

Pengembangan Instrumen Tes Keterampilan Proses Sains pada Topik Kesetimbangan Kimia: Analisis dengan Model Rasch

Hasnatul Hasanah, Faizah Qurrata Aini*
Universitas Negeri Padang, Padang, Indonesia

*Corresponding Author: faizah_qurrata@fmipa.unp.ac.id
Dikirim: 12-12-2024; Direvisi: 21-12-2024; Diterima: 25-12-2024

Abstrak: Proses penilaian di sekolah saat ini lebih fokus pada aspek pengetahuan daripada aspek keterampilan proses sains sehingga belum sesuai dengan capaian pembelajaran pada kurikulum merdeka. Tujuan dari penelitian ini adalah mengembangkan instrumen tes keterampilan proses sains pada topik kesetimbangan kimia menggunakan model Rasch. Terdapat enam aspek yang akan diukur pada tes KPS ini sesuai dengan capaian pembelajaran pada kurikulum merdeka yaitu mengamati, mempertanyakan dan memprediksi, merencanakan dan memilih metode, memproses, menganalisis, data dan informasi, mengevaluasi dan refleksi, serta mengomunikasikan hasil. Instrumen ini terdiri dari 18 soal pilihan ganda yang telah divalidasi oleh lima orang ahli. Hasil validitas logis dianalisis menggunakan program Minifac dan diperoleh penilaian ahli (*exact agreements*) sebesar 94,9% dan jika dibandingkan dengan penilaian yang diharapkan pada model Rasch (*expected agreement*) sebesar 94,5% sehingga soal yang telah dirancang dapat dikatakan valid secara logis. Data hasil uji coba kepada peserta didik dianalisis menggunakan model Rasch melalui program Ministep menunjukkan hasil yang sesuai untuk setiap butir soal karena memenuhi kriteria nilai *outfit* MNSQ, ZSTD, dan *Pt Mean Corr*. Nilai reliabilitas yang diperoleh sebesar 0,88 yang termasuk kriteria bagus. Empat klasifikasi butir soal yaitu sangat sulit, sulit, mudah, dan sangat mudah diperoleh dari nilai indeks kesukaran dan daya beda soal.

Kata Kunci: Instrumen Tes; Keterampilan Proses Sains; Kesetimbangan Kimia; Model Rasch

Abstract: The assessment process in schools currently focuses more on the knowledge aspect than the science process skills aspect so that it is not in accordance with the learning outcomes in the merdeka curriculum. The purpose of this research is to develop a science process skills test instrument on the topic of chemical equilibrium using the Rasch model. There are six aspects that will be measured in this KPS test in accordance with the learning outcomes of the merdeka curriculum, namely observing, questioning and predicting, planning and choosing methods, processing, analyzing, data and information, evaluating and reflecting, and communicating results. This instrument consists of 18 multiple choice questions that have been validated by five experts. The results of logical validity were analyzed using the Minifac program and obtained an expert assessment (*exact agreements*) of 94.9% and when compared to the expected assessment in the Rasch model (*expected agreement*) of 94.5% so that the questions that have been designed can be said to be logically valid. The data from the trial results to students analyzed using the Rasch model through the Ministep program showed appropriate results for each item because it met the criteria for the *outfit* MNSQ, ZSTD, and *Pt Mean Corr* values. The reliability value obtained is 0.88 which includes good criteria. Four classifications of question items, namely very difficult, difficult, easy, and very easy, are obtained from the difficulty index value and the differentiation of the question.

Keywords: Test Instrument; Science Process Skills; Chemical Equilibrium; Rasch Model

PENDAHULUAN

Ilmu kimia merupakan bidang ilmu pengetahuan alam yang membahas tentang struktur dan karakteristik materi (zat), perubahan materi serta energi yang menyertai perubahannya (Chang & Overby, 2011). Kimia dapat dibagi menjadi dua bagian yang berbeda yaitu kimia sebagai proses yang melibatkan kemampuan dan sikap para ilmuwan untuk mendapatkan dan mengembangkan pengetahuan, dan kimia sebagai hasil yang berbentuk fakta, konsep, hukum, dan hipotesis yang ditemukan oleh para ilmuwan. Kedua hal ini sangat erat kaitannya sehingga kimia sebagai produk dan proses harus sama-sama diperhatikan (Sudarmin, 2015). Keterampilan diperlukan dalam memahami kimia sebagai proses yang tidak terpisah dengan produknya. Keterampilan yang dimaksud merupakan Keterampilan Proses Sains (KPS) (Agustina et al., 2017).

Keterampilan proses sains adalah kemampuan dasar yang membantu peserta didik mempelajari sains dengan lebih mudah, menjadi lebih terlibat, bertanggung jawab dan dapat meningkatkan teknik penelitian dan pembelajaran (Gürses et al., 2015). Pentingnya keterampilan proses sains ini tertera dalam kurikulum merdeka yaitu pada capaian pembelajaran kimia yang terdiri dari elemen pemahaman kimia dan elemen keterampilan proses sains (Kemendikbudristek, 2022). Suatu proses pembelajaran akan bermakna ketika peserta didik mampu mengidentifikasi konsep-konsep dari materi yang diajarkan dan salah satunya dengan cara menilai KPS pada proses pembelajaran (Aditiyas & Kuswanto, 2024). Berdasarkan wawancara yang dilakukan kepada guru kimia di SMA Negeri 2 Padang, SMA Negeri 5 Padang, SMA Negeri 1 Padang Sago, dan SMA Pembangunan Laboratorium UNP yang mengungkapkan bahwa guru lebih fokus pada penilaian pengetahuan dari pada keterampilan dan menganggap keterampilan proses sains hanya bisa dinilai ketika peserta didik melakukan praktikum sehingga penilaian belum bisa dilakukan secara menyeluruh. Menurut Tosun (2019) penilaian KPS peserta didik dapat diukur dengan melibatkan materi didalamnya. Sementara itu, dalam proses penilaian KPS guru hanya melakukan pengamatan tanpa menggunakan instrumen khusus yang dapat mengukur KPS peserta didik. Dengan demikian, penilaian yang dilakukan menjadi kurang objektif karena dengan menggunakan metode pengamatan langsung guru tidak dapat mengamati semua peserta didik (Arini et al., 2017).

