



**Kasuari: Physics Education Journal (KPEJ)**  
**Universitas Papua**

Web: <http://jurnal.unipa.ac.id/index.php/kpej>



## **Analysis of Items in the Mid-Semester Assessment of Science Class VII SMP Yapis Manokwari**

**Eko Prayetno, Irfan Yusuf, Mujasam\***

Program Studi Pendidikan Fisika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Papua

\*Corresponding author: [m.mujasam@unipa.ac.id](mailto:m.mujasam@unipa.ac.id)

**Abstract:** *The purpose of this study was to analyze the test items in the midterm assessment using the Rasch model in terms of validity, reliability, level of difficulty, discriminating power, and the effectiveness of distractors in class VII science questions at SMP Yapis Manokwari. This research is a quantitative research that uses descriptive quantitative methods with data collection techniques using documentation. Respondent data were 60 students with purposive sampling. The items used in the Mid Semester Assessment are 40 questions. From the validity test, it was found that the suitability level of the items was 35 questions that were fit and 5 questions that were not fit. Student reliability 0.81 means good, item reliability item 0.92 means very good and reliability between students with item 0.83 means special. The level of difficulty of the questions is in the medium category, which means that the questions are good. The discriminating power of the items obtained is that 2 items (5%) are in the very good category, 30 items (75%) are in the good category, 4 items (10%) are in the sufficient category, and 4 items (10%) are bad category, from the items it has succeeded in identifying the distinguishing power of students divided into 3 groups in the category of above average ability, average ability, and below average ability. The effectiveness of the distractor of the 40 questions in the test generally functions well. The conclusion is that the analysis of the items using the Rasch model in the midterm assessment is stated to be valid, reliable, the level of difficulty of the items is very good, the discriminating power is very good, and the effectiveness of the distractors obtained is in the good category.*

**Keywords:** *Item Analysis, Learning Evaluation, Rasch Model*

## **Analisis Butir Soal Penilaian Tengah Semester pada Mata Pelajaran IPA Kelas VII SMP Yapis Manokwari**

**Abstrak:** Telah dilakukan penelitian dengan tujuan untuk menganalisis butir soal pada penilaian tengah semester dengan menggunakan pemodelan *Rasch*. Pemodelan *Rasch* digunakan untuk mencari nilai validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektifitas pengecoh pelajaran IPA kelas VII SMP Yapis Manokwari. Jenis penelitian ini merupakan penelitian kuantitatif menggunakan metode kuantitatif deskriptif dengan dokumentasi sebagai pengumpulan data. Responden sebanyak 60 peserta didik dengan teknik *purposive sampling*. Butir soal yang digunakan dalam Penilaian Tengah Semester sebanyak 40 soal. Dari Uji validitas diperoleh tingkat kesesuaian butir soal sebanyak 35 soal yang *fit* dan 5 soal yang tidak *fit*. Reliabilitas peserta didik 0,81 bermakna baik, reliabilitas butir soal butir soal 0,92 bermakna baik sekali dan reliabilitas antara peserta didik dengan butir soal 0,83 bermakna istimewa. Tingkat kesukaran soal masuk dalam kategori sedang, yang artinya soal tersebut sudah baik. Daya pembeda butir soal yang didapat bahwa 2 butir soal (5%) termasuk kategori baik sekali, 30 butir soal (75%) termasuk kategori baik, 4 butir soal (10%) termasuk kategori cukup, dan 4 butir soal (10%) termasuk kategori jelek, dari butir soal tersebut telah berhasil mengidentifikasi pembeda pada peserta didik terbagi menjadi 3 kelompok dalam kategori berkemampuan di atas rata-rata, berkemampuan rata-rata, dan berkemampuan di bawah rata-rata. Efektifitas pengecoh dari 40 soal pada pengujian pada umumnya berfungsi dengan baik, terdapat 2 butir soal (5%) yang tidak berfungsi dan 38 butir soal (95%) berfungsi dengan baik. Hasil tersebut dapat disimpulkan bahwa analisis butir soal menggunakan pemodelan *Rasch*

pada penilaian tengah semester dinyatakan valid, reliabel, tingkat kesukaran butir soal yang sangat baik, daya pembeda yang sangat baik, dan efektifitas pengecoh soal yang didapat masuk dalam kategori baik.

**Kata kunci:** Analisis Butir Soal, Evaluasi Belajar, Pemodelan *Rasch*.

## PENDAHULUAN

Peraturan pemerintah menjelaskan bahwa evaluasi hasil belajar sangatlah penting dilakukan untuk mengetahui sejauh mana kemampuan yang diperoleh peserta didik dalam hal prestasi atau tercapainya kompetensi dan pemahaman peserta didik setelah melalui proses pembelajaran. Peraturan tersebut sebagai acuan dalam proses penilaian hasil belajar dalam mengukur ketercapaian pembelajaran peserta didik. Sistem penilaian yang baik mendorong dan memotivasi guru dalam menentukan metode mengajar yang cocok untuk digunakan. Jika ditemukan metode mengajar yang baik untuk digunakan maka diharapkan dapat memberikan pengaruh terhadap keberhasilan peserta didik. Langkah yang digunakan untuk mengukur keberhasilan peserta didik yaitu dengan melakukan evaluasi pembelajaran.

Evaluasi dapat diartikan sebagai menilai dan juga mengukur. Menilai yaitu memberikan keputusan baik dan buruk terhadap obyek yang dinilai, sedangkan mengukur berarti membandingkan suatu yang diukur dengan acuan standarnya dan memberikan angka terhadap hasilnya. Hasil belajar harus dievaluasi dengan maksud untuk mengukur ketercapaian pembelajaran berbagai bidang ilmu yang ditentukan oleh kurikulum pembelajaran (Widyaningsih & Yusuf, 2018). Azwar (2014) menegaskan bahwa supaya sebuah keputusan dalam pendidikan menjadi berguna harus di dasari oleh informasi-informasi yang tepat, akurat, dan reliabel. Maka dari itu, evaluasi sangat membantu pendidik dalam mengukur progres belajar dari peserta didik dengan melakukan sebuah tes.

