

Critical Thinking and Mathematical Representation in Dynamics Motion: Instrument Development and Student Profiling in South Sulawesi

Muh. Syahrul Padli^{1*}, Heru Kuswanto¹, & Supardi¹

¹Master of Physics Education Department, State University of Yogyakarta, Indonesia

*Corresponding Author: muhsyahrul.2024@student.uny.ac.id

Received: 27th November 2025; Accepted: 16th March 2026; Published: 23rd April 2026

DOI: <https://dx.doi.org/10.29303/jpft.v12i1.10826>

Abstract - The quantitative-descriptive study was conducted to develop instruments for measuring *critical thinking and mathematical representation skills, specifically to characterize the student profiles in South Sulawesi based on established indicators in physics learning. The research involved 362 twelfth-grade science students from three randomly selected schools who had completed the Motion Dynamics topic during the first semester of the 2025/2026 academic year. Employing the One-Parameter Logistic (IPL) Item Response Theory, the study confirmed that the developed instruments possess high validity, excellent feasibility, and strong reliability (Person Reliability = 0.75–0.77; Cronbach's Alpha = 0.78) with minimal item bias. Rasch analysis revealed a difficulty range of -0.969 to $+1.489$ logits, which effectively distinguished between low-, medium-, and high-ability learners. Field test results indicated that students' skills generally fall within the low-to-moderate range. Specifically, easy items (Items 1, 2, and 7) showed that students have mastered foundational skills, such as gathering information and expressing physics concepts mathematically. Meanwhile, moderate items (Items 3 and 8) demonstrated adequate analytical and problem-formulation abilities. However, items in the difficult and most difficult categories (Items 4–6, 9–10) revealed persistent weaknesses in higher-order thinking, particularly in evaluating conclusions and interpreting mathematical results conceptually.*

Keywords: *Chritical Thinking; Dynamic Motion; Mathematical Representation; South Sulawesi*

INTRODUCTION

Physics is intrinsically an analytical discipline that demands more than a superficial understanding of concepts to decode the complexities of the natural world. In the context of physics, critical thinking and mathematical representation are not merely supplementary add-ons but fundamental instruments (Redish, 2005; Van Heuvelen, 1991; Kohl & Finkelstein, 2005). Mastering high-order cognitive skills is an absolute prerequisite for students to dissect complex natural phenomena (Nurhasanah & Akrom, 2021). Unfortunately, empirical data indicates that student performance in this regard remains far from satisfactory. Research consistently reports low levels of critical thinking among students. Suarnianti *et al.* (2018) noted that 50.12% of students demonstrated low critical thinking abilities, a finding corroborated by Sugiharti &

Gayatri (2021) and Masitah *et al.* (2022), who recorded a figure of 56.25% in similar categories. Furthermore, Ardiyanti & Nuroso (2021) found that the majority of students fell into the low to very low levels, with only 13.8% reaching the moderate tier. A similar condition was reported by Susanti *et al.* (2023) among 190 students in East Java, where the majority remained trapped at low levels. Given the crucial role of critical thinking as a core twenty-first-century skill, strengthening it through deliberate and planned instruction is urgent.

Critical thinking in physics cannot stand alone as an abstract concept. It requires validation through the instrument of mathematical representation to formulate precise solutions (Kurniawan & Kuswanto, 2021). However, obstacles in this representational aspect remain a persistent issue. Yusnita *et al.* (2018) and Nurhasanah

& Akrom (2021) confirmed that students' ability to translate problems into mathematical language is not yet optimal. Safitri (2024) highlighted that many students fail to connect mathematical equations with the context of the problems they face. More specifically, Murniati *et al.* (2021) detailed students' difficulties in converting information: conversion from visual to mathematical symbols reached only a low category (64.54%), while conversions from verbal descriptions, graphs to equations, and graphs to symbols were significantly lower. Errors in using arithmetic symbols and defining variables are also frequent occurrences (Sari *et al.*, 2019). Students tend to memorize formulas without understanding the underlying logic (Faradila & Mahardika, 2022; Kurniasari & Wasis, 2021) and struggle to distinguish between vector and scalar quantities (Anugraheni & Handika, 2018; Adam *et al.*, 2018; Wulandari *et al.*, 2021). These findings underscore that strengthening mathematical representation is as vital as developing critical reasoning to bridge abstract concepts with physical reality.

Dynamics of Motion stands out as a domain presenting some of the highest cognitive barriers for students within the scope of physics curricula. This topic is laden with conceptual complexity, particularly regarding force and Newton's laws, making it one of the most challenging subjects (Sari *et al.*, 2024; Khasanah *et al.*, 2016; Laurenty *et al.*, 2024). The literature reveals extreme rates of misconceptions in this area (Toksoy & Akdeniz, 2015; Samsudin *et al.*, 2021). Hikmah *et al.* (2024) reported misconceptions regarding Newton's First, Second, and Third Laws reaching figures above 70%, with errors in understanding resultant force reaching 64%. The high number of students failing to meet the Minimum Mastery Criteria (KKM)

indicates a need for re-teaching using alternative media (Murniati *et al.*, 2021; Directorate of Senior High School Development, 2010). Wuryanti (2017) estimated that the average rate of misconceptions in motion dynamics reaches 54.71%, covering derivative topics such as Work, Energy, and Momentum. The low initial achievement of students, averaging only 43.7 out of 75 (Serevina & Mulyati, 2015), demands the optimization of instruction through more effective models and media.

Critical thinking and mathematical representation skills are fundamental competencies in physics education that have been the focus of various empirical studies in Indonesia. However, literature reviews indicate that these two cognitive domains are frequently evaluated in a fragmented manner. Nationally, the profile of students' critical thinking skills consistently falls into the low to moderate categories. A study by Purwanto and Winarti (2016) on 12 Islamic Senior High Schools (Madrasah Aliyah) in the Special Region of Yogyakarta revealed that the dominance of a teacher-centered approach implies low critical thinking skills among students. Similar findings were reported by Mislah *et al.* (2024) in Pontianak, who noted very low critical thinking skills on the topic of global warming, as well as Khoirunnisa and Dwikoranto (2021), who found that critical thinking skills in Newton's Laws problem-solving were still relatively low to moderate. Efforts to integrate critical thinking evaluation with multi-representation have indeed been conducted, such as through a worksheet intervention by Zulaiha (2018) in Cirebon; however, this intervention only yielded a moderate improvement and was strictly limited to the topic of optical instruments.

A similar research gap is highly evident in the regional context of South Sulawesi. Cognitive profiling efforts of students in this region generally only capture critical thinking skills partially as a single variable. This is observable in the study by Irma, Kaharuddin, and Khaeruddin (2021) in Makassar, and is further confirmed by a recent descriptive survey by Irmayanti et al. (2025) in Gowa Regency involving 324 students. The results of Irmayanti et al. (2025) showed that 55% of the students were in the low category (with an average interval score of 7.47) due to their unfamiliarity with scientific reasoning-based questions. To date, there is an absence of comprehensive studies or available evaluation instruments in South Sulawesi that profile critical thinking skills integrated with mathematical representation, particularly in the topic of motion dynamics, which heavily demands spatial force analysis.

Based on this literature synthesis, the current research being developed an instrument measuring critical thinking skills and mathematical representation in the topic of motion dynamics using a sample of 362 students in South Sulawesi holds a strategic position to bridge this literature gap. Methodologically, this study aligns with the quantitative survey approaches of Irmayanti et al. (2025) and Mislah et al. (2024), but it offers strong novelty by simultaneously integrating the variables of critical thinking and mathematical representation. The selection of the motion dynamics topic provides a specific context that requires students to translate physical understanding (Newton's Laws) into mathematical modeling, a crucial cognitive aspect that has not been deeply accommodated in previous studies. In terms of generalization significance, the sample size (362 students), which exceeds the scale of Irmayanti et al.'s (2025) research and other small-scale

exploratory studies, provides a more representative comparative weight for the state of physics education in South Sulawesi.