Penelitian Amali & Firman (2024) menyatakan bahwa penilaian keterampilan proses sains ini masih belum dilakukan oleh tenaga pendidik dengan berbagai alasan, seperti sistem pembelajaran yang masih didominasi oleh sains sebagai produk, bukan sains sebagai proses. Selain itu, selama ini ada kecenderungan dalam proses pembelajaran kimia yang menganggap bahwa praktikum atau eksperimen adalah satu-satunya cara untuk mengukur KPS. Anggapan ini membuat para pendidik percaya bahwa KPS tidak dapat diterapkan dengan menggunakan materi pembelajaran yang tidak diberikan melalui kegiatan praktikum (Sriwahyuni, 2022). Akibatnya, guru lebih mementingkan penilaian pada aspek pengetahuan daripada menguji keterampilan proses sains. Oleh karena itu, diperlukan suatu instrumen tes untuk mengukur KPS peserta didik.

Beberapa penelitian terdahulu terkait instrumen tes ini telah dikembangkan salah satunya penelitian oleh Ilmiah et al, (2020), penelitian ini memiliki kelemahan yaitu teknik pengumpulan data pada instrumen penilaian diperoleh dari data observasi artinya guru tidak bisa mengamati seluruh peserta didik dalam satu kegiatan



pembelajaran. Penelitian lain juga telah dilakukan pada beberapa materi kimia seperti materi stoikiometri (Asmalia et al., 2015), hukum-hukum dasar kimia (Oktaviani et al., 2015), termokimia (Salmawati et al., 2023), hasil kali kelarutan (Nahadi et al., 2012), dan titrasi asam basa (Agustin et al., 2013).

Namun demikian, belum ditemukan instrumen untuk mengukur KPS materi kesetimbangan kimia khususnya pada kurikulum merdeka. Salah satu materi kimia yang banyak mengandung konsep, reaksi kimia, dan perhitungan adalah kesetimbangan kimia. Selain itu, selama proses pembelajaran peserta didik kesulitan untuk menghubungkan konsep kesetimbangan kimia dengan situasi dunia nyata yang membutuhkan keterampilan proses sains. Materi kesetimbangan kimia sangat erat kaitannya dengan keterampilan proses sains untuk menciptakan pengetahuan, memecahkan permasalahan, dan merumuskan hasil. Dengan adanya KPS, peserta didik dapat dilibatkan untuk menyelidiki konsep, fakta, teori dan hukum yang akan dipelajari, sehingga dapat meningkatkan aspek keterampilan proses (Sitorus & Dalimunthe, 2024). Terdapat enam aspek yang akan diukur pada tes KPS ini yaitu mengamati, mempertanyakan dan memprediksi, merencanakan dan memilih metode, memproses, menganalisis, data dan informasi, mengevaluasi dan refleksi, serta mengomunikasikan hasil.

Pemodelan Rasch akan digunakan dalam instrumen tes yang akan dikembangkan. Manfaat dari pemodelan Rasch antara lain adalah kemampuannya dalam mendeteksi tebakan, meramalkan skor dari data yang hilang, mengidentifikasi kesalahan respon, dan tidak bergantung pada kuantitas respon yang benar. Instrumen penilaian harus memenuhi sejumlah persyaratan sebelum dapat dikatakan baik, oleh karena itu untuk memenuhi syarat tersebut perlu dilakukan analisis validitas, reabilitas, daya beda soal, dan indeks kesukaran. Penelitian ini bertujuan untuk membuat instrumen tes yang dapat mengukur keterampilan proses sains pada topik kesetimbangan kimia yang ditinjau dari model Rasch.

METODE PENELITIAN

Jenis penelitian yang digunakan yaitu Research and Development (R&D) dengan menggunakan model Rasch. Penelitian ini dilakukan di kelas XII fase F SMA Negeri 1 Padang Sago pada semester ganjil 2024/2025. Sumber data pada penelitian ini diperoleh langsung dari 3 orang dosen Departemen Pendidikan Kimia UNP dan 2 orang guru kimia SMA Negeri 1 Padang Sago. Kemudian data dari peserta didik kelas XII fase F SMA Negeri 1 Padang Sago melalui uji coba instrumen tes. Penelitian ini dirancang menggunakan model Rasch yang terdiri dari 10 langkah yang telah dikembangkan oleh Liu (2010). Prosedur penelitian telah diubah dan disesuaikan dengan kebutuhan penelitian yaitu: pertama, menetapkan tujuan dan populasi, kedua menentukan konstruk, ketiga mengidentifikasi kinerja konstruk yang telah ditentukan, keempat melakukan uji coba, kelima menganalisis data menggunakan model Rasch, keenam meninjau statistik kecocokan *item*, ketujuh meninjau peta Wright, kedelapan mengulangi langkah 4-7 jika terdapat *item* yang tidak *fit*, kesembilan menetapkan klaim kualitas instrumen tes, dan yang terakhir yaitu mengembangkan pedoman penggunaan instrumen tes.

Pemodelan Rasch digunakan dalam analisis data karena dapat digunakan untuk memastikan atribut dan kualitas setiap butir soal. Validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal adalah empat kriteria yang akan diukur dalam



instrumen ini. Validitas logis diuji oleh pakar atau ahli yang bertujuan untuk melihat apakah instrumen tes ini layak untuk diuji cobakan kepada peserta didik. Hasil uji validitas logis akan dianalisis dengan bantuan program Minifac. Sedangkan validitas empiris dilakukan uji coba instrumen tes kepada 60 orang peserta didik untuk melihat kecocokan butir soal dengan tujuan pengukurannya. Penentuan item *fit* ditentukan berdasarkan kriteria berikut:

Tabel 1. Kriteria Nilai Validitas

Kriteria	Interval	Keterangan
<i>Outfit Mean Square</i>	0,5 < MNSQ < 1,5	Diterima
<i>Outfit Z-Standard</i>	-2,0 < ZSTD < +2,0	Diterima
<i>Pt. Mean Corr</i>	0,4 < Pt. M. C < 0,85	Diterima

Apabila salah satu dari tiga kriteria tersebut dapat terpenuhi, maka butir soalnya dapat dikatakan valid (Sumintono & Widhiarso, 2015). Reliabilitas yaitu sejauh mana pengukuran dilakukan berulang akan menghasilkan data atau hasil yang konsisten.

Tabel 2. Kriteria Nilai Reliabilitas

Interval	Keterangan
< 0,67	Lemah
0,67-0,8	Cukup
0,8-0,9	Bagus
0,91-0,94	Bagus sekali
> 0,94	Istimewa

Selanjutnya dengan melihat nilai *logit* dari setiap butir soal, indeks kesukaran soal dapat dilihat pada kolom *measure*.

