

# Analisis dan Pengembangan Instrumen Soal Laju Reaksi Berbasis HOTS dengan Menggunakan Rasch Model

Santri Angelia Damanik<sup>1\*</sup>, Inovayani Saragih<sup>2</sup>, dan Ayi Darmana<sup>3</sup>

<sup>1,2,3</sup> Program Magister Pendidikan Kimia Universitas Negeri Medan

\* E-mail: santriangelia@gmail.com

## Abstrak

Penelitian yang dilakukan adalah penelitian dan pengembangan (research & development) untuk memperoleh instrumen soal laju reaksi yang berbasis HOTS. Pengembangan instrument soal ini dilakukan terlebih dahulu dengan cara mengikuti langkah-langkah penyusunan soal HOTS yang tentunya juga memperhatikan karakteristik dari soal HOTS tersebut. Langkah selanjutnya adalah menganalisis item soal berdasarkan pola respon jawaban yang diberikan siswa menggunakan Rasch model berbantuan aplikasi WINSTEPS 3.73. Dalam penelitian ini hasil jawaban siswa dianalisis berdasarkan delapan (8) jenis *output tables*. Berdasarkan analisis menggunakan Rasch model diperoleh hasil sebagai berikut : 1) Hasil *Output tables item fit order* diperoleh nilai Outfit : MNSQ (0,70 – 1,61); ZSTD (-1,3 – 2,0) dan Pt Measure Corr (0,29 – 0,64). 2) *Summary statistics* : *Person Reliability* (0,74 & 0,76) kategori cukup; *Item Reliability* (0,71 & 0,73) kategori cukup, dan *Cronbach Alpha Reliability* (0,81) kategori bagus sekali. Hasil akhir diambil kesimpulan, sebanyak 1 soal dari 20 soal yang dikembangkan (5 %) masuk kedalam kategori *misfit*.

**Kata kunci:** Instrumen Soal Laju Reaksi, HOTS, Rasch Model.

## Abstract

*The research carried out was research and development to obtain instruments for HOTS-based reaction rates. The development of this item instrument is carried out first by following the steps for preparing HOTS questions which of course also pay attention to the characteristics of the HOTS questions. The next step is to analyze the item items based on the response pattern of the answers given by the students using the Rasch model assisted by the WINSTEPS 3.73 application. In this study the results of student answers were analyzed based on eight (8) types of output tables. Based on the analysis using the Rasch model, the following results are obtained: 1) Results of the Output tables item fit order obtained Outfit values: MNSQ (0.70 – 1.61); ZSTD (-1.3 – 2.0) and Pt Measure Corr (0.29 – 0.64). 2) Summary statistics: Person Reliability (0.74 & 0.76) enough category; Item Reliability (0.71 & 0.73) is sufficient category, and Cronbach Alpha Reliability (0.81) is very good category. The final result is concluded, as many as 1 question out of 20 questions developed (5%) fall into the misfit category.*

**Keywords:** Reaction Rate Question Instrument, HOTS, Rasch Model.

## PENDAHULUAN

Kita sedang berada di era baru, era industrialisasi digital dimana kegiatan industri terintegrasi melalui penggunaan teknologi wireless dan big data secara massif. Saat ini berbagai macam kebutuhan manusia telah banyak menerapkan dukungan internet dan dunia digital sebagai wahana interaksi dan transaksi. *Sharing economy, e-education, egovernment, cloud collaborative, marketplace, smart city* adalah wajah dunia saat ini yang semakin kompleks, begitu cepat berubah, dan

menantang sekaligus mengancam. Dunia pendidikan perlu menyiapkan peserta didik untuk menghadapi tantangan abad 21 yang semakin kompleks. Pendidikan tidak cukup hanya membekali peserta didik dengan pengetahuan dan proses berpikir sederhana seperti yang dikenal selama ini, tetapi juga perlu menyiapkan mereka untuk memiliki dan mampu mengembangkan kecakapan esensial abad ini (Kemendikbud; 2019)

Pendidikan abad 21 memfokuskan pada keterampilan mencipta dan membaharui (*creativity and innovation*)

*skills*). Hal ini dapat diwujudkan dengan cara memberi bekal pengetahuan pada tiap jenjang pendidikan serta melatihnya untuk memiliki kemampuan memecahkan masalah, berpikir kritis dan kreatif. Sifat-sifat ini akan tumbuh bila dilatih, peserta didik dibiasakan sejak kecil untuk melakukan eksplorasi, inkuiri, penemuan dan memecahkan masalah (Alam, S.; 2019). Dengan berpikir kritis siswa akan mampu mengidentifikasi, menganalisis serta mengkaji suatu permasalahan secara tajam sehingga mampu menemukan ruang untuk mencari solusi terbaik dari masalah tersebut. Hal ini tidak menutup kemungkinan siswa akan menemukan bahkan menciptakan solusi yang baru yang belum diketahui sebelumnya. Dengan demikian ide-ide baru akan selalu muncul dan membuat siswa menjadi lebih kreatif (Susilowati, Y., & Sumaji, S.; 2021).

Berpikir kritis merupakan salah satu keterampilan dari berpikir tingkat tinggi (*high order thinking skill*). Seperti yang sudah diungkapkan sebelumnya bahwa keterampilan ini akan tumbuh dan melekat dengan baik jika dijadikan sebagai suatu kebiasaan. Guru dapat berperan untuk membentuk kebiasaan berpikir tingkat tinggi siswa melalui berbagai kegiatan pembelajaran di dalam kelas, salah satu contohnya adalah dengan memfasilitasi siswa untuk berhadapan dengan soal-soal yang memiliki kategori *high order thinking skills* (HOTS). Keterampilan ini juga bisa dikembangkan secara kumulatif sebagai kemajuan siswa melalui kursus dan mata pelajaran mereka dan pengalaman lain yang mereka dapatkan institusi mereka. Selain itu, dengan mengikutsertakan mata pelajaran mereka melalui pemecahan masalah, berpikir kritis dan aktivitas pengambilan keputusan membantu siswa meningkatkan keterampilan berpikir tingkat tinggi mereka (Abosalem, Y.; 2016).