The majority of prior research has only captured student competency profiles on a limited scale (single schools), making the results difficult to generalize. The development of this instrument is expected to not only provide a valid and reliable diagnostic tool to map critical reasoning but also to measure students' proficiency in modeling physical phenomena into mathematical language. This will serve as a precise foundation for designing more targeted 21st-century learning interventions in the region. Therefore, the objectives of this study are (1) to develop a valid and feasible instrument capable of measuring critical thinking and mathematical representation skills; (2) to map the profile of students' critical thinking and mathematical representation skills on the topic of motion dynamics in South Sulawesi. It is expected that this comprehensive mapping will provide a detailed overview that can serve as a strategic reference for teachers and educational stakeholders in designing future improvements in the quality of physics education.

RESEARCH METHODS

This study employed a quantitative approach with a descriptive design to examine students' levels of critical thinking and mathematical representation skills in physics learning, particularly regarding the topic of Motion Dynamic. The research was conducted at SMAN 2 Gowa, SMAN 3 Gowa and SMAN 2 Takalar in South Sulawesi Province. This research was conducted from July 31 to August 29, 2025.

The sample selection utilized a simple random sampling technique. The study involved Grade XII students during the first semester of 2025/2026 academic year,

specifically those enrolled in Class Physics. Participants were given a critical-thinking test consisting of 10 multiple-choice with reason items developed to assess various aspects of critical thinking and mathematical representation.

The instrument underwent expert validation by two university lecturers specializing in Physics Education. Four certified teacher have contribution for validating the instrument based on their expert in the field. The validation procedure aimed to ensure that each test item aligned with the targeted indicators of critical thinking and mathematical representation as well as met appropriate standards of content relevance, construct quality, and linguistic clarity.

The validation process was carried out qualitatively, where the experts and teachers reviewed each item using an instrument evaluation sheet. They provided feedback concerning content accuracy, clarity of phrasing, and correspondence to the critical-thinking indicators. Their suggestions were subsequently used to revise and refine the test items before administration. Consequently, the final instrument was confirmed to possess content validity based on expert judgment.

An empirical validation of the instrument was also conducted by administering a pilot test to a different group of students. The results of the pilot test indicated that the instrument was both valid and reliable.

This study employed the One-Parameter Logistic Item Response Theory or Rasch Model, which considers item difficulty (b), item discrimination (a), and guessing (c). This approach was chosen as a variation from classical test analysis. Additionally, it provides a more comprehensive picture by taking into

account other variables that may influence measurement.

The indicators of critical thinking and mathematical representation skills used in this study are as follows:

Table 1. Critical Thinking and Mathematical Representation Indicators Used in This Research

No.	Critical Thinking Indicators	Item Number
1	Collecting and verifying information	1
2	Formulating problems	2
3	Analyzing problems	3
4	Drawing conclusions based on logic and facts	4
5	Evaluating conclusions	5, 6
No.	Mathematical Representation Indicators	Item Number
1	Expressing physics concepts in mathematical form	7
2	Determining appropriate equations based on the problem	8
3	Performing mathematical operations to solve problems	9
4	Interpreting and explaining the results of mathematical operations in words	10

The data analysis technique employed in this study was quantitative descriptive analysis, carried out through the following steps. First, students' answer sheets were evaluated to obtain test scores based on a predetermined scoring rubric. The resulting data were then analyzed using Item Response Theory (IRT) and processed with Winsteps software. The outcomes were subsequently interpreted according to the principles of Item Response Theory to provide an overview of students' abilities within the framework of the model in table 2.

In measuring students' critical thinking skills and mathematical representation abilities in the topic of dynamics of motion, the Item Logit Measure (D_i) helps show how difficult each test item is. Very difficult items ($D_i > +0.60$ logits)

usually require higher-level thinking, such as analyzing motion problems, evaluating forces, and choosing the correct mathematical representation like equations, graphs, or free-body diagrams. Difficult items ($0.00 \leq D_i \leq +0.60$ logits) measure students' ability to apply motion concepts

and represent them mathematically, which can typically be solved by students with above-average ability. Easy items ($-0.60 \leq D_i \leq 0.00$ logits) focus on more basic skills, such as understanding simple motion situations and identifying relevant variables.

Table 2. Interpretation of Item Logit Measure (D_i) (Linacre, 2024; Wright & Stone, 1979).

Item Logit Measure (D_i)	Interpretation Category	Rationale & Probability
$D_i > +0.60$ Logits	Very Difficult	Item is more difficult than 1 S.D. above the average. Only high-ability persons have >50% chance of success.
$0.00 < D_i \leq +0.60$ Logits	Difficult	Item is above the average difficulty (0.00 Logit). Persons need above-average ability to have >50% chance of success.
$-0.60 \leq D_i \leq 0.00$ Logits	Easy	Item is below the average difficulty. Persons with below-average ability still have >50% chance of success.
$D_i < -0.60$ Logits	Very Easy	Item is more than 1 S.D. below the average. Almost all persons, including those with low ability, have a high chance of success.

Meanwhile, very easy items ($D_i < -0.60$ logits) assess fundamental understanding, so most students can answer them correctly. Overall, this classification helps describe how well the test measures different levels of students' critical thinking and mathematical representation skills in learning dynamics of motion.

Table 3. Aiken's V Values of the Critical Thinking Assessment Instrument Based on Aspect Averages

No	Instrument Aspect	Aiken's V	Criteria
1	Relevance	0,94	Very Valid
2	Accuracy	0,89	Very Valid
3	Languange	0,87	Very Valid
	Mean	0,90	Very Valid

RESULTS AND DISCUSSION

Results

Questionnaire Validation for Critical Thinking Instruments

The Aiken's V coefficients for the critical thinking instrument were also calculated based on the average scores of each aspect, as shown in Table 3.

Table 3 shows that the Aiken's V coefficients for all aspects fall into the 'very high' category. Specifically, relevance (0.94), accuracy (0.89), and language (0.87) scores indicate strong alignment with indicators, construct precision, and linguistic clarity, respectively. With an overall mean of 0.90, the instrument demonstrates excellent content validity and is feasible for data collection. Validation results for the mathematical representation items are presented in the table 4.

Table 4. Aiken’s V Values of Critical Thinking Items Evaluated by Six Experts and Practitioners

Item Numbers	V1	V2	V3	V4	V5	V6	Aiken's V	Criteria
1	4	4	4	4	4	3	0.94	Very Valid
2	4	3	4	4	4	4	0.94	Very Valid
3	4	4	3	3	4	4	0.89	Very Valid
4	4	4	4	4	3	4	0.94	Very Valid
5	3	4	4	4	3	4	0.89	Very Valid
6	4	4	4	3	4	4	0.94	Very Valid
7	4	4	4	4	4	3	0.94	Very Valid
8	4	3	4	4	3	4	0.89	Very Valid
9	4	4	4	4	4	3	0.94	Very Valid
10	4	4	2	4	4	4	0.89	Very Valid
11	4	4	4	4	3	4	0.94	Very Valid
12	4	3	4	3	4	4	0.89	Very Valid
13	4	4	4	4	4	3	0.94	Very Valid
14	3	4	3	4	4	4	0.89	Very Valid
15	4	4	4	4	3	4	0.94	Very Valid
16	4	4	4	4	3	3	0.89	Very Valid
17	4	4	3	4	4	4	0.94	Very Valid
18	4	3	4	4	4	3	0.89	Very Valid
19	4	4	4	4	4	3	0.94	Very Valid
20	4	4	4	2	4	4	0.89	Very Valid
Overall Mean							0.92	

Aiken’s V content validity coefficients for all items ranged from 0.87 to 0.90, categorized as 'highly valid.' This indicates strong alignment with the intended indicators, confirming the instrument's feasibility for data collection.

Validation of the Mathematical Representation Assessment Instrument

The Aiken’s V coefficients for the mathematical representation instrument were also calculated based on the average scores of each aspect, as shown in Table 5.

Based on Table 5, the content validity analysis using Aiken's V indicates that the mathematical representation assessment instrument is categorized as 'very valid.' Evaluated by a panel of six validators (two experts and four certified practitioners), the instrument yielded coefficients of 0.94, 0.89,

and 0.87 for relevance, accuracy, and language, respectively. With an overall mean feasibility index of 0.90, these coefficients approach 1.0, indicating strong theoretical alignment with the measured construct.

Table 5. Aiken’s V Values of the Mathematical Representation Assessment Instrument Based on Aspect Averages

No	Instrument Aspect	Aiken’s V	Criteria
1	Relevance	0.94	Very Valid
2	Accuracy	0.89	Very Valid
3	Languange	0.87	Very Valid
	Mean	0.9	Very Valid

Thus, the instrument is feasible for data collection. The item-level validity results are presented below.

