Online Learning in Educational Research
Volume 6, Issue 1, 277-296.
e_ISSN: 2798-2351
https://www.journal.foundae.com/index.php/oler/index
DOI: 10.58524/oler.v6i1.1218

Mapping Students' Cognitive Profiles on Newtonian Mechanics: A Rasch-Based Analysis via Four-Tier Digital Diagnostic Assessment

Viyanti Viyanti*
Universitas Lampung, INDONESIA

Dinda Tara Cahyani
Universitas Lampung, INDONESIA

Fatkhur Rohman
Universitas Lampung, INDONESIA

Article Info

Article history:

Received: November 17, 2025
Revised: February 15, 2026
Accepted: March 10, 2026

Keywords:

Cognitive profile;
Diagnostic assessment;
Four-tier instrument;
Newtonian Mechanics;
Rasch Model.

Abstract

This study aims to develop a four-tier digital diagnostic instrument based on Google Forms and utilize it to map the cognitive ability profiles of high school students on Newton's Second Law across six levels of the Revised Bloom's Taxonomy. This study adopted a Research and Development design using the Research Design and Development model, and the resulting instrument was administered to 92 10th-grade students at a public high school in Bandar Lampung. Content validity was assessed by three experts with an average validity of 95.23%, while psychometric quality was analyzed using the Rasch Partial Credit Model through Ministep 5.10.0.0 based on 75 response records, resulting in a person reliability of 0.86 and a person separation of 2.47, sufficient to differentiate between student ability strata. Since each cognitive level is represented by only a few items, the overall psychometric quality of the instrument was evaluated using Rasch analysis, while the overall and Bloom-level cognitive profiles were mapped descriptively using achievement percentages. Profile mapping revealed high overall performance. Student achievement remained stable at the remembering, understanding, and applying levels, but declined at the analyzing and evaluating levels. C5 emerged as the lowest point, with significant student accumulation in the Sufficient category, particularly in reasoning about the role of friction and resultant force under balanced force conditions. This tiered profile allows teachers to design remediation that targets cognitive levels where students' understanding is weak, rather than simply re-teaching the material in its entirety.

To cite this article: Viyanti, V., Cahyani, D. T., & Rohman, F. (2026). Mapping students' cognitive profiles on newtonian mechanics: A rasch-based analysis via four-tier digital diagnostic assessment. *Online Learning in Educational Research, 6*(1), 277-296. https://doi.org/10.58524/oler.v6i1.1218

INTRODUCTION

Misconceptions about Newton's Second Law are not errors that disappear with regular practice; rather, they represent the most persistent cognitive barrier in physics learning and one of the main causes of poor problem-solving skills in STEM fields (Adimayuda et al., 2025; Arlianty et al., 2018; Bransford et al., 2000; Tan et al., 2023; Winarti et al., 2021). Quantitatively, this persistence is evident in 40% of students who exhibit active misconceptions, while 53.33% have knowledge gaps they are unaware of; moreover, these proportions worsen when the concept is applied to more complex phenomena (Liu & Fang, 2016; Mustofa et al., 2024; Suhandi et al., 2025). The root of this persistence is structural, students' understanding of the relationship between force, mass, and acceleration is shaped by everyday intuitions that are fundamentally at odds with Newton's principles, so their incorrect answers represent an alternative framework that is internally coherent but scientifically invalid (Bouchée et al., 2022; Maison et al., 2021; Novela et al., 2025; Resbiantoro & Setiani, 2022; Vicovaro, 2023). The consequences are pedagogical and immediate. Interventions

* **Corresponding Author**
Viyanti, Universitas Lampung, INDONESIA ✉ viyanti.1980@fkip.unila.ac.id

designed without first mapping what students believe and how strongly they hold those beliefs merely address the symptoms, not the underlying belief structures that produce them.

This structure of belief is, in fact, beyond the reach of diagnostic tools used in physics classrooms, such as conventional multiple-choice tests that merely distinguish between right and wrong and thus cannot differentiate between students who answer correctly because they understand the concept and those who answer correctly by guessing, as a result, conceptual uncertainty is misinterpreted as competence, leading to the entrenchment of misconceptions as the complexity of the material increases (Curtis et al., 2013; Foster et al., 2022; Grazziotin-Soares et al., 2021; Malaterre et al., 2023; Morton et al., 2008). Qualitative methods such as interviews and concept mapping yield richer diagnostic information, but they are not scalable to large classes and do not produce psychometrically measurable data (Chen, 2009; Kaltakci-Gurel et al., 2017). Multi-level diagnostic tests bridge this tension between depth and measurability. The addition of a reasoning layer in two-level tests and a confidence layer in three-level tests progressively enriches the information obtained (Chu et al., 2010; Taslidere, 2016; Yang & Sianturi, 2019; Tsui & Treagust, 2010), but only four-level tests close the final blind spot by assigning confidence not only to answers but also to reasoning, thereby allowing students to be simultaneously classified into scientific conceptions, misconceptions, knowledge gaps, and errors (Foster et al., 2022; Istiyono et al., 2023; Linn, 2008; Ma et al., 2025; Maharani et al., 2019; Soeharto et al., 2019; Uswatun & Mubarak, 2024). It is this ability to distinguish false beliefs from unconscious ignorance that qualitatively distinguishes four-level tests from lower-level instruments.

Four-level tests are typically administered in paper-based format, making data collection slow, error-prone, and unable to provide real-time feedback, which is precisely the reason for conducting the assessment (Choiriyah et al., 2025; Septiyani & Nanto, 2021; Sutrisno et al., 2023; Zainuri & Saepuloh, 2022). Google Forms-based administration eliminates these constraints due to its flexibility, the integration of visual stimuli, which are essential for physics questions, and the instant recording of responses (Gustina et al., 2025; Lubis et al., 2022; Wedana et al., 2025). Another challenge with four-level tests is psychometric in nature. The quality of four-level instruments is generally evaluated using Classical Test Theory, whose parameters depend on the sample and cannot be generalized to different student populations (Laliyo et al., 2019; Petrillo et al., 2015). The Rasch model overcomes these limitations by transforming ordinal responses into an interval scale that is invariant across samples and by placing student ability and item difficulty on a single logit continuum (Sari et al., 2023; Supurwoko et al., 2025; Zain et al., 2025; Viyanti et al., 2025). When Google Forms administration and Rasch measurement are integrated within a four-level framework, the most sophisticated diagnostic logic can finally be applied to real-world classroom scales without sacrificing measurement precision.

Four-level tests, digital delivery, and Rasch analysis have each been well developed, but they are typically used in isolation and rarely combined in a single study. Previous studies that developed four-level tests have proven reliable in identifying misconceptions on the topics of static fluids, harmonic motion, and Newton's laws; however, these instruments have rarely been subjected to Rasch-based psychometric evaluation (Janah et al., 2021; Maharani et al., 2019; Önder Çelikkanlı & Kızılcık, 2022; Putranta & Afifah, 2025). Studies that have shifted diagnosis to digital platforms have improved administrative efficiency but have treated digitization as a logistical issue and overlooked measurement precision (Nahadi et al., 2018; Wedana et al., 2025). Studies applying the Rasch Model in a diagnostic context stop at verifying item quality and have not yet translated logit scores into cognitive profile mapping (Fulmer et al., 2015; Istiyono et al., 2023; Mahmudi et al., 2026; Mardian et al., 2023; Romine et al., 2015). Even when four-level tests and the Rasch Model are combined, the Rasch Model is used only to ensure item quality. As a result, the diagnostic categories from the four-level test and the ability measures from the Rasch model have never been integrated into a single result that shows teachers at which cognitive level students' understanding begins to break down. This study addresses this gap by repositioning the Rasch Model from a mere validation tool to a psychometric framework for evaluating the overall quality of the diagnostic instrument and supporting the interpretation of students' cognitive profiles. The study aimed to develop a four-tier digital diagnostic instrument based on Google Forms and use it to map high school students' cognitive profiles based on Newton's Second Law across the six levels of the Revised Bloom's Taxonomy. Rasch calibration was used to establish the instrument's psychometric quality and unidimensionality, and the validated instrument was then used to characterize student achievement at each cognitive level. Because the instrument's measurement quality was assured through Rasch calibration, the resulting cognitive profiles can be reliably interpreted to indicate the cognitive levels at which students' mastery begins to decline. The resulting profiles can be read directly as a basis for differentiated remediation targeting specific cognitive levels, rather than the class as a whole. Therefore, its methodological contribution lies in demonstrating that a Rasch-validated four-tier instrument can produce pedagogically meaningful cognitive profiles.