Tes berperan sebagai alat ukur untuk menggambarkan keberhasilan dari suatu proses pembelajaran. Tes merupakan suatu metode pengumpulan informasi tentang ketercapaian tujuan tujuan pembelajaran (Wahyudi, 2010). Tes juga merupakan suatu langkah yang digunakan atau prosedur ditempuh dalam mengukur dan menilai di bidang pendidikan (Kadir, 2015). Tes yang baik yaitu tes yang dapat dipertanggungjawabkan kelayakannya, dimulai dari segi validitas, reliabilitas, dan tingkat kesukaran dari soal yang digunakan. Instrumen tes dibuat dalam bentuk pilihan ganda maupun esai dan diperiksa konstruksinya oleh para ahli seperti guru yang membidangi soal tersebut, kemudian di uji cobakan kepada peserta didik. Data hasil uji coba dapat dianalisis menggunakan pemodelan Rasch pada aplikasi Winstep® versi 3.73. Berdasarkan prinsip probabilitas, untuk mengukur validitas dan reliabilitas yang langsung dapat digunakan pemodelan Rasch (Tennant, dkk, 2004). Pemodelan Rasch berdasarkan pada pola respons yang sistematis mampu menduga data yang hilang; mampu mengukur tingkat kesalahan instrumen sehingga perhitungan menjadi tepat; dan mampu mengkalibrasi beberapa hal antara lain skala pengukuran, responden, dan butir soal (Sumintono & Widhiarso, 2015). Analisis Rasch juga memberikan pengembangan dan efisiensi terhadap instrumen tes pilihan ganda (Hofer, 2017). Pentingnya pemodelan Rasch ini dapat di aplikasikan di berbagai sekolah untuk membantu pendidik dalam menguji kelayakan soal penilaian tengah semester maupun penilaian akhir semester sebelum diujikan ke peserta didik.

Berdasarkan hasil observasi di SMP Yapis Manokwari, menunjukkan bahwa soal IPA yang digunakan adalah dibuat oleh guru IPA dan belum divalidasi oleh guru IPA lainnya. Soal tersebut tanpa divalidasi sudah di ujikan kepada peserta didik. Seharusnya bagi seorang pendidik harus memperhatikan tahapan-tahapan dalam pembuatan butir soal,

sehingga sesuatu yang ingin dicapai atau dikembangkan dapat berpengaruh dalam proses pembelajaran peserta didik.

Lembaga pendidikan di Papua Barat khususnya di Manokwari, ternyata masih ada Pendidik yang ketika membuat soal-soal ujian masih belum diuji kelayakan soal tersebut sebelum digunakan. Hasil observasi yang dilakukan di SMP Yapis Manokwari dengan mengumpulkan lembar soal, kunci jawaban, dan lembar jawaban peserta didik, telah ditemukan soal ujian penilaian tengah semester yang digunakan sebagai salah satu syarat kelulusan belum diuji kelayakannya. Soal ujian tersebut diujikan di kelas VII SMP pada empat kelas yang masing-masing kelas terdiri dari 30 peserta didik pada mata pelajaran IPA dengan jumlah soal sebanyak 40 butir.

## METODE PENELITIAN

Jenis penelitian yaitu penelitian kuantitatif dengan menggunakan metode kuantitatif deskriptif, yang bertujuan untuk menggambarkan suatu keadaan yang terjadi berdasarkan angka statistik. Metode deskriptif kuantitatif merupakan metode penelitian untuk menggambarkan situasi atau kejadian yang terjadi dengan menggunakan data statistik. Populasi dalam penelitian ini yaitu kelas VII SMP Yapis Manokwari dengan jumlah sampel 60 orang menggunakan teknik *purposive sampling*. Teknik statistik yang digunakan hanya untuk melihat kualitas tes yang dibuat berdasarkan pada hasil jawaban peserta didik. Jumlah butir soal sebanyak 40. Tabel 1 merupakan tahapan-tahapan yang dilakukan dalam penelitian.

**Tabel 1.** Tahapan Penelitian

No	Tahapan	Kegiatan
1	Tahap Pendahuluan	1. Persiapan literatur. 2. Pengumpulan dan pemilihan data awal dari subyek.
2	Tahap Analisis	1. Pengumpulan data meliputi lembar soal, kunci jawaban, dan lembar jawaban peserta didik. 2. Menganalisis data hasil jawaban peserta didik.
3	Tahap Simpulan	1. Pembahasan hasil analisis. 2. Kesimpulan.

Data dikumpulkan menggunakan metode dokumentasi yaitu mengambil hasil jawaban peserta didik di sekolah termasuk kunci jawaban. Dengan menggunakan metode dokumentasi akan mendapatkan data hasil belajar diantaranya yaitu: a) Lembaran Soal Penilaian Tengah Semester mata pelajaran IPA tahun ajaran 2019/2020, b) Lembaran jawaban peserta didik, dan c) lembaran kunci jawaban soal. Data yang telah didapatkan selanjutnya dianalisis tingkat validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektifitas pengecoh soal.

### 1. Validitas

Hasil dari analisis validitas menggunakan winsteps yaitu pada validitas konten yang meliputi tingkat kesesuaian soal pada *output tabels 10. Item (column): fit order output*. Tabel didapatkan dari menu utama *output table* kemudian pilih tabel 10. *Item (column): fit order*, sehingga tabel ini menampilkan informasi sejumlah butir soal dengan kriteria yang tidak *fit* berada pada bagian atas. Dengan banyaknya data dapat mempengaruhi nilai kesesuaian butir dan banyaknya sampel yang digunakan maka mempengaruhi tingkat kesesuaiannya. Menurut Boone, dkk (2014) bahwa nilai kriteria yang dipakai dalam melihat butir soal yang dipakai yaitu sebagai berikut:

- a. *Outfit Mean Square* (MNSQ) bernilai:  $0,5 < \text{MNSQ} < 1,5$
- b. *Outfit Z-standard* (ZSTD) bernilai:  $-2,0 < \text{MNSQ} < +2,0$
- c. *Point Measure Correlation* (Pt Mean Corr) bernilai:  $0,4 < \text{Pt Measure} < 0,85$ .