Tabel 3. Kriteria Nilai Indeks Kesukaran

Nilai Logit	Keterangan
> <i>logit</i> +1SD	Sangat sulit
0,0 <i>logit</i> +1SD	Sulit
0,0 <i>logit</i> -1SD	Mudah
< <i>logit</i> -1SD	Sangat mudah

Kemampuan sebuah butir soal untuk membedakan antara siswa yang memiliki kemampuan tinggi dan rendah dikenal sebagai daya pembeda. Pemisahan strata adalah rumus yang digunakan untuk memeriksa pengelompokkan secara lebih rinci:

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

Tabel 4. Kriteria Daya Beda Soal

Nilai H	Kategori Sebaran	Keterangan
0	Tidak memiliki daya beda	Sangat buruk
1	Mudah	Buruk
2	Sulit dan mudah	Cukup
3	Sulit, sedang, dan mudah	Baik
4	Sangat sulit, sulit, mudah, dan sangat mudah	Sangat baik

Daya beda soal atau dalam pemodelan Rasch disebut dengan daya diskriminasi butir soal yang menjelaskan seberapa baik tingkatan soal untuk mampu membandingkan individu yang mempunyai kemampuan tinggi dan rendah. Semakin



tinggi nilai daya beda butir soal, maka semakin baik pula instrumen yang dikembangkan (Sumintono & Widhiarso, 2015).

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah sebuah instrumen tes yang dirancang untuk mengukur keterampilan proses sains peserta didik pada topik kesetimbangan kimia. Instrumen tes ini terdiri dari 18 soal yang memuat tentang aspek-aspek KPS yang tertera dalam capaian pembelajaran kimia kurikulum merdeka. Pengembangan instrumen tes menggunakan model Rasch yang terdiri dari 10 tahapan, hasil penelitian pada setiap tahapan pengembangannya adalah sebagai berikut:

Menetapkan tujuan dan populasi

Tujuan pengembangan instrumen tes ini adalah untuk mengukur keterampilan proses sains peserta didik pada topik kesetimbangan kimia. Instrumen tes ini termasuk penilaian sumatif karena bertujuan untuk mengukur ketercapaian tujuan pembelajaran yang disesuaikan dengan aspek-aspek keterampilan proses sains. Penilaian sumatif merupakan penilaian yang diselesaikan pada akhir pengajaran atau tergantung pada peraturan yang ditetapkan dan dapat diselesaikan secara bersamaan untuk dua atau lebih tujuan pembelajaran (Mujiburrahman et al., 2023). Peserta didik kelas XII fase F SMA Negeri 1 Padang Sago dijadikan sebagai populasi pada penelitian ini.

Menentukan konstruk

Konstruk adalah atribut yang akan diukur oleh suatu alat ukur (Sumintono & Widhiarso, 2015). Konstruk bisa berupa kemampuan, keterampilan, sikap, maupun atribut lainnya yang diukur dari reaksi subjek terhadap objek tetapi tidak dapat diukur secara langsung. Konstruk pada instrumen tes ini adalah keterampilan proses sains yang diambil dari capaian pembelajaran kimia kurikulum merdeka.

Mengidentifikasi kinerja konstruk yang telah ditentukan

Tahapan ini terbagi menjadi beberapa bagian yaitu, pertama menentukan jumlah soal dan format tes, jumlah soal yang akan dibuat untuk instrumen tes ini yaitu sebanyak 18 butir soal yang disesuaikan dengan delapan belas tujuan pembelajaran topik kesetimbangan kimia berdasarkan aspek-aspek keterampilan proses sains pada capaian pembelajaran kurikulum merdeka. Format tes yang akan digunakan yaitu tes pilihan ganda dengan menggunakan kertas dan pena. Tes ini dipilih karena tes ini termasuk tes yang objektif dan kriteria penskorannya menghasilkan skor yang sama (Basuki et al., 2019). Pemilihan format tes menggunakan kertas dan pena dianggap lebih mudah dan terjangkau serta dapat membantu konsentrasi dan pemahaman peserta didik (Siegel, 2023). Kedua, penyusunan indikator soal yang dirangkum dalam tabel kisi-kisi soal.

Ketiga, merancang butir soal berdasarkan dengan kisi-kisi soal KPS dan membuat rubrik penilaian yang berisi aspek-aspek KPS, tujuan pembelajaran, dan skor. Keempat, menguji validasi logis terhadap instrumen tes. Validasi logis diuji oleh lima orang ahli yang telah memiliki pengalaman dalam menilai instrumen yang dikembangkan. Validasi logis terhadap instrumen tes dilakukan dengan pemberian angket berupa skala Guttman dengan pilihan “Ya atau Tidak” pada 12 kriteria



penilaian. Skala Guttman dipilih karena menghasilkan data interval atau rasio dan menampilkan skala pengukuran dengan dua respon yang pasti (Widana et al., 2022). Hasil validasi didapatkan bahwa sebagian besar dari butir soal dan aspek yang dinilai berada pada pilihan “Ya” dan sebagian kecil aspek butir soal berada dipilihan “Tidak”.

Data hasil validasi soal ini selanjutnya dianalisis dengan model Rasch melalui aplikasi Minifac. Hasil pengukuran oleh validator didapatkan skor *strata value* sebesar 2,73 mendekati 3, hal ini menunjukkan bahwa data yang diperoleh reliabel (Fahmina et al., 2019). Nilai reliabilitas yaitu sebesar 0,76 termasuk kategori cukup. Berikutnya penilaian ahli (*exact agreements*) sebesar 94,9% sedangkan penilaian yang diharapkan dengan menggunakan model Rasch (*expected agreement*) sebesar 94,5%. Dari data tersebut diketahui hasil penilaian validator dengan hasil diperkirakan tidak jauh berbeda sehingga soal yang telah dirancang dapat dikatakan valid (Sick, 2013).

Tabel 5. Hasil Penilaian Validasi Logis Oleh Pakar

<i>Strata Value</i>	<i>Reliability</i>	<i>Exact Agreement</i>	<i>Expect Agreement</i>
2,73	0,76	94,9%	94,5%

Kelima, melakukan perbaikan terhadap instrumen tes sesuai dengan saran validator untuk memastikan bahwa produk sudah sesuai untuk dilakukan uji coba.

Melakukan uji coba

Tahapan keempat yaitu melakukan uji coba kepada 9 orang peserta didik kelas XII fase F SMA Negeri 1 Padang Sago selama 60 menit. Hal ini sesuai dengan pendapat Borg & Gall (1983) bahwa jumlah minimal sampel untuk uji coba awal berada pada kisaran 6-12. Sebelum melakukan uji coba, peserta didik terlebih dahulu diingatkan kembali mengenai topik kesetimbangan kimia agar mereka siap untuk mengerjakan soal yang akan diuji cobakan.