METHOD

This study employed a Research and Development (R&D) design through the Design and Development Research (DDR) framework, encompassing the analysis, design, development, and evaluation phases (Richey & Klein, 2014). The development phase served as a psychometric standardization process to produce a precise cognitive lens (Aris et al., 2025), which served as the foundation before the instrument was used to map students' cognitive profiles based on Newton's Second Law. The research flow is presented in Figure 1.

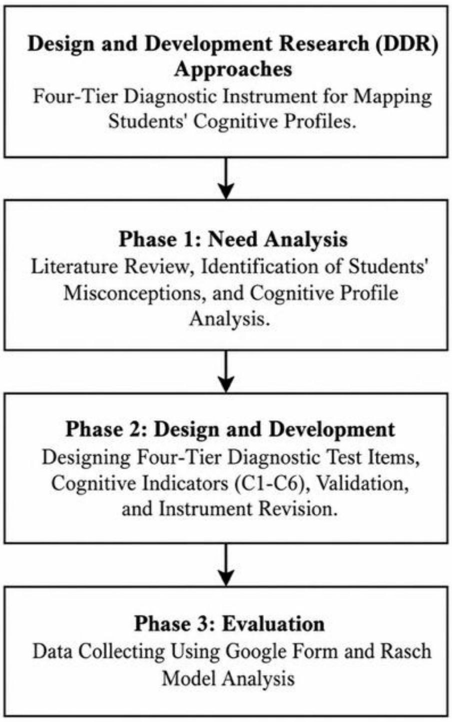


Figure 1. Research Flowchart

The study involved 92 tenth-grade students from a public high school in Bandar Lampung, selected through purposive sampling based on the criterion of having completed the unit on Newton's Second Law. The responses from all 92 students were used to map students' cognitive profiles across the six levels of the Revised Bloom's Taxonomy using descriptive achievement percentages. For the psychometric analysis, 75 response records were analyzed using the Rasch Partial Credit Model. This sample size was adequate for Rasch analysis. For both dichotomous and polytomous applications, a sample size of 64 to 144 respondents is sufficient to obtain stable item calibration within ±0.5 logit at the 95% confidence level (Linacre, 1994). Because the person and item estimates in the Rasch model are essentially independent of the sample, this size provides an adequate basis for the psychometric claims in this study, although the findings still need to be interpreted within the scope of the single-school sample and not generalized to the broader population.

The instrument developed was a four-tiered diagnostic test based on Google Forms to ensure accessibility and efficiency of data collection. The first level contains multiple-choice questions on the concept of Newton's Second Law, the second level records the student's level of confidence in the answer on a scale of Sure and Not Sure, the third level presents a choice of conceptual reasons underlying the first-level answer, and the fourth level records the student's level of confidence in the chosen reason. By combining answer accuracy, reason accuracy, and confidence, this format is essentially able to separate scientific conceptions, misconceptions, knowledge deficits, and errors, which distinguishes the four-level test from lower-level instruments that cannot separate mistaken beliefs from unconscious ignorance. The distribution of the twenty items across the six levels of the revised Bloom's Taxonomy, along with the corresponding learning indicators, is presented in the instrument grid in Table 1.

Table 1. Instrument Grid

No	Learning objectives	Cognitive Level	Items
1	Remembering the basic concepts, equations, and units of Newton's Second Law	C1 Remembering	1,2,3
2	Understanding the relationship between force, mass, and acceleration, and their applications	C2 Understanding	4,5,6
3	Applying Newton's Second Law to problem solving	C3 Applying	7,8,9
4	Analyzing the effect of force on the motion of an object	C4 Analyzing	10,11,12
5	Evaluating the motion of an object based on the forces acting on it	C5 Evaluating	13,14,15,16
6	Designing solutions based on Newton's Second Law	C6 Creating	17,18,19,20

Content validity was determined through the assessment of three experts who assessed the instrument on three aspects, namely construct, material, and language, using a rating scale for each indicator (Arikunto, 2010). The score for each aspect was converted into a percentage of the maximum achievable score, and the instrument validity was expressed as the average percentage of the three aspects. This procedure resulted in an average content validity of 95.23%, which is categorized as very high and indicates strong agreement among experts regarding the relevance of the items to the constructs being measured. After content validity was met, empirical responses were analyzed using a polytomous Rasch model, namely the Partial Credit Model, with the help of Ministep 5.10.0.0, because each item is scored on a graduated scale from 1 to 5. The Rasch model transforms ordinal responses into an interval logit scale and places student ability and item difficulty levels on the same continuum, so that the measurement is invariant across samples. The psychometric feasibility of the instrument was examined through the reliability and separation of persons and items, as well as through item suitability, with the MNSQ Infit and Outfit values in the range of $0.5 < MNSQ < 1.5$ as evidence that each item measures one dimension without significant bias (Bond, 2015; Boone et al., 2014).

The validated instrument was then used to map students' cognitive profiles. Because each Bloom's level is represented by only three to four items, this number is insufficient to support stable Rasch parameter estimates at the individual cognitive level, and thus, Rasch stratification by level would rely on unreliable individual estimates. Therefore, the overall psychometric quality of the instrument was evaluated using Rasch analysis, while the overall and Bloom-level cognitive profiles were mapped descriptively using achievement percentages. The percentage of each student's score relative to the maximum score was calculated and classified into five categories, followed by Very Good (81–100%), Good (61–80%), Fair (41–60%), Poor (21–40%), and Very Poor (below 21%). This combined strategy applied the Rasch model to its psychometrically robust areas, validating the instrument and characterizing overall ability, while simultaneously using descriptive analysis for the level profiles, where the limited number of items makes individual Rasch estimates unstable.

RESULTS AND DISCUSSION

The research that has been conducted is development research by producing a product in the form of a four-tier diagnostic instrument to identify students' cognitive profiles in the material on Newton's Second Law using Google Form.

Validity

The four-tier diagnostic instrument underwent expert validation testing to assess its feasibility, encompassing three assessment aspects: construct, content, and language, involving three experts as validators. The results of the expert validation test are presented in Table 2.

Table 2. Expert Validation Results of Four-Tier Diagnostic Instruments

No	Aspect	Number of Indicator	Max Score per Expert	Expert 1	Expert 2	Expert 3	Total	Validity (%)	Category
1	Construct	6	120	112	120	120	352	97.78%	Very Valid
2	Material	7	140	109	139	140	388	92.38%	Valid
3	Language	3	60	55	58	59	172	95.56%	Very Valid
Average percentage of assessment								95.24%	Very Valid

Based on the data in Table 2, it is confirmed that the questions compiled have a very high standard of suitability in terms of construction, material, and language aspects (Arikunto, 2010). Expert validation is a crucial step in developing digital assessments that effectively identify student thinking errors. Strong content validity is crucial for the instrument's ability to accurately identify student thinking errors and distinguish them from mere technical errors (Arief et al., 2012). After the instrument was declared valid by the three validators, an empirical validity test was conducted. The trial involved 92 10th-grade students at SMAN 1 Bandar Lampung. The responses from all 92 students were used for descriptive cognitive-profile mapping, while 75 response records were included in the Rasch analysis. The results of the empirical validity analysis are presented in Table 3.

Table 3. The Fit of Item to Rasch Models

MEANSURE	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR	Item
-0.23	0.84	-0.27	0.41	S2
0.10	1.09	0.40	0.66	S20
-0.23	0.48	-1.44	0.58	S3
-0.21	0.73	-0.61	0.54	S5
0.23	1.06	0.35	0.74	S16
0.12	1.15	0.63	0.71	S15
0.44	1.48	2.12	0.56	S12
0.07	0.83	-0.55	0.71	S17
0.06	1.45	1.52	0.65	S14
-0.14	1.13	0.45	0.53	S8
-0.07	1.19	0.65	0.60	S19
0.04	1.17	0.64	0.54	S13
-0.07	0.84	-0.39	0.64	S6
0.24	0.85	-0.63	0.47	S1
-0.05	0.50	-1.80	0.70	S4
0.18	0.77	-0.97	0.74	S10
-0.13	0.87	-0.24	0.65	S9
-0.09	0.73	-0.74	0.68	S7
-0.24	0.55	-1.16	0.69	S18
-0.02	0.59	-1.45	0.80	S11

The Rasch Model analysis presented in Table 3 shows that the majority of test items meet the fit criteria according to Boone (2016) the Rasch model parameters, with Outfit MNSQ values in the range of 0.5–1.5. Items with Outfit MNSQ values in the range of 0.5 to 1.5 indicate that the instrument is in accordance with the model used and does not contain bias (Li et al., 2023). In the special case of

Question 12, even though the Outfit ZSTD value is outside the range that indicates a cognitive barrier in the form of faulty logic, the item is retained because the MNSQ value (1.48) and its Point Measure Correlation are still within the statistical tolerance limits (Sumintono, 2015). Students with misconceptions tend to show incorrect but consistent answer patterns, often caused by excessive cognitive load when understanding the relationship between force and acceleration (Sweller, 1988). Furthermore, several items, such as S16, S15, and S10, demonstrated high PT-Measure Correlation values (above 0.70), indicating that they were effective in mapping students' cognitive profiles. This item clarity ensures that the instrument is not only statistically valid but also effective as a "cognitive lens" for detecting logic errors and cognitive doubts frequently experienced by students in mechanics (Novela et al., 2025).

