JIPFRI
https://journal.unuha.ac.id/index.php/JIPFRI

JIPFRI (Jurnal Inovasi Pendidikan Fisika dan Riset Ilmiah)
Vol. 10 No. 2, November 2026, hlm. 206–214
p-ISSN 2549-905X | e-ISSN 2549-9076
https://doi.org/10.30599/jipfri.v10i2.5461

Identifikasi Pemahaman Konsep Fisika Melalui Item Instrumen Tes Materi Usaha dan Energi Menggunakan Rasch Model

Putri Nabila Hudaimi* **dan Maimon Sumo**
Program Studi Pendidikan Fisika, Universitas Islam Madura, Pamekasan, Indonesia
***Penulis korespondensi:** putrinabilahudaimi@uim.ac.id

INFO ARTIKEL

Riwayat artikel:
Diterima: 24/06/2026
Direvisi: 14/11/2026
Disetujui: 14/11/2026
Terbit: 29/11/2026

Kata kunci:
Model Rasch
Pemahaman Konsep Fisika
Usaha dan Energi

Keywords:
Physics Concept Understanding
Rasch Model
Work and Energy

ABSTRAK

Penelitian ini bertujuan menganalisis kualitas butir dan memetakan profil penguasaan konsep peserta didik pada materi Usaha dan Energi menggunakan Model Rasch berbantuan Ministep versi 5.11.0. Metode deskriptif kuantitatif diterapkan pada 34 peserta didik kelas XI yang dipilih melalui purposive sampling. Data dikumpulkan secara daring melalui 25 soal pilihan ganda yang disusun berdasarkan matriks Taksonomi Bloom revisi (C1–C6). Analisis menghasilkan Cronbach's Alpha sebesar 0,78 (cukup), Person Reliability sebesar 0,68 (cukup; separation 1,47), dan Item Reliability sebesar 0,19 (lemah; separation 0,49). Analisis kesesuaian butir menunjukkan bahwa butir 23 (C5M) tidak memenuhi ketiga kriteria kesesuaian dengan korelasi poin-ukur yang sangat rendah (0,06), sedangkan butir 3 (C2K), 12 (C3K), dan 20 (C4K) melampaui batas atas Outfit Mean Square (> 1,5). Wright Map menunjukkan bahwa rata-rata kemampuan peserta didik (+1,29 logit) melebihi rata-rata kesulitan butir (0,00 logit), sehingga instrumen tergolong relatif mudah. Selanjutnya, butir pemahaman konseptual nomor 4 (C2K) merupakan butir tersulit (+1,00 logit), sedangkan butir nomor 5 (C1F), 21 (C4K), dan 17 (C4P) memiliki tingkat kesulitan terendah (−0,81 logit). Temuan ini menunjukkan perlunya mengganti butir 23, memperbaiki pengecoh pada butir 3, 12, dan 20, serta melakukan pengujian lanjutan dengan sekurang-kurangnya 100 responden.

ABSTRACT

This study aims to analyze item quality and map students' conceptual mastery profiles on the topic of Work and Energy using the Rasch Model assisted by Ministep version 5.11.0. A quantitative descriptive method was employed involving 34 eleventh-grade students selected through purposive sampling. Data were collected online via 25 multiple-choice questions constructed based on the revised Bloom's Taxonomy matrix (C1–C6). The analysis yielded a Cronbach's Alpha of 0.78 (fair), a Person Reliability of 0.68 (fair; separation 1.47), and an Item Reliability of 0.19 (weak; separation 0.49). Item fit analysis indicated that item 23 (C5M) failed all three fit criteria with a very low point-measure correlation (0.06), whereas items 3 (C2K), 12 (C3K), and 20 (C4K) exceeded the upper Outfit Mean Square threshold (> 1.5). The Wright Map revealed that mean student ability (+1.29 logit) exceeded mean item difficulty (0.00 logit), indicating a relatively easy instrument. Furthermore, conceptual understanding item 4 (C2K) was the most difficult (+1.00 logit), while items 5 (C1F), 21 (C4K), and 17 (C4P) had the lowest difficulty (-0.81 logit). These findings highlight the need to replace item 23, refine distractors in items 3, 12, and 20, and conduct follow-up testing with at least 100 respondents.

Cara mengutip: Hudaimi, P. N., & Sumo, M. (2026). Identifikasi pemahaman konsep fisika melalui item instrumen tes materi usaha dan energi menggunakan Rasch Model. *JIPFRI (Jurnal Inovasi Pendidikan Fisika dan Riset Ilmiah), 10*(2), 206–214. https://doi.org/10.30599/jipfri.v10i2.5461
Copyright © 2026 Penulis. Artikel akses terbuka di bawah lisensi Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

PENDAHULUAN

Penilaian dalam dunia pendidikan memiliki peran krusial dalam mengukur tingkat penguasaan konsep dan perkembangan kognitif peserta didik secara objektif (Black & Wiliam, 2018). Instrumen tes yang berkualitas harus memenuhi standar validitas dan reliabilitas psikometrik agar hasil pengukuran mencerminkan kemampuan peserta didik yang sebenarnya (Boone, 2016; Messick, 1995; Samsudin et al., 2016). Namun, dalam praktiknya, pengembangan instrumen evaluasi fisika masih sering bertumpu pada pendekatan *Classical Test Theory* (CTT). Keterbatasan utama CTT terletak pada ketergantungan parameter butir soal terhadap sampel (*sample-dependent*) serta ketidakmampuannya memberikan informasi mendalam mengenai interaksi spesifik antara tingkat kemampuan individu dan tingkat kesulitan

butir soal (Arfiani et al., 2025; Noorkholisoh et al., 2024).

Sebagai solusi atas keterbatasan tersebut, Model Rasch yang merupakan bagian dari *Item Response Theory* (IRT) hadir dengan mentransformasi data mentah dikotomi menjadi skala interval logit yang linear (Boone, 2016; Silaban et al., 2025). Berbagai studi terdahulu membuktikan bahwa penerapan Model Rasch mampu meningkatkan kualitas diagnostik instrumen tes dan memetakan karakteristik butir secara independen (Avinç & Doğan, 2024; Nurani et al., 2025). Meskipun demikian, sebagian besar kajian pengembangan instrumen fisika dengan Model Rasch masih berfokus pada uji kelayakan dasar. Pengembangan instrumen berbasis asesmen digital untuk memetakan profil pemahaman konsep fisika yang dikaitkan langsung dengan hierarki proses kognitif dan dimensi pengetahuan (Anderson & Krathwohl, 2001; Istiyono et al., 2014; Wilson, 2016) masih perlu dieksplorasi lebih lanjut, khususnya pada materi Usaha dan Energi yang kerap dianggap abstrak (Alatas et al., 2024; Rasyid et al., 2025).