Menganalisis data menggunakan model Rasch

Data skor jawaban peserta didik yang telah terkumpul kemudian dianalisis menggunakan model Rasch dengan bantuan aplikasi Ministep. Kualitas *item* yang akan ditinjau yaitu validitas, reliabilitas, indeks kesukaran, dan daya beda soal. Pada pemodelan Rasch uji validitas dilakukan menggunakan menu *Output Table: Item Fit Order*. Setiap butir soal harus memenuhi minimal satu kriteria dari tiga kriteria agar dikatakan valid. Nilai MNSQ berfungsi untuk menunjukkan kesesuaian data dengan model. ZSTD bertindak sebagai uji-t yang menunjukkan hipotesis kesesuaian data (Sumintono & Widhiarso, 2015). Nilai *Pt Mean Corr* dapat mengacu pada korelasi antara skor pada butir soal dengan kemampuan peserta didik, semakin tinggi nilai korelasinya maka semakin bagus instrumen tersebut (Wei et al., 2012).

Berdasarkan Gambar 1, kriteria validitas ditunjukkan pada kolom *outfit* dan *Pt-Measure* (kotak merah). Soal nomor S2 tidak memenuhi ketiga kriteria, maka dapat disimpulkan bahwa soal ini tidak valid. Butir soal yang tidak valid dalam pemodelan Rasch dapat dihilangkan atau diubah berdasarkan persyaratan alat ukur (Susilowati et al., 2023). Selanjutnya butir soal S14 tidak memenuhi nilai MNSQ (0,39) dan *Pt Mean Corr* (0,38), namun untuk kriteria ZSTD sudah terpenuhi. Dengan demikian, butir soal S14 ini dapat dipertahankan dan tidak perlu diubah. Butir Soal S14 berisi



tentang aspek mempertanyakan mengenai perubahan kesetimbangan yang terjadi akibat perubahan suhu.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item	
2	2	9	3.07	1.17	2.18	1.53	9.90	4.50	A	.14	.69	77.8	89.1	S2
11	5	9	-.11	.91	1.33	.81	1.10	.54	B	.57	.66	77.8	80.7	S11
6	1	9	4.53	1.29	1.31	.67	.39	-.10	C	.47	.53	77.8	89.9	S6
16	2	9	3.07	1.17	1.26	.58	.49	.03	D	.67	.69	77.8	89.1	S16
3	7	9	-1.75	.95	1.03	.21	.57	-.11	E	.48	.47	77.8	81.4	S3
5	7	9	-1.75	.95	1.03	.21	.57	.11	F	.48	.47	77.8	81.4	S5
7	7	9	-1.75	.95	.99	.13	.59	.13	G	.48	.47	77.8	81.4	S7
10	7	9	-1.75	.95	.99	.13	.59	-.13	H	.48	.47	77.8	81.4	S10
4	6	9	-.91	.90	.97	.05	.64	.18	I	.59	.57	77.8	79.3	S4
8	6	9	-.91	.90	.97	.05	.64	.18	i	.59	.57	77.8	79.3	S8
14	8	9	-2.81	1.15	.92	.10	.39	-.10	h	.38	.33	88.9	88.7	S14
15	5	9	-.11	.91	.79	-.36	.48	.01	g	.73	.66	77.8	80.7	S15
18	4	9	.77	.98	.70	-.42	.39	-.27	f	.81	.72	88.9	83.8	S18
1	8	9	-2.81	1.15	.68	-.34	.27	-.28	e	.44	.33	88.9	88.7	S1
13	8	9	-2.81	1.15	.68	-.34	.27	-.28	d	.44	.33	88.9	88.7	S13
9	5	9	-.11	.91	.45	-1.44	.29	-.25	c	.81	.66	100.0	80.7	S9
12	2	9	3.07	1.17	.34	-1.11	.15	-.52	b	.83	.69	100.0	89.1	S12
17	2	9	3.07	1.17	.34	-1.11	.15	-.52	a	.83	.69	100.0	89.1	S17
MEAN	5.1	9.0	.00	1.04	.94	-.04	.99	.20				84.0	84.6	
P. SD	2.3	.0	2.31	.13	.42	.71	2.17	1.07				8.5	4.1	

Gambar 1. Validitas Uji Coba Skala Kecil

Berikutnya S1, S6, S9, S12, S13, S15, S16, S17, dan S18 juga dapat dipertahankan karena memenuhi dua kriteria yaitu ZSTD dan Pt Mean Corr. Salah satu contoh soal yang memenuhi 2 kriteria ini yaitu S12 yang memuat aspek mengomunikasikan hasil dari data tetapan kesetimbangan melalui proses Haber-Bosch. Sementara itu, butir soal lainnya yaitu S3, S4, S5, S7, S8, S10, dan S11 memenuhi tiga kriteria yang diharapkan. Jika dilihat pada butir soal S5 membahas mengenai aspek mengevaluasi dan merefleksi tentang apa saja jenis-jenis reaksi kesetimbangan berdasarkan fasenya. Berdasarkan 3 contoh soal tersebut, dapat diambil kesimpulan bahwa hasil nilai validitas ini dipengaruhi oleh jawaban peserta didik yang beragam pada setiap butir soalnya.

SUMMARY OF 18 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	5.1	9.0	.00	1.04	.94	-.04	.99	.20
SEM	.6	.0	.56	.03	.10	.17	.53	.26
P. SD	2.3	.0	2.31	.13	.42	.71	2.17	1.07
S. SD	2.4	.0	2.38	.13	.43	.73	2.23	1.11
MAX.	8.0	9.0	4.53	1.29	2.18	1.53	9.90	4.50
MIN.	1.0	9.0	-2.81	.90	.34	-1.44	.15	-.52
REAL RMSE	1.12	TRUE SD	2.02	SEPARATION	1.81	Item RELIABILITY	.77	
MODEL RMSE	1.05	TRUE SD	2.06	SEPARATION	1.97	Item RELIABILITY	.79	
S. E. OF Item MEAN	= .56							

Gambar 2. Reliabilitas Uji Coba Skala Kecil

Reliabilitas menjelaskan apakah suatu instrumen memberikan informasi yang sama atau konsisten jika dilakukan pengukuran berulang kali. Analisis reliabilitas pada *Output Summary Statistic* dalam model Rasch memberikan informasi mengenai nilai *item reliability*. Skor *item reliability* sebesar 0,77 berada pada kategori cukup, ini berarti jika dilakukan uji coba berulang kali hasil yang didapatkan terhadap



masing-masing butir soal tidak akan jauh berbeda. Jika sebuah instrumen menghasilkan nilai yang konsisten atau sedikit variasi dalam hasil uji coba berulang kali pada subjek yang sama, maka instrumen tersebut dapat dikatakan reliabel (Arikunto, 2013).