Mapping cognitive profiles through Wright Map analysis indicates that students' abilities are uneven. This finding aligns with Maulina (2026), who stated that accuracy in identifying profiles is significantly influenced by students' self-efficacy in dealing with conceptual problems. The use of Google Forms as a diagnostic assessment platform is not only a media innovation but also a strategic step in providing real-time data for teachers in designing more targeted and personalized STEM learning (Csapó & Molnár, 2019).

Reliability

The developed diagnostic instrument was then subjected to a reliability test to ensure its reliability and suitability for use. The reliability of the four-level diagnostic test was analyzed using the Cronbach's Alpha equation. Cronbach's Alpha measures the interaction between person reliability and item reliability (Vaske et al., 2017). The results of the reliability analysis for the four-level diagnostic test items are shown in Tables 4 and 5.

Table 4. Person Reliability Analysis of Four-Tier Diagnostic Instruments

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	81.2	20.0	-1.45	0.22	0.93	0.01	0.91	-0.04
SEM	2.7	0.0	0.08	0.01	0.04	0.10	0.06	0.10
P.SD	23.4	0.0	0.65	0.07	0.37	0.87	0.52	0.88
S.SD	23.6	0.0	0.65	0.07	0.37	0.87	0.52	0.88
MAX.	210.0	20.0	0.40	0.35	3.11	4.96	4.56	4.75
MIN.	32.0	20.0	-3.05	0.05	0.38	-2.29	0.24	-2.29
REAL RMSE 0.24		TRUE SD 0.60		SEPARATION 2.47		PERSON RELIABILITY 0.86		
MODEL RMSE 0.23		TRUE SD 0.60		SEPARATION 2.57		PERSON RELIABILITY 0.87		
S.E. OF PERSON MEAN = 0.08								
PERSON RAW SCORE-TO-MEASURE CORRELATION = 0.75								
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = 0.94 SEM = 5.84								
STANDARDIZED (50 ITEM) RELIABILITY = 0.94								

Table 5. Item Reliability Analysis of the Four-Tier Diagnostic Instrument

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	304.4	75.0	0.00	0.10	1.32	0.93	0.91	-0.17
SEM	4.6	0.0	0.04	0.00	0.12	0.35	0.07	0.23
P.SD	20.2	0.0	0.18	0.00	0.52	1.51	0.28	0.99
S.SD	20.7	0.0	0.18	0.00	0.53	1.55	0.29	1.01
MAX.	331.0	75.0	0.44	0.10	2.55	3.62	1.48	2.12
MIN.	255.0	75.0	-0.24	0.09	0.54	-2.36	0.48	-1.80
REAL RMSE 0.11		TRUE SD 0.14		SEPARATION 1.23		ITEM RELIABILITY 0.60		
MODEL RMSE 0.10		TRUE SD 0.15		SEPARATION 1.58		ITEM RELIABILITY 0.71		
S.E. OF PERSON MEAN = 0.04								

Based on Table 4, the Cronbach's Alpha (KR-20) value reached 0.94, indicating very strong internal consistency between respondents and items (Sumintono, 2015). A Person Reliability value of 0.86 indicates the instrument's ability to consistently differentiate between various levels of student cognitive profiles, ranging from those with a good conceptual understanding to those with serious misconceptions. Table 5 shows a person reliability of 0.86 and a separation index of 2.47, demonstrating that the instrument consistently distinguishes between students' cognitive ability levels, ranging from those with strong conceptual mastery to those harboring misconceptions. The separation value of 2.47 indicates that students can be categorized into approximately two to three statistically distinct ability strata, making the instrument suitable for mapping cognitive profiles.

Meanwhile, the item reliability of 0.60 and separation of 1.23 are classified as moderate. This result can be attributed to several interrelated factors. The instrument comprises twenty items with a relatively narrow distribution of difficulty levels; consequently, the range of difficulty between items is insufficient to yield a high degree of separation. Furthermore, a large proportion of students achieved high scores on many items, indicating a ceiling effect and limited response variation, which reduced the information available for item separation. The limited sample size, involving only a single school, also constrained the stability of the item separation estimates (Payan-Carreira et al., 2022).

Nevertheless, the moderate item reliability does not invalidate the instrument's quality, provided the item fit statistics are met; as shown in Table 4, all items fell within the acceptable Infit and Outfit MNSQ ranges. These findings align with other research employing the Rasch approach for physics diagnostic instruments. Fratiwi et al. (2020) observed lower item reliability compared to person reliability in an instrument assessing Newton's Laws misconceptions, concluding that this pattern is common when the distribution of respondent ability is relatively homogeneous. This consistency suggests that high person reliability coupled with moderate item reliability is a characteristic often found in diagnostic instruments featuring a limited number of items per concept and samples drawn from relatively uniform populations.

Practicality

Instruments that met the criteria for validity and reliability were then tested for practicality. The practicality test aimed to determine the extent to which the instrument was easy for teachers to use and implement in the learning process (Viyanti et al., 2024). The results of the practicality assessment are presented in Table 6.

Table 6. Results of the Practicality Assessment of the Four-Tier Diagnostic

No.	Rated aspect	Score
1	Structure and Appearance	90
2	Ease of Use	100
3	Technical Practicality	100
	Final Score	96.66

Based on Table 6, the practicality assessment results achieved a final score of 96.66. The high scores for ease of use (100) and technical practicality (100) confirm that the integration of the four-tier diagnostic test into the Google Forms platform has minimized the administrative barriers often encountered with conventional paper-based tests (Maiyanti et al., 2025; Santosa & Jazuli, 2022). This practicality allows teachers to conduct mass diagnoses and obtain cognitive profile mapping results instantly (Ding et al., 2024).

The instrument's clear structure and intuitive interface (score of 90) support the creation of an assessment experience that does not overload students' cognitive capacity (cognitive load), allowing them to focus fully on scientific reasoning (Sweller, 1988). This digital flexibility also provides opportunities for educators to implement more responsive learning strategies (Yordan & Fahyuni, 2021). In line with findings (Sari et al., 2023), practical digital platforms like this are key to bridging teacher perceptions with students' actual needs in modern physics classrooms. With this instrument that has been proven valid, reliable, and practical, the analysis then continued with mapping the students' cognitive profiles to reveal the misconception map on Newton's Second Law material.

Cognitive Profile Mapping

Cognitive profiles are obtained through test results or measurements covering aspects of perception, reasoning, memory, and problem-solving, which serve to support the diagnostic process and learning planning (Marinda, 2020). According to Anderson's revised Bloom's taxonomy, there are six cognitive indicators: remembering, understanding, applying, analyzing, evaluating, and creating (Anderson & Krathwohl, 2001). Through a four-tier instrument, mapping is not only based on answer accuracy but also integrates the accuracy of the student's reasoning and level of confidence (Irianti, 2021). Based on the research results, the percentages of students' cognitive profiles are presented in Figure 2.

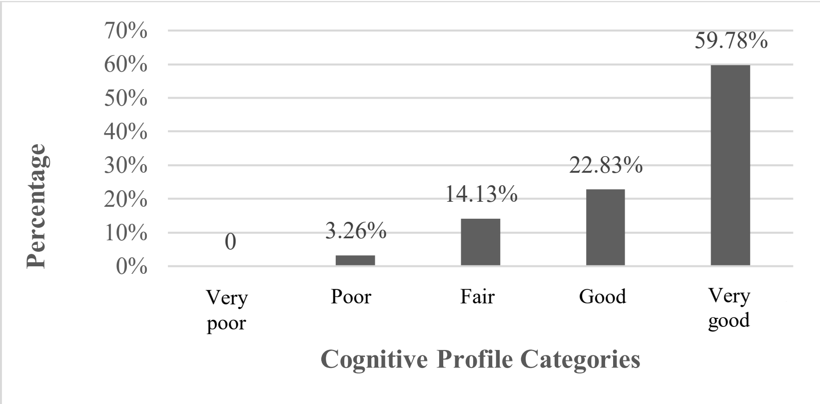


Figure 2. Percentage of Students' Cognitive Profiles

Visualization of the cognitive profile of students identified through the four-tier diagnostic instrument can be seen in Figure 2, the presence of students in the "Enough" (14.13%) and "Less" (3.26%) categories remains an important indicator of the presence of latent cognitive barriers that must be immediately intervened through more adaptive learning strategies (Nahak, 2024). These results indicate that the students' cognitive profiles are at a satisfactory level. This aligns with the analysis of six cognitive profile indicators, which indicate that students generally have relatively good cognitive abilities. The results of the analysis of students' cognitive profiles using the cognitive remembering indicator are presented in Figure 3.