**2. Reliabilitas**

Reliabilitas merupakan serapan dari kata *reliability* yang artinya dapat dipercaya (Arikunto, 2012). Koefisien reliabilitas merupakan angka tingkat kepercayaan suatu instrumen. Cerminan nilai reliabilitas yaitu dengan adanya hubungan antara dua distribusi skor yang dikenakan pada kelompok responden. Menurut Sumintono & Widhiarso (2015) memberikan hasil gambaran berupa:

- a. Nilai rata-rata peserta didik dan *item* ditunjukkan oleh *person* dan *item measure*, nilai terkecil adalah kemampuan yang dimiliki peserta didik dan soal jelek ditunjukkan pada logit 0,0.
- b. Nilai yang digunakan untuk mengukur reliabilitas adalah *nilai alpa crombach* yang merupakan hubungan antara *person* dan *item*, kriteria yang dipakai adalah  $< 0,5$ : buruk,  $0,5 - 0,6$ : jelek,  $0,6 - 0,7$ : cukup,  $0,7 - 0,8$ : baik,  $> 0,8$ : istimewa.
- c. Nilai reliabilitas peserta didik dan nilai reliabilitas *item* adalah *nilai person reliability* dan *item reliability*, kriteria yang dipakai yaitu:  $< 0,67$ : lemah,  $0,67 - 0,80$ : cukup,  $0,80 - 0,90$ : bagus,  $0,91 - 0,94$ : baik sekali,  $> 0,94$ : istimewa.

Separation merupakan pengelompokan *person* dan *item*. Kualitas instrumen semakin baik ditunjukkan dengan semakin besar nilai *separation* sebagaimana persamaan (1).

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

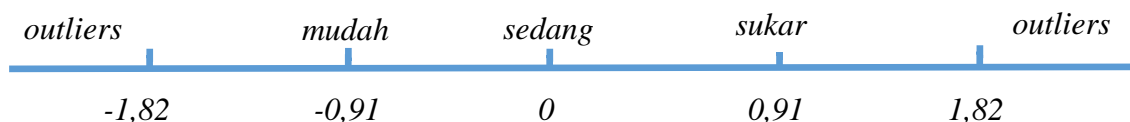
Penjelasan:

H : Nilai strata (pengelompokan)

Separation : Nilai *Separasi* hasil responden

**3. Tingkat kesukaran**

Tingkat kesukaran butir soal (*Item Measure*) ditunjukkan pada *output tables 13*. *Item Measure* yang menampilkan nilai *logit* butir soal dengan jelas. Informasi yang diberikan dari *output* tabel adalah butir soal yang bernilai *logit measure* tinggi ke rendah. Dari penelitian ini diperoleh kelompok soal sebagaimana Gambar 1.



**Gambar 1.** Kelompok Soal

Sumber: Sumintono & Widhiarso, 2015

Instrumen tes dikatakan baik apabila memenuhi tingkatan perbandingan yaitu mudah, sedang, dan sukar yang semuanya proporsional. Proporsional dalam hal dimaksud adalah 25% soal mudah, 50% soal sedang, 25% soal sukar. Proporsi yang lainnya yang dapat digunakan yaitu 3:5:2, 3:4:3 dan 2:5:3 (Ratumanan & Laurens, 2011). Tabel 2 menjelaskan pengelompokan soal berdasarkan tingkat kesukaran.

**Tabel 2.** Kategori Tingkat Kesukaran Soal

Kategori	Nilai
Outliner	$> 1,82$ dan $< (-1,82)$
Sukar	$0,91 - 1,82$
Sedang	$-0,91 - 0,91$
Mudah	$-1,82 - (-0,91)$

#### 4. Analisis daya pembeda

Untuk membedakan kualitas peserta tes maka seluruh peserta tes dibagi ke dalam kelompok pandai (atas) dan kelompok tidak pandai (bawah). Jika semua kelompok pandai dapat menjawab soal dengan benar dan semua kelompok tidak pandai menjawab salah, maka soal tersebut memiliki indeks diskriminasi paling besar yaitu 1,00. Sebaliknya jika semua kelompok pandai menjawab salah dan kelompok tidak pandai menjawab benar maka soal tersebut memiliki indeks diskriminasi yaitu -1,00. Apabila semua kelompok pandai dan tidak pandai sama-sama menjawab benar atau sama-sama menjawab salah, maka indeks diskriminasi soal yaitu 0,00 (tidak ada pembeda).

#### 5. Analisis Tingkat Keberfungsian Pengecoh/ Efektifitas Option

Pengecoh pada suatu soal berfungsi atau tidak dapat diketahui dari jawaban peserta didik. Jawaban benar merupakan pilihan kunci (*key option*), sedangkan pilihan lainnya merupakan pilihan pengecoh (*distractor*). Kunci jawaban dikatakan sudah efektif apabila:

- Pemilih pada kelompok atas lebih banyak dari pemilih pada kelompok bawah.
- Jumlah pemilih pada kelompok atas dan kelompok bawah hendaknya lebih dari 25% dan tidak lebih dari 75% peserta tes tersebut.

Selanjutnya pilihan pengecoh dikatakan efektif jika memenuhi kriteria sebagai berikut:

- Pemilih pada kelompok atas harus kurang dari pemilih pada kelompok bawah.
- Jumlah pemilih minimal 5% dari peserta tes dari kelompok atas dan kelompok bawah.
- Jika peserta tes tidak memilih salah satu pilihan pada butir tes (omit), maka tidak lebih dari 10% jumlah peserta tes pada kelompok atas dan kelompok bawah.

Pengecoh dikatakan tidak berfungsi paling tidak dipilih oleh 5% peserta tes dan banyak dipilih oleh peserta didik yang belum memahami materi (Marthunis M, *et.al*, 2015)