Analisis indeks kesukaran dilakukan menggunakan menu *Output Table: Item Measure*. Skor yang perlu diperhatikan terletak pada kolom *JMLE Measure* (kotak merah), yaitu *logit* yang menunjukkan urutan tingkat kesukaran soal dari yang tersulit sampai termudah sedangkan untuk kotak yang berwarna kuning menunjukkan nilai standar deviasi (SD). Dalam mengelompokkan indeks kesukaran soal patokan utama adalah nilai standar deviasi (SD). Instrumen penilaian ini memiliki SD sebesar 2,31.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXACT MATCH OBS% EXP%	Item				
6	1	9	4.53	1.29	1.31	.67	.39	-.10	.47	.53	77.8	89.9	S6
2	2	9	3.07	1.17	2.18	1.53	9.90	4.50	.14	.69	77.8	89.1	S2
12	2	9	3.07	1.17	.34	-1.11	.15	-.52	.83	.69	100.0	89.1	S12
16	2	9	3.07	1.17	1.26	.58	.49	.03	.67	.69	77.8	89.1	S16
17	2	9	3.07	1.17	.34	-1.11	.15	-.52	.83	.69	100.0	89.1	S17
18	4	9	.77	.98	.70	-.42	.39	-.27	.81	.72	88.9	83.8	S18
9	5	9	-.11	.91	.45	-1.44	.29	-.25	.81	.66	100.0	80.7	S9
11	5	9	-.11	.91	1.33	.81	1.10	.54	.57	.66	77.8	80.7	S11
15	5	9	-.11	.91	.79	-.36	.48	.01	.73	.66	77.8	80.7	S15
4	6	9	-.91	.90	.97	.05	.64	.18	.59	.57	77.8	79.3	S4
8	6	9	-.91	.90	.97	.05	.64	.18	.59	.57	77.8	79.3	S8
3	7	9	-1.75	.95	1.03	.21	.57	.11	.48	.47	77.8	81.4	S3
5	7	9	-1.75	.95	1.03	.21	.57	.11	.48	.47	77.8	81.4	S5
7	7	9	-1.75	.95	.99	.13	.59	.13	.48	.47	77.8	81.4	S7
10	7	9	-1.75	.95	.99	.13	.59	.13	.48	.47	77.8	81.4	S10
1	8	9	-2.81	1.15	.68	-.34	.27	-.28	.44	.33	88.9	88.7	S1
13	8	9	-2.81	1.15	.68	-.34	.27	-.28	.44	.33	88.9	88.7	S13
14	8	9	-2.81	1.15	.92	.10	.39	-.10	.38	.33	88.9	88.7	S14
MEAN	5.1	9.0	.00	1.04	.94	-.04	.99	.20			84.0	84.6	
P.SD	2.3	.0	2.31	.13	.42	.71	2.17	1.07			8.5	4.1	

Gambar 3. Indeks Kesukaran Uji Coba Skala Kecil

Hasil analisis menunjukkan bahwa butir soal yang memiliki *logit* > 2,31 termasuk kategori sangat sulit (S6, S2, S12, S16, dan S17). Butir soal dengan *logit* antara 0,0 s/d 2,31 merupakan soal sulit (S18) dan butir soal dengan *logit* antara 0,0 s/d -2,31 berada pada kategori mudah (S9, S11, S15, S4, S8, S3, S5, S7, dan S10). Butir soal dengan *logit* < -2,31 termasuk soal sangat mudah (S1, S13, dan S14).

Daya beda soal atau dalam pemodelan Rasch disebut dengan daya diskriminasi butir soal yang menjelaskan seberapa baik tingkatan soal untuk mampu membandingkan individu yang mempunyai kemampuan tinggi dan rendah (Sumintono & Widhiarso, 2015). Skor *separation* dapat digunakan untuk menentukan daya pembeda soal. Nilai *separation* ini dapat dilihat pada menu *summary statistic* (bagian yang ditandai kotak biru pada Gambar 2). Skor *separation* diperoleh sebesar 1,81 maka nilai ini dimasukkan ke persamaan strata H dan didapatkan nilai H sebesar 3. Berdasarkan skor ini dapat dinyatakan bahwa butir-butir soal dalam instrumen memiliki tiga macam daya beda yaitu soal sulit, sedang, dan mudah.

Meninjau statistik kecocokan item

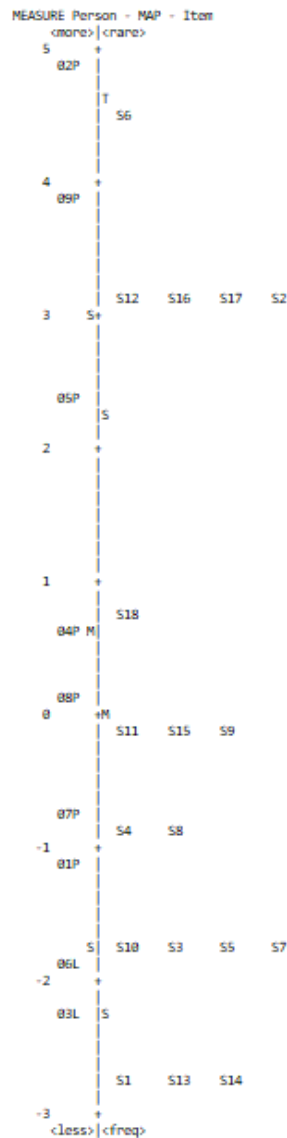
Pada tahap ini, dilakukan peninjauan kualitas butir soal berdasarkan hasil analisis model Rasch. Berdasarkan analisis validitas, terdapat satu soal yang tidak



valid karena tidak memenuhi aturan validitas yang sesuai dengan model Rasch. Soal yang tidak valid tersebut perlu diperbaiki dan tidak bisa dihilangkan karena satu soal mengukur satu aspek pada keterampilan proses sains. Dalam pemodelan Rasch jika suatu item atau butir soal dianggap tidak valid karena tidak memenuhi ketiga kriteria, maka butir soal tersebut perlu diubah atau diganti (Susilowati et al., 2023). Hasil analisis reabilitas dengan *item reliability* 0,77 artinya reliabilitas butir soal berada pada kategori cukup. Analisis indeks kesukaran memberikan informasi bahwa terdapat variasi soal yang sangat sulit, sulit, mudah, dan sangat mudah. Uji daya beda soal menghasikan tiga kelompok yang artinya instrumen penilaian ini mampu membedakan tiga kelompok butir soal (sulit, sedang, dan mudah) (Sumintono & Widhiarso, 2015).

Meninjau peta Wright

Peninjauan terhadap peta Wright dilakukan untuk melihat pendistribusian kemampuan peserta didik terhadap tingkat kesulitan soal.



