianalisis menggunakan model Rasch dengan bantuan *software* Minifacet dan data mentah uji coba dianalisis menggunakan *software* Ministep. Hasil analisis validitas logis instrumen didapatkan nilai *exact agreement* sebesar 96,2 % dan *expect agreement* sebesar 96,4% yang artinya terdapat kesesuaian (*fit*) antara penilaian dari para ahli dengan hasil rancangan model. Selanjutnya hasil dari uji coba didapatkan bahwa semua item telah memenuhi kriteria valid, reliabel, mempunyai indeks kesukaran dan daya beda soal yang baik.

Kata Kunci: Instrumen Penilaian; Keterampilan Proses Sains; Gaya Antarmolekul; Model Rasch

Abstract: Assessment instruments utilized by teachers predominantly emphasize knowledge over scientific process skills, rendering them misaligned with the chemistry learning outcomes outlined in the *Merdeka Curriculum*. This study aimed to develop an assessment instrument for evaluating scientific process skills within the topic of intermolecular forces. Type of this research is Research and Development (R&D) with Rasch measurement as its analytical framework. The study's participants include three chemistry lecturers from the Faculty of Mathematics and Natural Sciences at Universitas Negeri Padang, two high school chemistry teachers, and students from SMA Negeri 1 Padang Sago. The assessment instrument comprises 24 items tailored to match the learning objectives of intermolecular forces and the indicators of scientific process skills being measured, which are: (1) observing, (2) questioning and predicting, (3) selecting appropriate methods, (4) processing and analyzing data and information, (5) evaluating and reflecting, and (6) communicating results. Raw data from logical validity evaluations by chemistry experts were analyzed using the Rasch model with the assistance of Minifacet software, while field test data were analyzed using Ministep software. The logical validity analysis yielded an exact agreement value of 96.2% and an expected agreement value of 96.4%, indicating a fit between expert assessments and the proposed model design. Furthermore, field test results demonstrated that all items met criteria for validity, reliability, appropriate difficulty levels, and strong discriminatory index.

Keywords: Assessment Instrument; Science Process Skills; Intermolecular Forces; Rasch Model

PENDAHULUAN

Cabang ilmu yang mengkaji struktur, sifat, dan perubahan materi adalah ilmu kimia (Chang & Overby, 2011). Kimia sebagai proses dan sebagai produk adalah dua hal yang saling berkaitan. Kimia adalah produk dalam bentuk konsep, hukum, prinsip, fakta, dan teori. Kimia sebagai proses berbentuk kerja ilmiah, dimana peserta didik belajar dan mengembangkan pengetahuannya sendiri. Oleh karena itu, ciri-ciri kimia sebagai proses dan produk harus dipertimbangkan selama mempelajari kimia dan penilaian hasil belajar kimia.

Capaian Pembelajaran kimia kurikulum merdeka mencakup elemen pemahaman kimia dan keterampilan proses (Kemendikbudristek, 2024). Pemahaman kimia meliputi tentang materi-materi yang akan dipelajari sedangkan keterampilan proses meliputi proses ilmiah mulai dari mengamati hingga mengomunikasikan hasil. Mempelajari Keterampilan Proses Sains (KPS) membantu meningkatkan pemahaman konseptual dan kemampuan berpikir kritis. KPS berguna untuk pembelajaran karena dapat diterapkan dalam kehidupan sehari-hari (Ratnasari et al., 2017). Ketercapaian pembelajaran dapat diketahui dengan melakukan penilaian, yang bertujuan untuk mengukur seberapa baik peserta didik memenuhi persyaratan kompetensi kurikulum. Penilaian merupakan proses sistematis yang memerlukan instrumen penilaian yang efektif dan dapat dipercaya untuk mengukur seberapa baik peserta didik mencapai tujuan pembelajaran mereka (Arifin, 2014). Menurut Oktaviani et al., (2015) penilaian yang dilakukan di sekolah lebih menekankan pada aspek pengetahuan saja tanpa melatih KPS peserta didik.

Berdasarkan hasil wawancara dengan enam guru kimia diketahui bahwa instrumen penilaian yang digunakan lebih menekankan aspek pengetahuan daripada keterampilan proses sains. Hal tersebut dikarenakan guru beranggapan bahwa KPS itu hanya bisa diukur ketika peserta didik melakukan praktikum. Berdasarkan Jayali & Sriwahyuni (2022) mengatakan bahwa KPS itu tidak hanya diukur melalui kegiatan praktikum karena tidak semua topik pembahasan dalam pembelajaran kimia dapat dipraktikkan dan Tosun (2019) juga menjelaskan bahwa KPS itu seharusnya melibatkan materi. Sementara, beberapa guru yang sudah melaksanakan proses penilaian KPS masih secara umum yaitu dengan melakukan observasi selama proses pembelajaran berlangsung tanpa melibatkan materi di dalamnya. Salah satu materi kimia yang dipelajari adalah gaya antarmolekul. Gaya antarmolekul berhubungan dengan konsep terbentuknya suatu materi dikarenakan materi tersebut tersusun oleh atom, ion, dan molekul yang saling berinteraksi.

Penelitian terdahulu ditemukan tentang instrumen KPS materi termokimia oleh Salmawati et al., (2023). Selanjutnya oleh Ilmiah et al., (2020) pada materi asam basa. Penelitian lainnya yang mengembangkan instrumen KPS yaitu Oktaviani et al., (2015) pada materi hukum-hukum dasar kimia dan materi stoikiometri oleh Asmalia et al., (2015). Namun, belum ditemukan instrumen penilaian keterampilan proses sains materi gaya antarmolekul pada kurikulum merdeka.

Model Rasch adalah model yang dapat diterapkan untuk mengembangkan instrumen penilaian. Menurut Sumintono & Widhiarso (2014) Model Rasch merupakan pengukuran data yang dapat menghubungkan kemampuan individu



peserta didik dengan tingkat kesukaran soal sehingga diperoleh pengukuran dengan interval skor yang sama. Oleh sebab itu untuk menghasilkan instrumen penilaian yang berkualitas dari aspek validitas, realibilitas, daya beda soal, dan indeks kesukaran diperlukan model Rasch. Model ini dapat memberikan rincian mengenai karakteristik butir soal pada instrumen yang dikembangkan. Penelitian ini bertujuan untuk mengembangkan instrumen penilaian keterampilan proses sains pada materi gaya antarmolekul.