Kebaruan penelitian ini terletak pada eksplorasi empiris terhadap kesesuaian tingkat kesulitan butir soal materi Usaha dan Energi yang disusun berdasarkan matriks dimensi pengetahuan (faktual, konseptual, prosedural, dan metakognitif) serta proses kognitif Taksonomi Bloom revisi (C1–C6) (Anderson & Krathwohl, 2001; Wilson, 2016). Berbeda dengan penelitian sebelumnya yang sekadar menguji reliabilitas instrumen secara umum, studi ini memadukan analisis kesesuaian butir (*item fit*) dan pemetaan *Wright Map* pada Model Rasch untuk mengevaluasi konsistensi sebaran level kognitif teoretis terhadap tingkat kesulitan empiris yang dihadapi peserta didik selama asesmen digital berlangsung (Planinic et al., 2019).

Berdasarkan uraian tersebut, penelitian ini difokuskan untuk menjawab tiga rumusan masalah utama: (1) bagaimana reliabilitas, indeks pemisahan (*separation*), dan kesesuaian butir (*item fit*) instrumen tes pemahaman konsep pada materi Usaha dan Energi berdasarkan analisis Model Rasch; (2) bagaimana peta distribusi tingkat kemampuan peserta didik dibandingkan dengan tingkat kesulitan butir soal melalui *Wright Map*; serta (3) bagaimana keselarasan urutan tingkat kesulitan empiris butir soal terhadap hierarki proses kognitif Taksonomi Bloom revisi (C1–C6).

METODE

Metode yang digunakan dalam penelitian ini adalah deskriptif kuantitatif untuk menguraikan profil penguasaan konsep fisika peserta didik secara sistematis, faktual, dan akurat. Fokus utama penelitian adalah mengevaluasi kualitas psikometrik butir soal dan memetakan kemampuan kognitif peserta didik melalui pendekatan *Item Response Theory* (IRT) dengan pemodelan Rasch berbantuan aplikasi Ministep versi 5.11.0. Subjek penelitian melibatkan 34 peserta didik (N = 34) jenjang kelas XI fase F di SMAN 4 Pamekasan pada tahun ajaran 2025/2026 yang dipilih menggunakan teknik *purposive sampling*. Kriteria pemilihan sampel meliputi: (1) peserta didik yang telah menempuh pembelajaran materi Usaha dan Energi secara tuntas pada semester berjalan, (2) memiliki akses perangkat digital untuk mengikuti asesmen daring, serta (3) berasal dari kelas dengan variasi kemampuan akademik yang heterogen berdasarkan rekomendasi guru fisika. Ukuran sampel N = 34 untuk 25 butir soal tergolong terbatas bagi stabilitas estimasi kalibrasi parameter item; menurut Linacre (1994), sampel sekitar 30 responden hanya memberikan kestabilan estimasi kesulitan butir pada kisaran ±1 logit (tingkat kepercayaan 95%), sedangkan kestabilan tinggi pada kisaran ±0,5 logit memerlukan sekurang-kurangnya 100 responden (Boone, 2016; Linacre, 1994). Oleh karena itu, kajian ini diposisikan sebagai studi diagnostik awal dengan keterbatasan ukuran sampel yang diperhitungkan secara eksplisit dalam interpretasi hasil.

Instrumen yang digunakan berupa tes objektif pilihan ganda sebanyak 25 butir soal pada materi Usaha dan Energi yang didistribusikan melalui platform asesmen digital. Butir soal dirancang berdasarkan matriks dua dimensi Taksonomi Bloom revisi yang memadukan dimensi pengetahuan (faktual, konseptual, prosedural, metakognitif) dan proses kognitif mulai dari mengingat (C1), memahami (C2), menerapkan (C3), menganalisis (C4), mengevaluasi (C5), hingga mencipta (C6) (Anderson & Krathwohl, 2001; Wilson, 2016) yang mencakup enam sub-konsep utama materi Usaha dan Energi, yaitu: (1) konsep usaha oleh gaya konstan, (2) energi kinetik,

(3) energi potensial gravitasi dan pegas, (4) teorema usaha-energi, (5) hukum kekekalan energi mekanik, dan (6) daya. Perlu dicatat bahwa pemetaan level kognitif tiap butir disusun berdasarkan kesesuaian indikator pembelajaran dengan matriks Taksonomi Bloom revisi, namun belum melalui uji validasi isi kuantitatif oleh panel ahli (seperti indeks Aiken's V atau *Content Validity Ratio*), sehingga menjadi pertimbangan dalam menafsirkan hierarki kognitif butir. Rincian sub-konsep, indikator soal, dan sebaran level kognitif disajikan pada Tabel 1.

Tabel 1. Sub-konsep, indikator soal, dan level kognitif instrumen tes Usaha dan Energi

No	Sub-Konsep	Level	Indikator Soal
1	Konsep Usaha	(C1) Konseptual	Menjelaskan definisi fisis usaha berdasarkan hubungan gaya dan perpindahan
2	Konsep Usaha	(C2) Konseptual	Menafsirkan grafik hubungan gaya (F) terhadap perpindahan (s)
3	Energi Kinetik	(C2) Konseptual	Menjelaskan pengaruh perubahan kelajuan terhadap besar energi kinetik
4	Konsep Usaha	(C2) Konseptual	Menjelaskan syarat usaha bernilai nol pada gaya tegak lurus perpindahan
5	Energi Potensial	(C1) Faktual	Mengingat besaran dan satuan energi potensial gravitasi
6	Konsep Usaha	(C3) Konseptual	Menerapkan persamaan usaha oleh gaya membentuk sudut tertentu
7	Teorema Usaha-Energi	(C3) Prosedural	Menerapkan hubungan usaha total dengan perubahan energi kinetik
8	Energi Potensial	(C3) Prosedural	Menerapkan persamaan energi potensial gravitasi pada ketinggian tertentu
9	Daya	(C3) Prosedural	Menerapkan konsep daya mekanik dalam menghitung laju usaha
10	Kekekalan Energi	(C3) Konseptual	Menerapkan hukum kekekalan energi mekanik pada gerak jatuh bebas
11	Energi Potensial Pegas	(C3) Prosedural	Menerapkan persamaan energi potensial elastis pada sistem pegas
12	Teorema Usaha-Energi	(C3) Konseptual	Menerapkan teorema usaha-energi pada pengereman benda hingga berhenti
13	Kekekalan Energi	(C3) Konseptual	Menerapkan kekekalan energi mekanik pada lintasan melengkung licin
14	Kekekalan Energi	(C4) Konseptual	Menganalisis perubahan energi kinetik dan potensial pada bidang miring kasar
15	Teorema Usaha-Energi	(C4) Konseptual	Menganalisis hubungan gaya gesek dengan hilangnya energi mekanik sistem
16	Konsep Usaha	(C4) Konseptual	Menganalisis perbandingan usaha beberapa gaya pada lintasan berbeda
17	Kekekalan Energi	(C4) Prosedural	Menganalisis kelajuan benda pada titik terendah lintasan ayunan bandul
18	Energi Kinetik	(C4) Konseptual	Menganalisis perbandingan energi kinetik dua benda berbeda massa dan momentum
19	Daya	(C4) Konseptual	Menganalisis efisiensi dan daya keluaran mesin pengangkat beban vertikal
20	Energi Potensial Pegas	(C4) Konseptual	Menganalisis perubahan energi potensial pegas saat konstanta dan simpangan diubah
21	Teorema Usaha-Energi	(C4) Konseptual	Menganalisis jarak henti kendaraan berdasarkan variasi kelajuan awal dan gesekan
22	Kekekalan Energi	(C5) Konseptual	Mengevaluasi penerapan hukum kekekalan energi mekanik pada sistem non-konservatif
23	Teorema Usaha-Energi	(C5) Metakognitif	Mengevaluasi ketepatan strategi pemecahan masalah pada kasus gaya gesek berubah
24	Kekekalan Energi	(C6) Metakognitif	Merancang skenario lintasan roller coaster agar memenuhi ketinggian minimum
25	Daya & Efisiensi	(C5) Metakognitif	Mengevaluasi kelayakan desain sistem konversi energi berdasarkan efisiensi daya