Gambar 4. Peta Wright Uji Coba Skala Kecil

Berdasarkan Gambar 4, bagian sebelah kiri menunjukkan sebaran kemampuan peserta didik, sementara bagian sebelah kanan menunjukkan sebaran butir soal. Peserta didik dengan kode 02P memiliki kemampuan paling tinggi yang seharusnya mampu menjawab semua pertanyaan dengan benar, akan tetapi 02P salah pada soal S6 karena kurangnya tingkat ketelitian peserta didik 02P dalam membaca pilihan jawaban pada soal tersebut. Peserta didik dengan kode 03L memiliki kemampuan paling rendah dan hanya mampu menjawab soal yang benar sebanyak 4 dari 18 soal. Ketika diwawancara peserta didik 03L mengaku tidak serius dalam menjawab pertanyaan dan kebanyakan hanya menebak jawaban. Beberapa peserta didik lainnya juga ada yang hanya menerka jawaban sehingga sebaran butir soal yang diperoleh tidak terlalu bagus. Akibat adanya soal yang tidak valid serta penyebaran butir soal yang belum merata maka dilakukan uji coba kembali dengan memperbaiki soal terlebih dahulu supaya diperoleh hasil yang lebih baik.

Mengulagi langkah 4-7 jika terdapat item yang tidak fit

Berdasarkan data yang diperoleh dari langkah 4-7 terdapat satu butir soal yang tidak *fit* yaitu S2, sehingga soal ini perlu diperbaiki. Instrumen tes ini harus diuji cobakan kembali kepada 60 orang peserta didik kelas XII fase F agar sesuai dengan kriteria-kriteria yang diharapkan. Jumlah subjek yang digunakan disesuaikan dengan jumlah sampel pada model Rasch, untuk 60 orang memiliki tingkat kepercayaan 99% (Sumintono & Widhiarso, 2015). Analisis Rasch dilakukan kembali untuk melihat kualitas instrumen tes secara empiris.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXP.	OBS%	MATCH EXP%	Item
7	17	60	.47	.32	1.18	1.12	1.42	1.36	A .27	.43	76.7	76.6	S7
14	27	60	-.45	.29	1.18	1.50	1.30	1.50	B .32	.46	61.7	69.8	S14
13	45	60	-2.09	.33	1.29	1.62	1.09	.38	C .24	.42	66.7	78.7	S13
4	33	60	-.96	.29	1.24	2.03	1.18	.92	D .29	.46	56.7	69.8	S4
10	20	60	.17	.31	1.07	.52	1.19	.80	E .37	.44	78.3	74.0	S10
6	29	60	-.62	.29	.90	-.86	1.13	.74	F .50	.46	78.3	69.5	S6
11	16	60	.57	.33	1.07	.49	1.02	.17	G .38	.43	73.3	77.5	S11
8	20	60	.17	.31	1.05	.39	.99	.03	H .42	.44	71.7	74.0	S8
18	11	60	1.17	.37	1.03	.22	.81	-.33	I .39	.39	83.3	83.4	S18
3	8	60	1.62	.41	.98	.01	.97	.13	i .36	.36	90.0	87.5	S3
1	16	60	.57	.33	.97	-.15	.89	-.28	h .45	.43	80.0	77.5	S1
2	7	60	1.80	.44	.96	-.05	.89	.03	g .37	.34	88.3	88.8	S2
17	23	60	-.10	.30	.93	-.52	.92	-.32	f .50	.45	78.3	71.7	S17
5	28	60	-.54	.29	.89	-.96	.88	-.62	e .53	.46	73.3	69.4	S5
9	26	60	-.36	.29	.86	-1.21	.76	-1.25	d .57	.46	73.3	70.0	S9
15	38	60	-1.40	.30	.83	-1.36	.78	-.94	c .57	.45	80.0	71.9	S15
12	15	60	.68	.33	.82	-1.06	.64	-1.15	b .57	.42	78.3	78.5	S12
16	30	60	-.71	.29	.76	-2.32	.66	-1.98	a .64	.46	80.0	69.5	S16
MEAN	22.7	60.0	.00	.32	1.00	-.03	.97	-.04			76.0	75.5	
P.SD	10.0	.0	.98	.04	.15	1.12	.21	.90			8.1	6.0	

Gambar 5. Validitas Uji Coba Skala Besar

Berdasarkan hasil analisis validitas di atas, soal nomor S7, S14, S13, S4, S10, S11, S18, S3, dan S2 cenderung tidak valid karena memiliki nilai *Pt-Mean Corr* < 0,4. Namun 9 butir soal tersebut layak untuk dipertahankan karena skor MNSQ dan ZSTD yang dimiliki telah memenuhi kriteria. Butir soal lainnya telah memenuhi ketiga kriteria yang diharapkan, sehingga semua butir terbukti valid dan tidak ada



soal yang harus dihilangkan (Sumintono & Widhiarso, 2015). Artinya instrumen penilaian ini mampu mengukur apa yang seharusnya diukur yaitu KPS peserta didik.

Analisis reliabilitas dapat dilihat menggunakan menu *Output Table: Summary Statistic*. Komponen yang harus diperhatikan dalam menentukan reliabilitas instrumen, yaitu bagian yang ditandai dengan kotak merah (*Item Reliability*). Nilai reliabilitas yang diperoleh sebesar 0,88 artinya instrumen ini termasuk dalam kategori bagus dan jika dilakukan uji coba berulang kali hasil yang didapatkan terhadap masing-masing butir soal tidak akan jauh berbeda (Sumintono & Widhiarso, 2015).

SUMMARY OF 18 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	22.7	60.0	.00	.32	1.00	-.03	.97	-.04
SEM	2.4	.0	.24	.01	.04	.27	.05	.22
P.SD	10.0	.0	.98	.04	.15	1.12	.21	.90
S.SD	10.2	.0	1.01	.04	.15	1.16	.21	.93
MAX.	45.0	60.0	1.80	.44	1.29	2.03	1.42	1.50
MIN.	7.0	60.0	-2.09	.29	.76	-2.32	.64	-1.98
REAL RMSE	.34	TRUE SD	.92	SEPARATION 2.75	Item RELIABILITY .88			
MODEL RMSE	.33	TRUE SD	.93	SEPARATION 2.84	Item RELIABILITY .89			
S.E. OF Item MEAN = .24								