## HASIL DAN PEMBAHASAN

### 1. Validitas

Validitas soal yang dihasilkan berupa validitas konten. hasil analisis validitas konten diperlihatkan *output tables 10. item fit order* diberikan pada Gambar 2.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE CORR.	PURE EXP.	EXACT MATCH		Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	OBS%			EXP%		
24	10	60	1.72	.37	1.15	.7	1.65	1.7	A .10	.33	81.7	83.8	S24	
30	9	60	1.86	.38	1.24	1.0	1.64	1.6	B .03	.32	83.3	85.4	S30	
23	16	60	1.04	.32	1.27	1.6	1.64	2.5	C .03	.38	70.0	75.7	S23	
34	7	60	2.19	.42	1.20	.7	1.49	1.1	D .07	.30	86.7	88.4	S34	
38	15	60	1.14	.32	1.32	1.8	1.43	1.7	E .03	.37	71.7	76.9	S38	
37	17	60	.94	.31	1.20	1.3	1.36	1.6	F .15	.38	71.7	74.6	S37	
26	32	60	-.31	.28	1.22	2.1	1.17	1.3	G .18	.38	50.0	67.2	S26	
18	16	60	1.04	.32	.99	.0	1.20	.9	H .33	.38	76.7	75.7	S18	
25	19	60	.75	.30	1.11	.8	1.18	1.0	I .26	.39	68.3	72.6	S25	
5	22	60	.49	.29	1.10	.8	1.16	1.0	J .28	.39	68.3	70.1	S5	
14	17	60	.94	.31	1.14	.9	1.12	.6	K .25	.38	71.7	74.6	S14	
35	20	60	.66	.30	1.05	.4	1.13	.8	L .32	.39	73.3	71.6	S35	
11	36	60	-.63	.28	1.13	1.3	1.09	.6	M .25	.37	58.3	67.4	S11	
19	18	60	.84	.31	1.03	.3	1.10	.6	N .33	.38	78.3	73.6	S19	
4	24	60	.32	.29	1.07	.6	1.05	.4	O .33	.40	60.0	68.9	S4	
29	20	60	.66	.30	1.02	.2	1.06	.4	P .36	.39	70.0	71.6	S29	
31	39	60	-.88	.29	.99	-.1	1.04	.3	Q .35	.35	70.0	68.5	S31	
39	25	60	.24	.29	1.03	.3	1.03	.3	R .36	.40	65.0	68.3	S39	
15	30	60	-.16	.28	.98	-.1	1.01	.1	S .40	.39	71.7	67.2	S15	
28	33	60	-.39	.28	1.00	.1	1.00	.0	T .38	.38	66.7	67.2	S28	
17	29	60	-.08	.28	1.00	.0	.98	-.1	t .40	.39	65.0	67.3	S17	
22	31	60	-.23	.28	.95	-.5	.94	-.4	s .44	.39	70.0	67.2	S22	
7	52	60	-2.26	.39	.95	-.1	.70	-.6	r .33	.24	86.7	86.6	S7	
16	48	60	-1.74	.34	.91	-.5	.95	-.1	q .37	.28	80.0	79.9	S16	
40	5	60	2.60	.49	.94	.0	.56	-.7	p .38	.26	91.7	91.6	S40	
2	39	60	-.88	.29	.94	-.6	.87	-.7	o .43	.35	73.3	68.5	S2	
10	33	60	-.39	.28	.92	-.8	.87	-.9	n .47	.38	73.3	67.2	S10	
27	43	60	-1.23	.30	.91	-.7	.79	-.9	m .44	.33	71.7	72.1	S27	
32	34	60	-.47	.28	.90	-.9	.86	-1.0	l .48	.38	71.7	67.2	S32	
1	47	60	-1.63	.33	.89	-.6	.76	-.8	k .43	.29	78.3	78.3	S1	
9	18	60	.84	.31	.87	-.9	.89	-.5	j .51	.38	81.7	73.6	S9	
3	45	60	-1.42	.31	.88	-.8	.69	-1.2	i .47	.31	75.0	74.9	S3	
21	54	60	-2.60	.44	.88	-.3	.52	-.9	h .39	.21	90.0	90.0	S21	
13	32	60	-.31	.28	.85	-1.6	.86	-1.1	g .53	.38	73.3	67.2	S13	
36	27	60	.08	.28	.86	-1.4	.85	-1.2	f .54	.39	78.3	67.7	S36	
8	38	60	-.79	.29	.86	-1.4	.80	-1.2	e .51	.36	71.7	68.0	S8	
12	45	60	-1.42	.31	.84	-1.1	.67	-1.3	d .50	.31	75.0	74.9	S12	
6	33	60	-.39	.28	.84	-1.7	.79	-1.6	c .56	.38	76.7	67.2	S6	
33	32	60	-.31	.28	.80	-2.2	.75	-2.1	b .60	.38	73.3	67.2	S33	
20	26	60	.16	.28	.78	-2.1	.76	-2.0	a .62	.40	83.3	68.0	S20	
MEAN	28.4	60.0	.00	.31	1.00	-.1	1.01	.0			73.8	73.4		
S.D.	12.5	.0	1.15	.05	.14	1.0	.28	1.1			8.2	7.0		

Gambar 2. Analisis Validitas Konten

Gambar 2 merupakan hasil analisis tingkat kesesuaian butir soal, yang menjelaskan *item fit* atau tidak sesuai. Soal diterima apabila masuk dalam kriteria *outfit* MNSQ, ZSTD, *Pt-Mean corr.* Hasil analisis tersebut terdapat masing-masing 37 soal *fit* dan 3 soal tidak *fit* pada kategori MNSQ dan 2 soal tidak *fit* pada kategori ZSTD, sedangkan pada kategori *Pt-Mean corr* terdapat 17 soal *fit* dan 23 soal tidak *fit*. Hasil tersebut menunjukkan bahwa soal yang memenuhi kriteria yaitu 38 soal dengan 2 kriteria yaitu MNSQ dan ZSTD, sedangkan diperbaiki jika memenuhi 1 kriteria yakni terdapat 2 soal yang hanya memenuhi ZSTD. Selanjutnya, nilai minimum dari variansi data yang dijelaskan oleh ukuran yang mencapai 20% dianggap memenuhi unsur unidimensionalitas (Ubaidillah et al., 2022).

## 2. Reliabilitas

Hasil analisis reliabilitas menggunakan pemodelan *Rasch Winstep* ditunjukkan pada *output tabels 3.1. Summary Statistic* yang dilihat pada nilai *separation* dan nilai *alpha crombach* Gambar 3.