METODE PENELITIAN

Jenis penelitian ini yaitu penelitian *Research and Development* (R&D) dengan menggunakan model Rasch. Penelitian ini akan merancang instrumen penilaian KPS pada materi gaya antarmolekul. Subjek penelitian ini adalah tiga orang dosen kimia FMIPA UNP dan dua orang guru kimia sebagai validator serta peserta didik SMAN 1 Padang Sago. Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2024/2025.

Pengembangan instrumen penilaian ini menggunakan 10 tahapan model Rasch menurut Liu (2010) yaitu: Pertama, menentukan tujuan dan populasi. Pada tahapan ini menentukan tujuan mengembangkan instrumen penilaian dan populasi penelitian. Kedua, menentukan konstruk. Pada tahap ini menentukan konstruk yang akan diukur, yaitu keterampilan proses sains. Selanjutnya pada tahapan ini merumuskan tujuan pembelajaran materi gaya antarmolekul yang disesuaikan dengan indikator KPS pada capaian pembelajaran kimia kurikulum merdeka. Ketiga, identifikasi kinerja konstruk yang ditentukan, yaitu menentukan jumlah soal, format tes, menyusun indikator soal, menyusun butir soal dan rubrik penilaian, dan menguji validitas logis instrumen penilaian oleh lima orang pakar/ahli kimia. Hasil validitas logis dianalisis menggunakan model Rasch dengan bantuan Program Minifacet. Keempat, melakukan uji coba. Kelima, melakukan analisis Rasch. Pada tahapan ini data mentah yang didapatkan dari uji coba dianalisis menggunakan model Rasch dengan bantuan program Ministep. Keenam, meninjau statistik kecocokan. Ketujuh, meninjau peta Wright. Kedelapan, mengulangi langkah 4-7 jika terdapat item yang tidak *fit*. Kesembilan, menetapkan klaim kualitas instrumen penilaian. Kesepuluh, mengembangkan pedoman penggunaan instrumen.

Teknik analisis data yang dilakukan adalah validitas, reliabilitas, indeks kesukaran dan daya beda soal. Dalam pemodelan Rasch validitas dikenal *item fit*. Menurut Boone et al., (2014) *item fit* diukur berdasarkan tiga kriteria: *Outfit Mean Square* (MNSQ), *Outfit Z-Standard* (ZSTD) dan *Pt Measure Correlation* (*Pt Mean Corr*). Item termasuk *fit* bila memenuhi satu dari ketiga kriteria pada Tabel 1 (Sumintono & Widhiarso, 2015).

Tabel 1. Kriteria *Item Fit*

Kriteria	Interval	Keterangan
<i>Outfit Mean Square</i>	$0,5 < \text{MNSQ} < 1,5$	Diterima
<i>Outfit Z-Standard</i>	$-2,0 < \text{ZSTD} < +2,0$	Diterima
<i>Point Measure Correlation</i>	$0,4 < \text{Pt Measure Corr} < 0,85$	Diterima

Nilai reliabilitas dilihat dari menu *summary statistic*, indeks kesukaran soal dianalisis dengan menggunakan menu tabel *output* yaitu *Item Measure* dengan

melihat nilai *logit*. Pengelompokkan indeks kesukaran menurut model Rasch pada Tabel 2.

Tabel 2. Kategori Indeks Kesukaran

Nilai <i>Logit</i>	Kategori
Lebih besar dari +1SD	Sangat Sulit
0,0 <i>logit</i> +1SD	Sulit
0,0 <i>logit</i> -1SD	Mudah
Lebih kecil dari -1SD	Sangat Mudah

Daya bedasoal dianalisis dari nilai *separation* pada menu *summary statistic*. Pemisahan strata adalah persamaan lain yang dapat digunakan untuk memahami bagaimana objek dikelompokkan secara lebih lengkap. Persamaan strata sebagai berikut:

$$H = \frac{[(4 \times SEPARATION) + 1]}{3}$$

HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan instrumen penilaian yang terdiri dari 24 butir soal yang disesuaikan dengan jumlah tujuan pembelajaran dan indikator keterampilan proses sains. Pengembangan instrumen penilaian dilakukan dalam sepuluh tahapan yang mendapatkan hasil sebagai berikut.

Menentukan tujuan dan populasi. Tujuan membuat instrumen ini adalah untuk mengukur KPS peserta didik pada materi gaya antarmolekul. Instrumen ini akan digunakan sebagai penilaian sumatif, yaitu penilaian yang dilakukan untuk memastikan ketercapaian tujuan pembelajaran. Penilaian sumatif ini akan dilakukan secara bersamaan untuk dua atau lebih tujuan pembelajaran atau pada akhir cakupan materi (Kemendikbudristek, 2024). Peserta didik SMAN 1 Padang Sago sebagai populasi. Pemilihan sekolah ini didasarkan padasalah satu sekolah yang telah menerapkan kurikulum merdeka.

Menentukan konstruk, konstruk adalah konsep yang akan diukur (Sumintono & Widhiarso, 2015). Konstruk pada instrumen ini adalah keterampilan proses sains yang diambil dari capaian pembelajaran kimia kurikulum merdeka. Indikator KPS yang akan diukur dapat dilihat pada Tabel 3.