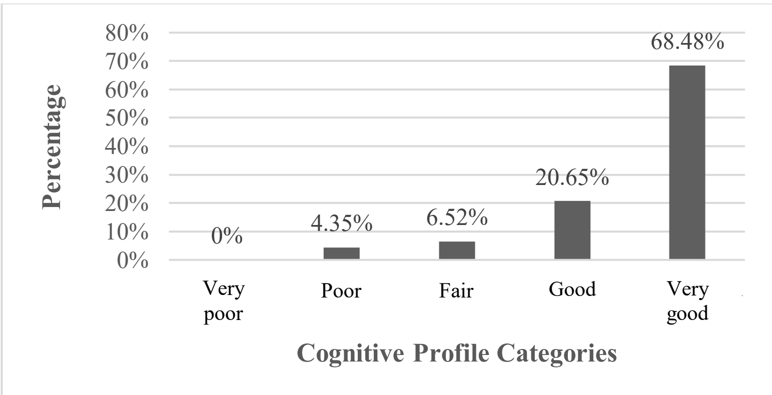
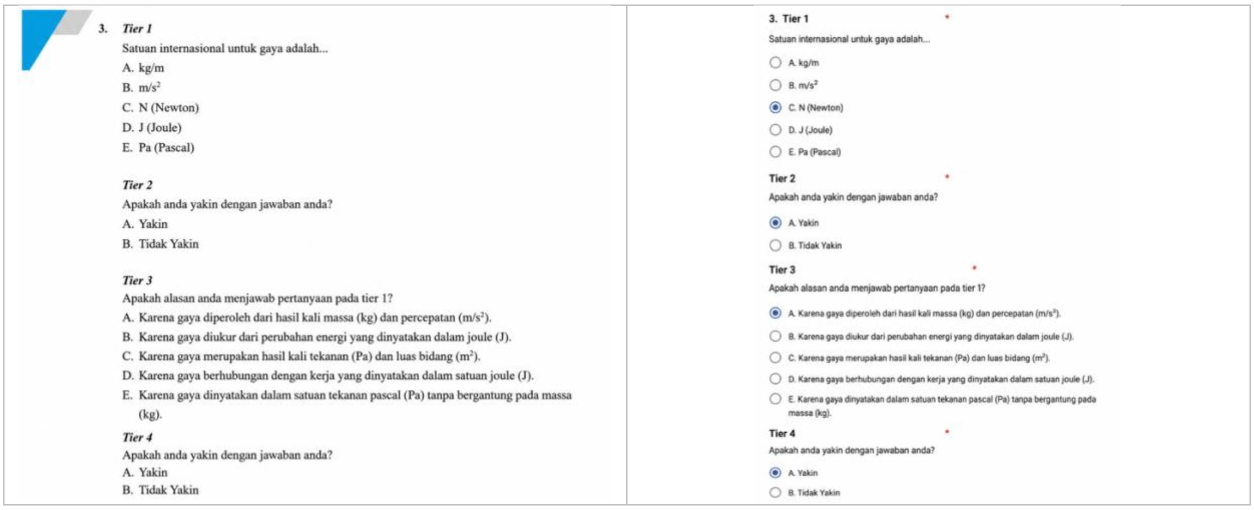


Figure 3. Cognitive Profile of Remembering Indicator (C1)

In Figure 3, 68.48% of students were in the "Very Good" category. This indicates that 63 students did not experience difficulties in the memory or retention aspect of the basic definition of Newton's Second Law. This is evidenced by the students' responses to Question Number 3, which requires the ability to identify the International System of Units (SI) of the force presented in Figure 4.



3. *Tier 1*
Satuan internasional untuk gaya adalah...
A. kg/m
B. m/s²
C. N (Newton)
D. J (Joule)
E. Pa (Pascal)

Tier 2
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 3
Apakah alasan anda menjawab pertanyaan pada tier 1?
A. Karena gaya diperoleh dari hasil kali massa (kg) dan percepatan (m/s²).
B. Karena gaya diukur dari perubahan energi yang dinyatakan dalam joule (J).
C. Karena gaya merupakan hasil kali tekanan (Pa) dan luas bidang (m²).
D. Karena gaya berhubungan dengan kerja yang dinyatakan dalam satuan joule (J).
E. Karena gaya dinyatakan dalam satuan tekanan pascal (Pa) tanpa bergantung pada massa (kg).

Tier 4
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Figure 4. Questions and Examples of Student Answers for Question Number 3

Figure 4 shows that recalling the definition, symbol, and unit of Newton's Second Law constitutes declarative knowledge, knowledge that can be acquired through repetition and rote memorization without requiring an understanding of the interrelationships between concepts. Consequently, the high level of achievement at the C1 level is expected and aligns with findings that students generally master the lowest cognitive level regarding Newton's Laws (Zhai, 2022). Students' success across all four levels in providing accurate scientific reasoning and trusting their own reasoning indicates that their foundational factual and conceptual knowledge is firmly established (Anderson & Krathwohl, 2001). However, some students still exhibit inaccuracies in conceptual understanding, particularly at the third level (reasoning). These inaccuracies are characterized by a tendency to associate force with other physical quantities, such as energy or pressure, leading students to select units like Joules (J) or Pascals (Pa). The emergence of uncertainty at the fourth level suggests that the students' understanding remains rooted in rote memorization and has not yet evolved into a comprehensive, functional understanding. This phenomenon of cognitive uncertainty underscores the need for early identification through diagnostic assessment to prevent the accumulation of misconceptions as students' progress to higher levels of thinking (Arief et al., 2012; Kuo et al., 2018; Novela et al., 2025). Nevertheless, mastery at this level does not guarantee conceptual understanding; the ability to state a formula differs fundamentally from the ability to grasp its meaning, a distinction that becomes apparent at higher cognitive levels. Figure 5 illustrates the cognitive profile for the "understanding" indicator (C2).

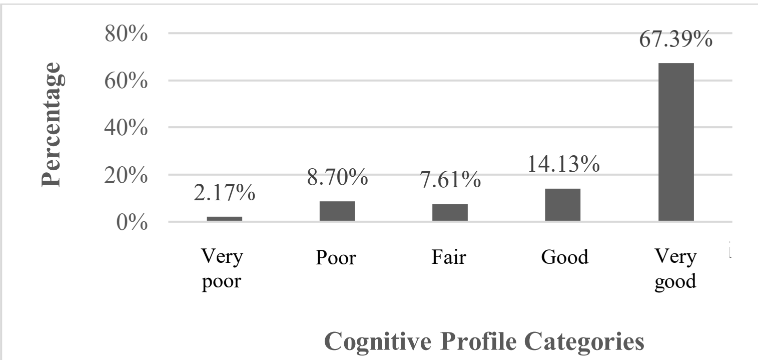


Figure 5. Cognitive Profile of the Understanding Indicator (C2)

Figure 5 shows that at the understanding level, the Very Good achievement decreased slightly, to 67.39, although it is still considered high. This slight decrease is because at the understanding level, students are required not only to memorize formulas, but also to explain the relationship between force, mass, and acceleration. Some students began to lose accuracy when having to translate formulas into physical meaning, for example, understanding that acceleration is inversely proportional to mass when the force is constant (Suwasono et al., 2025). This initial difficulty is in line with the view that conceptual understanding in mechanics requires students to build correct mental models of the relationships between quantities, not simply recognize symbols (Maulina et al., 2026; Novela et al., 2025; Yulia et al., 2022). The decrease from C1 to C2 marks the starting point where memorization begins to be inadequate. The shift in students' cognitive profiles is increasingly visible at a more complex level of thinking, namely in the cognitive ability to apply (C3), as presented in Figure 6.