SUMMARY OF 60 MEASURED Person

	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	18.9	40.0	-.13	.38	1.00	-.1	1.01	.0
S.D.	6.5	.0	.90	.02	.20	1.1	.37	1.1
MAX.	35.0	40.0	2.38	.51	1.57	3.0	2.22	2.5
MIN.	10.0	40.0	-1.37	.36	.63	-2.3	.34	-1.7
REAL RMSE	.39	TRUE SD	.80	SEPARATION	2.05	Person	RELIABILITY	.81
MODEL RMSE	.38	TRUE SD	.81	SEPARATION	2.15	Person	RELIABILITY	.82
S.E. OF Person MEAN = .12								

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00  
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .83

SUMMARY OF 40 MEASURED Item

	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	28.4	60.0	.00	.31	1.00	-.1	1.01	.0
S.D.	12.5	.0	1.15	.05	.14	1.0	.28	1.1
MAX.	54.0	60.0	2.60	.49	1.32	2.1	1.65	2.5
MIN.	5.0	60.0	-2.60	.28	.78	-2.2	.52	-2.1
REAL RMSE	.33	TRUE SD	1.11	SEPARATION	3.38	Item	RELIABILITY	.92
MODEL RMSE	.32	TRUE SD	1.11	SEPARATION	3.49	Item	RELIABILITY	.92
S.E. OF Item MEAN = .18								

UMEAN=.0000 USCALE=1.0000  
 Item RAW SCORE-TO-MEASURE CORRELATION = -.99  
 2400 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 2540.24 with 2301 d.f. p=.0003  
 Global Root-Mean-Square Residual (excluding extreme scores): .4201  
 Capped Binomial Deviance = .2298 for 2400.0 dichotomous observations

**Gambar 3.** Analisis Reliabilitas Butir Soal

Berdasarkan pada Gambar 3 didapatkan reliabilitas peserta didik yaitu 0,81 yang berarti masuk pada kriteria baik (> 0,80) dan reliabilitas soal yaitu 0,92 yang berarti termasuk pada kriteria baik sekali (0,91 - 0,94). *Alpha crombach* yaitu 0,83 masuk pada kategori istimewa. Sumintono & Widhiarso (2015) yang menyatakan bahwa nilai separasi dari instrumen semakin besar menjadikan kualitas instrumen pada keseluruhan responden dan butir makin baik, karena dapat mengidentifikasi kelompok responden dan butir. Pengelompokkan siswa dan butir soal dapat ditentukan dengan persamaan (2).

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

$$H_{siswa} = \frac{[(4 \times 2,05) + 1]}{3} = \frac{[9,2]}{3} = 3,067$$

$$H_{soal} = \frac{[(4 \times 3,38) + 1]}{3} = \frac{[14,52]}{3} = 4,84 \tag{2}$$

Dari hasil tersebut didapatkan nilai  $H_{siswa}$  sebanyak 3,067, dan  $H_{soal}$  sebanyak 4,84 dibulatkan menjadi 5. Hal ini dapat dikatakan bahwa *separation* pada peserta didik bernilai baik karena dapat membagi kemampuan peserta didik dalam 3 kelompok, demikian halnya separasi pada soal juga menjadi baik karena mengelompokkan soal dalam 5 kriteria yaitu sangat sukar, sukar, sedang, mudah dan sangat mudah (Rozali et al., 2022).



### 3. Tingkat kesukaran

Berdasarkan hasil analisis data pada *output tables 13 item measure* pada program aplikasi *Winstep*, didapatkan nilai tingkat kesukaran dengan rentang dari -2,3 sampai +2,3 dengan kategori sukar (1,15 – 2,3), sedang (-1,15 - +1,15), dan mudah (-1,15 - -2,3). Gambar 4 dan 5 menjelaskan analisis tingkat kesukaran soal.