Data respons peserta didik dianalisis menggunakan Model Rasch berdasarkan tiga parameter utama: (1) Reliabilitas *person*, *item*, indeks pemisahan (*separation*), dan *Cronbach's Alpha*, untuk mengukur konsistensi jawaban responden, daya pemisah tingkat kesulitan butir, dan keandalan instrumen; (2) Kesesuaian butir (*Item Fit*), ditinjau melalui tiga kriteria berdasarkan Sumintono dan Widhiarso (2013) serta Boone et al. (2014), yaitu *Outfit Mean Square* (MNSQ) pada rentang 0,5 < MNSQ < 1,5, *Outfit Z-Standard* (ZSTD) pada rentang −2,0 < ZSTD < +2,0, serta *Point Measure Correlation* (*Pt-Measure Corr*) pada rentang 0,32 < *Pt-Measure Corr* < 0,85 (Bond &

Fox, 2015; Sumintono & Widhiarso, 2013). Dalam pengambilan keputusan kualitas butir, diterapkan aturan berjenjang: butir dinyatakan *misfit* kuat dan wajib diganti atau direvisi total apabila melanggar ketiga kriteria sekaligus (atau memiliki *Pt-Measure Corr* mendekati nol/negatif), sedangkan butir yang hanya melanggar salah satu kriteria (misalnya *Outfit* MNSQ sedikit di atas 1,5 dengan ZSTD tetap dalam batas toleransi) dikategorikan sebagai *misfit* ringan yang memerlukan perbaikan redaksi atau pengecoh; dan (3) *Wright Map*, digunakan untuk memvisualisasikan perbandingan antara tingkat kemampuan peserta didik dan tingkat kesulitan butir soal pada skala logit yang sama. Interpretasi kategori reliabilitas merujuk pada kriteria Sumintono dan Widhiarso (2013) pada Tabel 2.

Tabel 2. Interpretasi nilai reliabilitas (Sumintono & Widhiarso, 2013)

Nilai Reliabilitas	Interpretasi
> 0,94	Istimewa
0,91 – 0,94	Bagus sekali
0,81 – 0,90	Bagus
0,67 – 0,80	Cukup
< 0,67	Lemah

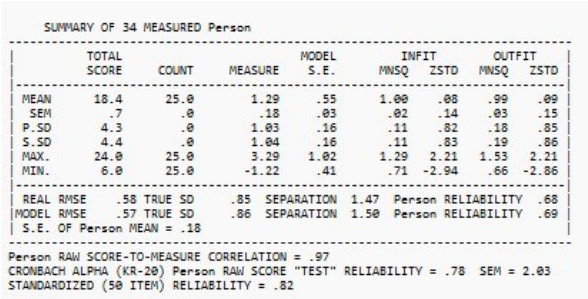
HASIL DAN PEMBAHASAN

Hasil

Analisis data hasil pengukuran penguasaan konsep peserta didik pada materi Usaha dan Energi dilakukan melalui parameter objektif Model Rasch berbantuan aplikasi Ministep versi 5.11.0. Berdasarkan ringkasan statistik yang diperoleh, nilai *Cronbach's Alpha* (KR-20) menunjukkan angka sebesar 0,78 (kategori cukup). Jika ditinjau secara terpisah, nilai *Person Reliability* berada pada angka 0,68 (*Real RMSE*) atau 0,69 (*Model RMSE*) dengan *Person Separation* sebesar 1,47 (*Real RMSE*) atau 1,50 (*Model RMSE*). Sementara itu, nilai *Item Reliability* berada pada angka 0,19 (*Real RMSE*) atau 0,25 (*Model RMSE*) dengan indeks *Item Separation* sebesar 0,49 (*Real RMSE*) atau 0,58 (*Model RMSE*). Ringkasan statistik parameter *person* dan *item* disajikan pada Tabel 3 serta ditampilkan secara visual pada Gambar 1 dan Gambar 2.

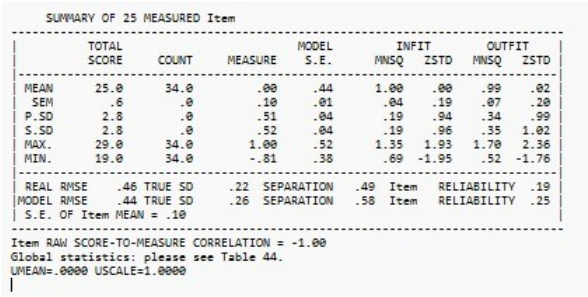
Tabel 3. Ringkasan statistik pengukuran Person dan Item

Parameter Statistik	Person (N = 34)	Item (25 Butir)
Mean Total Score	18,4 (dari 25)	25,0 (dari 34)
Mean Measure (Logit)	+1,29	0,00
Standard Deviation (P.SD)	1,03	0,51
Mean Model S.E. (Logit)	0,55	0,44
Rentang Measure	−1,22 s.d. +3,29	−0,81 s.d. +1,00
Mean Infit MNSQ / ZSTD	1,00 / 0,08	1,00 / 0,00
Mean Outfit MNSQ / ZSTD	0,99 / 0,09	0,99 / 0,02
Separation (Real / Model)	1,47 / 1,50	0,49 / 0,58
Reliability (Real / Model)	0,68 / 0,69	0,19 / 0,25
Cronbach's Alpha (KR-20)	0,78 (Cukup)	—



```
SUMMARY OF 34 MEASURED Person
-----------------------------------------------------------------------------
|          TOTAL                        MODEL        INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|---------------------------------------------------------------------------|
| MEAN      18.4      25.0        1.29     .55      1.00    .08    .99    .09 |
|  SEM        .7        .0         .18     .03       .02    .14    .03    .15 |
| P.SD       4.3        .0        1.03     .16       .11    .82    .18    .85 |
| S.SD       4.4        .0        1.04     .16       .11    .83    .19    .86 |
| MAX.      24.0      25.0        3.29    1.02      1.29   2.21   1.53   2.21 |
| MIN.       6.0      25.0       -1.22     .41       .71  -2.94    .66  -2.86 |
|---------------------------------------------------------------------------|
| REAL RMSE    .58 TRUE SD     .85  SEPARATION  1.47  Person RELIABILITY  .68 |
|MODEL RMSE    .57 TRUE SD     .86  SEPARATION  1.50  Person RELIABILITY  .69 |
| S.E. OF Person MEAN = .18                                                 |
-----------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .97
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78  SEM = 2.03
STANDARDIZED (50 ITEM) RELIABILITY = .82
```