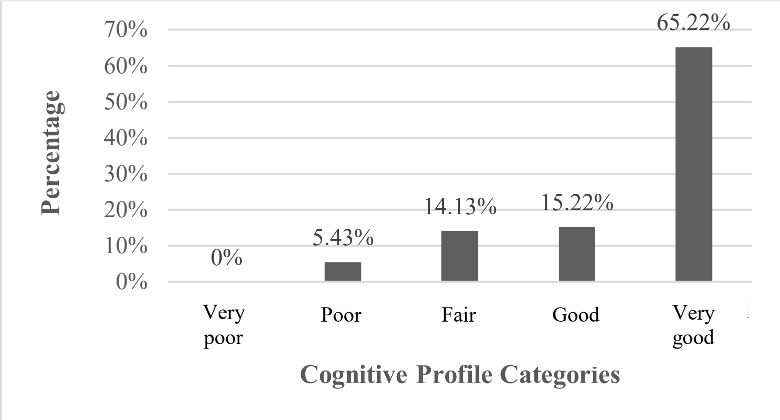


Figure 6. Cognitive Profile Applying (C3)

At the application level, Very Good achievement was recorded at 65.22%, still considered good but continuing the downward trend. At this level, students are required to use Newton's Second Law in calculations, for example, determining acceleration from known force and mass. The fact that C3 achievement is still relatively high indicates that students are quite capable of applying formulas procedurally (Weber & Wilhelm, 2024). The ability to apply formulas procedurally is not always supported by a strong conceptual understanding, because students can obtain correct answers through number substitution without understanding why the formula applies. This condition is consistent with the finding that students are often able to perform mathematical calculations without a deep understanding of physics (Sutopo, 2016). Figure 7 explains about analyzing Cognitive Profile (C4).

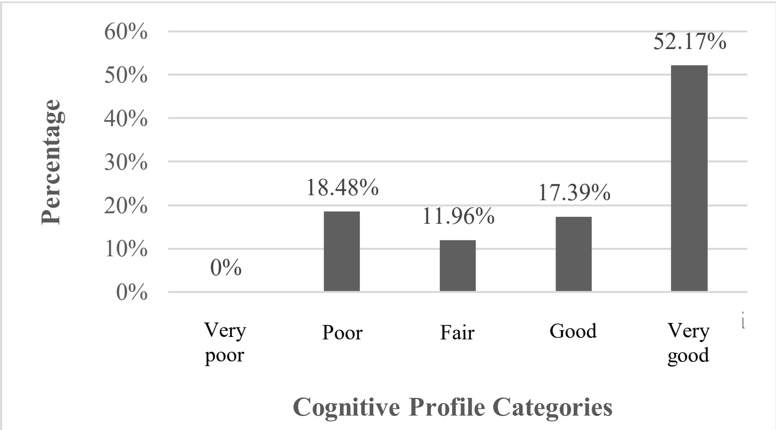
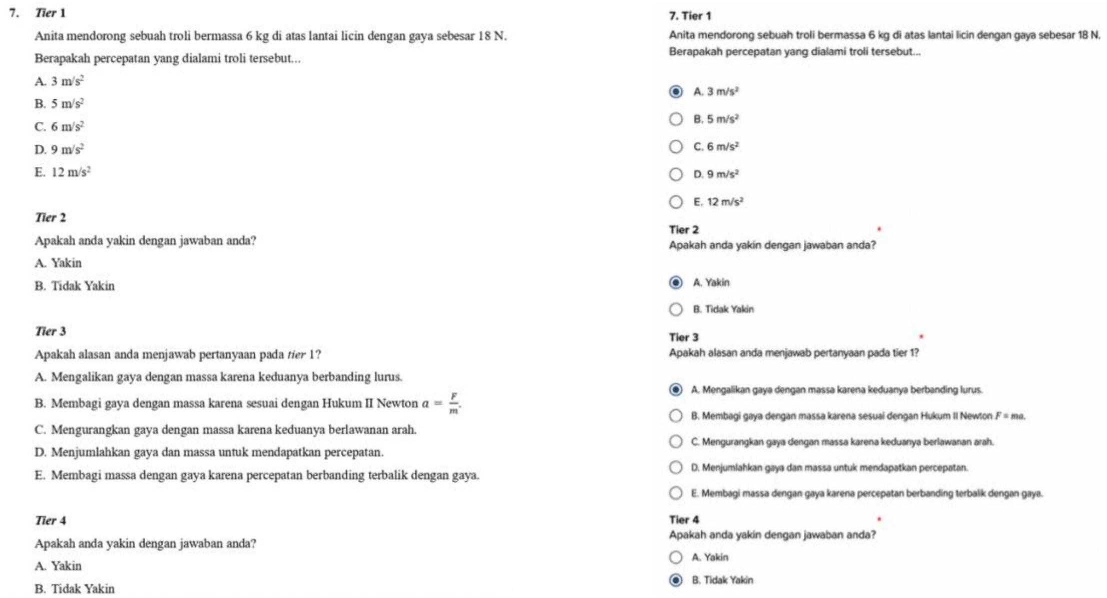


Figure 7. Analyzing Cognitive Profile (C4)

Based on Figure 7, cognitive challenges occur when students begin to struggle to integrate the concept of Newton's Second Law into various physical situations. The conceptual fragility hidden behind procedural success at C3 begins to be revealed at the analysis level. Achievements from "Very Good" dropped sharply, and more significantly, 18.48% of students fell into the "Poor" category. This decline reflects fundamentally different cognitive demands (Nahak, 2024; Sweller, 1988). At the analysis level, students no longer simply need to substitute formulas; instead, they must analyze the forces acting on an object and assess their effect on motion. This is where fundamental misconceptions about the relationship between force and motion begin to undermine reasoning, as students struggle to analyze situations where multiple forces act but the object remains at rest (Suwasono et al., 2025; Weber & Wilhelm, 2024). This difficulty in analyzing forces at equilibrium explains the surge in the proportion of students categorized as "Poor" at this level. This question requires students to use Newton's Second Law directly in the context of a simple calculation, namely determining the magnitude of the trolley's acceleration based on the known force and mass information as presented in Figure 8.


7. *Tier 1*
Anita mendorong sebuah troli bermassa 6 kg di atas lantai licin dengan gaya sebesar 18 N. Berapakah percepatan yang dialami troli tersebut...
A. 3 m/s^2
B. 5 m/s^2
C. 6 m/s^2
D. 9 m/s^2
E. 12 m/s^2

Tier 2
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 3
Apakah alasan anda menjawab pertanyaan pada *tier* 1?
A. Mengalikan gaya dengan massa karena keduanya berbanding lurus.
B. Membagi gaya dengan massa karena sesuai dengan Hukum II Newton $a = \frac{F}{m}$.
C. Mengurangkan gaya dengan massa karena keduanya berlawanan arah.
D. Menjumlahkan gaya dan massa untuk mendapatkan percepatan.
E. Membagi massa dengan gaya karena percepatan berbanding terbalik dengan gaya.

Tier 4
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

7. Tier 1
Anita mendorong sebuah troli bermassa 6 kg di atas lantai licin dengan gaya sebesar 18 N. Berapakah percepatan yang dialami troli tersebut...
(•) A. 3 m/s^2
( ) B. 5 m/s^2
( ) C. 6 m/s^2
( ) D. 9 m/s^2
( ) E. 12 m/s^2

Tier 2 *
Apakah anda yakin dengan jawaban anda?
(•) A. Yakin
( ) B. Tidak Yakin

Tier 3 *
Apakah alasan anda menjawab pertanyaan pada tier 1?
(•) A. Mengalikan gaya dengan massa karena keduanya berbanding lurus.
( ) B. Membagi gaya dengan massa karena sesuai dengan Hukum II Newton $F = ma$.
( ) C. Mengurangkan gaya dengan massa karena keduanya berlawanan arah.
( ) D. Menjumlahkan gaya dan massa untuk mendapatkan percepatan.
( ) E. Membagi massa dengan gaya karena percepatan berbanding terbalik dengan gaya.

Tier 4 *
Apakah anda yakin dengan jawaban anda?
( ) A. Yakin
(•) B. Tidak Yakin

Figure 8. Questions and Examples of Student Answers for Question Number 7

Based on Figure 8, to determine the acceleration of the trolley, students must understand the relationship between force, mass, and acceleration according to Newton's Second Law, namely that acceleration is directly proportional to force and inversely proportional to mass. Mapping of cognitive profiles at more complex cognitive levels, namely evaluating (C5) in Figure 9.