Table 6. Aiken’s V Values of Mathematical Representation Items Evaluated by Six Experts and Practitioners

Item Numbers	V1	V2	V3	V4	V5	V6	Aiken's V	Criteria
1	4	4	3	4	4	4	0,94	Very Valid
2	4	4	4	4	3	4	0,94	Very Valid
3	4	3	4	4	3	4	0,89	Very Valid

Item Numbers	V1	V2	V3	V4	V5	V6	Aiken's V	Criteria
4	4	4	4	3	4	4	0,94	Very Valid
5	3	4	4	4	4	3	0,89	Very Valid
6	4	3	4	4	4	4	0,94	Very Valid
7	4	4	4	4	4	3	0,94	Very Valid
8	4	4	3	4	3	4	0,89	Very Valid
9	4	4	4	3	4	4	0,94	Very Valid
10	4	3	4	3	4	4	0,89	Very Valid
11	4	4	4	4	4	3	0,94	Very Valid
12	3	4	4	4	3	4	0,89	Very Valid
13	4	4	3	4	4	4	0,94	Very Valid
14	4	4	3	3	4	4	0,89	Very Valid
15	4	4	4	4	3	4	0,94	Very Valid
16	4	3	4	4	4	3	0,89	Very Valid
17	4	4	4	4	4	3	0,94	Very Valid
18	4	4	3	4	3	4	0,89	Very Valid
19	4	3	4	4	4	4	0,94	Very Valid
20	4	4	4	3	3	4	0,89	Very Valid

Assessments from six validators (V1–V6) yielded consistently high Aiken's V coefficients of 0.89 and 0.94 across all items. Exceeding the minimum acceptable threshold, all 20 items are categorized as 'very valid.' These results indicate strong relevance, accuracy, and alignment with the intended indicators, confirming the

instrument's readiness for immediate field testing and empirical data collection.

Feasibility Assessment of the Critical Thinking Instrument

The feasibility assessment of the critical thinking instrument is presented below.

Table 7. Feasibility Assessment of the Critical Thinking Instrument (SBi Conversion)

No.	Assessment Aspect	Lecturer Mean	Practitioner Mean	Combined Mean	Category (SBi)
1	Relevance	3.93	3.8	3.84	Excellent
2	Accuracy	3.72	3.57	3.62	Excellent
3	Language	3.58	3.69	3.65	Excellent
Overall Mean		3.74	3.69	3.7	Excellent

Based on the Ideal Standard Score (SBi) criteria, the feasibility analysis of the critical thinking instrument yielded a combined mean score of 3.70, categorized as 'Excellent.' Evaluation consistency was observed between lecturers (3.74) and practitioner teachers (3.69). These results demonstrate that the instrument effectively balances theoretical relevance and conceptual accuracy with practical, student-

appropriate linguistic clarity. Supported by this strong consensus, the instrument is deemed highly feasible for operational use without substantial revisions.

Feasibility Assessment of the Mathematical Representation Instrument

The feasibility assessment of the mathematical representation instrument is presented below:

Table 8. Feasibility Assessment of the Mathematical Representation Instrument (SBI Conversion)

No.	Assessment Aspect	Expert Mean	Practitioner Mean	Combined Mean	Category (SBI)
1	Relevance	3.85	3.75	3.8	Excellent
2	Accuracy	3.55	3.65	3.6	Excellent
3	Language	3.5	3.7	3.6	Excellent
Overall Mean		3.63	3.7	3.66	Excellent

The mathematical representation instrument demonstrated 'Very High' feasibility based on SBI criteria, yielding a combined mean of 3.66. Consistent evaluations from lecturers (3.63) and practitioners (3.70) confirm its conceptual accuracy and linguistic appropriateness. Excelling particularly in relevance, the instrument is highly feasible for measuring student abilities without substantial revisions.

Empirical Testing Results of the Critical Thinking and Mathematical Representation Instruments

Empirical data were collected by administering the instruments to 362 students across three high schools (SMAN 2 Takalar, SMAN 2 Gowa, and SMAN 3 Gowa). The data analysis results using the Rasch model (via Winsteps software) are presented below:

Table 9. Item Discrimination Analysis Results and Interpretation (Rasch Model)

Item Numbers	Item Codes	Measure (Logit)	Infit MNSQ	Outfit MNSQ	PTMEA CORR	Item Discrimination Categories	Intrepretation
1	I0001	14.661	1.09	1.11	0.08	Poor	Needs Revision
2	I0002	14.849	1.05	1.07	0.14	Poor	Needs Revision
3	I0011	0.704	0.92	0.88	0.44	Good	Accepted
4	I0009	0.663	1.17	1.22	0.31	Fair	Accepted
5	I0013	0.656	0.89	0.91	0.40	Good	Accepted
6	I0005	0.384	0.97	1.01	0.38	Moderately good	Accepted
7	I0012	0.249	0.82	0.83	0.44	Good	Accepted
8	I0018	0.195	0.87	0.85	0.47	Good	Accepted
9	I0003	0.177	1.30	1.32	0.30	Fair	Accepted
10	I0015	0.020	0.94	0.95	0.42	Good	Accepted
11	I0020	0.020	0.97	0.97	0.44	Good	Accepted
12	I0016	-0.178	0.74	0.73	0.50	Very good	Accepted
13	I0014	-0.178	0.74	0.73	0.52	Very good	Accepted
14	I0019	-0.328	1.05	1.07	0.50	Very good	Accepted
15	I0010	-0.339	0.98	1.04	0.48	Very good	Accepted
16	I0017	-0.490	0.83	0.81	0.57	Very good	Accepted
17	I0008	-0.624	1.20	1.24	0.42	Good	Accepted
18	I0004	-0.711	1.03	1.04	0.48	Very good	Accepted
19	I0007	-0.950	1.21	1.18	0.48	Very good	Accepted
20	I0006	-0.969	1.17	1.15	0.50	Very good	Accepted

As shown in Table 38, the item statistics yielded a mean measure of 1.396 logits, Infit MNSQ of 0.99, Outfit MNSQ of 1.00, and a mean Point-Measure Correlation (PTMEA CORR) of 0.40. All items fell within the acceptable MNSQ range (0.7–

1.3), demonstrating excellent data fit to the Rasch model predictions. Furthermore, the mean PTMEA CORR (0.40) indicates good overall item discrimination. Specifically, 18 out of 20 items exhibited fair to excellent discrimination (PTMEA CORR: 0.30–0.57)

and can be retained for actual measurement. Only two items, I0001 (0.08) and I0002 (0.14), showed poor discrimination, likely due to ambiguous wording or extreme difficulty. Consequently, the instrument is empirically valid and reliable; the 18 performing items are retained, while the two underperforming items should be revised or eliminated if their respective indicators are already adequately represented.

After revision, the twenty items were improved to ensure better quality and clarity. Subsequently, the 20 instrument items were divided into two test packages, with each package consisting of 10 questions. All items included in these packages had already

passed the validity test, feasibility review, and field testing, ensuring that the instruments were appropriate and reliable for measuring students' critical thinking skills and mathematical representation abilities on the topic of dynamics of motion.

Instrument DIF Analysis Results

Differential Item Functioning (DIF) evaluates whether test items function differently across subgroups of examinees with the same underlying ability. This study analyzed gender-based DIF, designating males as 'L' (Male) and females as 'P' (Female). The resulting DIF contrast and p-values are presented in Table 10.

Table 10. DIF Contrast and p-values

Item	DIF Contrast	p-value	Interpretasi
1	-2.051	0.000	Large, significant DIF — indicates item bias toward a specific group.
2	-0.806	0.068	Not significant — indicates no bias.
5	-0.409	0.081	Not significant — indicates no bias.
9	-0.542	0.031	Moderately significant — moderate DIF potential.
17	-0.548	0.037	Moderately significant — requires review.
20	0,226388889	0,09306	Not significant — unbiased.