Gambar 1. Ringkasan statistik reliabilitas person



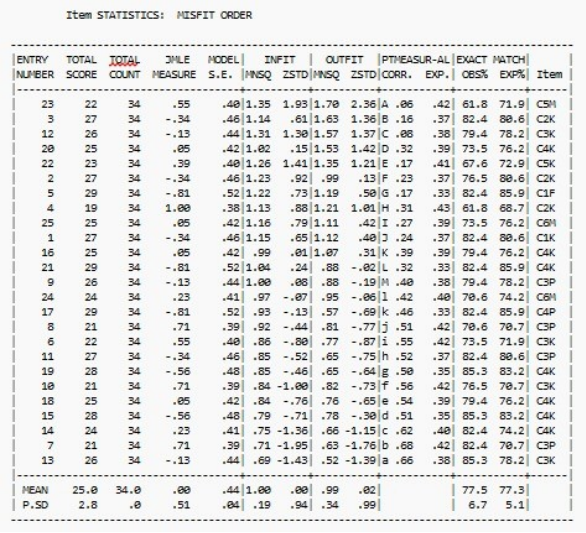
```
SUMMARY OF 25 MEASURED Item
-----------------------------------------------------------------------------
|          TOTAL                        MODEL        INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|---------------------------------------------------------------------------|
| MEAN      25.0      34.0         .00     .44      1.00    .00    .99    .02 |
|  SEM        .6        .0         .10     .01       .04    .19    .07    .20 |
| P.SD       2.8        .0         .51     .04       .19    .94    .34    .99 |
| S.SD       2.8        .0         .52     .04       .19    .96    .35   1.02 |
| MAX.      29.0      34.0        1.00     .52      1.35   1.93   1.70   2.36 |
| MIN.      19.0      34.0        -.81     .38       .69  -1.95    .52  -1.76 |
|---------------------------------------------------------------------------|
| REAL RMSE    .46 TRUE SD     .22  SEPARATION   .49  Item   RELIABILITY  .19 |
|MODEL RMSE    .44 TRUE SD     .26  SEPARATION   .58  Item   RELIABILITY  .25 |
| S.E. OF Item MEAN = .10                                                   |
-----------------------------------------------------------------------------
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Gambar 2. Ringkasan statistik reliabilitas item

Tabel 3 menunjukkan bahwa rata-rata kemampuan peserta didik berada pada +1,29 logit (*P.SD* = 1,03), dengan nilai minimum −1,22 logit (skor 6/25; 24%) dan maksimum +3,29 logit (skor 24/25; 96%), serta tidak ditemukan peserta didik dengan skor sempurna 25/25 (0%) maupun skor nol 0/25 (0%). Di sisi lain, tingkat kesulitan butir soal dikalibrasi pada rata-rata 0,00 logit (rentang −0,81 hingga +1,00 logit) dengan rata-rata kesalahan baku pengukuran (*Model S.E.*) sebesar 0,44 logit (rentang 0,38 hingga 0,52 logit tiap butir). Hasil kalibrasi dan kesesuaian butir (*item fit*) lengkap untuk ke-25 soal disajikan pada Tabel 4 dan Gambar 3.

Tabel 4. Statistik kalibrasi dan kesesuaian butir lengkap 25 soal (urutan misfit)

No	Kode	Skor	Logit	SE	Outfit MNSQ	Outfit ZSTD	PTMEA	Status
23	C5M	22	+0,55	0,40	1,70*	2,36*	0,06*	Misfit Kuat
3	C2K	27	−0,34	0,46	1,63*	1,36	0,16*	Misfit Ringan
12	C3K	26	−0,13	0,44	1,57*	1,37	0,08*	Misfit Ringan
20	C4K	25	+0,05	0,42	1,53*	1,42	0,32	Misfit Ringan
22	C5K	23	+0,39	0,40	1,35	1,21	0,17*	PTMEA < 0,32
2	C2K	27	−0,34	0,46	0,99	0,13	0,23*	PTMEA < 0,32
5	C1F	29	−0,81	0,52	1,19	0,50	0,17*	PTMEA < 0,32
4	C2K	19	+1,00	0,38	1,21	1,01	0,31*	PTMEA < 0,32
25	C5M	25	+0,05	0,42	1,11	0,42	0,27*	PTMEA < 0,32
1	C1K	27	−0,34	0,46	1,12	0,40	0,24*	PTMEA < 0,32
16	C4K	25	+0,05	0,42	1,07	0,31	0,39	Fit
21	C4K	29	−0,81	0,52	0,88	−0,02	0,32	Fit
9	C3P	26	−0,13	0,44	0,88	−0,19	0,40	Fit
24	C6M	24	+0,23	0,41	0,95	−0,06	0,42	Fit
17	C4P	29	−0,81	0,52	0,57	−0,69	0,46	Fit
8	C3P	21	+0,71	0,39	0,81	−0,77	0,51	Fit
6	C3K	22	+0,55	0,40	0,77	−0,87	0,55	Fit
11	C3P	27	−0,34	0,46	0,65	−0,75	0,52	Fit
19	C4K	28	−0,56	0,48	0,65	−0,64	0,50	Fit
10	C3K	21	+0,71	0,39	0,82	−0,73	0,56	Fit
18	C4K	25	+0,05	0,42	0,76	−0,65	0,54	Fit
15	C4K	28	−0,56	0,48	0,78	−0,30	0,51	Fit
14	C4K	24	+0,23	0,41	0,66	−1,15	0,62	Fit
7	C3P	21	+0,71	0,39	0,63	−1,76	0,68	Fit
13	C3K	26	−0,13	0,44	0,52	−1,39	0,66	Fit



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.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
23	22	34	.55	.40	1.35	1.93	1.70	2.36	A .06	.42	61.8	71.9	C5M
3	27	34	-.34	.46	1.14	.61	1.63	1.36	B .16	.37	82.4	80.6	C2K
12	26	34	-.13	.44	1.31	1.30	1.57	1.37	C .08	.38	79.4	78.2	C3K
20	25	34	.05	.42	1.02	.15	1.53	1.42	D .32	.39	73.5	76.2	C4K
22	23	34	.39	.40	1.26	1.41	1.35	1.21	E .17	.41	67.6	72.9	C5K
2	27	34	-.34	.46	1.23	.92	.99	.13	F .23	.37	76.5	80.6	C2K
5	29	34	-.81	.52	1.22	.73	1.19	.50	G .17	.33	82.4	85.9	C1F
4	19	34	1.00	.38	1.13	.88	1.21	1.01	H .31	.43	61.8	68.7	C2K
25	25	34	.05	.42	1.16	.79	1.11	.42	I .27	.39	73.5	76.2	C6M
1	27	34	-.34	.46	1.15	.65	1.12	.40	J .24	.37	82.4	80.6	C1K
16	25	34	.05	.42	.99	.01	1.07	.31	K .39	.39	79.4	76.2	C4K
21	29	34	-.81	.52	1.04	.24	.88	-.02	L .32	.33	82.4	85.9	C4K
9	26	34	-.13	.44	1.00	.08	.88	-.19	M .40	.38	79.4	78.2	C3P
24	24	34	.23	.41	.97	-.07	.95	-.06	l .42	.40	70.6	74.2	C6M
17	29	34	-.81	.52	.93	-.13	.57	-.69	k .46	.33	82.4	85.9	C4P
8	21	34	.71	.39	.92	-.44	.81	-.77	j .51	.42	70.6	70.7	C3P
6	22	34	.55	.40	.86	-.80	.77	-.87	i .55	.42	73.5	71.9	C3K
11	27	34	-.34	.46	.85	-.52	.65	-.75	h .52	.37	82.4	80.6	C3P
19	28	34	-.56	.48	.85	-.46	.65	-.64	g .50	.35	85.3	83.2	C4K
10	21	34	.71	.39	.84	-1.00	.82	-.73	f .56	.42	76.5	70.7	C3K
18	25	34	.05	.42	.84	-.76	.76	-.65	e .54	.39	79.4	76.2	C4K
15	28	34	-.56	.48	.79	-.71	.78	-.30	d .51	.35	85.3	83.2	C4K
14	24	34	.23	.41	.75	-1.36	.66	-1.15	c .62	.40	82.4	74.2	C4K
7	21	34	.71	.39	.71	-1.95	.63	-1.76	b .68	.42	82.4	70.7	C3P
13	26	34	-.13	.44	.69	-1.43	.52	-1.39	a .66	.38	85.3	78.2	C3K