*Higher Order Thinking Skills* (HOTS) merupakan kemampuan berpikir yang tidak sekedar mengingat (*recall*), menyatakan kembali (*restate*), atau merujuk tanpa melakukan pengolahan (*recite*). HOTS pada

konteks asesmen mengukur kemampuan: 1) Transfer satu konsep ke konsep lainnya, 2) Memproses dan menerapkan informasi 3) Mencari kaitan dari berbagai informasi yang berbeda-beda, 4) Menggunakan informasi untuk menyelesaikan masalah, dan 5) Menelaah ide dan informasi secara kritis. (Ismafitri, R., Alfani, M., & Kusumaningrum, S. R.; 2022).

Secara spesifik, Brookhart (2010) menyatakan bahwa tes berbasis HOTS adalah tes yang diarahkan agar peserta didik pada saat merespons atau menjawab butir-butir tes mampu:

- a. mentransfer satu konsep ke konsep lainnya;
- b. memproses dan menerapkan informasi;
- c. mencari kaitan dari berbagai informasi yang berbeda-beda;
- d. menggunakan informasi untuk menyelesaikan masalah;
- e. menelaah ide dan informasi secara kritis.

Karakteristik tes tersebut mengisyaratkan bahwa pengembangan tes berbasis HOTS tidak dapat dilakukan serampangan, tetapi memerlukan perencanaan yang matang dengan memperhatikan indikator-indikator penting penting dari kemampuan berpikir tingkat tinggi.

Indikator kemampuan berpikir tingkat tinggi yaitu:

1. Transfer pembelajaran

Transfer pembelajaran merupakan kemampuan peserta didik untuk memanfaatkan pengetahuan dan keterampilan yang telah dipelajari sebelumnya untuk diterapkan untuk memecahkan masalah pada konteks atau situasi pembelajaran baru.

2. Berpikir kritis

Berpikir kritis mengacu pada proses intelektual yang dimiliki oleh peserta didik untuk secara aktif dan terampil dalam mengonseptualisasi, menerapkan, menganalisis, menyintesis dan mengevaluasi berbagai informasi yang dikumpulkan melalui observasi, pengalaman, refleksi, penalaran dan komunikasi.

3. Pemecahan masalah

Keterampilan pemecahan masalah mengacu pada kemampuan peserta didik untuk memecahkan masalah secara efektif, dan tepat waktu. Keterampilan ini melibatkan kemampuan peserta didik untuk mengidentifikasi dan mendefinisikan masalah, menghasilkan solusi alternatif, mengevaluasi dan memilih solusi alternatif yang masuk akal terhadap suatu permasalahan tertentu.

#### 4. Berpikir logis

Sama seperti berpikir kritis, berpikir logis juga membutuhkan keterampilan bernalar untuk mempelajari suatu masalah secara objektif, yang memungkinkan peserta didik mampu membuat kesimpulan rasional.

Dari karakter dan indikator yang sudah dipaparkan di atas, tentunya jenis soal HOTS ini akan mampu melatih siswa untuk berpikir kritis dan kreatif.

## METODE PENELITIAN

Penelitian ini termasuk jenis *research & development*. Pengembangan instrument soal Laju Reaksi berbasis HOTS terlebih dahulu mengikuti langkah-langkah sebagai berikut : (Kemendikbud; 2019)

#### 1. Menentukan kompetensi dasar dan materi yang akan dinilai

Peneliti menganalisis proses kognitif, dimensi pengetahuan, dan materi pada kompetensi dasar dalam kurikulum yang memungkinkan dapat dibuatkan soal keterampilan berpikir tingkat tinggi.

#### 2. Menyusun kisi-kisi

Peneliti memastikan seluruh komponen yang terdapat dalam kisi-kisi konsisten, selaras, dan dapat dibuatkan soal keterampilan berpikir tingkat tinggi.

#### 3. Merumuskan indikator soal

Untuk menghasilkan soal yang mengukur keterampilan berpikir tingkat tinggi, rumusan indikator perlu memenuhi prinsip penilaian pada keterampilan ini yaitu perlunya stimulus, konteks baru, dan proses berpikir tingkat tinggi. Peneliti menggunakan stimulus yang berkenaan dengan kehidupan nyata sehari-hari dan sesuai dengan tingkat

perkembangan kognitif peserta didik. Stimulus yang kontekstual akan memudahkan peserta didik untuk mentransfer hal-hal yang telah dipelajari sehingga timbul sikap positif dan mengapresiasi hal-hal yang telah dipelajari. Stimulus dengan konteks yang tidak sesuai dengan perkembangan peserta didik akan sulit dicerna sehingga tidak mendukung berkembangnya keterampilan berpikir tingkat tinggi.

#### 4. Menulis soal sesuai dengan kaidah penulisan soal

Untuk memastikan kualitas soal sehingga memberi informasi yang valid, soal perlu memenuhi kaidah penulisan soal dari aspek konstruksi, substansi, dan bahasa.

Pengembangan soal HOTS yang sesuai dengan langkah tersebut di atas dapat kita terapkan menggunakan model Rasch. Salah satu kegunaan paling penting dari model Rasch dalam penelitian pendidikan adalah untuk membantu memandu penyusunan instrumen tes serta mengevaluasi fungsinya. Konstruksi instrumen pengukuran dengan model Rasch adalah "suatu proses sistematis di mana item sengaja dibangun sesuai dengan teori dan diuji secara empiris melalui model Rasch untuk menghasilkan satu set item yang mendefinisikan skala pengukuran linier ." Liu (2020) menyertakan langkah-langkah berikut