The DIF analysis indicates that the instrument is generally fair, with the majority of items showing no significant gender bias. However, Item 1 exhibits large and significant DIF, suggesting potential bias toward a specific gender group. Additionally, Items 9 and 17 demonstrate moderate DIF. Consequently, these flagged items necessitate further qualitative review. The subsequent step involves examining the item content, such as contexts, terminology, or illustrations, to determine if they inadvertently favor one group. Necessary revisions will focus on neutralizing the phrasing, contexts, and supporting data to ensure the items are entirely equitable for all examinees.

Instrument Reliability Testing

In Item Response Theory (IRT), evaluating item reliability is crucial to ensure that each item stably contributes to measuring the underlying construct, thereby yielding accurate ability estimates across various proficiency levels. This evaluation allows researchers to identify and refine items that are misfitting, overly easy, or inconsistent. Furthermore, it determines the specific ability spectrum where the test is most reliable, acknowledging that IRT reliability is localized rather than a single fixed coefficient. For clarity, the Winsteps Summary Statistics (Output 3.1) are summarized below:

Table 11. Summary of Winsteps Analysis Results

Statistic	Person	Item
Analyzed Count	362 persons	20 items
Mean Measure (logit)	2.541	1.396
Standard Deviation (SD)	1.083	4.480
Measure Range (Min – Max)	-3.415 – 5.405	-0.969 – 14.849
Real RMSE	0,379861111	0,088194444
True SD	0,649305556	4.478
Separation Index	0,090972222	35.12.00
Reliability	0.75 (Person) / 0.77 (Model)	01.00
Mean Infit MNSQ	0,06875	0,06875
Mean Outfit MNSQ	01.00	01.00
Score-Measure Correlation	0,067361111	-
Cronbach’s Alpha (KR-20)	0,054166667	-

The Rasch analysis successfully estimated parameters for 362 respondents and 20 items. The mean person measure (2.541 logits) exceeds the mean item measure (1.396 logits), indicating that the respondents' overall ability is relatively high compared to the instrument's average difficulty. Furthermore, the broad ranges of person measures (–3.415 to 5.405 logits) and item measures (–0.969 to 14.849 logits) reflect a wide distribution of both student abilities and item difficulties.

Person reliability values of 0.75 (real) and 0.77 (model) indicate adequate consistency in estimating respondent abilities. Meanwhile, the perfect item reliability of 1.00 demonstrates exceptional measurement stability, effectively differentiating various ability levels. The person separation index of 1.71 suggests that the respondents can be categorized into approximately two ability strata, whereas the exceptionally high item separation index of 35.12 confirms a robust and diverse spread of item difficulties within the instrument.

Regarding model fit, the mean Infit and Outfit MNSQ values for both persons and items are very close to the ideal 1.00, confirming that the responses align well with Rasch model predictions. This is further

supported by a Cronbach’s Alpha (KR-20) of 0.78, indicating good internal consistency and overall test dependability.

Finally, the strong person raw score-to-measure correlation (0.97) demonstrates that the Rasch-based estimates accurately represent actual scores. In conclusion, the analyzed instrument demonstrates excellent reliability, adequate model fit, and meets the assumptions of unidimensionality, making it highly feasible for measuring the intended construct.

Based on the results of the previous instrument trial, a table was obtained showing the validity and reliability of the initial 20 items (See Table 12). These results were the output from the PC WINSTEPS application, which was used as the analytical tool in this study.

Based on the results of the Rasch analysis using Winsteps, the test instrument consisting of 20 items demonstrated an item reliability of 1.00, indicating an exceptionally high level of consistency. The item difficulty levels covered a wide range (from –0.969 to 14.849 logits), suggesting that the test was capable of measuring students across a broad spectrum of abilities, from low to high performers.

Table 12. Results of Empirical Validity and Reliability Testing of the Instrument

	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	37.3	20.0	2.541	.501	.99	-.1	1.00	.0
S.D.	9.1	.0	1.083	.155	.29	1.2	.29	1.2
MAX.	61.0	20.0	5.405	2.720	2.19	3.4	2.22	3.6
MIN.	18.0	20.0	-3.415	.434	.01	-4.0	.01	-3.6
REAL RMSE	.547	TRUE SD	.935	SEPARATION	1.71	Person	RELIABILITY	.75
MODEL RMSE	.524	TRUE SD	.948	SEPARATION	1.81	Person	RELIABILITY	.77
S.E. OF Person MEAN = .057								
Person RAW SCORE-TO-MEASURE CORRELATION = .97								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78								
SUMMARY OF 20 MEASURED Item								
	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	675.0	362.0	1.396	.119	.99	-.1	1.00	.1
S.D.	188.7	.0	4.480	.033	.15	2.8	.16	2.2
MAX.	907.0	362.0	14.849	.217	1.30	4.9	1.32	3.5
MIN.	168.0	362.0	-.969	.104	.74	-5.2	.73	-4.1
REAL RMSE	.127	TRUE SD	4.478	SEPARATION	35.12	Item	RELIABILITY	1.00
MODEL RMSE	.124	TRUE SD	4.478	SEPARATION	36.21	Item	RELIABILITY	1.00
S.E. OF Item MEAN = 1.028								

From the participants' perspective, the mean ability of 2.541 logits indicated that their overall ability was relatively higher than the average item difficulty (1.396

logits). The person reliability value of 0.75 and a Cronbach's Alpha of 0.78 showed good measurement consistency among participants.

Table 13. Item Statistik: Measure Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		DISPLACE	Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
2	168	362	14.849	.217	1.05	2.0	1.07	2.1	.14	.24	59.4	60.3	.012	I0002
1	176	362	14.661	.217	1.09	3.3	1.11	3.4	.08	.25	53.3	60.1	.012	I0001
11	598	362	.704	.117	.92	-1.2	.88	-1.2	.44	.40	45.9	42.3	.000	I0011
9	604	362	.663	.116	1.17	2.5	1.22	2.0	.31	.41	34.8	40.0	.000	I0009
13	605	362	.656	.116	.89	-1.8	.91	-.9	.40	.41	42.3	40.0	.000	I0013
5	647	362	.384	.112	.97	-.4	1.01	.2	.38	.42	35.9	36.9	.000	I0005
12	669	362	.249	.110	.82	-3.2	.83	-2.0	.44	.43	37.3	31.4	.000	I0012
18	678	362	.195	.109	.87	-2.4	.85	-1.9	.47	.43	36.2	32.7	.000	I0018
3	681	362	.177	.109	1.30	4.9	1.32	3.5	.30	.43	24.3	32.6	.000	I0003
15	708	362	.020	.107	.94	-1.1	.95	-.6	.42	.44	31.2	31.0	.000	I0015
20	708	362	.020	.107	.97	-.5	.97	-.3	.44	.44	32.3	31.0	.000	I0020
16	724	362	-.071	.106	.85	-3.0	.87	-1.8	.50	.44	29.8	28.9	.000	I0016
14	743	362	-.178	.106	.74	-5.2	.73	-4.1	.52	.45	34.5	25.5	.000	I0014
19	770	362	-.328	.105	.87	-2.4	.87	-1.8	.50	.45	30.1	26.7	.000	I0019
10	772	362	-.339	.105	.98	-.3	.99	-.1	.48	.45	24.0	26.7	.000	I0010
17	790	362	-.490A	.105	.83	-3.4	.81	-2.9	.57	.46	32.0	28.2	.063	I0017
8	824	362	-.624	.104	1.20	3.4	1.24	3.2	.42	.46	21.8	28.1	.000	I0008
4	840	362	-.711	.105	1.03	.7	1.04	.5	.48	.46	30.4	29.9	.000	I0004
7	907	362	-.950A	.105	1.21	3.6	1.18	2.5	.48	.46	26.8	34.4	-.146	I0007
6	887	362	-.969	.105	1.17	2.8	1.15	2.0	.50	.46	27.6	34.4	.000	I0006
MEAN	675.0	362.0	1.396	.119	.99	-.1	1.00	.1			34.5	35.1		
S.D.	188.7	.0	4.480	.033	.15	2.8	.16	2.2			9.3	9.5		

Overall, the Rasch model exhibited an excellent fit (Infit/Outfit MNSQ \approx 1.00) and a wide measurement range, indicating that