MEAN	25.0	34.0	.00	.44	1.00	.00	.99	.02			77.5	77.3	
P.SD	2.8	.0	.51	.04	.19	.94	.34	.99			6.7	5.1	

Gambar 3. Peta kesesuaian butir (Item Misfit Order)

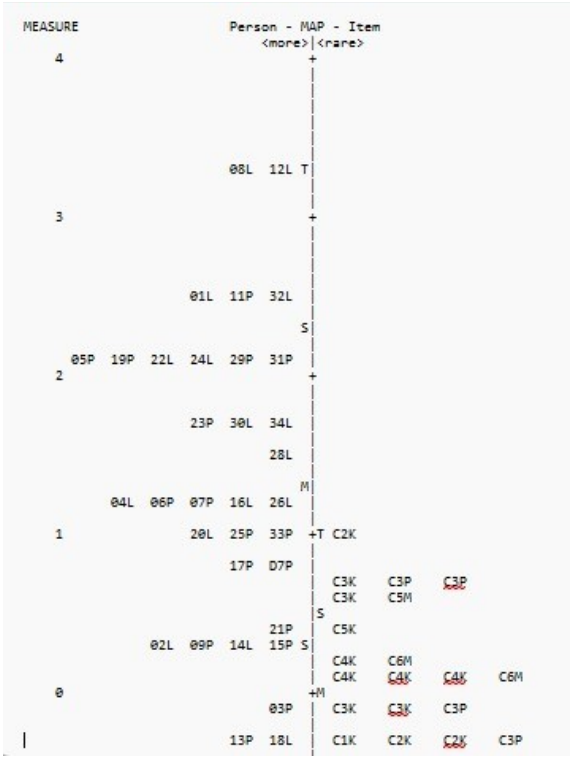
Berdasarkan Tabel 4 dan Gambar 3, butir soal nomor 23 (C5M) terdeteksi sebagai satu-satunya item *misfit* kuat karena gagal memenuhi ketiga kriteria kesesuaian sekaligus (*Outfit* MNSQ = 1,70 > 1,5; *Outfit* ZSTD = 2,36 > +2,0; dan *Pt-Measure Corr* = 0,06 < 0,32). Tiga butir melampaui batas atas *Outfit* MNSQ (> 1,5) dengan nilai ZSTD tetap dalam batas toleransi (< +2,0), yakni nomor 3 (C2K; MNSQ = 1,63; PTMEA = 0,16), nomor 12 (C3K; MNSQ = 1,57; PTMEA = 0,08), dan nomor 20 (C4K; MNSQ = 1,53; PTMEA = 0,32), sehingga dikategorikan sebagai *misfit* ringan. Selain itu, terdapat 6 butir lain (nomor 22, 2, 5, 4, 25, dan 1) yang memiliki nilai MNSQ dan ZSTD ideal tetapi nilai *Pt-Measure Corr* di bawah 0,32 (berkisar 0,17–0,31). Jika batas kriteria *Pt-Measure Corr* ≥ 0,32 diterapkan secara kaku tanpa mempertimbangkan ukuran sampel, sebanyak 9 dari 25 butir (36%) memiliki daya beda rendah, sementara 15 butir lainnya (60%) memenuhi seluruh kriteria *fit* secara utuh.

Person STATISTICS: MISFIT ORDER

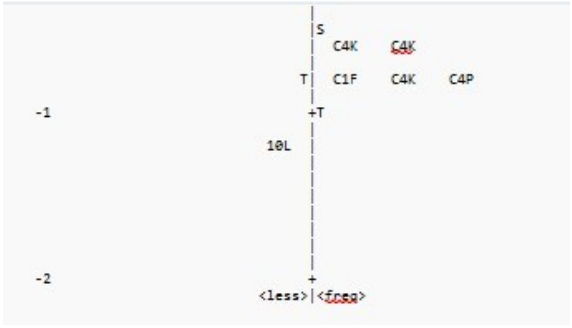
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
31	22	25	2.09	.62	1.13	.42	1.53	1.02	A	-.22	.16	88.0	88.0	31P
27	17	25	.80	.44	1.29	1.60	1.43	1.96	B	-.38	.23	64.0	68.4	D7P
2	14	25	.25	.42	1.25	2.21	1.28	2.21	C	-.25	.24	44.0	61.5	02L
7	19	25	1.22	.48	1.06	.31	1.12	.48	D	.07	.21	76.0	76.0	07P
9	14	25	.25	.42	1.10	.96	1.11	.95	E	.04	.24	52.0	61.5	09P
22	22	25	2.09	.62	1.00	.14	1.11	.37	F	.12	.16	88.0	88.0	22L
13	11	25	-.26	.42	1.10	.95	1.10	.89	G	.05	.24	60.0	61.0	13P
33	18	25	1.00	.46	1.10	.55	1.09	.45	H	.04	.22	68.0	72.0	33P
23	21	25	1.74	.55	.96	.02	1.08	.33	I	.20	.18	84.0	84.0	23P
21	15	25	.43	.42	1.06	.50	1.07	.52	J	.13	.24	60.0	63.2	21P
12	24	25	3.29	1.02	1.02	.33	1.06	.41	K	.02	.10	96.0	96.0	12L
11	23	25	2.55	.74	1.00	.21	1.05	.30	L	.11	.14	92.0	92.0	11P
29	22	25	2.09	.62	1.05	.25	1.05	.27	M	.07	.16	88.0	88.0	29P
18	11	25	-.26	.42	1.03	.31	1.01	.16	N	.19	.24	60.0	61.0	18L
24	22	25	2.09	.62	1.03	.21	1.02	.21	O	.11	.16	88.0	88.0	24L
6	19	25	1.22	.48	1.00	.07	1.02	.18	P	.20	.21	76.0	76.0	06P
26	19	25	1.22	.48	1.02	.16	.97	-.01	Q	.20	.21	76.0	76.0	26L
1	23	25	2.55	.74	1.01	.23	.91	.11	q	.13	.14	92.0	92.0	01L
5	22	25	2.09	.62	1.01	.18	.98	.13	p	.15	.16	88.0	88.0	05P
32	23	25	2.55	.74	1.01	.22	.96	.18	o	.12	.14	92.0	92.0	32L
4	19	25	1.22	.48	1.00	.08	.99	.06	n	.21	.21	76.0	76.0	04L
28	20	25	1.46	.51	1.00	.09	.90	-.17	m	.24	.20	80.0	80.0	28L
8	24	25	3.29	1.02	.98	.28	.66	-.01	l	.22	.10	96.0	96.0	08L
10	6	25	-1.22	.48	.91	-.32	.98	.04	k	.35	.21	76.0	75.9	10L
30	21	25	1.74	.55	.98	.07	.87	-.17	j	.26	.18	84.0	84.0	30L
34	21	25	1.74	.55	.98	.07	.88	-.14	i	.25	.18	84.0	84.0	34L
3	12	25	-.09	.41	.95	-.49	.96	-.36	h	.33	.24	68.0	60.3	03P
19	22	25	2.09	.62	.91	-.05	.72	-.39	g	.39	.16	88.0	88.0	19P
20	18	25	1.00	.46	.89	-.50	.84	-.60	f	.45	.22	76.0	72.0	20L
25	18	25	1.00	.46	.89	-.50	.84	-.63	e	.45	.22	76.0	72.0	25P
15	14	25	.25	.42	.88	-1.09	.88	-.98	d	.46	.24	68.0	61.5	15P
17	17	25	.80	.44	.85	-.86	.86	-.64	c	.50	.23	72.0	68.4	17P
16	19	25	1.22	.48	.78	-.88	.68	-1.12	b	.65	.21	76.0	76.0	16L
14	14	25	.25	.42	.71	-2.94	.70	-2.86	a	.79	.24	84.0	61.5	14L
MEAN	18.4	25.0	1.29	.55	1.00	.08	.99	.09				77.5	77.3	
P.SD	4.3	.0	1.03	.16	.11	.82	.18	.85				12.5	11.4	

Gambar 4. Peta kesesuaian responden (Person Misfit Order)

Data visual pada Gambar 4 mengindikasikan siswa dengan kode 31P (skor 22/25; *measure* +2,09 logit; *Outfit* MNSQ = 1,53; ZSTD = 1,02; *Pt-Measure Corr* = −0,22), kode D7P (skor 17/25; *measure* +0,80 logit; *Outfit* MNSQ = 1,43; ZSTD = 1,96; *Pt-Measure Corr* = −0,38), dan kode 02L (skor 14/25; *measure* +0,25 logit; *Outfit* MNSQ = 1,28; ZSTD = 2,21; *Pt-Measure Corr* = −0,25) berada pada urutan teratas yang mendeteksi adanya pola jawaban tidak biasa (*unusual pattern*). Meskipun nilai *Outfit* MNSQ pada D7P dan 02L masih di bawah 1,5 (hanya 31P yang sedikit melampaui batas 1,5), korelasi poin-ukur ketiganya yang bernilai negatif mengindikasikan respons yang tidak konsisten, seperti siswa berkemampuan tinggi (31P) keliru menjawab soal mudah akibat kurang teliti, atau siswa berkemampuan sedang/rendah cenderung menebak (*guessing*) dengan benar pada soal sulit selama pelaksanaan tes daring. Ketiga responden tersebut tetap dipertahankan dalam analisis karena jumlahnya kecil (8,8% dari total sampel), pengeluarannya justru akan memperkecil ukuran sampel estimasi, serta bertujuan memotret kondisi alami respons kelas pada tahap studi diagnostik awal. Rekonstruksi interaksi kemampuan peserta didik dan kesulitan soal disajikan pada Gambar 5 dan Gambar 6.



Gambar 5. Peta distribusi kemampuan person dan kesulitan item (Wright Map) bagian atas