Gambar 6. Reliabilitas Uji Coba Skala Besar

Indeks kesukaran butir soal berpatokan pada nilai Standar Deviasi (SD). Nilai standar deviasi yang diperoleh sebesar 0,98 sehingga diperoleh penyebaran indeks kesukaran soal seperti yang terlihat pada Gambar 7.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT MATCH OBS%	EXP%	Item
2	7	60	1.80	.44	.96	-.05	.89	.03	.37	.34	88.3	88.8	S2
3	8	60	1.62	.41	.98	.01	.97	.13	.36	.36	90.0	87.5	S3
18	11	60	1.17	.37	1.03	.22	.81	-.33	.39	.39	83.3	83.4	S18
12	15	60	.68	.33	.82	-1.06	.64	-1.15	.57	.42	78.3	78.5	S12
1	16	60	.57	.33	.97	-.15	.89	-.28	.45	.43	80.0	77.5	S1
11	16	60	.57	.33	1.07	.49	1.02	.17	.38	.43	73.3	77.5	S11
7	17	60	.47	.32	1.18	1.12	1.42	1.36	.27	.43	76.7	76.6	S7
8	20	60	.17	.31	1.05	.39	.99	.03	.42	.44	71.7	74.0	S8
10	20	60	.17	.31	1.07	.52	1.19	.80	.37	.44	78.3	74.0	S10
17	23	60	-.10	.30	.93	-.52	.92	-.32	.50	.45	78.3	71.7	S17
9	26	60	-.36	.29	.86	-1.21	.76	-1.25	.57	.46	73.3	70.0	S9
14	27	60	-.45	.29	1.18	1.50	1.30	1.50	.32	.46	61.7	69.8	S14
5	28	60	-.54	.29	.89	-.96	.88	-.62	.53	.46	73.3	69.4	S5
6	29	60	-.62	.29	.90	-.86	1.13	.74	.50	.46	78.3	69.5	S6
16	30	60	-.71	.29	.76	-2.32	.66	-1.98	.64	.46	80.0	69.5	S16
4	33	60	-.96	.29	1.24	2.03	1.18	.92	.29	.46	56.7	69.8	S4
15	38	60	-1.40	.30	.83	-1.36	.78	-.94	.57	.45	80.0	71.9	S15
13	45	60	-2.09	.33	1.29	1.62	1.09	.38	.24	.42	66.7	78.7	S13
MEAN	22.7	60.0	.00	.32	1.00	-.03	.97	-.04			76.0	75.5	
P.SD	10.0	.0	.98	.04	.15	1.12	.21	.90			8.1	6.0	

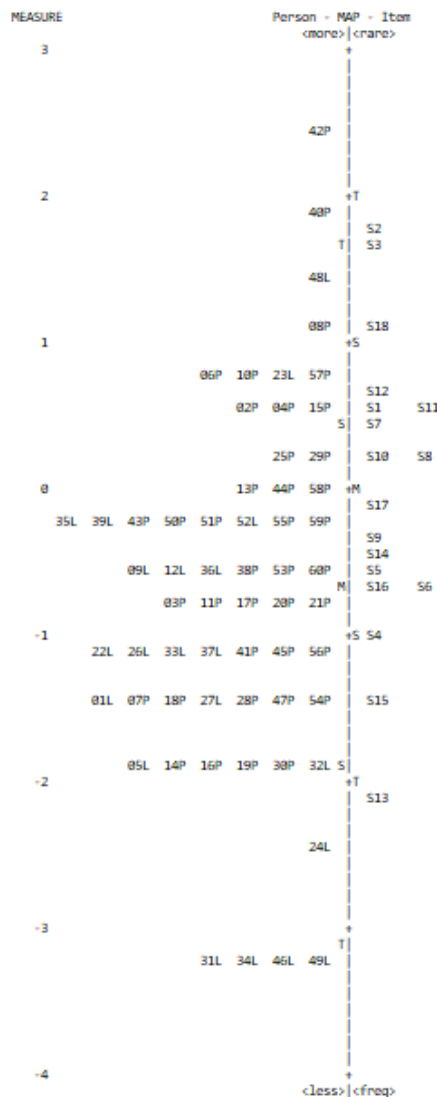
Gambar 7. Indeks Kesukaran Uji Coba Skala Kecil

Berdasarkan hasil analisis pada *item measure* (kotak merah) instrumen penilaian keterampilan proses sains ini telah memiliki variasi tingkat kesulitan soal yang merata seperti yang terlihat pada Tabel 6.

Tabel 6. Hasil Uji Coba Tingkat Kesukaran Butir Soal

Kriteria Soal	Measure	Butir Soal
Sangat sulit	Logit > 0,98	S2,S3, dan S18
Sulit	Logit 0,0 s/d 0,98	S12, S1, S11, S7, S8, dan S10
Mudah	Logit 0,0 s/d -0,98	S17, S9, S14, S5, S6, S16, dan S4
Sangat mudah	Logit < 0,98	S15 dan S13

Selanjutnya dilakukan peninjauan terhadap peta Wright untuk melihat pendistribusian kemampuan peserta didik terhadap tingkat kesulitan soal. Salah satu karakteristik dari model Rasch adalah analisis peta Wright yang dapat menghasilkan peta yang menggambarkan tingkat kesulitas soal serta kemampuan peserta didik dari tinggi ke rendah. Gambar 9 menampilkan hasil analisis peta Wright.



Gambar 9. Peta Wright Uji Coba Skala Besar

Peta Wright sebelah kiri merupakan sebaran kemampuan peserta didik. Peserta didik dengan kode 42P memiliki kemampuan paling tinggi dengan nilai *logit* > +2 yang memungkinkan dapat menjawab semua soal dengan benar, bahkan soal S2 yang terindikasi sebagai soal yang paling sulit. Sedangkan peserta didik dengan

kemampuan paling rendah yaitu 31L, 34L, 46L, dan 49L dengan nilai logit < -3 dan jumlah jawaban benar hanya 1 dari 18 soal. Peta Wright sebelah kanan merupakan sebaran tingkat kesukaran butir soal. Berdasarkan analisis peta Wright, ditemukan bahwa soal S2 memiliki tingkat kesukaran yang paling tinggi. Semua butir soal dinilai telah valid dan memiliki variasi tingkat kesukaran yang merata. Beberapa soal ada yang berjejer, tetapi soal tersebut tidak perlu diperbaiki karena mengukur aspek yang berbeda.

Menetapkan klaim kualitas produk

Berdasarkan data yang diuji cobakan kembali pada langkah kedelapan semua butir soal dapat diklaim valid karena telah memenuhi kriteria. Nilai reliabilitas butir soal termasuk kategori bagus dengan nilai sebesar 0,88, sesuai dengan acuan kriteria instrumen dalam model Rasch. Terdapat empat tingkat soal yang berbeda sesuai dengan indeks kesukaran dan daya pembeda soal. Berdasarkan analisis Rasch instrumen tes ini dapat digunakan untuk mengukur KPS peserta didik.