1. Tentukan konstruk yang dapat dicirikan oleh atribut linier.
2. Identifikasi perilaku yang sesuai dengan level yang berbeda dari konstruk yang ditentukan.
3. Tentukan ruang hasil dari perilaku (kumpulan item).
4. Uji lapangan dengan sampel yang representatif dari populasi sasaran. Uji coba terhadap 36 orang siswa kelas XI MIPA SMA Eka Prasetya Medan, pada hari Senin, 17 April 2023 menggunakan link google form.
5. Melakukan pemodelan Rasch. Berhubung soal yang dikembangkan ini berbentuk pilihan ganda dengan skor 1 untuk jawaban benar dan 0 untuk jawaban

salah maka digunakan model Rasch dikotomi. Langkah selanjutnya adalah menganalisis item soal berdasarkan pola respon jawaban yang diberikan siswa menggunakan Rasch model berbantuan aplikasi WINSTEPS 3.73. Berdasarkan kriteria Boone et al. (2014), kriteria yang digunakan untuk memeriksa kesesuaian butir soal dan individu yang tidak sesuai (outliers atau misfits) adalah: nilai Outfit mean square (MNSQ) yang diterima:  $0,5 < \text{MNSQ} < 1,5$ ; Nilai ZSTD ( $-2,0 < \text{ZSTD} < +2,0$ ), dan Nilai Point Measure Correlation (Pt Mean Corr) adalah  $0,4 < \text{Pt. Measure Corr} < 0,85$ . Selain itu deteksi adanya soal bias melalui variabel jenis kelamin (laki-laki dan perempuan) sampel dalam penelitian. Suatu butir soal dikatakan bias jika probabilitas butirnya berada di bawah 5% (0,05). Namun dalam penelitian ini hasil jawaban siswa dianalisis berdasarkan delapan (8) jenis *output tables*.

6. Tinjau statistik kecocokan item dan revisi item jika itu diperlukan
7. Meninjau hasil Wright map. Peta Wright memberikan gambaran soal ujian pilihan ganda dengan menempatkan tingkat kesulitan soal-soal ujian pada skala pengukuran yang sama dengan kemampuan para kandidat. Ini memberi pengguna perbandingan kandidat dan item, untuk lebih memahami seberapa tepat tes diukur, sehingga item dapat ditambahkan atau dihapus. (Andrich, D.; 2010).
8. Ulangi (4) sampai (7) sampai satu set item sesuai dengan Rasch
9. Tetapkan klaim validitas dan reliabilitas untuk instrumen pengukuran.
10. Kembangkan dokumentasi untuk pengukuran dalam instrumen.

Setelah semua tahapan tersebut selesai, maka instrument soal pada materi Laju Reaksi berbasis HOTS siap digunakan.

## HASIL DAN PEMBAHASAN

Kompetensi dasar dan materi yang dinilai dalam pengembangan instrument ini adalah materi Laju Reaksi dengan KD :

3.6 Menjelaskan factor-faktor yang memengaruhi laju reaksi menggunakan teori tumbukan; dan

3.7. Menentukan orde reaksi dan tetapan laju reaksi berdasarkan data hasil percobaan.

Penyusunan kisi-kisi soal dilakukan secara konsisten, selaras, dan dapat dibuatkan soal keterampilan berpikir tingkat tinggi. Terdapat 11 indikator soal dari kedua kompetensi dasar, disertai stimulus yang berkenaan dengan kehidupan nyata sehari-hari dan sesuai dengan tingkat perkembangan kognitif peserta didik. Selanjutnya dikembangkan soal HOTS sebanyak 20 butir soal pilihan ganda. Untuk memastikan kualitas soal sehingga memberi informasi yang valid, soal perlu memenuhi kaidah penulisan soal dari aspek materi, konstruksi, substansi, dan bahasa.

Validasi dilakukan oleh 4 orang validator ahli (2 orang dosen kimia dan 2 orang guru kimia). Hasil validasi dari ahli selanjutnya dievaluasi menggunakan formula Aiken. Aiken (1985) merumuskan formula Aiken's V untuk menghitung content-validity coefficient yang didasarkan pada hasil penilaian dari panel ahli sebanyak n orang terhadap suatu aitem dari segi sejauh mana aitem tersebut mewakili konstruk yang diukur. Formula yang diajukan oleh Aiken adalah sebagai berikut:

$$V = \sum s / [n(c-1)] \text{ (dalam Hendryadi; 2014)}$$

$$S = r - l_o$$

$l_o$  = angka penilaian validitas yang terendah

$C$  = angka penilaian validitas tertinggi

$R$  = angka yang diberikan oleh penilai

Item dinilai oleh 4 rater dengan 4 pilihan skala 4, maka jika kita merujuk pada tabel nilai V minimal yang diterima dengan taraf kesalahan 5% adalah 0,92.

**HASIL**

Hasil analisis validasi disajikan dalam tabel 1.  
Tabel 1. Analisis Validasi (Aiken's V)

No Item	V	Keterangan
1	0,94	Valid
2	0,96	Valid
3	0,96	Valid
4	0,96	Valid
5	0,96	Valid
6	0,96	Valid
7	0,94	Valid
8	0,96	Valid
9	0,97	Valid
10	0,95	Valid
11	0,95	Valid
12	0,94	Valid
13	0,94	Valid
14	0,94	Valid
15	0,97	Valid
16	0,97	Valid
17	0,92	Valid
18	0,93	Valid
19	0,95	Valid
20	0,97	Valid

Dikarenakan seluruh soal sudah dinyatakan valid, maka dilakukan uji lapangan dengan sampel yang representatif dari populasi sasaran. Uji coba terhadap 36 orang siswa kelas XI MIPA SMA Eka Prasetya Medan, pada hari Senin, 17 April 2023 menggunakan link google form. Hasil jawaban siswa selanjutnya dianalisis menggunakan pemodelan Rasch dengan hasil sebagai berikut:

**1. Summary statistics**

SUMMARY OF 36 MEASURED (EXTREME AND NON-EXTREME) Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MISQ	ZSTD	MISQ	ZSTD	
MEAN	12,9	20,0	.87	.61					
S.D.	4,3	.0	1,33	.24					
MAX.	20,0	20,0	4,54	1,84					
MIN.	5,0	20,0	-1,25	-.48	.65	-2,2	.30	-1,9	
REAL RMSE	.68	TRUE SD	1,14	SEPARATION	1,69	Person RELIABILITY	.74		
MODEL RMSE	.66	TRUE SD	1,15	SEPARATION	1,76	Person RELIABILITY	.76		
S.E. OF Person MEAN	= .22								
Person RAW SCORE-TO-MEASURE CORRELATION = .97									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .81									
SUMMARY OF 20 MEASURED (NON-EXTREME) Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MISQ	ZSTD	MISQ	ZSTD	
MEAN	23,2	36,0	.00	.41	1,01	.1	.94	-.1	
S.D.	5,0	.0	.80	.02	.14	.8	.26	.8	
MAX.	28,0	36,0	1,95	.44	1,32	1,6	1,61	2,0	
MIN.	11,0	36,0	-.80	.39	.77	-1,4	.67	-1,3	
REAL RMSE	.43	TRUE SD	.67	SEPARATION	1,58	Item RELIABILITY	.71		
MODEL RMSE	.42	TRUE SD	.68	SEPARATION	1,64	Item RELIABILITY	.73		
S.E. OF Item MEAN	= .18								