Gambar 6. Peta distribusi kemampuan person dan kesulitan item (Wright Map) bagian bawah

Pembahasan

Berdasarkan hasil analisis statistik deskriptif menggunakan Model Rasch, indeks *Cronbach's Alpha* sebesar 0,78 menunjukkan bahwa instrumen tes digital yang dikembangkan memiliki reliabilitas interaksi keseluruhan dalam kategori cukup (Sumintono & Widhiarso, 2013). Namun, terdapat disparitas yang mencolok antara *Person Reliability* (0,68; kategori cukup) dan *Item Reliability* (0,19; kategori lemah) dengan *Item Separation* sebesar 0,49. Rendahnya reliabilitas item ini terutama dipengaruhi oleh faktor keterbatasan ukuran sampel (*N* = 34), yang menyebabkan tingginya kesalahan baku pengukuran (*Model S.E.* rata-rata 0,44 logit) sehingga kalibrasi tingkat kesulitan antarbutir belum stabil, di samping rentang sebaran

tingkat kesulitan soal yang relatif sempit (−0,81 hingga +1,00 logit). Sebagaimana dijelaskan oleh Boone (2016) serta Linacre (1994), ukuran sampel yang kecil dalam analisis Rasch menghasilkan estimasi parameter item yang kurang presisi dan menekan indeks pemisahan serta reliabilitas item, sehingga diperlukan ukuran sampel yang lebih besar (minimal 100–250 responden) dan bank soal dengan variasi tingkat kesulitan yang lebih heterogen agar kalibrasi item menjadi stabil. Temuan ini sejalan dengan studi serupa oleh Rasyid et al. (2025) yang mengembangkan tes pemahaman konsep fisika pada materi gelombang menggunakan Model Rasch; dalam penelitian tersebut, penggunaan sampel yang jauh lebih besar dengan sebaran tingkat kesulitan butir yang terkalibrasi secara luas mampu menghasilkan indeks *Item Reliability* yang sangat tinggi (kategori istimewa), berbeda dengan studi diagnostik awal bersampel terbatas yang masih menghasilkan estimasi parameter butir kurang stabil.

Analisis *Wright Map* pada Gambar 5 menunjukkan bahwa rata-rata kemampuan peserta didik (+1,29 logit) berada lebih tinggi dibandingkan rata-rata tingkat kesulitan soal (0,00 logit), yang mengindikasikan paket instrumen tes cenderung relatif mudah bagi kelompok responden. Berdasarkan distribusi skor mentah, tidak ditemukan peserta didik yang memperoleh skor sempurna (25 dari 25; 0%) maupun skor nol (0 dari 25; 0%). Namun, terdapat 2 peserta didik (5,9%, yakni kode 08L dan 12L) yang mencapai skor ekstrem tinggi sebesar 24 dari 25 (96%; +3,29 logit), serta 9 peserta didik (26,5%) yang berada pada rentang logit +2,00 hingga +3,29 di area atas tanpa diimbangi keberadaan butir soal di sisi kanan, sementara 1 peserta didik (2,9%, kode 10L) berada pada skor terendah sebesar 6 dari 25 (24%; −1,22 logit). Kondisi ini mengindikasikan kurangnya butir soal yang menantang (*high-order thinking items*) pada level mengevaluasi (C5) dan mencipta (C6) sesuai Taksonomi Bloom revisi (Anderson & Krath-wohl, 2001; Wilson, 2016). Meskipun demikian, mengingat nilai *Item Reliability* masih tergolong lemah (0,19) sehingga urutan dan posisi kalibrasi kesulitan butir belum sepenuhnya stabil, perbedaan posisi logit antara kemampuan peserta didik dan kesulitan soal tersebut dimaknai sebagai indikasi deskriptif awal (bukan kesimpulan signifikansi statistik) yang memerlukan konfirmasi lebih lanjut pada sampel yang lebih besar.

Temuan paling menarik dalam penelitian ini adalah terungkapnya ketidaksesuaian pada hierarki taksonomi kognitif teoretis. Secara ideal, tingkat kesulitan soal berpola linier mendaki dari C1 hingga C6 (Anderson & Krathwohl, 2001). Namun secara empiris, butir soal nomor 4 dengan kode level kognitif rendah (C2K/Pemahaman Konseptual) justru menempati posisi paling sulit (+1,00 logit; dijawab benar oleh 19 dari 34 siswa atau 55,9%). Pada butir soal nomor 4 ini, siswa diminta menelaah konsep usaha ketika gaya bekerja tegak lurus terhadap arah perpindahan benda pada bidang datar, dengan 15 siswa (44,1%) memilih opsi pengecoh yang menganggap usaha tetap bernilai positif selama gaya dan perpindahan tidak nol. Fenomena tingginya tingkat kesulitan butir nomor 4 mengindikasikan adanya hambatan pemahaman konseptual di kalangan siswa mengenai konsep dasar abstrak Usaha dan Energi, seperti visualisasi arah gaya terhadap perpindahan atau syarat terjadinya usaha fisis. Hal ini sejalan dengan studi Docktor & Mestre, (2014), Alatas et al. (2024), serta Rasyid et al. (2025) yang menegaskan bahwa hambatan konseptual dalam fisika sering kali membuat soal konseptual yang tampak sederhana menjadi jauh lebih sulit diselesaikan daripada soal komputasi mekanis. Meskipun demikian, karena instrumen ini berupa tes pilihan ganda satu tingkat tanpa disertai analisis pengecoh secara mendalam maupun tingkat keyakinan jawaban (*certainty of response index*), dugaan adanya miskonsepsi tersebut masih bersifat indikasi awal dan belum dapat dibedakan secara mutlak dari ketidaktahuan konsep. Selain itu, selisih tingkat kesulitan butir nomor 4 sebesar +1,00 logit dari rata-rata kesulitan soal (0,00 logit) memiliki nilai galat baku (*Model S.E.*) sebesar 0,38 logit akibat ukuran sampel terbatas (N = 34), sehingga posisi kesulitan relatif butir ini masih mengandung ketidakpastian estimasi yang perlu diuji kembali pada sampel yang lebih besar.