Tabel 3. Indikator KPS

No	Indikator Keterampilan Proses Sains
1	Mengamati
2	Mempertanyakan dan memprediksi
3	Memilih metode yang sesuai
4	Memproses, menganalisis data dan informasi
5	Mengevaluasi dan refleksi
6	Mengomunikasikan hasil

Identifikasi kinerja konstruk yang ditentukan. Pada tahapan ini dilakukan beberapa langkah, yang pertama menentukan jumlah soal dan format tes. Jumlah soal yang dikembangkan didasari pada jumlah tujuan pembelajaran materi gaya antarmolekul dan indikator keterampilan proses sains. Terdapat 4 tujuan pembelajaran materi gaya antarmolekul yang telah dirumuskan, setiap satu tujuan pembelajaran akan menghasilkan enam butir soal yang disesuaikan dengan enam indikator KPS yang akan diukur sehingga jumlah soal pada instrumen ini sebanyak

24 butir soal. Tes pilihan ganda adalah format yang digunakan dengan menggunakan kertas dan pena. Dalam pemeriksaan atau penskoran tes pilihan ganda tidak ada unsur subjektif yang memengaruhi, baik itu dari segi guru ataupun peserta didiknya (Arifin, 2014). Pemilihan format tes menggunakan kertas dan pena dianggap lebih mudah dan terjangkau serta dapat membantu konsentrasi dan pemahaman peserta didik (Siegel, 2023).

Selanjutnya menyusun indikator soal yang dirangkum dalam tabel kisi-kisi soal yang di dalamnya terdapat tujuan pembelajaran, indikator KPS, indikator soal dan rancangan soal beserta kunci jawabannya. Kisi-kisi berguna sebagai pedoman dalam membuat soal menjadi perangkat tes (Arifin, 2014). Terdapat 24 indikator soal yang disesuaikan dengan jumlah tujuan pembelajaran materi gaya antarmolekul. Berdasarkan indikator soal yang telah dirancang dihasilkan 24 butir soal pilihan ganda yang sesuai dengan jumlah tujuan pembelajaran pada materi gaya antarmolekul. Selanjutnya rubrik penilaian instrumen KPS berisi nomor soal, indikator KPS yang diukur, tujuan pembelajaran dan skor.

Validitas logis instrumen dilakukan oleh tiga orang dosen kimia FMIPA UNP dan dua orang guru kimia SMAN 1 Padang Sago. Validasi logis yang dilakukan menggunakan skala Guttman. Menurut Sugiyono (2006) alat ukur yang akan menghasilkan respons yang pasti adalah skala Guttman. Validitas logis instrumen penilaian dilakukan dengan pemberian angket yang terdapat 12 kriteria penilaian dari aspek materi, konstruk, kebahasaan. Arikunto (2012) menyatakan bahwa jika sebuah instrumen memenuhi standar yang valid berdasarkan hasil penalaran, maka instrumen tersebut dianggap valid secara logis. Diketahui hasil validitas logis bahwa mayoritas dari butir soal dan aspek yang dinilai mendapatkan pilihan “Ya” dan sebagian kecil “Tidak”. Hasil dari validitas logis dianalisis dengan menggunakan model Rasch dengan bantuan program minifacet. Hasil analisis validitas logis pada Tabel 4.

Tabel 4. Hasil Validitas Logis

<i>Strata Value</i>	<i>Reliability</i>	<i>Exact Agreement</i>	<i>Expect Agreement</i>
3,35	0,84	96,2%	96,4%

Hasil analisis penilaian pakar terhadap instrumen penilaian yang dianalisis menggunakan model Rasch bisa ditinjau dari empat indikator penentu. Kolom pertama menunjukkan nilai strata yang diperoleh yaitu sebesar 3,35 yang menunjukkan penilaian ahli reliabel. Adapun indeks reliabilitas validator/ahli diperoleh sebesar 0,84 dengan kategori baik sekali. Kolom ketiga menggambarkan nilai *exact agreement* (persetujuan validator) dengan persentase sebesar 96,2%, yang mana ini tidak jauh berbeda dari *expect agreement* (perkiraan model) dengan persentase 96,4%. Artinya terdapat kesesuaian (*fit*) antara hasil penilaian pakar dengan hasil yang diperkirakan oleh model sehingga instrumen penilaian bisa dikatakan valid (Sick, 2013).

Melakukan uji coba, yang bertujuan untuk meninjau kualitas instrumen penilaian secara empiris. Produk yang telah divalidasi oleh pakar/ahli diuji cobakan kepada 9 orang peserta didik kelas XII Fase F SMAN 1 Padang Sago. 9 orang peserta didik ini terdiri kemampuan tinggi, sedang dan rendah. Tujuan memvariasikan kemampuan peserta didik tersebut untuk mewakili semua kategori kemampuan. Sebelum melakukan uji coba, terlebih dahulu peserta didik akan diingatkan kembali mengenai materi gaya antarmolekul. Di saat kesiapan peserta

didik sudah matang, peserta didik diminta mengerjakan 24 butir soal pilihan ganda dalam waktu 60 menit secara individu. Setelah soal selesai diuji cobakan kepada peserta didik, selanjutnya data jawaban peserta didik akan dianalisis.

Melakukan analisis Rasch, data mentah yang didapatkan dianalisis dengan menggunakan aplikasi perangkat lunak Ministep dan model Rasch. Pada pemodelan Rasch uji validitas dikenal dengan *item fit*, yaitu item yang digunakan untuk mengevaluasi apakah instrumen mampu mengukur apa yang dimaksudkan untuk diukur (Sumintono & Widhiarso, 2015).