Gambar 1. Summary statistics

Dari hasil summary statistics tersebut kita dapat melihat tiga (3) jenis reliabilitas, *person reliability*, *item reliability* dan *Cronbach Alpha* dengan ketentuan sebagai berikut :

Tabel 2. Kategori *Person reliability* dan *Item reliability*

No	Nilai <i>Person reliability</i> dan <i>Item reliability</i>	Kategori
1	< 0,67	Lemah
2	0,67 – 0,80	Cukup
3	0,80 – 0,90	Bagus
4	0,91 – 0,94	Bagus sekali
5	>0,94	Istimewa

Sedangkan untuk ketentuan *Cronbach Alpha* dapat kita lihat pada tabel 3.

Tabel 3. Kategori Nilai *Cronbach Alpha* (reliabilitas) butir soal

No	Nilai Cronbach Alpha (reliabilitas)	Kategori
1	< 0,50	Buruk
2	0,50 – 0,60	Jelek
3	0,60 – 0,70	Cukup
4	0,70 – 0,80	Bagus
5	>0,80	Bagus sekali

Berdasarkan kriteria tersebut maka dapat kita nyatakan bahwa :

- Person Reliability* 0,74 dan 0,76 kategori **cukup**
- Item Reliability* 0,71 dan 0,73, kategori **cukup**
- Cronbach Alpha Reliability* 0,81, kategori **bagus sekali**

**2. Item Fit Order**

Dari hasil yang ditunjukkan oleh Item Fit Order dapat kita lihat Tingkat kesesuaian butir soal (validitas) yang digunakan untuk menjelaskan apakah butir soal berfungsi normal melakukan pengukuran atau tidak.

Menurut Boone et al (2014), kriteria yang digunakan untuk memeriksa kesesuaian butir soal yang tidak sesuai (outlier atau misfit) adalah :

- Nilai Outfit mean square (MNSQ) yang diterima :  $0,5 < MNSQ < 1,5$
- Nilai Outfit Z-standar (ZSTD) yang diterima

: -2,0 < ZSTD < + 2,0

- c. Nilai Point Measure Correlation (Pt Measure Corr) yang diterima :0,4 < Pt Measure Corr < 0,85

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT [MISQ]	OUTFIT [ZSTD]	PT-MEASURE [MISQ]	EXACT MATCH [ZSTD]	CORR.	EXP.	OBS%	EXP%	Item	
17	14	36	1.44	.40	1.32	1.6	1.61	2.0	A	.29	.52	71.4	73.7	IT17
6	21	36	.37	.39	1.19	1.1	1.47	1.7	B	.35	.49	71.4	71.7	IT6
7	23	36	.06	.40	1.25	1.4	1.28	1.0	C	.31	.47	60.0	72.4	IT7
18	11	36	1.95	.42	1.22	1.1	1.21	.7	D	.38	.51	62.9	75.8	IT18
16	27	36	-.61	.43	1.12	.6	.93	.0	E	.36	.41	65.7	76.5	IT16
8	23	36	.06	.40	1.06	.4	1.10	.4	F	.43	.47	65.7	72.4	IT8
20	27	36	-.61	.43	1.05	.3	.91	-.1	G	.40	.41	77.1	76.5	IT20
14	25	36	-.26	.41	1.02	.2	.95	.0	H	.44	.44	74.3	74.1	IT14
11	26	36	-.43	.42	1.00	.1	.96	.0	I	.43	.43	77.1	75.1	IT11
5	22	36	.22	.39	.98	-.1	.88	-.4	J	.51	.48	74.3	71.7	IT5
13	28	36	-.80	.44	.98	.0	.73	-.5	J	.44	.39	77.1	78.4	IT13
10	28	36	-.80	.44	.96	-.1	.78	-.3	I	.43	.39	82.9	78.4	IT10
1	28	36	-.80	.44	.95	-.2	.67	-.6	H	.46	.39	77.1	78.4	IT1
15	28	36	-.80	.44	.94	-.2	.77	-.4	G	.44	.39	77.1	78.4	IT15
3	28	36	-.80	.44	.92	-.3	.67	-.6	F	.47	.39	77.1	78.4	IT3
2	18	36	.83	.39	.89	-.6	.82	-.8	E	.59	.51	74.3	72.6	IT2
19	16	36	1.13	.39	.88	-.7	.78	-.9	D	.61	.52	77.1	73.1	IT19
4	27	36	-.61	.43	.87	-.6	.71	-.6	C	.50	.41	82.9	76.5	IT4
12	23	36	.06	.40	.83	-.1	.77	-.8	B	.58	.47	82.9	72.4	IT12
9	21	36	.37	.39	.77	-.1	.4	.70	A	.64	.49	82.9	71.7	IT9
MEAN	23.2	36.0	.00	.41	1.01	.1	.94	-.1				74.6	74.9	
S.D.	5.0	.0	.80	.02	.14	.8	.26	.8				6.5	2.5	

Gambar 2. Item Fit Order

Sesuai dengan kriteria yang dikemukakan oleh Boone et al, maka dapat kita lihat pada tabel berikut, beberapa item soal yang kurang memenuhi kriteria:

Tabel 4. Interpretasi Item Fit Order

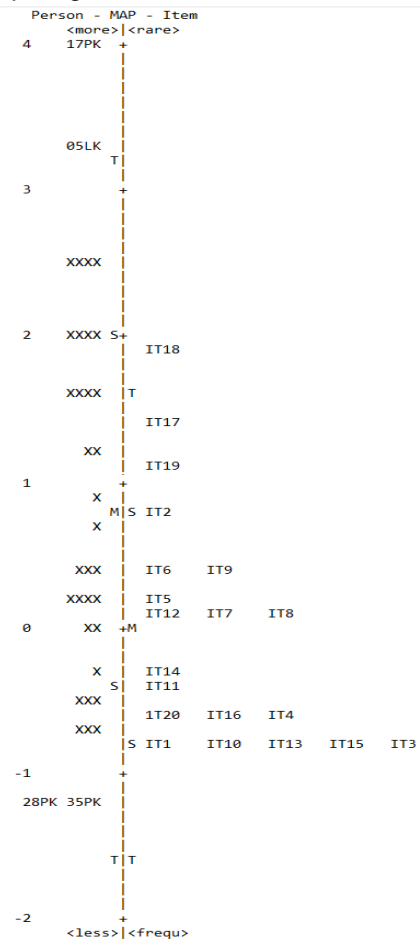
No	MNSQ	ZSTD	Pt Measure Corr
17	Tidak memenuhi	Memenuhi	Tidak memenuhi
6	Memenuhi	Memenuhi	Tidak memenuhi
7	Memenuhi	Memenuhi	Tidak memenuhi
18	Memenuhi	Memenuhi	Tidak memenuhi
16	Memenuhi	Memenuhi	Tidak memenuhi

### 3. Wright Map

Hasil yang ditunjukkan oleh *wright map* adalah sebagai berikut :

- Bagian kanan wright map di atas merupakan sebaran tingkat kesulitan butir soal, sedangkan bagian kiri merupakan sebaran abilitas/kemampuan siswa.
- Soal nomor 18 adalah soal yang memiliki tingkat kesulitan tinggi, sedangkan item soal nomor 1, 10, 13, 15 dan 3 memiliki tingkat kesulitan yang sama dan meupak soal yang mudah.
- Siswa dengan abilitas paling tinggi adalah

17PK, sementara siswa dengan abilitas paling rendah adalah 28PK dan 35PK.



Gambar 3. Wright map

### 4. Item Measure (tingkat kesulitan butir soal)

Untuk mengetahui tingkat kesulitan butir soal (item measure) dilihat dari nilai logit tiap butir soal yang terdapat pada kolom measure. Nilai logit yang tinggi menunjukkan tingkat kesulitan soal yang paling tinggi.

Tabel 5. Kategori kelompok soal berdasarkan tingkat kesulitannya

No	Nilai Logit	Kategori
1	>+1,37 SD	Sangat sulit
2	0,0 logit +1,37 SD	Sulit
3	0,0 logit -1,37 SD	Sedang
4	< -1,37 SD	Mudah
5	Nilai Logit	Kategori

Berdasarkan tabel 5, maka dapat dikategorikan item soal sebagai berikut:

Tabel 6. Tingkat Kesulitan Soal

No	Score	Count	Measure	Category
18	11	36	1.95	Sangat sulit
17	14	36	1.44	Sangat sulit
19	16	36	1.13	Sulit
2	18	36	0.83	Sulit
6	21	36	0.37	Sulit
9	21	36	0.37	Sulit
5	22	36	0.22	Sulit
7	23	36	0.06	Sulit
8	23	36	0.06	Sulit
12	23	36	0.06	Sulit
14	25	36	-0.26	Sedang
11	26	36	-0.43	Sedang
4	27	36	-0.61	Sedang
16	27	36	-0.61	Sedang
20	27	36	-0.61	Sedang
1	28	36	-0.80	Sedang
3	28	36	-0.80	Sedang
10	28	36	-0.80	Sedang
13	28	36	-0.80	Sedang
15	28	36	-0.80	Sedang

26	17	20	1,95	Tinggi
32	17	20	1,95	Tinggi
1	16	20	1,56	Sedang
8	16	20	1,56	Sedang
13	16	20	1,56	Sedang
27	16	20	1,56	Sedang
15	15	20	1,23	Sedang
33	15	20	1,23	Sedang
19	14	20	0,94	Sedang
2	13	20	0,68	Sedang
7	12	20	0,43	Sedang
18	12	20	0,43	Sedang
25	12	20	0,43	Sedang
9	11	20	0,20	Sedang
16	11	20	0,20	Sedang
20	11	20	0,20	Sedang
29	11	20	0,20	Sedang
22	10	20	-0,03	Sedang
30	10	20	-0,03	Sedang
24	9	20	-0,26	Sedang
3	8	20	-0,49	Sedang
21	8	20	-0,49	Sedang
31	8	20	-0,49	Sedang
12	7	20	-0,72	Sedang
34	7	20	-0,72	Sedang
36	7	20	-0,72	Sedang
28	5	20	-1,25	Sedang
35	5	20	-1,25	Sedang

### 5. Person measure (abilitas siswa)

*Person measure* ini digunakan untuk mengidentifikasi tingkat kemampuan siswa dalam menjawab soal. Tingkat abilitas siswa diurutkan dari yang tertinggi hingga terendah berdasarkan nilai logit tiap person. Nilai standar deviasi (SD) pada *person measure* ini adalah 1,80. Kriteria penentuannya dapat kita lihat pada tabel 7.

Tabel 7. Nilai Logit Abilitas Siswa

No	Nilai logit abilitas siswa	Kategori
1	>1,80	Tinggi
2	-1,29 – 1,80	Sedang
3	< -1,29	Rendah

Berdasarkan kriteria kategori kemampuan siswa tersebut, maka diperoleh hasil pengelompokan siswa sebagai berikut :

Tabel 8. Kategori Abilitas Siswa

No	Total Score	Total Count	Measure	Category
17	20	20	4,54	Tinggi
5	19	20	3,26	Tinggi
4	18	20	2,46	Tinggi
10	18	20	2,46	Tinggi
14	18	20	2,46	Tinggi
23	18	20	2,46	Tinggi
6	17	20	1,95	Tinggi
11	17	20	1,95	Tinggi

### 6. Person Fit Order

*Person fit order* ini digunakan untuk mendeteksi adanya individu yang pola responnya tidak sesuai (berbeda), artinya ada ketidaksesuaian jawaban yang diberikan berdasarkan abilitasnya dibandingkan model ideal.