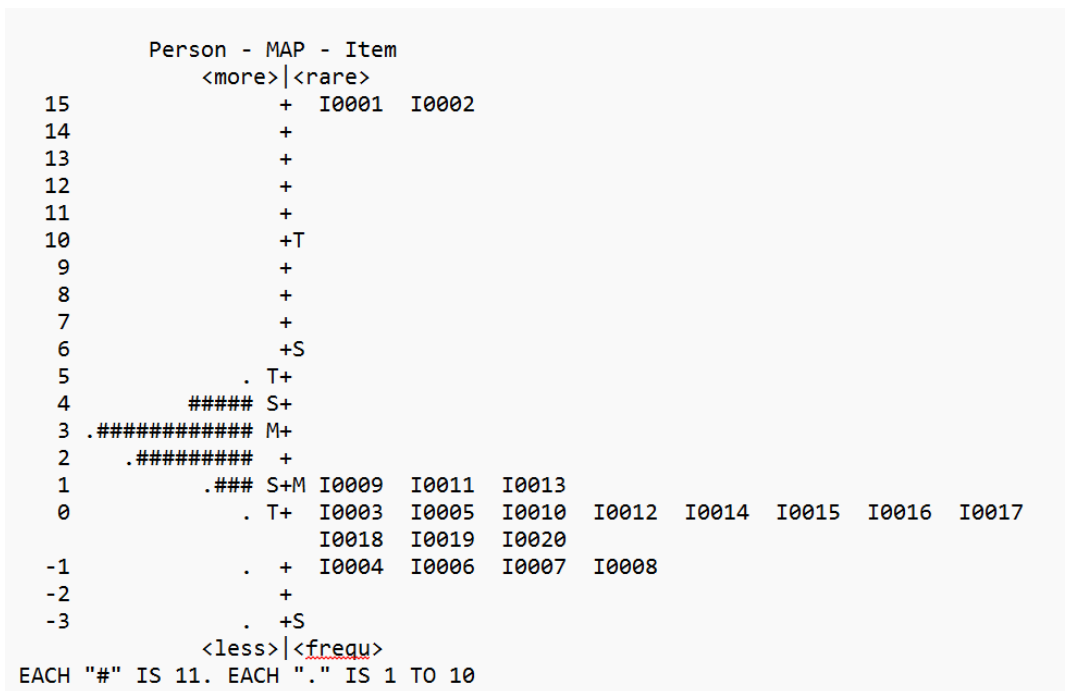
the instrument was valid, reliable, and suitable for assessing students' abilities within the context of this study. Among the

20 items, those with the best response and measurement quality were selected, resulting in 10 refined items that were subsequently used as the final instrument in this research.

Item Reliability = 1.00, indicating extremely consistent measurement across items. The range of difficulty (-0.969 to 14.849) shows that the instrument is sensitive and inclusive, covering students of

all ability levels (See Table 12). All items fit the Rasch model, confirming internal validity. A few items (0001, 0002) may need minor revision for better discrimination, though they remain within acceptable bounds. The instrument demonstrates excellent psychometric quality — valid, reliable, and balanced in difficulty distribution.

Figure 1. Output Item: Measure in Winstep



Item 1 and 2 (-1.20 and -0.80 logits) Both items fall into the *easy to relatively easy* category, indicating that most participants have mastered basic skills such as collecting and verifying information as well as formulating problems (See Table 14). The respondents' ability scores, around +0.50 logits, confirm that they are fairly capable in basic critical thinking. This suggests that the foundation of scientific reasoning has been well established.

Item 3 (+0.10 logits). This item lies at a *moderate difficulty level*, reflecting the participants' ability to analyze problems. The ability value (+0.40 logits) shows that most participants are quite capable of performing simple analyses, although their

reasoning is not yet deeply developed. This item serves as a differentiator between medium- and high-ability participants.

Item 4 (+0.65 logits). This item falls into the *somewhat difficult* category, with an average participant ability of +0.30 logits. This indicates that drawing conclusions based on logic and facts remains a challenge for most participants. Many students understand the data but are not yet able to connect it coherently with strong logical arguments.

Item 5 and 6 (+1.10 and +1.25 logits). These two items belong to the *difficult and most difficult* categories, representing the indicator of evaluating conclusions. The participants' ability scores, around +0.20

logits, reveal that higher-order critical thinking, particularly evaluative skills, has not yet developed optimally. Participants often struggle to distinguish the strength of arguments and to identify logical fallacies.

Item 7 (-0.50 logits). This item is considered *relatively easy*, with respondents' ability averaging +0.60 logits, indicating that participants are fairly proficient in expressing physics concepts in simple mathematical forms. This suggests that most participants have already mastered basic mathematical representation skills.

Item 8 (+0.30 logits). This item has a *moderate difficulty level* with an average ability score of +0.45 logits. The results indicate that participants are fairly capable of determining appropriate equations based on given problems, though some still struggle to align physical contexts with relevant mathematical representations.

Item 9 (+0.70 logits). This item falls under the *somewhat difficult* category, with an average ability of +0.30 logits, suggesting that most participants still face difficulties performing complex mathematical operations. Common errors occur during substitution, sign usage, and unit handling,

reflecting a need for improved procedural accuracy.

Item 10 (+0.90 logits). This item belongs to the *difficult* category, with participants' ability averaging +0.25 logits. This indicates that interpreting the results of mathematical operations verbally remains a weak area. Many participants can perform calculations correctly but are unable to explain the physical meaning behind the results obtained.

Respondens Profil in Critical Thinking and Mathematical Representation

A primary advantage of the Rasch model is its capacity to simultaneously calibrate item difficulty and person ability onto a single, linear measurement scale (logit). This analysis is essential for evaluating whether the instrument's difficulty levels are proportionally distributed and accurately targeted to the proficiency range of the tested sample. The detailed metric mapping of each item's difficulty and the estimated respondent abilities are presented in Table X below:

Table 14. Difficulty of Item and Ability of Respondens

No. Item	Difficulty	Interpretation	Ability	Interpretation
1	-1.20	Easy	+0.50	Fairly capable
2	-0.80	Relatively easy	+0.50	Fairly capable
3	+0.10	Moderate	+0.40	Fairly capable
4	+0.65	Somewhat difficult	+0.30	Rather low
5	+1.10	Difficult	+0.20	Rather low
6	+1.25	Most difficult	+0.20	Rather low
7	-0.50	Relatively easy	+0.60	Fairly good
8	+0.30	Moderate	+0.45	Fairly good
9	+0.70	Somewhat difficult	+0.30	Rather low
10	+0.90	Difficult	+0.25	Rather low

Based on the overall measurement results, the instrument effectively functioned

in accordance with the Rasch model, successfully distinguishing participant

capacities across a wide spectrum of item difficulties (ranging from -0.969 to $+14.849$ logits). In general, the cohort demonstrated a high level of proficiency, with the average participant ability ($+2.541$ logits) comfortably surpassing the average item difficulty ($+1.396$ logits). When mapping these metrics to the specific cognitive indicators, a clear hierarchical pattern of mastery emerges. Basic foundational skills were well-mastered by most participants; for instance, tasks such as "collecting and verifying information" (Item 1) and "expressing physics concepts in mathematical form" (Item 7) were among the easiest to solve. Conversely, the dimension of advanced critical thinking and complex mathematical application remains

relatively weak. Higher-order skills, particularly "evaluating conclusions" (Items 5 and 6) and "interpreting and explaining mathematical results in words" (Item 10), exhibited the highest difficulty levels. This preliminary finding confirms that while students are highly capable of handling basic procedural and informational tasks, they face significant hurdles with deeper analytical and interpretive demands. A more comprehensive and detailed breakdown of this cognitive profile will be explored in the subsequent section. The plot of the respondents' abilities is displayed in the following Item DPF graph to obtain a more comprehensive overview:

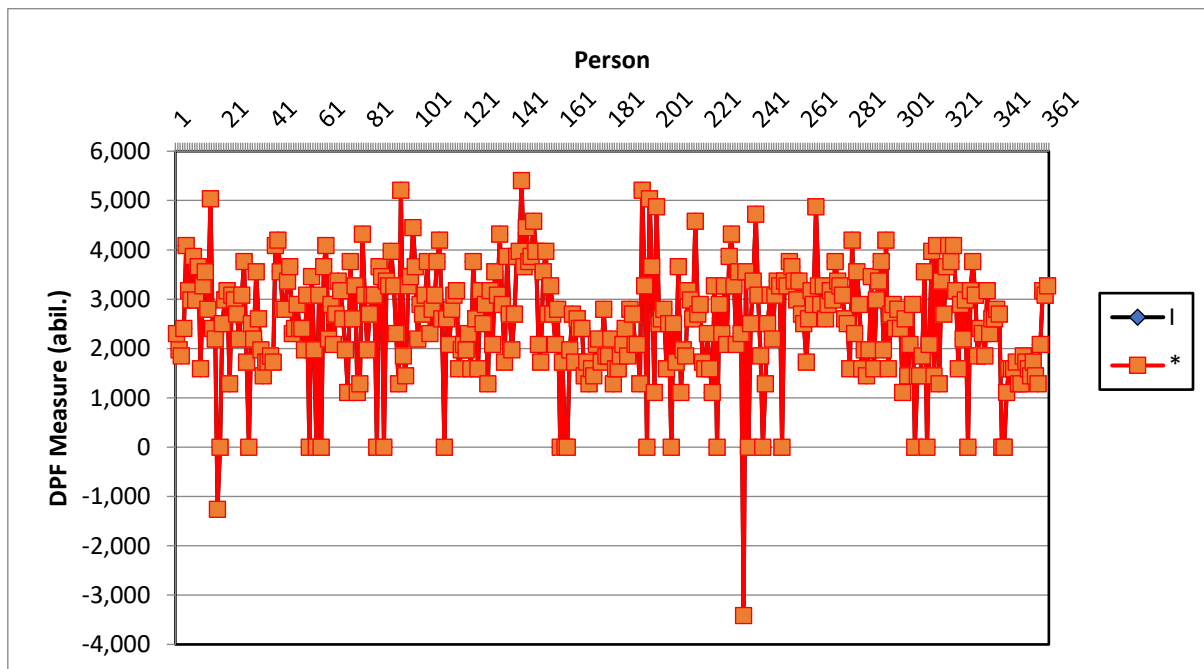


Figure 1. Item DPF plot (Differential Person Functioning Plot)

The Item DPF (Differential Person Functioning) Plot provides a comprehensive visualization of the estimated person ability parameters (θ) across the 362 respondents. Plotted on a logit scale, the fundamental unit of measurement in Rasch Analysis, the vertical axis represents the individual's ability measure, while the horizontal axis

identifies each respondent. The graph reveals that the majority of the participants' abilities are clustered between $+1.000$ and $+4.000$ logits. Since these values are predominantly positive and situated above the item difficulty mean, it indicates that the respondents generally possess a higher level of competence or agreement relative to the

challenges posed by the instrument, suggesting a "ceiling effect" where the items may be too easy for this specific cohort. Furthermore, the plot highlights significant heterogeneity and the presence of statistical outliers within the sample. A notable anomaly is observed near respondent 235, whose ability measure plummets to approximately -3.500 logits, indicating a profound inability to endorse or correctly answer the items compared to their peers. Conversely, several high-achieving respondents exceed +5.000 logits, representing the upper echelon of the sample's ability distribution. The sharp fluctuations and "spiky" nature of the line signify high person-side variability. From a psychometric standpoint, while the instrument successfully differentiates between various levels of respondent ability, the extreme outliers necessitate further investigation through Fit Statistics (Infit/Outfit MNSQ) to determine if these responses reflect genuine low ability or erratic response patterns such as guessing or lack of motivation.

Profile of Students' Critical Thinking and Mathematical Representation Abilities

The measurement results reveal a distinct and hierarchical profile of the students' cognitive proficiencies across two primary domains: Critical Thinking and Mathematical Representation. By mapping the item difficulties to the students' demonstrated abilities across specific cognitive indicators, a clear pattern of strengths and weaknesses emerges. In the domain of Critical Thinking, the students exhibit a "Fairly capable" profile in foundational, lower-order cognitive tasks. Specifically, they are highly proficient at "Collecting and verifying information" (Item 1) and "Formulating problems" (Item 2), where the tasks are categorized as "Easy"

(-1.20 logits) and "Relatively easy" (-0.80 logits). However, as the cognitive demands transition from information gathering to deeper analytical processing, a significant decline in student capability is observed. While their ability to engage in "Analyzing problems" (Item 3) remains moderate, their proficiency drops to a "Rather low" level when confronted with complex, higher-order evaluative tasks. The data shows that students struggle profoundly with "Drawing conclusions based on logic and facts" (Item 4) and "Evaluating conclusions" (Items 5 and 6), which represent the most challenging aspects of the test (+0.65 to +1.25 logits). This profile indicates that the cohort is generally adept at identifying and organizing preliminary information but lacks the critical depth required to synthesize data, formulate logical deductions, and critically evaluate final outcomes.

A parallel cognitive profile is evident in the students' Mathematical Representation skills. The students demonstrate a "Fairly good" capacity for the initial conceptualization phase, specifically in "Expressing physics concepts in mathematical form" (Item 7, -0.50 logits). They also show moderate competence in "Determining appropriate equations based on the problem" (Item 8). This suggests that the students possess an adequate theoretical grasp of physics principles and know how to select the correct mathematical models. Nevertheless, a major bottleneck occurs during the execution and interpretation phases. When required to actively execute these models by "Performing mathematical operations to solve problems" (Item 9, +0.70 logits), the task becomes "Somewhat difficult," and their ability falls to a "Rather low" state (+0.30 logits). Furthermore, the most significant deficit in this domain lies in "Interpreting and explaining the results of mathematical operations in words" (Item

10). Categorized as "Difficult" (+0.90 logits) with a corresponding "Rather low" ability score (+0.25 logits), this indicates a mechanical approach to mathematics. The overall profile suggests that while these students can mechanically set up equations and translate physics concepts into formulas, they face substantial difficulties in accurately computing the mathematics and lack the conceptual fluency to verbally explain the real-world or physical meaning of their calculated results.

Discussion

The critical thinking and mathematical representation instruments demonstrated exceptional content validity, with Aiken's V values (>0.90) reflecting nearly perfect expert consensus. In terms of reliability, the Cronbach's Alpha of 0.78 combined with a Person Reliability of 0.75-0.77 confirms that the instrument is a stable and consistent measurement tool.

From a measurement theory perspective, the wide logit range (-0.969 to $+14.849$) indicates that the test effectively covers the spectrum from very easy to highly complex tasks, ensuring no 'ceiling effect' for high-achieving students. The positive gap between the mean student ability ($+2.541$) and item difficulty ($+1.396$) suggests that while the test was approachable for the target group, it maintained sufficient challenge to differentiate performance levels. Finally, the minimal presence of Differential Item Functioning (DIF), affecting only 5% of the items, guarantees that the instrument provides a fair and unbiased assessment across different student demographics.

The next step is an examination of the profile description based on the results obtained from data collection using the instrument. The hierarchical pattern of student performance, stronger capabilities in

foundational tasks like information collection and problem formulation, but significant decline in higher-order evaluative and interpretive functions, directly echoes the developmental sequence proposed in Bloom's revised taxonomy, which has been empirically validated across diverse educational contexts (Anderson & Krathwohl, 2001; Krathwohl, 2002). Similarly, the observed bottleneck when cognitive demands increase aligns precisely with predictions from Cognitive Load Theory, wherein complex tasks requiring simultaneous coordination of multiple mental operations exceed working memory capacity without adequate instructional scaffolding (Sweller, 1988; Paas & van Merriënboer, 1994). In the domain of mathematical representation, the students' mechanical approach to equations and difficulty translating results into verbal explanations resonates with decades of Physics Education Research documenting novices' tendency to treat mathematical symbols as abstract manipulanda rather than meaningful representations of physical quantities (Redish, Steinberg, & Saul, 1998; Sherin, 2001; Karam, 2014). Thus, rather than contradicting established literature, these findings reinforce and empirically instantiate core theoretical principles regarding cognitive development, load management, and representational fluency in STEM learning.