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Item
6	7	9	-1.67	1.08	1.58	1.08	2.10	1.07	.41	.60	77.8	86.0	S6
12	7	9	-1.67	1.08	1.58	1.08	2.10	1.07	.41	.60	77.8	86.0	S12
24	3	9	2.36	1.00	1.81	1.50	1.05	.50	.53	.67	55.6	83.2	S24
15	1	9	4.85	1.32	1.78	1.08	.61	.14	.36	.48	77.8	91.5	S15
2	8	9	-3.02	1.28	1.58	.93	.56	.10	.37	.46	77.8	90.6	S2
20	8	9	-3.02	1.28	1.58	.93	.56	.10	.37	.46	77.8	90.6	S20
21	8	9	-3.02	1.28	1.58	.93	.56	.10	.37	.46	77.8	90.6	S21
23	7	9	-1.67	1.08	1.27	.63	.69	.23	.55	.60	77.8	86.0	S23
14	3	9	2.36	1.00	1.21	.56	.68	.21	.65	.67	77.8	83.2	S14
10	4	9	1.40	.99	1.01	.20	.58	.12	.73	.71	77.8	83.9	S10
1	7	9	-1.67	1.08	.89	-.01	.39	-.10	.64	.60	77.8	86.0	S1
22	7	9	-1.67	1.08	.89	-.01	.39	-.10	.64	.60	77.8	86.0	S22
13	4	9	1.40	.99	.83	-.16	.44	-.03	.77	.71	77.8	83.9	S13
16	4	9	1.40	.99	.83	-.16	.44	-.03	.77	.71	77.8	83.9	S16
4	3	9	2.36	1.00	.73	-.45	.37	-.13	.74	.67	77.8	83.2	S4
5	3	9	2.36	1.00	.51	-1.09	.26	-.29	.78	.67	100.0	83.2	S5
19	3	9	2.36	1.00	.51	-1.09	.26	-.29	.78	.67	100.0	83.2	S19
3	7	9	-1.67	1.08	.47	-1.02	.22	-.38	.71	.60	100.0	86.0	S3
9	6	9	-.59	1.02	.42	-1.20	.22	-.37	.81	.68	100.0	84.9	S9
7	8	9	-3.02	1.28	.40	-.88	.14	-.56	.58	.46	100.0	90.6	S7
17	8	9	-3.02	1.28	.40	-.88	.14	-.56	.58	.46	100.0	90.6	S17
18	1	9	4.85	1.32	.34	-.89	.12	-.61	.60	.48	100.0	91.5	S18
MEAN	5.6	9.0	-.39	1.18	1.01	.05	.58	.01			83.8	86.5	
P. SD	2.5	.0	2.76	.26	.50	.87	.53	.44			12.0	3.1	

Gambar 1. Item Fit Uji Coba Skala Kecil

Berdasarkan Gambar 1, terdapat 12 item yang tidak memenuhi kriteria MNSQ (kotak merah) yaitu S1, S22, S13, S16, S4, S5, S19, S3, S9, S7, S17, dan S18 karena memiliki nilai MNSQ kecil dari 0,5 serta item S6 dan S12 karena memiliki nilai MNSQ besar dari 1,5. Namun, 14 item tersebut tetap dikategorikan *fit* karena memenuhi nilai ZSTD dan *Pt Mean Corr*. Kemudian, terdapat 4 item yang tidak memenuhi kriteria *Pt Mean Corr* (kotak biru) yaitu S15, S2, S20 dan S21 karena memiliki nilai kecil dari 0,4. Tetapi soal tersebut tetap dipertahankan karena memenuhi kriteria MNSQ dan ZSTD. Sementara itu, item S24, S23, S14 dan S10 memenuhi 3 kriteria. Sebanyak 24 item dalam instrumen yang dikembangkan hanya 22 item yang *fit* (sesuai), dua item lainnya tidak *fit* yaitu item S8 dan S11 dikarenakan item tersebut *minimum measure* (terlalu mudah).

Nilai MNSQ menunjukkan kesesuaian data dengan model. Semakin mendekati nilai 1,00 maka semakin sesuai butir soal tersebut. Nilai ZSTD bertindak sebagai uji-t yang menunjukkan hipotesis kesesuaian data. Nilai ZSTD yang diharapkan adalah mendekati (Sumintono & Widhiarso, 2015). *Pt Mean Corr* mengacu pada korelasi antara skor peserta didik dan kemampuan, semakin tinggi dan semakin positif korelasinya maka semakin baik (Wei et al., 2012).

Reliabilitas merujuk pada konsistensi dan kestabilan hasil dari suatu alat ukur. Reliabilitas adalah ukuran seberapa konsisten pengukuran yang diperoleh dari waktu ke waktu memberikan hasil (Sumintono & Widhiarso, 2014).

SUMMARY OF 24 MEASURED (EXTREME AND NON-EXTREME) Item

	TOTAL		MEASURE	MODEL	INFIT		OUTFIT	
	SCORE	COUNT		S.E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	5.6	9.0	-.39	1.18				
SEM	.5	.0	.58	.05				
P.SD	2.5	.0	2.76	.26				
S.SD	2.5	.0	2.82	.27				
MAX.	9.0	9.0	4.85	1.95				
MIN.	1.0	9.0	-4.67	.99				
REAL RMSE	1.33	TRUE SD	2.42	SEPARATION	1.83	Item	RELIABILITY	.77
MODEL RMSE	1.21	TRUE SD	2.48	SEPARATION	2.05	Item	RELIABILITY	.81
S.E. OF Item MEAN = .58								