Menurut boone et al (2014), kriteria yang digunakan untuk melihat tingkat kesesuaian individu (*person fit*) adalah :

- Nilai Outfit mean square (MNSQ) yang diterima :  $0,5 < MNSQ < 1,5$
- Nilai Outfit Z-standar (ZSTD) yang diterima :  $-2,0 < ZSTD < + 2,0$
- Nilai Point Measure Correlation (Pt Measure Corr) yang diterima :  $0,4 < Pt Measure Corr < 0,85$

Hasil *Person fit order* ditunjukkan oleh gambar 4.

Person STATISTICS: MISFIT ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXACT MATCH	Person				
18	12	20	.43	.49	1.53	2.5	1.60	2.5	A-.26	.36	50.0	68.2	18PK
31	8	20	-.49	.48	1.17	1.1	1.46	1.8	B-.07	.52	50.0	64.7	31PD
35	5	20	-1.25	.54	1.10	.5	1.46	1.1	C-.07	.66	75.0	74.9	35PK
27	16	20	1.56	.60	1.31	.9	1.44	1.0	D-.03	.34	75.0	80.9	27PK
15	15	20	1.23	.55	1.24	.9	1.43	1.2	E-.04	.36	75.0	77.2	15LK
1	16	20	1.56	.60	1.32	.9	1.21	.6	F-.04	.34	75.0	80.9	01LD
7	12	20	.43	.49	1.25	1.3	1.28	1.3	G-.07	.36	60.0	68.2	07LD
29	11	20	.20	.48	1.16	.9	1.25	1.3	H-.15	.36	60.0	65.9	29LK
3	8	20	-.49	.48	1.22	1.4	1.13	.6	I-.10	.32	40.0	64.7	03LK
25	12	20	.43	.49	1.18	1.0	1.19	.9	J-.16	.36	60.0	68.2	25PK
36	7	20	-.72	.49	1.03	.3	1.18	.7	K-.23	.31	70.0	65.8	36PD
20	11	20	.20	.48	1.10	.6	1.10	.5	L-.25	.36	60.0	65.9	20LK
5	19	20	3.26	1.05	1.09	.4	.95	.4	M-.11	.21	95.0	95.0	05LK
33	15	20	1.23	.55	1.08	.4	1.06	.3	N-.28	.36	75.0	77.2	33PK
21	8	20	-.49	.48	1.05	.4	1.01	.1	O-.27	.32	60.0	64.7	21PK
16	11	20	.20	.48	1.04	.3	1.03	.2	P-.32	.36	60.0	65.9	16LK
13	16	20	1.56	.60	1.01	.1	.89	-.1	Q-.36	.34	85.0	80.9	13PD
4	18	20	2.46	.78	1.00	.2	.79	.0	R-.32	.28	90.0	89.9	04LD
34	7	20	-.72	.49	.98	.0	.89	-.3	q-.35	.31	70.0	65.8	34PK
12	7	20	-.72	.49	.94	-.3	.86	-.4	p-.40	.31	70.0	65.8	12LK
32	17	20	1.95	.66	.92	.0	.68	-.4	o-.46	.32	85.0	84.9	32PK
2	13	20	.68	.50	.91	-.3	.89	-.4	n-.47	.37	75.0	71.1	02LK
30	10	20	-.03	.48	.85	-.1	.87	-.6	m-.52	.35	80.0	64.9	30PK
26	17	20	1.95	.66	.86	-.2	.67	-.4	i-.50	.32	85.0	84.9	26PK
8	16	20	1.56	.60	.84	-.4	.67	-.6	k-.55	.34	85.0	80.9	08LK
28	5	20	-1.25	.54	.83	-.6	.67	-.8	j-.50	.26	75.0	74.9	28PK
6	17	20	1.95	.66	.83	-.3	.58	-.6	i-.55	.32	85.0	84.9	06LD
24	9	20	-.26	.48	.81	-.1	.47	-.6	h-.57	.34	75.0	64.6	24PK
9	11	20	.20	.48	.80	-.1	.31	-.7	g-.60	.36	80.0	65.9	09LK
19	14	20	.94	.52	.78	-.9	.75	-.8	f-.62	.36	85.0	74.0	19PK
11	17	20	1.95	.66	.75	-.5	.58	-.6	e-.61	.32	85.0	84.9	11PK
10	18	20	2.46	.78	.72	-.4	.35	-.8	d-.64	.28	90.0	89.9	10LK
22	10	20	-.03	.48	.70	-.2	.67	-.9	c-.71	.35	90.0	64.9	22PD
14	18	20	2.46	.78	.65	-.5	.30	-.1	b-.71	.28	90.0	89.9	14LK
23	18	20	2.46	.78	.65	-.5	.30	-.1	a-.71	.28	90.0	89.9	23PK
MEAN	12.9	20.0	.87	.61	.99	.1	.94	.0			74.6	74.9	
S.D.	4.3	.0	1.33	.24	.21	.9	.33	.9			13.3	9.6	

Gambar 4. Person Fit Order

Kriteria pemeriksaan person fit hampir sama dengan kriteria pemeriksaan item fit order. Pada pemeriksaan ini, pola respon siswa dinyatakan fit apabila sudah memenuhi paling tidak dua dari tiga kriteria, yakni MNSQ, ZSTD dan Pt. Measure Corr. Jika ditemukan pola respon siswa yang tidak fit, maka selanjutnya dapat diperiksa pada scalogram.

### 7. Scalogram

Scalogram yang dikembangkan oleh Louis Guttman. Guttman memperkenalkan pemeringkatan skala sikap dari yang terendah ke yang tertinggi, yang dikembangkan menjadi suatu matriks tertentu. Hasil scalogram penelitian ini ditunjukkan oleh gambar 5.