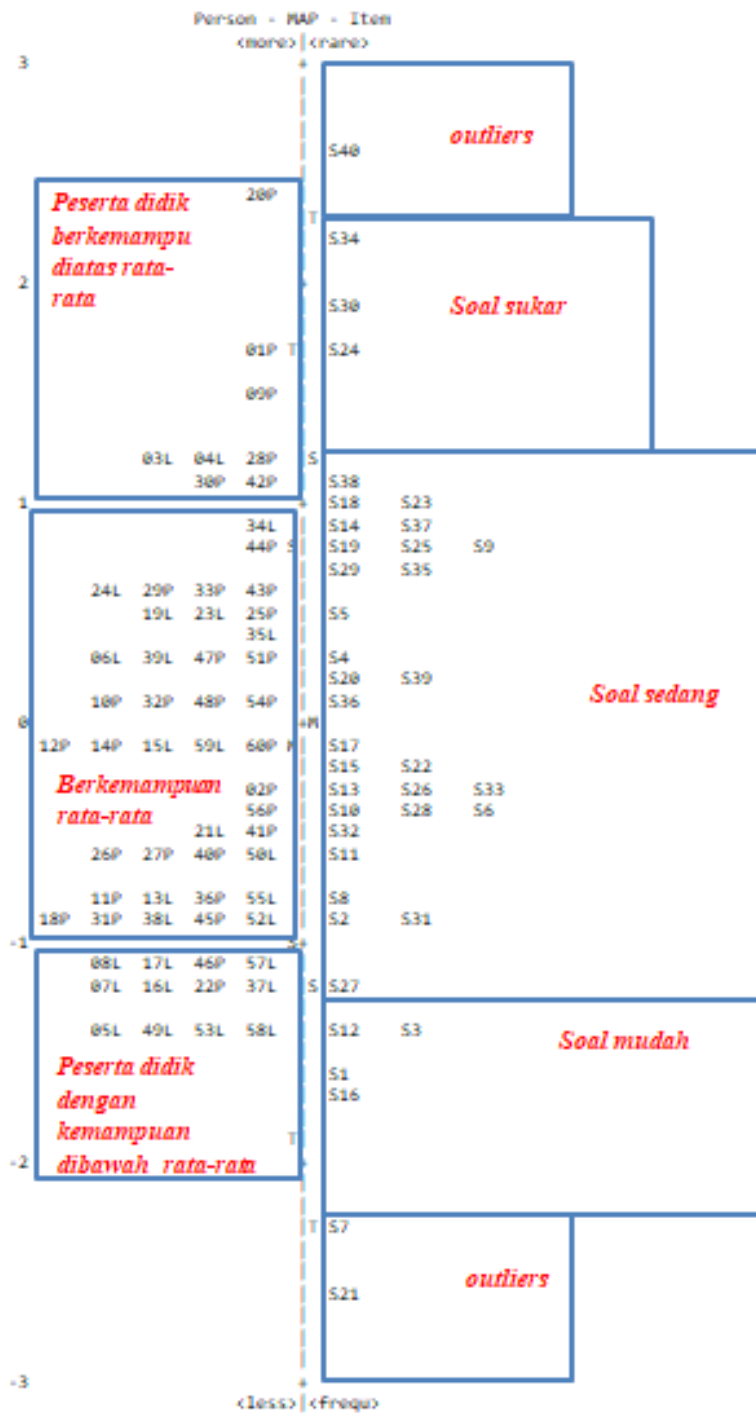
Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item
				S. E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
40	5	60	2.60	.49	.94	.0	.56	-.7	.38	.26	91.7	91.6	S40	
34	7	60	2.19	.42	1.20	.7	1.49	1.1	.07	.30	86.7	88.4	S34	
30	9	60	1.86	.38	1.24	1.0	1.64	1.6	.03	.32	83.3	85.4	S30	
24	10	60	1.72	.37	1.15	.7	1.65	1.7	.10	.33	81.7	83.8	S24	
38	15	60	1.14	.32	1.32	1.8	1.43	1.7	.03	.37	71.7	76.9	S38	
18	16	60	1.04	.32	.99	.0	1.20	.9	.33	.38	76.7	75.7	S18	
23	16	60	1.04	.32	1.27	1.6	1.64	2.5	.03	.38	70.0	75.7	S23	
14	17	60	.94	.31	1.14	.9	1.12	.6	.25	.38	71.7	74.6	S14	
37	17	60	.94	.31	1.20	1.3	1.36	1.6	.15	.38	71.7	74.6	S37	
9	18	60	.84	.31	.87	-.9	.89	-.5	.51	.38	81.7	73.6	S9	
19	18	60	.84	.31	1.03	.3	1.10	.6	.33	.38	78.3	73.6	S19	
25	19	60	.75	.30	1.11	.8	1.18	1.0	.26	.39	68.3	72.6	S25	
29	20	60	.66	.30	1.02	.2	1.06	.4	.36	.39	70.0	71.6	S29	
35	20	60	.66	.30	1.05	.4	1.13	.8	.32	.39	73.3	71.6	S35	
5	22	60	.49	.29	1.10	.8	1.16	1.0	.28	.39	68.3	70.1	S5	
4	24	60	.32	.29	1.07	.6	1.05	.4	.33	.40	60.0	68.9	S4	
39	25	60	.24	.29	1.03	.3	1.03	.3	.36	.40	65.0	68.3	S39	
20	26	60	.16	.28	.78	-2.1	.76	-2.0	.62	.40	83.3	68.0	S20	
36	27	60	.08	.28	.86	-1.4	.85	-1.2	.54	.39	78.3	67.7	S36	
17	29	60	-.08	.28	1.00	.0	.98	-.1	.40	.39	65.0	67.3	S17	
15	30	60	-.16	.28	.98	-.1	1.01	.1	.40	.39	71.7	67.2	S15	
22	31	60	-.23	.28	.95	-.5	.94	-.4	.44	.39	70.0	67.2	S22	
13	32	60	-.31	.28	.85	-1.6	.86	-1.1	.53	.38	73.3	67.2	S13	
26	32	60	-.31	.28	1.22	2.1	1.17	1.3	.18	.38	50.0	67.2	S26	
33	32	60	-.31	.28	.80	-2.2	.75	-2.1	.60	.38	73.3	67.2	S33	
6	33	60	-.39	.28	.84	-1.7	.79	-1.6	.56	.38	76.7	67.2	S6	
10	33	60	-.39	.28	.92	-.8	.87	-.9	.47	.38	73.3	67.2	S10	
28	33	60	-.39	.28	1.00	.1	1.00	.0	.38	.38	66.7	67.2	S28	
32	34	60	-.47	.28	.90	-.9	.86	-1.0	.48	.38	71.7	67.2	S32	
11	36	60	-.63	.28	1.13	1.3	1.09	.6	.25	.37	58.3	67.4	S11	
8	38	60	-.79	.29	.86	-1.4	.80	-1.2	.51	.36	71.7	68.0	S8	
2	39	60	-.88	.29	.94	-.6	.87	-.7	.43	.35	73.3	68.5	S2	
31	39	60	-.88	.29	.99	-.1	1.04	.3	.35	.35	70.0	68.5	S31	
27	43	60	-1.23	.30	.91	-.7	.79	-.9	.44	.33	71.7	72.1	S27	
3	45	60	-1.42	.31	.88	-.8	.69	-1.2	.47	.31	75.0	74.9	S3	
12	45	60	-1.42	.31	.84	-1.1	.67	-1.3	.50	.31	75.0	74.9	S12	
1	47	60	-1.63	.33	.89	-.6	.76	-.8	.43	.29	78.3	78.3	S1	
16	48	60	-1.74	.34	.91	-.5	.95	-.1	.37	.28	80.0	79.9	S16	
7	52	60	-2.26	.39	.95	-.1	.70	-.6	.33	.24	86.7	86.6	S7	
21	54	60	-2.60	.44	.88	-.3	.52	-.9	.39	.21	90.0	90.0	S21	
MEAN	28.4	60.0	.00	.31	1.00	-.1	1.01	.0			73.8	73.4		
S. D.	12.5	.0	1.15	.05	.14	1.0	.28	1.1			8.2	7.0		