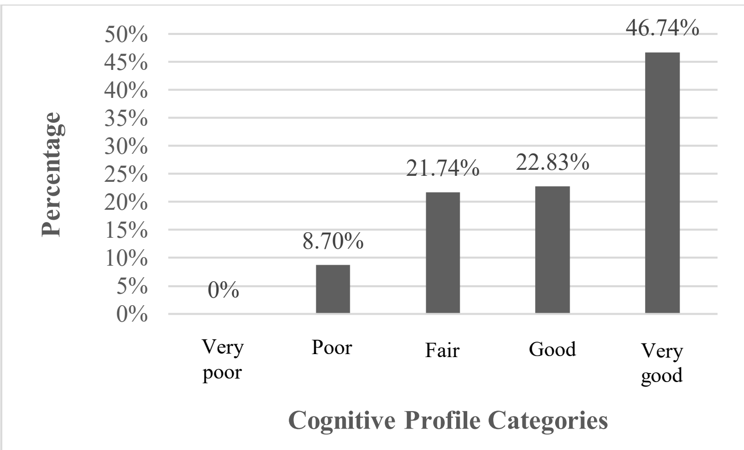


Figure 9. Evaluating Cognitive Profile (C5)

The lowest point of all cognitive profiles is at the evaluation level, with only 46.74% achieving Very Good, accompanied by a concentration of students in the Sufficient category of 21.74%. The finding is that the evaluation level is the weakest point. At the evaluation level, students are required to assess the truth of a statement about the conditions of motion based on the resultant force, for example, assessing why a cupboard that is pushed remains at rest. Such a task touches directly on the core of the most resistant misconception in mechanics, namely the belief that the presence of force always means motion. Rather than concluding that the pushing force and friction produce a zero resultant, students tend to conclude that there is no force at all, because for them, a stationary object is force-free (Bussotti, 2025). This belief is rooted in pre-Newtonian intuitive reasoning that has long been documented to persist even after formal instruction (Segado, 2025).

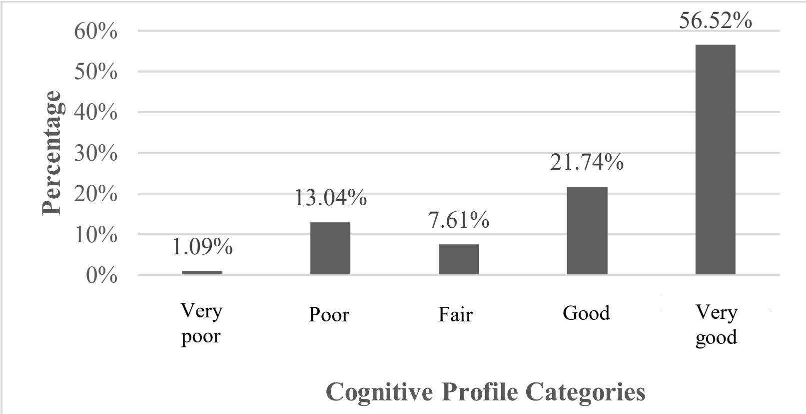
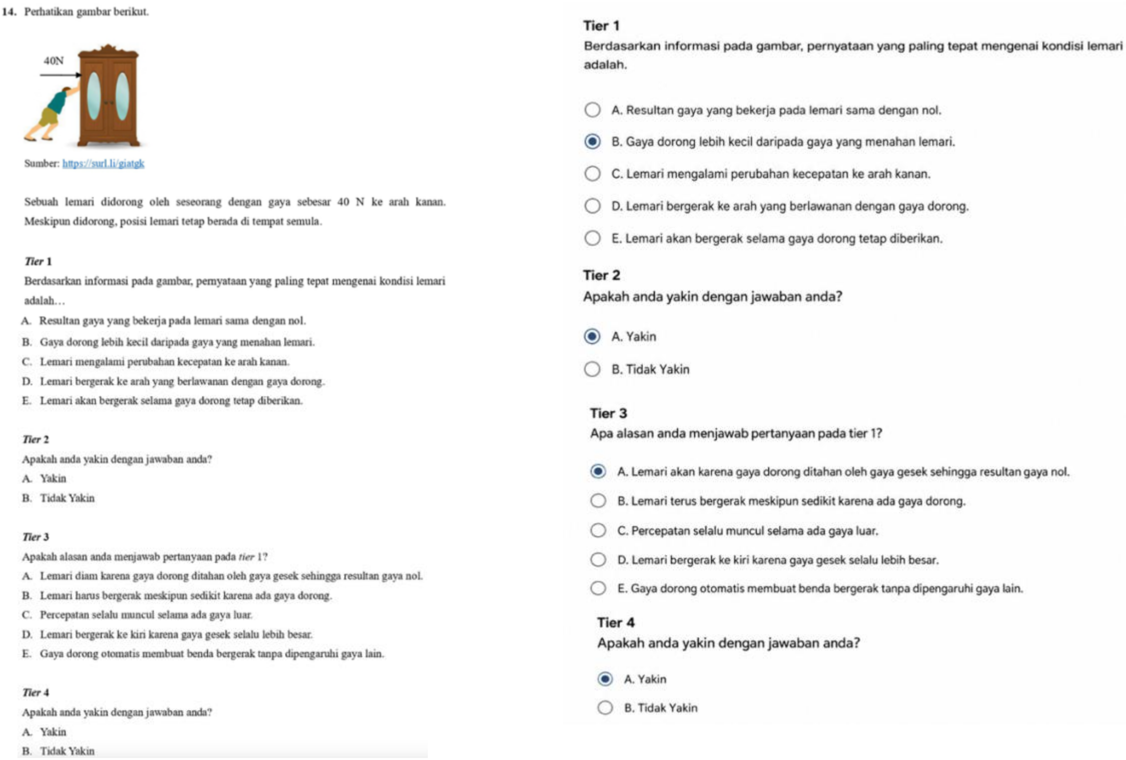


Figure 10. Cognitive Profile of Creation Indicators (C6)

The findings in Figure 10 show that student achievement at the creation level did not decline but rather recovered slightly, with the Very Good category at 56.52 percent, higher than the analysis and evaluation levels. Students had more mastery of the highest level because some of the problems at the creation level required students to design strategies based on the relationship between mass and acceleration that, when presented in a more structured context, could be solved by reapplying familiar reasoning patterns, thus not relying entirely on conceptual evaluation, which was a weak point in C5 (Chen, 2025; Sagatbek et al., 2024). Some students in the Poor category continued to show difficulty in constructing new solutions, as seen in the problem of preventing an object from falling from an accelerating cart, where students mistakenly chose a strategy that actually increased acceleration. This pattern confirms that the recovery at C6 reflects the different nature of the tasks between evaluating and creating. The study by Horcajo et al. (2022) confirmed that low self-efficacy directly contributes to students' uncertainty in completing complex cognitive tasks. This is confirmed through the analysis of Problem Number 14 presented in Figure 11, where students are asked to evaluate the equilibrium conditions of a static object subjected to a pushing force.



14. Perhatikan gambar berikut.

40N

Sumber: https://surl.li/giatgk

Sebuah lemari didorong oleh seseorang dengan gaya sebesar 40 N ke arah kanan. Meskipun didorong, posisi lemari tetap berada di tempat semula.

Tier 1
Berdasarkan informasi pada gambar, pernyataan yang paling tepat mengenai kondisi lemari adalah…
A. Resultan gaya yang bekerja pada lemari sama dengan nol.
B. Gaya dorong lebih kecil daripada gaya yang menahan lemari.
C. Lemari mengalami perubahan kecepatan ke arah kanan.
D. Lemari bergerak ke arah yang berlawanan dengan gaya dorong.
E. Lemari akan bergerak selama gaya dorong tetap diberikan.

Tier 2
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 3
Apakah alasan anda menjawab pertanyaan pada *tier* 1?
A. Lemari diam karena gaya dorong ditahan oleh gaya gesek sehingga resultan gaya nol.
B. Lemari harus bergerak meskipun sedikit karena ada gaya dorong.
C. Percepatan selalu muncul selama ada gaya luar.
D. Lemari bergerak ke kiri karena gaya gesek selalu lebih besar.
E. Gaya dorong otomatis membuat benda bergerak tanpa dipengaruhi gaya lain.

Tier 4
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 1
Berdasarkan informasi pada gambar, pernyataan yang paling tepat mengenai kondisi lemari adalah.
A. Resultan gaya yang bekerja pada lemari sama dengan nol.
B. Gaya dorong lebih kecil daripada gaya yang menahan lemari.
C. Lemari mengalami perubahan kecepatan ke arah kanan.
D. Lemari bergerak ke arah yang berlawanan dengan gaya dorong.
E. Lemari akan bergerak selama gaya dorong tetap diberikan.

Tier 2
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 3
Apa alasan anda menjawab pertanyaan pada tier 1?
A. Lemari akan karena gaya dorong ditahan oleh gaya gesek sehingga resultan gaya nol.
B. Lemari terus bergerak meskipun sedikit karena ada gaya dorong.
C. Percepatan selalu muncul selama ada gaya luar.
D. Lemari bergerak ke kiri karena gaya gesek selalu lebih besar.
E. Gaya dorong otomatis membuat benda bergerak tanpa dipengaruhi gaya lain.