Person	Item
	111 1211 1 111
	13035460147825692978
17	+11111111111111111111 17PK
5	+11111111111110111111 05LK
4	+11111111111011111101 04LD
10	+11111111111111111010 10LK
14	+11111111111111111100 14LK
23	+11111111111111111100 23PK
6	+111111111111111100110 06LD
11	+11111111110111111100 11PK
26	+11111111111110111010 26PK
32	+11111111101111111001 32PK
1	+11111111011010011111 01LD
8	+11111111110111011010 08LK
13	+11111111011011001110 13PD
27	+00111100011111110111 27PK
15	+1110001111111110101 15LK
33	+11011111101110110101 33PK
19	+11111011111110110100 19PK
2	+1101111101101100100 02LK
7	+00110110110011111100 07LD
18	+1010100000111111101 18PK
25	+0011101111100111010 25PK
9	+11101101110110100000 09LK
16	+1100110110011010100 16LK
20	+1101101101001010100 20LK
29	+11001101101100110001 29LK
22	+01111111111100000000 22PD
30	+11111011101000000010 30PK
24	+01111101011100000000 24PK
3	+00100100100111110000 03LK
21	+01010101011100001000 21PK
31	+00011001110100100000 31PD
12	+10101101000001100000 12LK
34	+11100010000111000000 34PK
36	+10110010011000000010 36PD
28	+11001010100000000000 28PK
35	+00010011001000000010 35PK
	111 1211 1 111
	13035460147825692978

Gambar 5. Scalogram

Dari gambar 5 dapat kita setiap butir memiliki urutan yang secara sistematis dapat dijadikan peringkat dari yang rendah ke peringkat yang tinggi dan juga diurutkan dari soal yang paling mudah, menuju paling susah. Tujuannya adalah untuk menganalisis, memberikan penjelasan serta memprediksi secara sekaligus kemampuan individu sekaligus tingkat kesulitan butir soal (Sumintono, B., & Widhiarso, W.; 2015).

### 8. Item DIF

Salah satu ciri pengukuran instrument yang valid adalah butir soal tidak mengandung bias. Butir soal disebut bias jika didapati bahwa salah satu individu dengan karakteristik tertentu lebih diuntungkan dibandingkan individu dengan karakteristik yang lain. Untuk memeriksa apakah butir soal dalam penelitian dapat kita lihat dari gambar 6.



Person CLASSES	SUMMARY DIF		BETWEEN-CLASS		Item Number Name	
	CHI-SQUARE	D.F.	MEAN-SQUARE	t=ZSTD		
2	1.5799	1	.2088	.6498	.1875	1 IT1
2	.0006	1	.9806	.0001	-1.5378	2 IT2
2	1.5799	1	.2088	.6498	.1875	3 IT3
2	.0027	1	.9586	.0017	-1.3973	4 IT4
2	2.6760	1	.1019	1.1686	.5845	5 IT5
2	.4086	1	.5227	.1559	-.5083	6 IT6
2	.5672	1	.4514	.2184	-.3723	7 IT7
2	3.2840	1	.0700	1.4848	.7702	8 IT8
2	.4086	1	.5227	.1559	-.5083	9 IT9
2	.5154	1	.4728	.1969	-.4158	10 IT10
2	.0241	1	.8765	.0085	-1.2169	11 IT11
2	.0237	1	.8776	.0090	-1.2095	12 IT12
1	.0000	0	1.0000	.0000	.0000	13 IT13
2	1.4994	1	.2208	.6372	.1756	14 IT14
2	.0861	1	.7692	.0329	-.9704	15 IT15
2	.8060	1	.3693	.3188	-.2008	16 IT16
2	.6664	1	.4143	.2713	-.2766	17 IT17
2	.3558	1	.5508	.1464	-.5320	18 IT18
2	1.5373	1	.2150	.6458	.1837	19 IT19
2	.0027	1	.9586	.0017	-1.3973	20 IT20

Gambar 6. Item Person DIF

Butir soal terdeteksi DIF apabila memiliki nilai probabilitas kurang dari 5%. Hasil analisis dari gambar 6 menunjukkan bahwa tidak ada nilai probabilitas yang kurang dari 0,05 (5%), yang berarti dapat disimpulkan bahwa kelas DIF tidak terdapat bias.

## PEMBAHASAN

Berdasarkan hasil yang diperoleh, seluruh item soal dinyatakan reliabel, soal tersebut memiliki konsistensi yang sesuai. Berdasarkan interpretasi hasil Item Fit Order ditemukan beberapa item soal yang kurang sesuai dengan kriteria, meskipun demikian jika ditemukan salah satu butir soal dimana nilai MNSQ dan Pt Measure Corr tidak memenuhi kriteria akan tetapi nilai ZSTD memenuhi kriteria maka butir soal tersebut masih dianggap fit, artinya butir soal tersebut masih bisa dipertahankan. Namun berhubung soal yang dikembangkan berbasis HOTS maka alangkah lebih baik jika item soal nomor 17 diganti. Hal ini juga hampir sesuai dengan nilai analisis pada formula Aiken's, dimana soal nomor 17 memiliki nilai  $V = 0,92$ , yang merupakan nilai batas bawah yang dapat diterima pada formula tersebut.

Berdasarkan hasil yang kita peroleh dari wright map, seluruh item soal berada dalam standar deviasi (T), dimana tidak ada soal yang terlalu mudah, maupun soal yang terlalu sulit. Siswa 17PK, 05PK, 04LD, 10LK, 14LK dan 23PK, memiliki abilitas diatas tingkat kesulitan soal nomor 18, yang artinya siswa tersebut mampu menjawab soal nomor 18 dengan

benar, dan juga mereka masih memiliki kemampuan untuk menjawab soal yang lebih sulit dari soal nomor 18.

Item measure yang ditunjukkan pada tabel 6, berkaitan dengan wright map yang sudah dibahas sebelumnya. Tingkat kesulitan butir soal (item measure) diurutkan dari tingkat kesukaran tinggi hingga rendah, dari yang sangat sulit hingga mudah. Butir soal nomor 18 adalah butir soal yang sangat sulit dikerjakan oleh siswa, hal ini dapat juga kita lihat dari jumlah siswa yang dapat menjawab benar soal tersebut sebanyak 11 orang (30,56 %). Nilai logit yang sama pada setiap butir soal menunjukkan tingkat kesulitan masing-masing soal yang tidak jauh berbeda.

Nilai logit menunjukkan kemampuan siswa dalam menjawab soal. Siswa yang memiliki kemampuan menjawab soal tertinggi adalah siswa nomor 17 dengan nilai logit sebesar 4,54. Siswa yang memiliki kemampuan menjawab soal terendah adalah siswa nomor 35 dengan nilai logit sebesar -1,25.