Kejanggalan struktural berikutnya terlihat pada *Wright Map* bagian bawah (Gambar 6), di mana butir soal analisis konseptual (C4K nomor 21) dan analisis prosedural (C4P nomor 17) justru jatuh terperosok di area logit bawah (−0,81 logit; dijawab benar oleh 29 dari 34 siswa atau 85,3%), sejajar dengan soal mengingat faktual (C1F nomor 5).

Kondisi ini mengindikasikan tiga kemungkinan: pertama, adanya potensi ketidaktepatan klasifikasi level kognitif pada penyusunan butir soal karena instrumen ini belum melalui tahap uji validasi isi kuantitatif oleh panel ahli (seperti indeks Aiken's V atau *Content Validity Ratio*) untuk memverifikasi kesesuaian stimulus soal dengan indikator proses kognitif C4 sebelum ditafsirkan sebagai penyimpangan hierarki. Kedua, konstruk stimulus pada butir soal C4 tersebut masih bersifat rutin sehingga belum menuntut penalaran analitis tingkat tinggi yang sebenarnya. Ketiga, terjadi kecenderungan mekanisasi belajar (*rote problem solving*), di mana siswa terbiasa menghafal algoritma rumus dan langkah pengerjaan matematis secara prosedural tanpa memahami esensi fisis dari fenomena Usaha dan Energi itu sendiri, sehingga mereka mahir menyelesaikan hitungan berpola serupa (Docktor & Mestre, 2014). Di samping itu, untuk butir nomor 24 yang dirancang pada level mencipta (C6 Metakognitif; *measure* +0,23 logit), format soal pilihan ganda memiliki keterbatasan dalam mengukur kemampuan mencipta secara autentik karena peserta didik hanya memilih opsi rancangan yang tersedia alih-alih menghasilkan produk atau solusi orisinal secara terbuka, sehingga pada revisi instrumen mendatang butir C6 disarankan menggunakan format uraian terstruktur atau tugas kinerja.

Dari aspek kelayakan item, keberadaan item *misfit* kuat pada nomor 23 (C5M) dengan nilai *Outfit* MNSQ (1,70), ZSTD (2,36), dan *Pt-Measure Corr* (0,06) yang melampaui batas ideal serta gejala *misfit* ringan pada nomor 3 (C2K), nomor 12 (C3K), dan nomor 20 (C4K) menegaskan adanya distorsi pengukuran. Butir-butir ini berpotensi memicu bias informasi diagnostik karena pola jawabannya yang tidak konsisten. Secara keseluruhan, penerapan Model Rasch mampu memberikan informasi diagnostik yang rinci mengenai kalibrasi tingkat kesulitan dan kesesuaian masing-masing butir soal pada skala logit yang sama, melengkapi informasi dasar yang umumnya diperoleh melalui pendekatan teori tes klasik (CTT) (Arfiani et al., 2025; Nurani et al., 2025). Implikasi pedagogis dari penelitian ini menekankan pentingnya penataan ulang bank soal fisika secara bertahap dengan menambah proporsi soal HOTS berdaya beda tinggi pada rentang logit atas, disertai penguatan pembelajaran konseptual di kelas untuk mengatasi hambatan pemahaman dasar peserta didik.

SIMPULAN

Berdasarkan hasil analisis menggunakan Model Rasch pada instrumen tes materi Usaha dan Energi, dapat disimpulkan bahwa instrumen memiliki konsistensi internal keseluruhan dalam kategori cukup (*Cronbach's Alpha* = 0,78) dan *Person Reliability* sebesar 0,68 (cukup; *separation* 1,47), namun memiliki *Item Reliability* yang tergolong lemah (0,19) serta *Item Separation* rendah (0,49) akibat keterbatasan ukuran sampel (*N* = 34) yang menghasilkan galat baku (*Model S.E.*) butir sebesar 0,38–0,52 logit dan sempitnya variasi kesulitan soal. Analisis kesesuaian butir (*item fit*) mendeteksi butir soal nomor 23 (C5M) sebagai item *misfit* kuat yang melanggar ketiga kriteria *fit* (*Outfit* MNSQ = 1,70; ZSTD = 2,36; *Pt-Measure Corr* = 0,06), serta butir nomor 3 (C2K), nomor 12 (C3K), dan nomor 20 (C4K) sebagai *misfit* ringan. Pemetaan melalui *Wright Map* menunjukkan rata-rata kemampuan peserta didik (+1,29 logit) berada di atas rata-rata kesulitan soal (0,00 logit), serta memperlihatkan ketidaksesuaian urutan kesulitan empiris di mana butir C2K nomor 4 menempati posisi tersulit (+1,00 logit) dan butir C1F nomor 5, C4K nomor 21, serta C4P nomor 17 berada pada posisi termudah (−0,81 logit), yang mengindikasikan perlunya peninjauan ulang validitas isi butir dan potensi hambatan pemahaman konsep dasar. Sebagai rekomendasi konkret bagi pengembang instrumen dan guru fisika: (1) butir nomor 23 wajib dibuang atau direvisi total, butir nomor 3, 12, dan 20 direvisi pada bagian pengecoh dan redaksi soal, serta 6 butir dengan *Pt-Measure Corr* < 0,32 (nomor 1, 2, 4, 5, 22, dan 25) ditinjau ulang daya bedanya; (2) dilakukan validasi isi kuantitatif oleh ahli terhadap pemetaan Taksonomi Bloom serta penyesuaian format soal uraian untuk level mencipta (C6) pada butir nomor 24; dan (3) dilakukan uji coba lanjutan dengan sampel minimal 100 hingga 250 peserta didik agar kalibrasi parameter item stabil.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Bapak Dosen Pembimbing di Program Studi Pendidikan Fisika Universitas Islam Madura atas bimbingan dan arahan akademis yang diberikan