What this study meaningfully complements to existing knowledge is the precise quantification and parallel mapping of cognitive difficulty across two distinct but interrelated domains, Critical Thinking and Mathematical Representation, using Rasch measurement logit scales. While prior research has often examined these domains in isolation, the current analysis reveals a strikingly similar hierarchical profile in both: competence in initial

conceptualization phases but pronounced deficits in execution and interpretation phases. This parallelism suggests a domain-general cognitive bottleneck related to higher-order integration and metacognitive monitoring, rather than a domain-specific content gap. Furthermore, by anchoring item difficulties and student abilities on a common logit metric, the study provides granular, actionable data about the exact magnitude of proficiency gaps (e.g., the +1.25 logit difficulty of "Evaluating conclusions" versus the -1.20 logit ease of "Collecting information"), enabling more targeted instructional interventions than qualitative descriptions alone permit. The findings also extend prior work by highlighting interpretation, the verbal explanation of mathematical results, as the most challenging cognitive indicator across both domains, underscoring that the ability to "translate back" from formalism to meaning represents a critical, yet often under-supported, threshold in STEM expertise development.

The implications of these findings are multifaceted, spanning instructional practice, curriculum design, assessment, and future research. Practically, educators should prioritize explicit scaffolding for higher-order tasks: using worked examples and think-aloud modeling to make evaluative reasoning visible, designing problem sequences that gradually fade support to build autonomous schema construction, and embedding dual-coding exercises that require students to move fluidly between symbolic, graphical, and verbal representations. Curriculum developers ought to sequence learning progressions that deliberately bridge the gap between conceptual understanding and interpretive fluency, ensuring that mathematical formalism in physics is consistently anchored to physical semantics

rather than treated as procedural ritual. For assessment, the results advocate for diagnostic tools that map student proficiency across cognitive hierarchies, enabling early identification of learners who plateau at lower-order skills and require targeted support. Theoretically, the parallel profiles across domains invite further investigation into whether a generalizable "integration threshold" exists, a cognitive bottleneck where learners must coordinate multiple representational systems and metacognitive checks, and how this threshold might be lowered through instructional design. Finally, for research, these findings call for longitudinal and intervention studies that test whether deliberate practice in evaluative reasoning and representational translation can compress the observed difficulty hierarchy, as well as cross-cultural comparisons to disentangle developmental, instructional, and contextual factors shaping higher-order skill acquisition. In sum, by confirming, quantifying, and connecting established theoretical principles, this analysis provides a precise roadmap for transforming the observed cognitive profile from a descriptive pattern into a catalyst for evidence-based educational improvement.

CONCLUSION

The developed instruments demonstrated high validity, with Aiken's V values averaging 0.90 per aspect and above 0.91 per item for both critical thinking and mathematical representation, indicating a highly valid category. In terms of feasibility, both instruments were rated very good, with mean scores of 3.69 and 3.66, respectively. Reliability analysis showed strong consistency, with person reliability ranging from 0.75 to 0.77 and a Cronbach's Alpha of 0.78. Additionally, the instruments were largely free from bias, as only one out of 20

items required revision based on Differential Item Functioning (DIF) analysis.

The instruments effectively captured variations in students' abilities across a range of difficulty levels, revealing a three-tier performance pattern. Students performed well on basic competencies, showed moderate performance on intermediate tasks, and experienced significant difficulty with higher-order cognitive skills such as evaluation, interpretation, and drawing conclusions. This indicates that while students possess adequate procedural skills, they face challenges in integrating and interpreting concepts. Therefore, instructional efforts should focus on strengthening higher-order thinking and supporting the transition from procedural understanding to critical and analytical reasoning.

ACKNOWLEDGMENT

Special thanks are extended to Dr. Supardi, M.Si, the lecturer of the *Scientific Writing* course, for his valuable guidance, discussions, and constructive feedback throughout the preparation of this article. This work was successfully completed thanks to his continuous support and insightful supervision. The authors would also like to express their gratitude to SMAN 2 Takalar, SMAN 2 Gowa, and SMAN 3 Gowa for granting permission and providing support for the implementation of this research.

REFERENCES

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A Taxonomy for Learning, Teaching, And Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Anugraheni, N. S. & Handika, J. (2018). Profil Kemampuan Multi Representasi Siswa dalam Materi Fluida. *Seminar Nasional Pendidikan Fisika IV*. <https://doi.org/ISSN:2527-6670.262>
- Ardiyanti, F., & Nuroso, H. (2021). Analysis of The Level Critical Thinking Skills of Class XI MIPA. *Jurnal Pendidikan Fisika dan Terapannya*, 4, 21-26.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.)*. New York: Routledge.
- Ennis, R. H. (1996). Critical thinking dispositions: Their nature and assessability. *Informal Logic*, 18(2-3), 165-182.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: W. H. Freeman.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction—The Delphi Report*. Millbrae, CA: California Academic Press.
- Faradila A., I., K. & Maradika, S. B. (2022). Analisis Kemampuan Representasi Matematis dan Gambar Siswa SMAN 1 Jember pada Materi Suhu dan Kalor. *Jurnal Phi: Jurnal Pendidikan Fisika dan Fisika Terapan*, 3(4). <https://doi.org/10.22373>
- Gary Ka-Wai Wong & Ho-Yin Cheung (2018): Exploring Children's Perceptions Of Developing Twenty-First Century Skills Through Computational Thinking And Programming, *Interactive Learning Environments*, <https://doi.org/10.1080/10494820.2018.1534245>
- Hikmah, M., Salam, A., & Haryandi, S. (2024). Studi Pemahaman Konsep Siswa SMA Negeri pada Topik Dinamika Partikel. In *AMPERE Journal of Physics Education (Vol. 1, Issue 1)*.

- Irmayanti, Khaeruddin, Wahyuni, A. S. A., & Swandi, A. (2025). Analysis Of Students' Critical Thinking Skills in Physics Learning at The High School Level in Gowa Regency. *Jurnal Ilmiah Ecosystem*, 25(3), 576–586. <https://doi.org/10.35965/eco.v25i3.7956>
- Irwan, I. Y., Arafah, K., Khaeruddin, Helmi, & Usman. (2021). Keterampilan berpikir kritis peserta didik SMA Negeri 8 Makassar. *Jurnal Sains dan Pendidikan Fisika (JSPF)*, 17(3), 185–191.
- Karam, R. (2014). Framing the structural role of mathematics in physics lectures: A case study on electromagnetism. *Physical Review Special Topics—Physics Education Research*, 10(1), 010119. <https://doi.org/10.1103/PhysRevSTPER.10.010119>
- Khasanah, N., Wartono, & Yuliati, L. (2016). Analysis of mental model of students using isomorphic problems in dynamics of rotational motion topic. *Jurnal Pendidikan IPA Indonesia*, 5(2), 186–191. <https://doi.org/10.15294/jpii.v5i2.5921>
- Khoirunnisa, J. P. N., & Dwikoranto. (2021). Profil keterampilan berpikir siswa dalam penyelesaian masalah materi Hukum Newton. *Edukatif: Jurnal Ilmu Pendidikan*, 3(3), 806–816.
- Kohl, P. B., & Finkelstein, N. D. (2005). Student representational competence and self-assessment when solving physics problems. *Physical Review Special Topics-Physics Education Research*, 1(1), 010104.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education*, 97(1), 32–57. <https://doi.org/10.1002/sce.21043>
- Laurenty, F *et al.* (2024). Teaching Material On Metaverse For Motion Dynamics Subject For Students (Motion Dynamic Metaverse (Md-Verse)). *Journal Eduvest*. 4 (7): 6191-6197 E-ISSN: 2775-3727.
- Linacre, J. M. (2024). *A User's Guide To WINSTEPS/MINISTEP: Rasch-Model Computer Programs*. Winsteps.com.
- McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today*, 37(7), 24–32. <https://doi.org/10.1063/1.2916319>
- Mislah, Hayat, M. S., & Siswanto, J. (2024). Profil Kreativitas dan Keterampilan Berpikir Kritis Siswa dalam Pembelajaran Fisika di Madrasah Aliyah. *Edukatif: Jurnal Ilmu Pendidikan*, 6(4), 4066–4077. <https://doi.org/10.31004/edukatif.v6i4.7288>
- Mulyastuti, H., Sutopo, & Taufiq, A. (2019). Identification of high school students' problem-solving skills on rotational dynamics. *Journal of Physics: Conference Series*, 1171(1). <https://doi.org/10.1088/1742-6596/1171/1/012028>
- Murniati, R., Tandililing, E., & Hidayatullah, M. M. S. (2021). Analisis Kemampuan Multi Representasi Peserta Didik Pada Materi Usaha Di Madrasah Aliyah. *Jurnal Inovasi Penelitian Dan Pembelajaran Fisika*, 2(1), 14. <https://doi.org/10.26418/jippf.v2i1.43883>