Mengembangkan pedoman penggunaan instrumen

Pedoman penggunaan instrumen dibuat untuk memberikan informasi kepada pengguna dalam mengaplikasikan instrumen yang telah dikembangkan (Sabekti & Khoirunnisa, 2018). Pedoman penggunaan instrumen terdiri dari cover instrumen penilaian, petunjuk penggunaan instrumen, kisi-kisi soal, petunjuk pengerjaan soal, lembar soal, dan rubrik penilaian. Petunjuk penggunaan instrumen tes bagi guru yang bertujuan untuk memudahkan guru dalam menggunakan instrumen. Tujuan penyusunan kisi-kisi soal yaitu sebagai petunjuk dalam menulis soal sehingga mencakup semua materi secara proporsional (Nofriyandi & Effendi, 2019). Tujuan dari petunjuk pengerjaan soal bagi peserta didik adalah untuk memberikan panduan kepada peserta didik saat mengerjakan soal. Terakhir rubrik penilaian dapat digunakan sebagai petunjuk pemberian skor terhadap suatu pertanyaan (Ali & Khaeruddin, 2012).

KESIMPULAN

Dari penelitian ini dihasilkan sebuah instrumen tes keterampilan proses sains pada topik kesetimbangan kimia dengan analisis model Rasch. Instrumen tes ini terdiri dari 18 butir soal pilihan ganda yang teruji valid, reliabel, memiliki indeks kesukaran, dan daya beda.

DAFTAR PUSTAKA

- Aditiyas, S. E., & Kuswanto, H. (2024). Analisis Implementasi Keterampilan Proses Sains Di Indonesia Pada Pembelajaran Fisika: Literatur Review. *Jurnal Penelitian Pembelajaran Fisika*, 15(2), 155–167. <https://doi.org/10.26877/jp2f.v15i2.15912>
- Agustin, R. R., Siswaningsih, W., & Dwiyantri, G. (2013). Pengembangan Tes Keterampilan Proses Siswa SMA Kelas XI Pokok Bahasan Titrasi Asam Basa. *Jurnal Pengajaran MIPA*, 18(2), 240–244.



- Agustina, U., Fadiawati, N., & Tania, L. (2017). Keterampilan Proses Sains (KPS) Siswa pada Materi Laju Reaksi Menggunakan LKS Berorientasi KPS. *Jurnal Pendidikan dan Pembelajaran Kimia*, 6(3), 479–492.
- Ali, S., & Khaeruddin. (2012). *Evaluasi Pembelajaran*. Makassar: Badan Penerbit UNM.
- Amali, N., & Firman, H. (2024). A Framework Design for Developing and Validating Virtual Test to Assess Science Process Skills in Chemistry. *IKnE Social Sciences*, 2024, 269–279. <https://doi.org/10.18502/kss.v9i8.15557>
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. (R. Damayanti, Ed.) (2 ed.). Jakarta: PT Bumi Aksara.
- Arini, F. Z. R., Susilaningsih, E., & Dewi, N. K. (2017). Pengembangan Instrumen Penilaian Proses untuk Mengukur Keterampilan Sains dan Aktivitas Siswa SMP. *Journal of Innovative Science Education*, 6(2), 170–178.
- 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.
- Basuki, F. R., Jufrida, Kurniawan, W., Devi, I. P., & Fitaloka, O. (2019). Tes Keterampilan Proses Sains: Multiple Choice Format. *Jurnal Pendidikan Sains*, 7(2), 101–111.
- Borg, W. R., & Gall, M. D. (1983). *Educational Research: An Intrduction* (4 ed.). New York: Longman,Inc.
- Chang, R., & Overby, J. (2011). *General Chemistry The Essential Concepts* (6 ed.). New: McGraw-Hill Companies, Inc.
- Fahmina, S. S., Masykuri, M., Ramadhani, D. G., & Yamtinah, S. (2019). Content Validity Uses Rasch Model on Computerized Testlet. *AIP Conference Proceedings*, 2194(December), 1–8.
- Gürses, A., Çetinkaya, S., Do, Ç., & Elif, Ş. (2015). Determination Of Levels Of Use Of Basic Process Skills Of High School Students, *191*, 644–650.
- 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.
- Kemendikbudristek. (2022). Keputusan Kepala Badan Standar, Kurikulum, dan Assesmen Pendidikan. *Kemendikbudristek Nomor 033/H/KR/2022*.
- Liu, X. (2010). *Using and Developing Measurement Instruments in Science Education: A Rasch Modeling Approach* (2 ed.). IAP Press: Charlotte, NC.
- Mujiburrahman, Kartiani, B. S., & Parhanuddin, L. (2023). Asesmen Pembelajaran Sekolah Dasar dalam Kurikulum Merdeka. *Jurnal Pendidikan Sekolah Dasar*, 1(1), 39–48. <https://doi.org/10.33830/penaanda.v1i1.5019>
- Nahadi, Siswaningsih, W., & Watiningsih, E. (2012). Pengembangan Penilaian Keterampilan Proses Sains Berbasis Kelas pada Pembelajaran Kimia. *Jurnal Pengajaran MIPA*, 17(1), 116–121.



- Nofriyandi, & Effendi, L. A. (2019). Workshop Penyusunan Kisi-Kisi Soal Bagi Guru-Guru SMA PGRI Pekanbaru. *Community Education Engagement Joournal*, 1(1), 73–79.
- 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.
- 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 journa*, 4(December 2022), 1–10.
- Sitorus, D. M., & Dalimunthe, M. (2024). Pengembangan Instrumen Tes Diagnostik Five-Tier Multiple Choice untuk Mengidentifikasi Miskonsepsi Siswa pada Materi Kesetimbangan Kimia. *Jurnal Pendidikan Kimia*, 9(1), 55–72.
- Sriwahyuni, E. (2022). Penggunaan Flashcard Sistem Periodik Unsur Terhadap Keterampilan Proses Sains Dasar Peserta Didik Kelas X SMA. *ORBITAL: Jurnal Pendidikan Kimia*, 6(December), 136–146.
- Sudarmin. (2015). *Model Pembelajaran Inovatif Kreatif (Model PAIKEM dalam Konteks Pembelajaran dan Penelitian Sains Bermuatan Karakter)*. Semarang: Swadaya Manunggal.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata.
- Susilowati, N. I., Liliawati, W., & Rusdiana, D. (2023). Science Process Skills Test Instruments in The New Indonesian Curriculum (Merdeka): Physics Subject in Renewable Energy Topic. *Indonesian Journal of Teaching in Science*, 3(2), 121–132.
- 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.
- Widana, I. D. K. K., Prakoso, B., Sukendro, A., Kurniadi, A., & Wilopo. (2022). *Metodologi Penelitian*. (I. D. K. K. Widana, Ed.). Yogyakarta: CV Budi Utama.