selama proses penelitian hingga penyusunan artikel ilmiah ini. Penulis juga menyampaikan apresiasi kepada pimpinan SMAN 4 Pamekasan, guru mata pelajaran fisika, orang tua/wali, serta seluruh 34 peserta didik kelas XI yang telah memberikan persetujuan dan berpartisipasi secara sukarela sebagai responden sehingga data penelitian ini dapat terkumpul dengan baik.

PERNYATAAN

Kontribusi Penulis: Putri Nabila Hudaimi: Konseptualisasi, Metodologi, Investigasi, Kurasi Data, Analisis Formal, dan Penulisan Draf Awal; Maimon Sumo: Supervisi, Validasi, dan Penulisan & Penyuntingan. Seluruh penulis telah membaca dan menyetujui naskah akhir.

Pendanaan: Penelitian ini tidak menerima pendanaan eksternal.

Konflik Kepentingan: Para penulis menyatakan tidak ada konflik kepentingan.

Pernyataan Penggunaan Kecerdasan Artifisial (AI): Para penulis tidak menggunakan alat kecerdasan artifisial generatif dalam pengumpulan data, analisis statistik, maupun pembuatan gambar; bantuan penyuntingan bahasa telah ditinjau dan diverifikasi sepenuhnya oleh para penulis.

Persetujuan Etik: Penelitian ini disetujui oleh pihak administrasi SMAN 4 Pamekasan. Persetujuan (assent) dari seluruh 34 peserta didik yang berpartisipasi dan persetujuan orang tua/wali diperoleh sebelum pengumpulan data.

Ketersediaan Data: Data yang mendukung temuan penelitian ini tersedia dari penulis korespondensi atas permintaan yang wajar.

REFERENSI

Alatas, S. A., Jahja, M., & Yusuf, M. (2024). Assessment of students' science literacy skills using diagnostic items on the concept of energy with Rasch modeling. *World Journal of Advanced Research and Reviews*, *21*(1), 1356–1362. https://doi.org/10.30574/wjarr.2024.21.1.2697

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Arfiani, Y., Yulianti, D., Masturi, M., & Nugroho, S. E. (2025). Feasibility analysis of scientific reasoning skills instruments for prospective science teachers using the Rasch model. *Unnes Science Education Journal*, 14(2), 405–416. https://doi.org/10.15294/usej.v14i2.25493

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the Rasch model. *Education and Information Technologies*, *29*(17), 22895–22941. https://doi.org/10.1007/s10639-024-12662-7

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, 15(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics - Physics Education Research*, 10(2), 020119. https://doi.org/10.1103/PhysRevSTPER.10.020119

Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (PhysTHOTS) peserta didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 1–12. https://doi.org/10.21831/pep.v18i1.2120

Linacre, J. M. (1994). Sample size and item calibration or person measure stability. *Rasch Measurement Transactions*, *7*(4), 328. https://www.rasch.org/rmt/rmt74m.htm

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Noorkholisoh, L., Yustiana, Y. R., Budiman, N., & Suryana, D. (2024). Validity and reliability analysis using the Rasch model in developing creativity tests instruments for elementary school students. *Jurnal Ilmiah Bimbingan Konseling Undiksha*, *15*(1), 128–135. https://doi.org/10.23887/jibk.v15i1.81249

Nurani, N. S., Rokhman, F., & Sarwi. (2025). Measuring the unseen: Digitalization of physics diagnostic assessment based on Rasch model. *Journal of Educational Sciences*, *9*(6), 4950–4963. https://doi.org/10.31258/jes.9.6.p.4950-4963

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Rasyid, F., Istiyono, E., Gunawan, C. W., & Kijambu, J. B. (2025). Developing and analyzing items of a physics conceptual understanding test on wave topics for high school students using the Rasch model. *REID (Research and Evaluation in Education)*, *11*(1), 1–16. https://doi.org/10.21831/reid.v11i1.75575

Samsudin, A., Suhandi, A., Rusdiana, D., Kaniawati, I., & Coştu, B. (2016). Investigating the effectiveness of an active learning based-interactive conceptual instruction (ALBICI) on electric field concept. *Asia-Pacific Forum on Science Learning and Teaching*, *17*(1), Article 1. https://www.eduhk.hk/apfslt/v17_issue1/samsudin/index.htm

Silaban, O. R., Kusnadi, & Wulan, A. R. (2025). Enhancing the five-tier diagnostic test on cell concepts through Rasch model analysis. *REID (Research and Evaluation in Education)*, *11*(1), 45–58. https://doi.org/10.21831/reid.v11i1.87212

Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. Trim Komunikata.

Wilson, L. O. (2016). *Anderson and Krathwohl – Bloom's taxonomy revised: Understanding the new version of Bloom's taxonomy*. The Second Principle.