Gambar 2. Item Reliability Uji Coba Skala Kecil

Berdasarkan Gambar 2, nilai reliabilitas pengukuran uji coba (kotak merah) didapatkan sebesar 0,77 yang termasuk kategori cukup. Hal ini menunjukkan bahwa hasil pengukuran dapat diandalkan dan konsisten (Sumintono & Widhiarso, 2015). Selanjutnya indeks kesukaran, nilai yang perlu diperhatikan pada analisis indeks kesukaran pada model Rasch terletak pada kolom JMLE *Measure* (kotak biru pada Gambar 3) yaitu *logit* yang menunjukkan urutan tingkat kesulitan soal dari tersulit hingga termudah.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE	MODEL	INFIT		OUTFIT		PTMEASUR-AL		EXACT	MATCH	Item
			MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
15	1	9	4.85	1.32	1.78	1.08	.61	.14	.36	.48	77.8	91.5	S15
18	1	9	4.85	1.32	.34	-.89	.12	-.61	.60	.48	100.0	91.5	S18
4	3	9	2.36	1.00	.73	-.45	.37	-.13	.74	.67	77.8	83.2	S4
5	3	9	2.36	1.00	.51	-1.09	.26	-.29	.78	.67	100.0	83.2	S5
14	3	9	2.36	1.00	1.21	.56	.68	.21	.65	.67	77.8	83.2	S14
19	3	9	2.36	1.00	.51	-1.09	.26	-.29	.78	.67	100.0	83.2	S19
24	3	9	2.36	1.00	1.81	1.50	1.05	.50	.53	.67	55.6	83.2	S24
10	4	9	1.40	.99	1.01	.20	.58	.12	.73	.71	77.8	83.9	S10
13	4	9	1.40	.99	.83	-.16	.44	-.03	.77	.71	77.8	83.9	S13
16	4	9	1.40	.99	.83	-.16	.44	-.03	.77	.71	77.8	83.9	S16
9	6	9	-.59	1.02	.42	-1.20	.22	-.37	.81	.68	100.0	84.9	S9
1	7	9	-1.67	1.08	.89	-.01	.39	-.10	.64	.60	77.8	86.0	S1
3	7	9	-1.67	1.08	.47	-1.02	.22	-.38	.71	.60	100.0	86.0	S3
6	7	9	-1.67	1.08	1.58	1.08	2.10	1.07	.41	.60	77.8	86.0	S6
12	7	9	-1.67	1.08	1.58	1.08	2.10	1.07	.41	.60	77.8	86.0	S12
22	7	9	-1.67	1.08	.89	-.01	.39	-.10	.64	.60	77.8	86.0	S22
23	7	9	-1.67	1.08	1.27	.63	.69	.23	.55	.60	77.8	86.0	S23
2	8	9	-3.02	1.28	1.58	.93	.56	.10	.37	.46	77.8	90.6	S2
7	8	9	-3.02	1.28	.40	-.88	.14	-.56	.58	.46	100.0	90.6	S7
17	8	9	-3.02	1.28	.40	-.88	.14	-.56	.58	.46	100.0	90.6	S17
20	8	9	-3.02	1.28	1.58	.93	.56	.10	.37	.46	77.8	90.6	S20
21	8	9	-3.02	1.28	1.58	.93	.56	.10	.37	.46	77.8	90.6	S21
8	9	9	-4.67	1.95	MINIMUM MEASURE				.00	.00	100.0	100.0	S8
11	9	9	-4.67	1.95	MINIMUM MEASURE				.00	.00	100.0	100.0	S11
MEAN	5.6	9.0	-.39	1.18	1.01	.05	.58	.01			83.8	86.5	
P.SD	2.5	.0	2.76	.26	.50	.87	.53	.44			12.0	3.1	

Gambar 3. Item Measure Uji Coba Skala Kecil

Berdasarkan Gambar 3, nilai Standar Deviasi (SD) sebesar 2,76 sehingga pengelompokan tingkat kesulitan soal dapat dilihat pada Tabel 5.

Tabel 5. Pengelompokan Indeks Kesukaran Soal

Tingkat Kesulitan	Nilai <i>Measure</i>	Butir Soal
Sangat sulit	> 2,76	S15, S18
Sulit	0,0 – 2,76	S4, S5, S14, S19, S24, S10, S13, S16
Mudah	-2,76 – 0,0	S9, S1, S3, S6, S12, S22, S23
Sangat mudah	< - 2,76	S2, S7, S17, S20, S21, S8, S11

Nilai separasi menunjukkan bagaimana item-item dikelompokkan. Karena dapat mengidentifikasi kelompok-kelompok responden, semakin besar nilai separasi, semakin baik instrumen tersebut (Sumintono & Widhiarso, 2015). Nilai separasi yang didapatkan adalah 1,83 sehingga diperoleh nilai strata sebagai berikut.

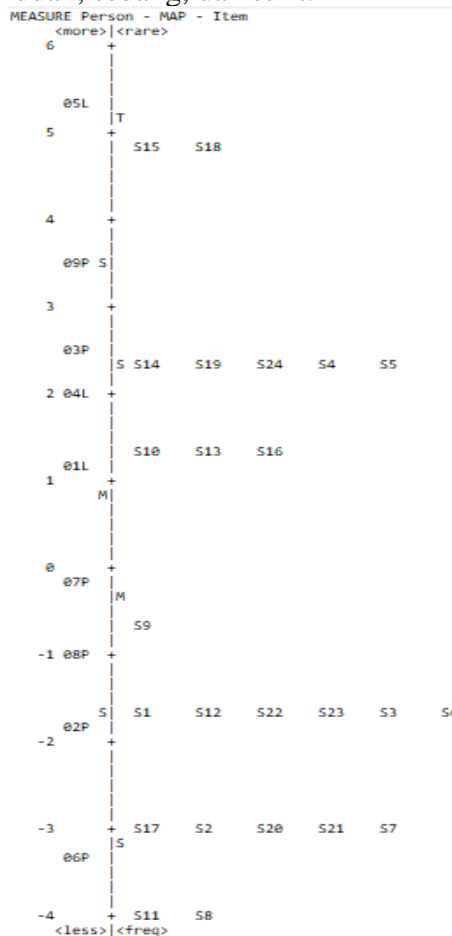
$$H = \frac{[(4 \times SEPARATION) + 1]}{3}$$

$$H = \frac{[(4 \times 1,83) + 1]}{3}$$

$$H = \frac{8,32}{3} = 2,77$$

Nilai strata sebesar 2,77 jika dibulatkan menjadi 3, hal ini mengindikasikan bahwa soal-soal dibagi ke dalam tiga kategori: soal mudah, sedang, dan sulit (Sumintono & Widhiarso, 2015).

Meninjau statistik kecocokan item, berdasarkan hasil analisis Rasch didapatkan bahwa 22 item dinyatakan valid karena telah memenuhi kriteria *fit*. Kemudian terdapat 2 item yang tidak valid dikarenakan soal tersebut *minimum measure* (terlalu mudah) yaitu S8 dan S11 sehingga perlu diperbaiki. Nilai realibilitas didapatkan sebesar 0,77 dengan kategori cukup. Analisis indeks kesukaran terdapat empat kategori yaitu sangat sulit, sulit, mudah dan sangat mudah. Selain itu, analisis daya beda soal menghasilkan nilai strata 3, yang mengindikasikan bahwa butir soal dibagi ke dalam tiga kategori: mudah, sedang, dan sulit.



Gambar 4. Peta Wright Uji Coba Skala Kecil

Meninjau peta Wright, peninjauan dilakukan untuk melihat pendistribusian abilitas/kemampuan subjek terhadap tingkat kesulitan soal. Peta Wright menampilkan distribusi tingkat kesulitan soal di sisi kanan dan distribusi kemampuan peserta didik di sisi kiri. Peserta didik yang berada di posisi paling atas artinya memiliki kemampuan dan pemahaman tinggi terhadap materi gaya antarmolekul. Item atau butir soal yang berada paling atas memiliki tingkat kesulitan soal tinggi, begitu juga sebaliknya (Sumintono & Widhiarso, 2015).