Tier 4
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

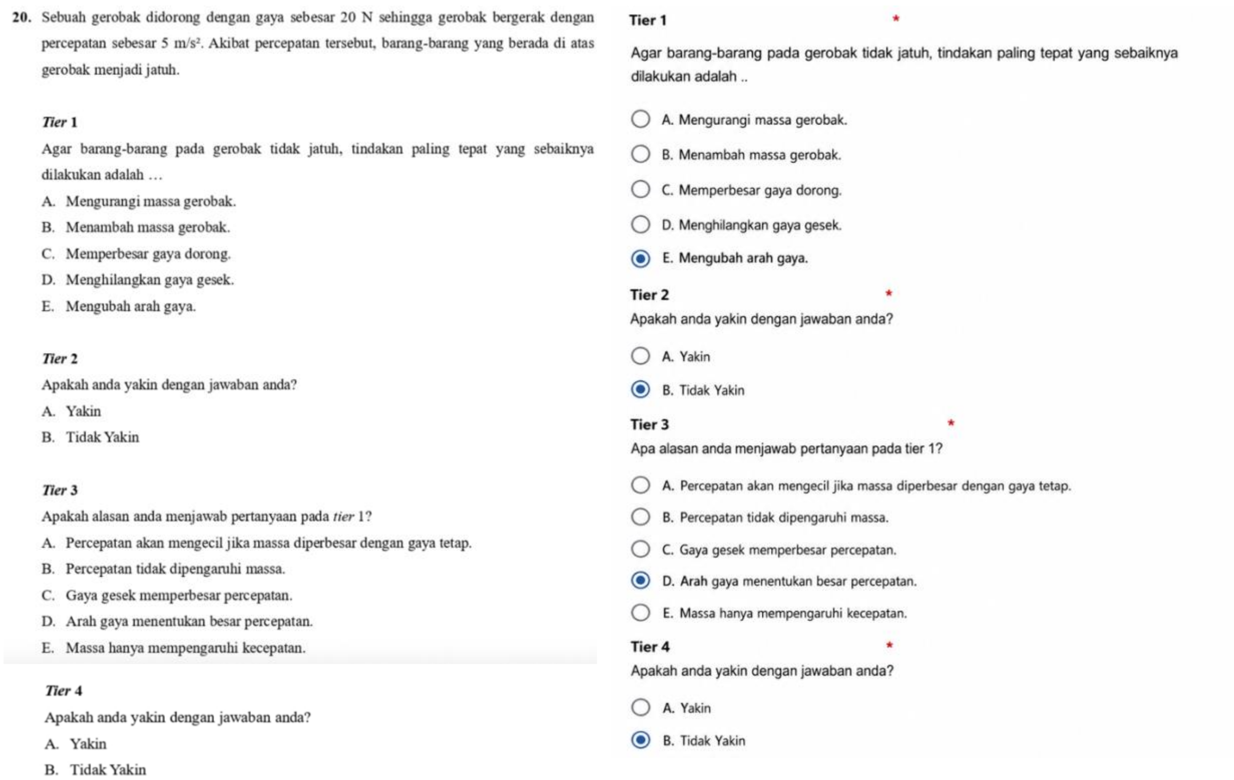
Figure 11. Questions and Examples of Student Answers for Question Number 14

Based on Figure 11, the analysis results show that 13 students in the sufficient ability category obtained an average of 2, which indicates that students' evaluation abilities (C5) are still not optimal in assessing the accuracy of an action and determining the alternative answers that best suit the given physics concept. Question number 14 presents a cupboard that is pushed with a force of 40 N to the right, but remains in its original position. To answer this question correctly, students must be able to evaluate the conditions of the forces acting on the cupboard and conclude that the resultant force on the cupboard is zero due to the friction force that is equal in magnitude and opposite in direction to the pushing force. The results of students' answers in Figure 11 show that students in the sufficient category have an initial understanding of the concept of force and motion, but have not been able to conduct a comprehensive evaluation.

Student errors are generally seen in the inability to distinguish between the presence of a force and its effect on an object's motion. Some students still assume that the presence of a pushing force always causes an object to move or accelerate, without considering the role of friction as a balancing force. This finding is in line with Nasution et al. (2021), which states that misconceptions about force and motion, especially in conditions of balanced force, are still frequently found among students.

Thus, an average score of 2 on the four-tier diagnostic instrument indicates that students' cognitive capacity is still developing and they are highly susceptible to high cognitive load when having to compare multiple physical variables simultaneously. This emphasizes the urgency of using a four-tier instrument capable of distinguishing whether student errors stem from a lack of knowledge or a misconception believed to be true.

The peak of cognitive challenges is seen in the "Less" category, which is found at the highest cognitive level, namely creating (C6). Question 20, presented in Figure 11, reveals that students still experience significant difficulty in constructing new solutions or integrating Newton's laws into situations they have never encountered before.



20. Sebuah gerobak didorong dengan gaya sebesar 20 N sehingga gerobak bergerak dengan percepatan sebesar 5 m/s². Akibat percepatan tersebut, barang-barang yang berada di atas gerobak menjadi jatuh.

Tier 1
Agar barang-barang pada gerobak tidak jatuh, tindakan paling tepat yang sebaiknya dilakukan adalah …
A. Mengurangi massa gerobak.
B. Menambah massa gerobak.
C. Memperbesar gaya dorong.
D. Menghilangkan gaya gesek.
E. Mengubah arah gaya.

Tier 2
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 3
Apakah alasan anda menjawab pertanyaan pada *tier* 1?
A. Percepatan akan mengecil jika massa diperbesar dengan gaya tetap.
B. Percepatan tidak dipengaruhi massa.
C. Gaya gesek memperbesar percepatan.
D. Arah gaya menentukan besar percepatan.
E. Massa hanya mempengaruhi kecepatan.

Tier 4
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 1 *
Agar barang-barang pada gerobak tidak jatuh, tindakan paling tepat yang sebaiknya dilakukan adalah ..
A. Mengurangi massa gerobak.
B. Menambah massa gerobak.
C. Memperbesar gaya dorong.
D. Menghilangkan gaya gesek.
E. Mengubah arah gaya.

Tier 2 *
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Tier 3 *
Apa alasan anda menjawab pertanyaan pada tier 1?
A. Percepatan akan mengecil jika massa diperbesar dengan gaya tetap.
B. Percepatan tidak dipengaruhi massa.
C. Gaya gesek memperbesar percepatan.
D. Arah gaya menentukan besar percepatan.
E. Massa hanya mempengaruhi kecepatan.

Tier 4 *
Apakah anda yakin dengan jawaban anda?
A. Yakin
B. Tidak Yakin

Figure 12. Questions and Examples of Student Answers for Question Number 20

Based on Figure 12, question 20 presents the problem of a cart being accelerated so that an object on it falls. To prevent the object from falling, students must understand that the acceleration of the cart is directly proportional to the force and inversely proportional to the mass, so the correct solution is to increase the mass of the cart to reduce the acceleration. However, the analysis results in Figure 11 indicate that students in the less category are not yet able to choose the correct answer and reason. Student errors are evident in choosing options such as changing the direction of the force or reducing the mass, which actually increases the acceleration. A score of 1 on the four-tier diagnostic instrument indicates that students' confidence in their answer is not supported by a correct understanding of the concept. Students' knowledge is still based on memorizing formulas, not yet reaching the conceptual understanding required for higher-order thinking skills (C6). This finding confirms that higher-order thinking skills, especially in the indicator of creating (C6), are still not optimally developed.

Overall, the sequence of achievement from C1 to C6 illustrates a pattern where students are relatively strong in remembering, understanding, and applying, but weaken sharply in analyzing and evaluating, before improving slightly in creating. This pattern of weakening mastery as we move from lower to higher cognitive levels aligns with the findings of Patricia & Sabani (2025), who also found that students excelled at lower levels but declined at higher levels of thinking. This study contributes beyond simply confirming this pattern. By placing the lowest point at the evaluating level, rather than at the highest level as is commonly assumed, it demonstrates that students' difficulties do not simply increase with increasing cognitive levels, but are concentrated at the level where reasoning about the relationship between force and motion is most demanding. This finding corroborates the findings of (Suwasono et al., 2025) that force-motion reasoning is the most persistent misconception in mechanics, and also reveals in more detail at which cognitive level these misconceptions are most detrimental to students' mastery.

A key difference between this study and most previous research lies in the instrument used. While many studies rely on reasoned multiple-choice tests or the Force Concept Inventory, this study used a four-tier diagnostic test based on confidence levels, effectively separating students with active misconceptions from those simply lacking knowledge. This diagnostic distinction is important because the two conditions require different learning responses. Students with active misconceptions require conceptual restructuring, while those lacking knowledge require concept building from scratch (Fadillah & Istikomah, 2021; Istiyono et al., 2023).