Gambar 4. Analisis Kesukaran Butir Soal



INPUT: 60 Person 40 Item REPORTED: 60 Person 40 Item 2 CATS MINSTEPS 3.73



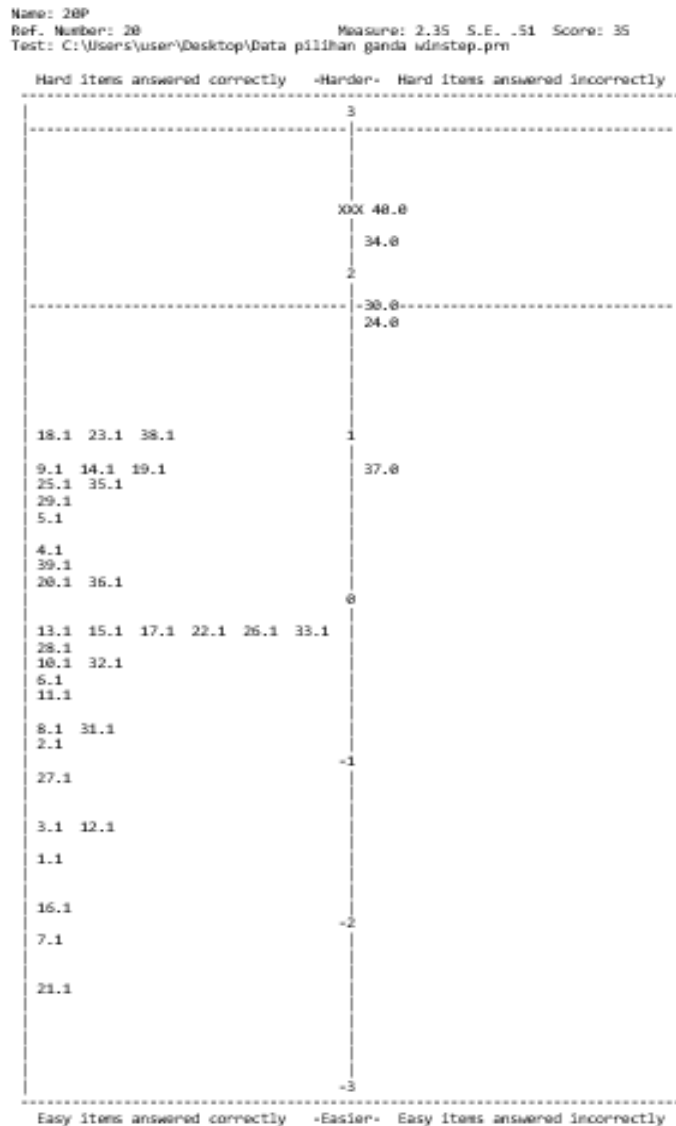
Gambar 5 Analisis Person Dan Item

Berdasarkan hasil analisis pada Gambar 4 dan 5 didapatkan pengelompokan tingkat kesulitan butir soal pada tabel bagian kiri yang meliputi outliers, sukar, sedang dan mudah. Analisis tingkat kesukaran didapatkan bahwa paling sukar sebanyak 2 soal, sedang sebanyak 30 soal dan mudah sebanyak 4 soal serta ada 4 soal yang outlier. Tingkat kesulitan soal didasarkan pada nilai logit, yaitu probabilitas soal akan dijawab oleh peserta didik (Mulyanti et al., 2022). Berdasarkan hasil analisis pada Gambar 5

didapatkan pengelompokan tingkat kemampuan peserta didik pada kolom sebelah kanan dan pengelompokan tingkat kesukaran butir soal pada kolom bagian kiri. Hasil analisis didapatkan pengelompokan peserta didik dengan kemampuan diatas rata-rata dengan kode 20P, 01P, 09P, 03L, 04L, 28P, 30P, dan 42P, kemampuan rata-rata dengan kode 34L, 44P, 24L, 29P, 33P, 43P, 19L, 23L, 25P, 35L, 06L, 39L, 47P, 51P, 10P, 32P, 48P, 54P, 12P, 14P, 15L, 59L, 60P, 02P, 56P, 21L, 41P, 26P, 27P, 40P, 50L, 11P, 13L, 36P, 55L, 18P, 31P, 38L, 45P, 52L, dan berkemampuan dibawah rata-rata dengan kode 08L, 17L, 46P, 57L, 07L, 16L, 22P, 37L, 05L, 49L, 53L dan 58L.

**4. Daya pembeda**

Analisis daya pembeda soal dilakukan untuk menganalisis kemampuan butir soal membedakan antara peserta didik yang mampu menjawab dan kurang mampu menjawab. Selain itu daya pembeda digunakan untuk membedakan antara peserta didik mampu menjawab dengan peserta didik yang belum mampu menjawab. Kemampuan peserta didik dapat kita lihat pada hasil output analisis tingkat kemampuan masing-masing peserta didik yang dianalisis menggunakan *Winstep* pada *output tabel 36. Person diagnostic PKMAPS* yang tabel outputnya dapat dilihat pada Gambar 6.



Gambar 6. Analisis PKMAPS

Gambar 6 memperlihatkan kemampuan peserta didik di atas rata-rata, kemampuan rata-rata dan di bawah rata-rata dengan melihat nilai *measure*, jika nilai *measure* diatas 0,00 berarti peserta didik tersebut berkemampuan diatas rata-rata dan jika nilai *measure* dibawah 0,00 berarti peserta didik tersebut berkemampuan dibawah rata-rata atau untuk melihat kemampuan masing-masing berdasarkan nilai SD person yang dapat membagi kelompok kemampuan berdasarkan *variabel map*. Batas maksimum peserta didik dalam mengerjakan soal dapat diketahui dengan cara nilai *measure* + *S.E* (*Standar Error*) dan nilai minimum peserta didik yaitu dengan cara nilai *measure* - *S.E* (*Standar Error*). Nilai batas maksimum dan nilai batas minimum ditandai dengan garis putus-putus seperti yang ada pada *output tabel 36. Person diagnostic PKMAPS*.

Peserta didik dengan kode 20P memiliki kemampuan di atas rata-rata jika dilihat pada *output tabel 36. Person diagnostic PKMAPS* dikarenakan mempunyai nilai *measure* 2,35 yang berarti masih masuk dalam lingkup kemampuan diatas rata-rata yang mempunyai nilai SD person sebesar (+0,90)-(+1,80) maka peserta didik ini dikatakan berkemampuan diatas rata-rata. Nilai *S.E* (*standar error*) yaitu 0,51 dan nilai *Score* nya 35 artinya peserta tersebut menjawab benar sebanyak 35 butir soal dari 40 soal yang diujikan. Dari *output tabel 36. Person diagnostic PKMAPS* peserta didik menjawab benar soal nomor 21, 7, 16, 1, 3, 12, 27, 2, 8, 31, 11, 6, 10, 32, 28, 13, 15, 17, 22, 26, 33, 20, 36, 39, 4, 5, 29, 25, 35, 9, 14, 19, 18, 23, 38 karena masih dalam lingkungannya. Soal nomor 37, 24, 30, 34, 40 peserta didik tidak dapat menjawab dengan benar padahal seharusnya dapat dijawab dengan benar karena masih dalam lingkungannya, mungkin faktor yang mempengaruhi jawaban salah tersebut adalah berfungsinya efektifitas pengecoh soal dengan baik atau kurang teliti dalam menjawab.

### 5. Efektifitas pengecoh soal

Pengecoh atau *distractor* merupakan pilihan jawaban soal yang sengaja dibuat sebagai pengecoh pada pilihan jawaban. Option dikatakan berfungsi apabila telah dipilih oleh sekurang-kurangnya 5% orang. Tabel 3 menunjukkan keberfungsian pengecoh soal.