Berdasarkan Gambar 4, penyebaran tingkat kesulitan item masih belum secara merata karena terdapat interval kesulitan item yang tidak diisi oleh item. Penyebab kosongnya interval tersebut adalah faktor kebetulan, sehingga ketika ukuran sampel diperbesar maka akan banyak peserta didik yang mengisi interval tersebut (Sumintono & Widhiarso, 2015).

Mengulangi langkah 4-7 jika terdapat item yang tidak *fit*. Berdasarkan langkah 4 sampai 7 yang telah dilakukan terdapat item yang tidak sesuai (*fit*) dan perlu direvisi. Instrumen yang telah direvisi harus diberikan kepada sampel baru dan mengulangi kembali langkah 4-7 hingga didapatkan item yang *fit*. Uji coba tahap kedua dilakukan kepada peserta didik kelas XI Fase F sebanyak 60 orang. Data mentah yang didapatkan dari uji coba skala besar dianalisis menggunakan model Rasch dan didapatkan hasil sebagai berikut.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT		OUTFIT		PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Item
					MNSQ	ZSTD	MNSQ	ZSTD					
14	14	60	1.27	.33	1.08	.50	1.40	1.47	.23	.36	81.7	79.1	S14
11	16	60	1.07	.31	1.23	1.40	1.31	1.33	.13	.37	68.3	76.5	S11
6	31	60	-.18	.28	1.26	2.85	1.28	1.87	.09	.36	50.0	64.2	S6
4	42	60	-1.07	.30	1.06	.51	1.17	.79	.23	.32	75.0	72.2	S4
21	44	60	-1.25	.31	1.14	.96	1.05	.29	.18	.31	68.3	74.8	S21
22	20	60	.70	.30	1.13	1.01	1.13	.75	.24	.37	66.7	71.5	S22
20	39	60	-.81	.29	1.10	.95	1.06	.35	.23	.33	65.0	68.7	S20
23	22	60	.53	.29	1.06	.54	1.09	.58	.31	.37	70.0	69.3	S23
19	24	60	.36	.28	1.04	.44	1.07	.51	.32	.37	65.0	67.5	S19
12	23	60	.44	.29	1.04	.42	.96	-.20	.35	.37	61.7	68.4	S12
17	28	60	.05	.28	1.03	.36	.96	-.24	.35	.36	56.7	64.7	S17
15	22	60	.53	.29	1.01	.13	1.00	.07	.36	.37	70.0	69.3	S15
18	12	60	1.50	.35	.98	-.03	.99	.06	.37	.36	81.7	82.0	S18
5	17	60	.97	.31	.96	-.21	.98	-.03	.40	.37	75.0	75.2	S5
1	25	60	.28	.28	.95	-.49	.95	-.32	.42	.37	70.0	66.6	S1
24	33	60	-.34	.28	.95	-.52	.90	-.69	.41	.35	65.0	64.8	S24
3	39	60	-.81	.29	.93	-.64	.86	-.71	.42	.33	65.0	68.7	S3
7	45	60	-1.35	.31	.92	-.51	.83	-.57	.39	.30	78.3	76.1	S7
16	21	60	.61	.29	.91	-.78	.89	-.65	.47	.37	71.7	70.3	S16
9	36	60	-.57	.28	.88	-1.28	.81	-1.20	.48	.34	73.3	66.4	S9
13	18	60	.87	.30	.87	-.95	.79	-1.08	.52	.37	76.7	73.9	S13
10	27	60	.13	.28	.85	-1.72	.81	-1.47	.53	.37	75.0	65.2	S10
2	48	60	-1.67	.34	.84	-.81	.64	-1.15	.47	.28	80.0	80.3	S2
8	44	60	-1.25	.31	.80	-1.47	.68	-1.34	.53	.31	81.7	74.8	S8
MEAN	28.8	60.0	.00	.30	1.00	.03	.98	-.06			70.5	71.3	
P. SD	10.8	.0	.89	.02	.12	1.01	.18	.89			7.8	5.1	

Gambar 5. Item Fit Uji Coba Skala Besar

Berdasarkan Gambar 5, terdapat 14 item tidak memenuhi kriteria *Pt Measure Corr* yaitu S14, S11, S6, S4, S21, S22, S20, S23, S19, S12, S17, S15, S18 dan S7 karena memiliki nilai kecil dari 0,4. Namun, 13 item tersebut termasuk item yang *fit* karena memenuhi nilai MNSQ dan ZSTD. Sementara itu, item lainnya memenuhi seluruh kriteria yang diharapkan sehingga semua item atau butir soal pada instrumen terbukti valid. Hal ini mengindikasikan bahwa alat ini mengukur sesuatu yang perlu diukur (Sumintono & Widhiarso, 2015).

SUMMARY OF 24 MEASURED Item

	TOTAL		MEASURE	MODEL S. E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	28.8	60.0	.00	.30	1.00	.03	.98	-.06
SEM	2.2	.0	.19	.00	.02	.21	.04	.19
P. SD	10.8	.0	.89	.02	.12	1.01	.18	.89
S. SD	11.0	.0	.91	.02	.12	1.03	.19	.91
MAX.	48.0	60.0	1.50	.35	1.26	2.85	1.40	1.87
MIN.	12.0	60.0	-1.67	.28	.80	-1.72	.64	-1.47
REAL RMSE	.31	TRUE SD	.84	SEPARATION	2.75	Item	RELIABILITY	.88
MODEL RMSE	.30	TRUE SD	.84	SEPARATION	2.82	Item	RELIABILITY	.89
S. E. OF Item MEAN = .19								