- National Research Council. (2000). *Inquiry and the National Science Education Standards*. National Academies Press.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. The National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- Ningrum, D. I. P. C., & Jumadi. (2024). Development of Physics Websites Based on STEM Assisted of Google Sites on Momentum and Impulse
- Niedderer, H. (n.d.). *Physics Learning as Cognitive Development*. <https://www.researchgate.net/publication/241084418>
- Ningsih, S., & Sopyan, A. (2012). Unnes Physics Education Journal IMPLEMENTASI MODEL PEMBELAJARAN PROCESS ORIENTED GUIDED INQUIRY LEARNING (POGIL) UNTUK MENINGKATKAN KEMAMPUAN BERPIKIR KRITIS SISWA. UPEJ, 1(2). <http://journal.unnes.ac.id/sju/index.php/upej>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Nurhasanah, F., & Akrom, M. (2021). Profile of Senior High School Students' Mathematical Reasoning Ability Based on Guardian Personality Type. In *International Online Journal of Education and Teaching (IOJET)* (Vol. 8, Issue 1).
- Novarensa, L. A., (2024). Pengembangan E-book Praktikum Inkuiri Terbimbing Berbasis Liveworksheet Berbantuan Analisis Software Tracker pada Materi Modulus Elastisitas Bahan untuk Meningkatkan Kemampuan Berpikir Kritis. *Unnes Physics Education Journal* 13 (1) (2024)
- OECD. (2019). *PISA 2021 mathematics framework*. OECD Publishing.
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351–371. <https://doi.org/10.1007/BF02213420>
- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.
- Paivio, A. (1990). *Dual Coding Theory*. In *Mental Representations: A Dual Coding Approach (Pp. 583–605)*. Oxford: Oxford University Press.
- Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. *Journal of Educational Psychology*, 86(1), 122–133.
- Purwanto, J., & Winarti. (2016). Profil pembelajaran fisika dan kemampuan berpikir kritis siswa Madrasah Aliyah se-DIY. *Jurnal Penelitian Pembelajaran Fisika*, 7(1), 8–18. <http://e-jurnal.upgrismg.ac.id/index.php/JP2F>
- Rahmawati, I., Sutopo, & Zulaikah, S. (2017). Analysis of students' difficulties about rotational dynamics based on resource theory. *Jurnal Pendidikan IPA Indonesia*, 6(1), 95–102. <https://doi.org/10.15294/jpii.v6i1.9514>
- Rafika, R., & Syuhendri, S. (2021). Students' misconceptions on rotational and rolling motions. *Journal of Physics: Conference Series*, 1816(1), 012016.

- <https://doi.org/10.1088/1742-6596/1816/1/012016>
- Redish, E. F. (2005). *Teaching Physics with the Physics Suite*. John Wiley & Sons.
- Redish, E. F. (1999). Millikan Award Lecture (1998): Building a science of teaching physics. *American Journal of Physics*, 67(7), 564–573. <https://doi.org/10.1119/1.19337>
- Redish, E. F., & Kuo, E. (2015). Language of physics, language of math: Disciplinary culture and dynamic epistemology. *Science & Education*, 24(5–6), 561–590. <https://doi.org/10.1007/s11191-015-9749-7>
- Safitri, R. (2024). Pengembangan E-book Fisika Berbasis STEM-PBL untuk Meningkatkan Kemampuan Berpikir Kreatif dan Representasi Matematis Peserta Didik Kelas XI SMA. Tesis, Program Magister Pendidikan Fisika. Yogyakarta: Universitas Negeri Yogyakarta.
- Samsudin, A., Afif, N. F., Nugraha, M. G., Suhandi, A., Fratiwi, N. J., Aminudin, A. H., Adimayuda, R., Linuwih, S., & Costu, B. (2021). Reconstructing Students' Misconceptions on Work and Energy through the PDEODE*E Tasks with Think-Pair-Share. *Journal of Turkish Science Education*, 18(1), 118–144. <https://doi.org/10.36681/tused.2021.56>
- Sari, D. P., Darhim & Rosjanuardi, R. (2018). Errors of Students Learning with React Strategy in Solving Problems of Mathematical Representation Ability. *Journal on Mathematics Education*, 9(1). <https://doi.org/10.22342/jme.9.1.4378.121-128>.
- Sari, I. K., Handayani, R. D., & Supriadi, B. (2024). Students' Creative Thinking Skills on The Material of Motion Dynamics Through STEM-PjBL. *Jurnal Riset dan Kajian Pendidikan Fisika*, 11(1), 41–48. <https://doi.org/10.12928/jrkpf.v11i1.656>
- Serevina, V., & Mulyati, D. (2015). Peningkatan Hasil Belajar Siswa pada Materi Dinamika Gerak Partikel Dengan Menerapkan Model Pembelajaran Project Based Learning. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 01(1), 61–68. <https://doi.org/10.21009/1.01109>
- Sherin, B. L. (2001). How students understand physics equations. *Cognition and Instruction*, 19(4), 479–541. https://doi.org/10.1207/S1532690XCI1904_3
- Suarnianti, N. W., Hidayah, N. and Handarini, D. M. (2018). The Development of Learning Tools to Improve Students' Critical Thinking Skills in Vocational High School. *In IOP Conference Series: earth and Environmental Science*, Vol. 175 (No. 1. P. 012095). <https://doi.org/10.1088/1755-1315/175/1/012095>
- Sugiharti, N., & Gayatri, Y. (2021). Profil Kemampuan Berpikir Kritis Siswa SMA Muhamaddiyah Kota Surabaya Pada Pembelajaran Biologi. *Pedago Biologi: Jurnal Pendidikan dan Pembelajaran Biologi*, 9(1), 34-40.
- Susanti, E., Septiana, S., Meilinda, S., & Rosa, I. M. (2023). Effectiveness of Using Google Sites-Based E-Modules to Optimize Critical Thinking Skills: Student Perceptions Analysis. *Jurnal Penelitian Pendidikan IPA*, 9(12), 10555–10561. <https://doi.org/10.29303/jppipa.v9i12.5887>

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Toksoy, S. E., & Akdeniz, A. R. (2015). Determining Student Difficulties in Solving Problems Related to Force and Motion Units Via Hint Cards. *Education and Science*, 40(180), 343–362. <https://doi.org/10.15390/EB.2015.3817>
- Van Heuvelen, A. (1991). Learning to Think Like A Physicist: A Review of Research-Based Instructional Strategies. *American Journal of Physics*, 59(10), 891–897.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: MESA Press.
- Wulandari, W. N., Ardhanariswari, W. E., Khalidah, H. I., & Setiaji, B. (2021). Analysis of High School Students' Learning Difficulties in Understanding The Mechanics Concept. *IMPULSE. Journal of Research and Innovation in Physics Education*, 1(2), 72–79. <http://ejournal.uin-suka.ac.id/tarbiyah/impulse>
- Wuryanti, S. (2017). Analisis Miskonsepsi Siswa Pada Materi Dinamika Gerak Menggunakan Tes Diagnostik Pilihan Ganda Tiga Tingkat. *Jurnal Geliga Sains (Vol. 5, Issue 2)*.
- Yusnita Fitrianna, A., Dinia, S., & Yuliani Nurhafifah, A. (2018). Mathematical Representation Ability of Senior High School Students: An Evaluation from Students' Mathematical Disposition. *Journal of Research and Advances in Mathematics Education (Vol. 3, Issue 1)*. <http://journals.ums.ac.id/index.php/jramathedu>
- Zulaiha, F. (2018). Profil Keterampilan Berpikir Kritis Siswa dalam Penggunaan Worksheet dan Problemsheet Menggunakan Multi Modus Representasi. *JPFS: Jurnal Pendidikan Fisika dan Sains*, 1(1), 28–32. <http://journal.unucirebon.ac.id/index.php/jpfs>
- Zulaiha, F. (2018). Profil Keterampilan Berpikir Kritis Siswa dalam Penggunaan Worksheet dan Problemsheet Menggunakan Multi Modus Representasi. *Jurnal Pendidikan Fisika dan Sains (JPFS)*, 1(1), 27–31.