**Tabel 3.** Analisis Hasil Keberfungsian Pengecoh Soal

Kriteria dipilih oleh	Jumlah Option	Persentase
Berfungsi >5%	112	70%
Tidak Berfungsi <5%	8	5%
Kunci jawaban	40	25%
Jumlah	160	100%

Berdasarkan hasil analisis yang dilakukan terhadap soal yang dibuat didapatkan hasil dari 40 butir soal terdapat 160 pilihan jawaban yang digunakan dimana 40 pilihan jawaban lainnya merupakan kunci jawaban dari soal. Pilihan jawaban lainnya terdapat 120 pilihan jawaban yang digunakan sebagai *distractor* sehingga dihasilkan 8 jawaban yang tidak berfungsi yang dipilih oleh < 5% peserta didik, sedangkan 112 pilihan jawaban yang lain dipilih oleh > 5% peserta didik. *Distractor* yang berhasil berarti dapat menghasilkan soal yang lebih sulit. *Distractor* juga mengurangi tebakan acak jawaban untuk meningkatkan performa soal (Isnaeni et al., 2022). Dengan demikian dapat diketahui bahwa *distractor* pada soal penilaian tengah semester pada mata pelajaran IPA kelas VII SMP Yapis Manokwari sebagian besar berfungsi dengan baik. Kondisi ini terlihat dari banyaknya *distractor* yang bekerja dengan baik.

## KESIMPULAN DAN SARAN

Kesimpulan berdasarkan hasil penelitian dan pembahasan adalah soal valid berjumlah 35 butir soal (87,5%) dan tidak valid yaitu 5 butir soal (12,5%). Dari segi reliabilitas didapatkan hasil analisis butir soal pada peserta didik sebesar 0,81 (baik) dan *item* soal sebesar 0,92 (baik sekali). Dari kedua reliabilitas pada peserta didik dan item soal mempunyai  $\alpha$  *Crombach* sebesar 0,83 (istimewa), yang berarti reliabilitasnya bagus. Tingkat kesukaran butir soal didapatkan bahwa 4 butir soal (10%) termasuk dalam kategori mudah, 30 butir soal (75%) termasuk kategori sedang, 2 butir soal (5%) termasuk kategori sukar dan 4 butir soal termasuk kategori *outliers*. Daya pembeda soal didapatkan bahwa 2 butir soal (5%) termasuk dalam kategori baik sekali, 30 butir soal (75%) termasuk kategori baik, 4 butir soal (10%) termasuk kategori cukup, 4 butir soal (10%) termasuk kategori jelek atau kategori butir soal yang harus dibuang. Daya pembeda butir soal tersebut juga telah mengidentifikasi daya beda pada peserta didik yang terbagi menjadi 3 kategori yaitu dengan kategori berkemampuan diatas rata-rata, berkemampuan rata-rata, dan berkemampuan dibawah rata-rata. Efektivitas pengecoh mendapatkan berfungsi dengan baik.

Adapun saran terkait dengan hasil penelitian yang diperoleh yaitu soal dengan kualitas baik sebaiknya disimpan sebagai soal yang siap pakai. Untuk butir soal yang kualitasnya belum baik dapat diperbaiki pada bagian mana yang menjadi penyebabnya dan soal yang tidak baik diganti dengan soal baru. Butir soal yang sudah diperbaiki dan yang diganti dilakukan pengujian kembali untuk mengetahui kualitasnya.

## DAFTAR PUSTAKA

- Arikunto, S. (2012). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Azwar, S. (2014). *Tes Prestasi: "Fungsi dan Pengembangan Pengukuran Prestasi Belajar"*. Edisi II. Yogyakarta: Pustaka Pelajar.
- Boone, W. J., Staver, R. J., & Yal, S. M. (2014). *Rasch Analysis in the Human Sciences*. London: Springer.
- Hofer, S. I. (2017). The test of basic Mechanics Conceptual understanding (bMCU): Using Rasch Analisis to Develop and Evaluate an Efficient Multiple Choice Test on Newton's Mechanics. *International Journal of STEM Education*, 4(18), 1-20.
- Isnaeni, B., Aprilia, N., & Saifuddin, M. F. (2022). Rasch model: Quality of final semester assessment items for class x on biology subject. *Biosfer: Jurnal Pendidikan Biologi*, 15(2), 355–367.
- Kadir, A. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Al-Ta'dib: Jurnal Kajian Ilmu Pendidikan*, 8(2), 70-81.
- Mulyanti, S., Sukmawati, W., & Tarkin, N. E. H. (2022). Development of items in Acid-Base Identification Experiments Using Natural Materials: Validity Test with Rasch Model Analysis. *Phenomenon: Jurnal Pendidikan MIPA*, 12(1), 17–30. <https://doi.org/10.21580/phen.2022.12.1.10703>
- Ratumanan G. T. & Laurens, T. (2011). *Penilaian Hasil Belajar Pada Tingkat Satuan Pendidikan*. Ambon: Unesa University Press.
- Rozali, M. Z., Puteh, S., Yunus, F. A. N., Hamdan, N. H., & Latif, H. F. M. (2022). Reliability and Validity of Instrument on Academic Enhancement Support for Student-Athlete Using Rasch Measurement Model. *Asian Journal of University Education*, 18(1), 290–299. <https://doi.org/10.24191/ajue.v18i1.17199>

- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assasment Pendidikan*. Bandung: Trim Komunikata.
- Tennant, A., Mckenna, S. P., & Hagell, P. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health*, **7**, S22-S26.
- Ubaidillah, M., Marwoto, P., Wiyanto, Rusilowati, A., Subali, B., Mindyarto, B. N., & Isnaeni, W. (2022). Development of Habits of Mind Instruments in the Context of Basic Physics Practicum: EFA and Rasch Model. *Journal of Educational, Cultural and Psychological Studies*, *2022(26)*, 23–49. <https://doi.org/10.7358/ecps-2022-026-ubai>
- Wahyudi, W. (2010). Assesment Pembelajaran Berbasis Portofolio di Sekolah”. *Jurnal Visi Ilmu Pendidikan*, *2(1)*, 288-296.
- Widyaningsih, S. W., & Yusuf, I. (2018). Analisis Soal Modul Laboratorium Fisika Sekolah I Menggunakan Rasch Model. *Gravity: Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, *4(1)*, 33-46.