Gambar 6. Nilai Reliability Uji Coba Skala Besar

Berdasarkan Gambar 6, nilai item *reliability* sebesar 0,88 (kotak merah) artinya reliabilitas pengukuran yang dihasilkan bagus. Ini menandakan bahwa jika dilakukan uji coba ulang berkali-kali dalam rentang waktu yang lama, hasil yang didapatkan terhadap masing-masing butir soal tidak akan jauh berbeda (Sumintono & Widhiarso, 2015). Oleh karena itu, dapat disimpulkan bahwa instrumen yang dihasilkan reliabel.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		Item
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
18	12	60	1.50	.35	.98	-.03	.99	.06	.37	.36	81.7	82.0	S18
14	14	60	1.27	.33	1.08	.50	1.40	1.47	.23	.36	81.7	79.1	S14
11	16	60	1.07	.31	1.23	1.40	1.31	1.33	.13	.37	68.3	76.5	S11
5	17	60	.97	.31	.96	-.21	.98	-.03	.40	.37	75.0	75.2	S5
13	18	60	.87	.30	.87	-.95	.79	-1.08	.52	.37	76.7	73.9	S13
22	20	60	.70	.30	1.13	1.01	1.13	.75	.24	.37	66.7	71.5	S22
16	21	60	.61	.29	.91	-.78	.89	-.65	.47	.37	71.7	70.3	S16
15	22	60	.53	.29	1.01	.13	1.00	.07	.36	.37	70.0	69.3	S15
23	22	60	.53	.29	1.06	.54	1.09	.58	.31	.37	70.0	69.3	S23
12	23	60	.44	.29	1.04	.42	.96	-.20	.35	.37	61.7	68.4	S12
19	24	60	.36	.28	1.04	.44	1.07	.51	.32	.37	65.0	67.5	S19
1	25	60	.28	.28	.95	-.49	.95	-.32	.42	.37	70.0	66.6	S1
10	27	60	.13	.28	.85	-1.72	.81	-1.47	.53	.37	75.0	65.2	S10
17	28	60	.05	.28	1.03	.36	.96	-.24	.35	.36	56.7	64.7	S17
6	31	60	-.18	.28	1.26	2.85	1.28	1.87	.09	.36	50.0	64.2	S6
24	33	60	-.34	.28	.95	-.52	.90	-.69	.41	.35	65.0	64.8	S24
9	36	60	-.57	.28	.88	-1.28	.81	-1.20	.48	.34	73.3	66.4	S9
3	39	60	-.81	.29	.93	-.64	.86	-.71	.42	.33	65.0	68.7	S3
20	39	60	-.81	.29	1.10	.95	1.06	.35	.23	.33	65.0	68.7	S20
4	42	60	-1.07	.30	1.06	.51	1.17	.79	.23	.32	75.0	72.2	S4
8	44	60	-1.25	.31	.80	-1.47	.68	-1.34	.53	.31	81.7	74.8	S8
21	44	60	-1.25	.31	1.14	.96	1.05	.29	.18	.31	68.3	74.8	S21
7	45	60	-1.35	.31	.92	-.51	.83	-.57	.39	.30	78.3	76.1	S7
2	48	60	-1.67	.34	.84	-.81	.64	-1.15	.47	.28	80.0	80.3	S2
MEAN	28.8	60.0	.00	.30	1.00	.03	.98	-.06			70.5	71.3	
P. SD	10.8	.0	.89	.02	.12	1.01	.18	.89			7.8	5.1	

Gambar 7. Item Measure Uji Skala Besar

Berdasarkan Gambar 7, nilai standar deviasi (SD) uji coba skala besar yaitu 0,89 sehingga Pengelompokan indeks kesukaran soal dapat dilihat pada Tabel 6.

Tabel 6. Pengelompokan Indeks Kesukaran Uji Coba Skala Besar

Tingkat Kesulitan	Nilai Measure	Butir Soal
Sangat sulit	> 0,89	S18, S14, S11, S5
Sulit	0,0 – 0,89	S13, S22, S16, S15, S23, S12, S19, S1, S10, S17
Mudah	-0,89 – 0,0	S6, S24, S9, S3, S20
Sangat mudah	< - 0,89	S4, S8, S21, S7, S2

Soal yang tidak terlalu mudah dan tidak terlalu sulit adalah soal yang baik (Arikunto, 2012). Dalam model Rasch tidak ada soal dengan kategori sedang, semua item dikategorikan menjadi sangat sulit, sulit, mudah dan sangat mudah. Pada hakikatnya tingkat kesulitan soal itu dibagi menjadi tiga yaitu sulit, sedang dan



mudah sehingga soal yang kategori sangat sulit dalam model Rasch merupakan soal sulit, soal kategori sulit dan mudah termasuk soal sedang, dan soal kategori sangat mudah merupakan soal mudah.

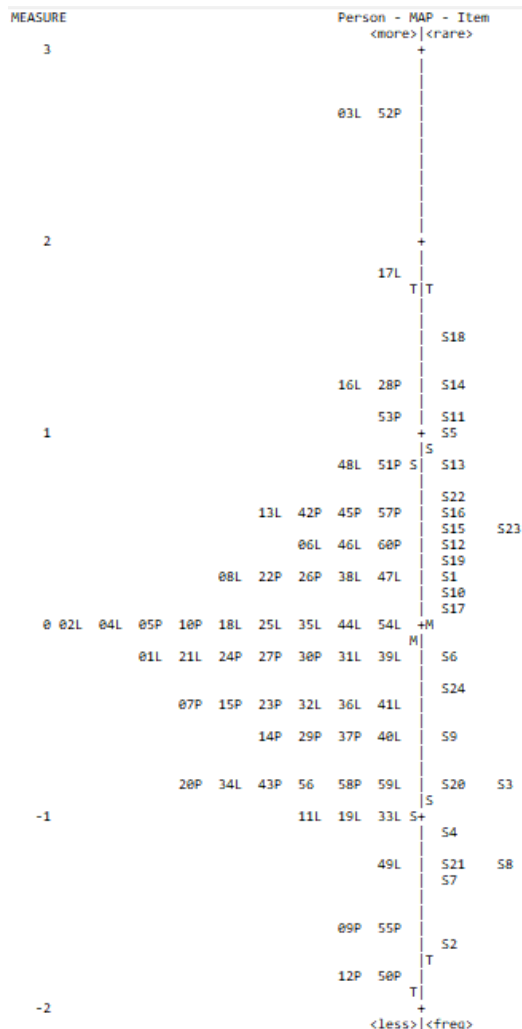
Pengelompokkan butir soal secara lebih teliti dapat menggunakan persamaan pemisahan strata dengan menggunakan nilai separation. Nilai separation yang didapatkan pada uji coba tahap kedua sebesar 2,75 sehingga nilai stratanya sebagai berikut.

$$H = \frac{[(4 \times SEPARATION) + 1]}{3}$$

$$H = \frac{[(4 \times 2,75) + 1]}{3}$$

$$H = \frac{12}{3} = 4$$

Nilai strata sebesar 4 maka terdapat 4 kelompok butir soal. Semakin tinggi nilai pemisahan, semakin tinggi kualitas instrumen (Sumintono & Widhiarso, 2015).