Hasil Person Fit Order (gambar 4) menunjukkan bahwa siswa 10LK, 14LK, 23PK, dan 18PK tidak memenuhi kriteria MNSQ. Siswa 18PK tidak memenuhi kriteria ZSTD. Siswa 18PK, 31PD, 35PK, 27PK, 15LK, 01LD, 07LD, 29LK, 03LK, 25PK, 36PD, 20LK, 05LK, 33PK, 21PK, 16LK, 13PD, 04LD, 34PK tidak memenuhi kriteria Pt. Measure Corr. Dari hasil tersebut maka dinyatakan 1 orang siswa (5%) yakni siswa 18PK memiliki pola respon yang tidak fit.

Pola respon yang tidak fit dari siswa 18PK dapat kita analisis pada hasil scalogram dan disesuaikan kembali dengan posisi siswa tersebut dalam wright map. Pola respon dinyatakan tidak fit karena siswa 18PK mampu menjawab soal sulit yang berada diatas kemampuannya dengan benar, namun siswa tersebut tidak bisa menjawab dengan benar soal mudah yang tingkat kesulitannya berada dibawah abilitas siswa tersebut.

Selain itu dari hasil scalogram juga kita dapat mengidentifikasi jika ada terdapat kecurangan, misalnya siswa menyontek. Hal pertama yang dapat kita lihat adalah jika terdapat nilai person logit yang sama,

selanjutnya dapat kita tinjau dari pola hasil jawaban siswa tersebut.

## PENUTUP

Hasil akhir dari penelitian ini diambil kesimpulan sebagai berikut :

1. Pengembangan instrument soal ini dilakukan dengan mengikuti empat (4) langkah pengembangan soal HOTS yang dikombinasikan dengan 10 langkah pengembangan soal HOTS menggunakan model Rasch.
2. Berdasarkan *Summary statistics* : *Person Reliability* (0,74 & 0,76) kategori cukup; *Item Reliability* (0,71 & 0,73) kategori cukup, dan *Cronbach Alpha Reliability* (0,81) kategori bagus sekali.
3. Hasil *Output tables item fit order* diperoleh nilai *Outfit* : *MNSQ* (0,70 – 1,61); *ZSTD* (-1,3 – 2,0) dan *Pt Measure Corr* (0,29 – 0,64). Berdasarkan kriteria *item fit order* tersebut diambil kesimpulan, sebanyak 1 soal (5%) masuk kedalam kategori misfit.
4. Hasil analisis *Item Person DIF* menunjukkan bahwa tidak ada nilai probabilitas yang kurang dari 0,05 (5%), yang berarti dapat disimpulkan bahwa kelas *DIF* tidak terdapat bias.
5. Hasil pengembangan menghasilkan produk 19 instrumen soal pilihan ganda berbasis HOTS pada materi Laju Reaksi sudah dapat digunakan sebagai instrument tes.

## REFERENCES

- Andrich, D. (2010). *Rasch models for measurement*. Sage.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.
- Boone, W. J., Staver, J. R., Yale, M. S., Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Person reliability, item reliability, and more. *Rasch analysis in the human sciences*, 217-234.
- Boone, W. J., Staver, J. R., Yale, M. S., Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Understanding person measures. *Rasch analysis in the human sciences*, 69-92.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Ascd.
- Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N. (2021). Analysis of RASCH model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia*, 9(3), 329-345.
- Fahreza, R., Yudha, S., Purwanto, R., Suyanti, R. D., & Darmana, A. (2022). Used of RASCH Model for Analysis of Students' Critical Thinking Skills Test Instruments on Thermochemical Topics. *Jurnal Pendidikan dan Pembelajaran Kimia*, 11(2).
- Harmurni, L. (2019). *Instrumen Penilaian & Validasinya*. Uwaiz Inspirasi Indonesia.
- Hartik, S., Utaminingsih, S., & Madjidi, A. H. (2021, March). A Need Assessment of Integrated Science Teaching Material Based Higher Order Thinking Skills (HOTS). In *Journal of Physics: Conference Series* (Vol. 1823, No. 1, p. 012078). IOP Publishing.
- Hendryadi, H. (2014). Content Validity (Validitas isi). *International Encyclopedia of the Social & Behavioral Sciences*, 1, 774-777.
- Kemendikbud, P. P. P., & Abduh, M. (2019). Panduan penulisan soal HOTS-higher order thinking skills.
- Laliyo, L. A. R., Sumintono, B., & Panigoro, C. (2022). Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model. *Heliyon*, 8(3), e09126.
- Liu, X. (2020). *Using and developing measurement instruments in science education: a Rasch modeling approach 2nd edition*. IAP.
- PUTRA, B. R. PENGEMBANGAN BUKU TEKS BERMUATAN HIGHER ORDER

THINKING SKILLS (HOTS) PADA  
MATERI LAJU REAKSI KIMIA.

- Risdiana, A., Erna, M., & Holiwarni, B. (2022). Pengembangan Soal HOTS (Higher Order Thinking Skills) pada Materi Asam-Basa untuk Kelas XI SMA/MA Sederajat. *Jurnal Inovasi Pendidikan Kimia*, 16(2), 111-117.
- Rodríguez-Mora, F., Cebrián-Robles, D., & Blanco-López, Á. (2021). An assessment using Rubrics and the Rasch Model of 14/15-year-old students' difficulties in arguing about bottled water consumption. *Research in Science Education*, 1-17.
- Shahat, M. A., Boone, W. J., Ambusaidi, A. K., Al Bahri, K., & Ohle-Peters, A. (2022). Use of Rasch Analysis to Develop an Arabic Language Survey (STPLTS) to Measure Omani Science Teachers' Views towards the Classroom Application of Pedagogical Learning Theories. *Journal of Baltic Science Education*, 21(3), 513-527.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.
- Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, 16(2), 92-96.
- Werner, S. M., Chen, Y., & Stieff, M. (2021). Examining the psychometric properties of the chemistry self-concept inventory using rasch modeling. *Journal of Chemical Education*, 98(11), 3412-3420.
-