Gambar 8.Peta Wright Uji Coba Skala Besar

Berdasarkan Gambar 8, sebaran tingkat kesulitan item menyebar dengan baik sehingga tidak banyak interval yang kosong.

Menetapkan klaim kualitas instrumen penilaian. Validitas dari keseluruhan butir soal dapat diklaim valid karena telah memenuhi paling sedikit 2 dari 3 kriteria *fit*. Reliabilitas pengukuran dapat diklaim reliabel dengan nilai reliabilitas item = 0,88 (bagus). Demikian juga dengan indeks kesukaran yang memiliki 4 tingkatan. Kemudian daya beda soal yang diperoleh dari nilai *separation* dan persamaan strata memiliki 4 kelompok butir soal. Instrumen ini dapat digunakan karena sudah teruji valid, reliabel, memiliki indeks kesukaran dan daya beda soal yang baik.

Mengembangkan pedoman penggunaan instrumen, pedoman penggunaan instrumen harus dibuat untuk memberikan petunjuk kepada guru dan peserta didik tentang penggunaan instrumen penilaian dan untuk mengumpulkan informasi yang lebih rinci tentang fitur-fitur instrumen (Sabekti & Khoirunnisa, 2018). Pedoman penggunaan instrumen terdiri dari cover, petunjuk penggunaan instrumen, kisi-kisi soal, petunjuk pengerjaan soal, lembar soal dan rubrik penilaian.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh bahwa 24 butir soal pada instrumen penilaian yang dikembangkan teruji valid, reliabel dengan nilai 0,88 (bagus), mempunyai empat indeks kesukaran yaitu sangat sulit, sulit, mudah dan sangat mudah serta mempunyai daya beda soal yang baik menurut model Rasch.

DAFTAR PUSTAKA

- Arifin, Z. (2014). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2012). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT Rineka Cipta.
- Asmalia, I., Fadiawati, N., & Kadaritna, N. (2015). Pengembangan Instrumen Asesmen Berbasis Keterampilan Proses Sains pada Materi Stoikiometri. *Jurnal Pendidikan dan Pembelajaran Kimia*, 4(1), 299–311.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.
- Chang, R., & Overby, J. (2011). *General Chemistry The Essential Concepts* (6 ed.). New: McGraw-Hill Companies, Inc.
- Ilmiah, I., Anwar, M., & Herawati, N. (2020). Pengembangan Tes Keterampilan Proses Sains (KPS) pada Materi Asam Basa Kelas XI SMA/MA. *Chemistry Education Review*, 4(1), 64–70. <https://doi.org/https://doi.org/10.26858/cer.v4i1.13315>
- Jayali, A. M., & Sriwahyuni, E. (2022). Analisis Butir Soal Instrumen Tes Keterampilan Proses Sains pada Topik Sistem Periodik Unsur. *Journal on Teacher Education*, 4(2), 1581–1591. Diambil dari <http://journal.universitaspahlawan.ac.id/index.php/jote/article/view/10800%0A> <http://journal.universitaspahlawan.ac.id/index.php/jote/article/download/10800/8354>
- Kemendikbudristek. (2024). Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 032/H/KR/2024 (Issue 021).



- Liu, X. (2010). *Using and Developing Measurement Instruments in Science Education: A Rasch Modeling Approach* (2nd ed.). America: IAP Press: Charlotte, NC.
- Oktaviani, E., Fadiawati, N., & Kadaritna, N. (2015). Pengembangan Instrumen Asesmen Berbasis Keterampilan Proses Sains pada Materi Hukum-Hukum Dasar Kimia. *Jurnal Pendidikan dan Pembelajaran Kimia*, 4(1), 324–338.
- Ratnasari, D., Sukarmin, S., Suparmi, S., & Aminah, N. S. (2017). Students' Conception on Heat and Temperature toward Science Process Skill. *Journal of Physics: Conference Series*, 895(1). <https://doi.org/10.1088/1742-6596/895/1/012044>
- Sabekti, W. A., & Khoirunnisa, F. (2018). Penggunaan Rasch Model untuk Mengembangkan Instrumen Pengukuran Kemampuan Berpikir Kritis Siswa pada Topik Ikatan Kimia. *Zarah*, 6(2), 68–75.
- Salmawati, L., Siswaningsih, W., & Rahmawati, T. (2023). Pengembangan Tes Keterampilan Proses Sains Kelas XI Pada Materi Termokimia. *Jurnal Riset dan Praktik Pendidikan Kimia*, 11(2), 173–183.
- Sick, J. (2013). Rasch measurement and inter-rater reliability. *Shiken Research Bulletin*, 17(2), 23–26.
- Siegel, J. (2023). Pen and paper or computerized notetaking? L2 English students' views and habits. *Computers and Education Open journal*, 4(December 2022), 1–10.
- Sugiyono. (2006). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ILmu Sosial*. (B. Trim, Ed.). Cimahi: Trim Komunikata.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata.
- Tosun, C. (2019). Scientific Process Skills Test Development within the Topic “Matter and Its Nature” and the Predictive Effect of Different Variables on 7th and 8th Grade Students' Scientific Process Skill Levels. *The Royal Society of Chemistry*, 1–16. <https://doi.org/10.1039/C8RP00071A>
- Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using Rasch Measurement To Develop a Computer Modeling-Based Instrument To Assess Students' Conceptual Understanding of Matter. *J. Chem. Educ*, 89, 335–345.