The findings of this study indicate that cognitive profile mapping based on Bloom's Taxonomy does not always follow a linear pattern of decline from low to high levels. The weakest point is actually at the evaluation level, rather than the highest level. Students' difficulties in mechanics are determined more by the location of conceptual misconceptions than by cognitive level alone. These findings enhance understanding of how force-motion misconceptions are distributed across cognitive levels and emphasize that misconception diagnosis needs to be linked to specific cognitive levels, not simply to topical content.

The instruments and profiles generated by this study can be used by teachers as a basis for designing more targeted learning. The instrument's ability to separate students with active misconceptions from those simply lacking knowledge allows teachers to provide differentiated remediation because the instrument is Google Forms-based; its use is also practical in classrooms with limited resources and allows teachers to obtain a diagnostic picture efficiently.

LIMITATIONS

This study still has several limitations, although it provides insight into mapping students' cognitive profiles. It lacks generalizability due to its limited sample size, involving only students from one school. It may also exhibit bias due to its reliance on self-reported confidence levels in a four-tier diagnostic assessment.

CONCLUSION

This study successfully developed a four-tiered diagnostic instrument based on Google Forms, which has been psychometrically tested as an accurate "lens" for mapping students' cognitive profiles related to Newton's Second Law. The results of the Rasch Model analysis indicate that this instrument has excellent reliability (Cronbach's Alpha 0.94) and a very high level of practicality (96.66), making it highly effective for use on a large class scale. The cognitive profile mapping revealed that although the majority of students were in the "Very Good" category at the remembering (C1) and understanding (C2) levels, significant cognitive difficulties were most evident at the evaluating level (C5), which showed the lowest achievement among the six cognitive levels, while performance increased again at the creating level (C6). These findings suggest that students' difficulties in Newtonian mechanics are determined more by the location of conceptual misconceptions, particularly in reasoning about the relationship between force and motion, than solely by cognitive level. The main contribution of this study lies in combining the diagnostic power of the four-tiered format with the precision of Rasch measurement to produce a pedagogically actionable cognitive profile. The resulting profiles not only indicate whether students understand Newton's Second Law, but also indicate at what cognitive level their understanding begins to weaken, thus providing a basis for differentiated remediation. Future research is recommended to test the effectiveness of learning interventions designed based on these cognitive profiles and to expand the sample size to multiple schools to make the resulting profiles more generalizable.

AUTHOR CONTRIBUTIONS

VV conceived the study, designed the research framework, and led the manuscript writing and revision. DTC conducted data collection, administered the diagnostic instrument, and performed initial data processing. FR contributed to Rasch model analysis, results interpretation, and manuscript review. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the principal, teachers, and students of SMAN 1 Bandar Lampung for their cooperation and participation in this study. The authors also thank the expert validators whose constructive feedback significantly contributed to the quality of the diagnostic instrument developed in this research.

REFERENCES

Adimayuda, R., Suhandi, A., Samsudin, A., Suhendi, E., Setiawan, A., & Fratiwi, N. J. (2025). Breaking misconceptions: Technology-integrated MORE model for meaningful learning of momentum and impulse. *Online Learning in Educational Research, 5*(1), 25–40. https://doi.org/10.58524/oler.v5i1.606

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

Arief, M. K., Handayani, L., & Dwijananti, P. (2012). Identifikasi kesulitan belajar fisika pada siswa RSBI: Studi kasus di RSMABI se Kota Semarang. *UPEJ Unnes Physics Education Journal*, *1*(2), 1-10.

Arikunto, S. (2010). *Prosedur penelitian suatu pendekatan praktik*. Rineka Cipta.

Aris, N. M., Ibrahim, N. H., & Halim, N. (2025). Design and development research (DDR) approach in designing design thinking chemistry module to empower students' innovation competencies. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *44*(1), 55–68. https://doi.org/10.37934/araset.44.1.5568

Arlianty, W. N., Febriana, B. W., Diniaty, A., & Fauzi'ah, L. (2018). Student profile in completing questions based on cognitive level of bloom's taxonomy by Anderson and Krathwohl. *AIP Conference Proceedings, 2026*(1), 1-8. https://doi.org/10.1063/1.5065023

Bond, T. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge.

Boone, L., Soenens, B., & Luyten, P. (2014). When or why does perfectionism translate into eating disorder pathology? A longitudinal examination of the moderating and mediating role of body dissatisfaction. *Journal of Abnormal Psychology*, *123*(2), 412–418. https://doi.org/10.1037/a0036254

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, *15*(4), 1-7. https://doi.org/10.1187/cbe.16-04-0148

Bouchée, T., de Putter-Smits, L., Thurlings, M., & Pepin, B. (2022). Towards a better understanding of conceptual difficulties in introductory quantum physics courses. *Studies in Science Education*, *58*(2), 183–202. https://doi.org/10.1080/03057267.2021.1963579

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. National academy press.

Bussotti, P. (2025). The concept of force: an educational itinerary with a historical support. *Proceedings of the 6th International Baltic Symposium on Science and Technology Education (BalticSTE2025), Šiauliai, 16–19 June, 2025* (pp. 47–70). Scientia Socialis. https://doi.org/10.33225/BalticSTE/2025.47

Chen, S. (2009). Shadows: Young Taiwanese children's views and understanding. *International Journal of Science Education*, *31*(1), 59–79. https://doi.org/10.1080/09500690701633145

Chen, Y.C. (2025). Cultivating a higher level of student agency in collective discussion: Teacher strategies to navigate student scientific uncertainty to develop a trajectory of sensemaking. *International Journal of Science Education*, *47*(3), 440–473. https://doi.org/10.1080/09500693.2024.2333714

Choiriyah, S., Ramadhan, S., Nugroho, A., Pembangunan, H. R. P., & Muharom, F. (2025). Artificial intelligence-driven learning assessment in faculties of education: An exploratory study. *Munaddhomah: Jurnal Manajemen Pendidikan Islam,* *6*(3), 482–495. https://doi.org/10.31538/munaddhomah.v6i3.1937

Chu, H. C., Hwang, G. J., Tsai, C. C., & Tseng, J. C. (2010). A two-tier test approach to developing location-aware mobile learning systems for natural science courses. *Computers & Education*, *55*(4), 1618–1627. https://doi.org/10.1016/j.compedu.2010.07.004

Csapó, B., & Molnár, G. (2019). Online Diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, *10*(7), 1-14. https://doi.org/10.3389/fpsyg.2019.01522

Curtis, D. A., Lind, S. L., Boscardin, C. K., & Dellinges, M. (2013). Does student confidence on multiple-choice question assessments provide useful information? *Medical Education*, *47*(6), 578–584. https://doi.org/10.1111/medu.12147

Ding, K., Cai, Z., & Zheng, Z. (2024). Advanced diagnostic framework for analyzing student emotional and cognitive patterns: A big data approach to personalized teaching. *Journal of Computational Methods in Sciences and Engineering*, *1*(1), 1-10. https://doi.org/10.1177/14727978251322746

Fadillah, D. P., & Istikomah, I. (2021). The strategy of school literacy culture in elementary school. *Nazhruna: Jurnal Pendidikan Islam, 4*(3), 503–517. https://doi.org/10.31538/nzh.v4i3.1614

Foster, C., Woodhead, S., Barton, C., & Clark-Wilson, A. (2022). School students' confidence when answering diagnostic questions online. *Educational Studies in Mathematics*, *109*(3), 491–521. https://doi.org/10.1007/s10649-021-10084-7

Fratiwi, N. J., Samsudin, A., Ramalis, T. R., Saregar, A., Diani, R., Irwandani, Rasmitadila, & Ravanis, K. (2020). Developing MeMoRI on newton's laws: for identifying students' mental models. *European Journal of Educational Research*, *9*(2), 699–708. https://doi.org/10.12973/eu-jer.9.2.699

Fulmer, G. W., Chu, H.-E., Treagust, D. F., & Neumann, K. (2015). Is it harder to know or to reason? Analyzing two-tier science assessment items using the rasch measurement model. *Asia-Pacific Science Education*, *1*(1), 1–16. https://doi.org/10.1186/s41029-015-0005-x

Grazziotin-Soares, R., Blue, C., Feraro, R., Tochor, K., Ardenghi, T. M., Curtis, D., & Ardenghi, D. M. (2021). The interrelationship between confidence and correctness in a multiple-choice assessment: Pointing out misconceptions and assuring valuable questions. *BDJ Open*, *7*(1), 1-8. https://doi.org/10.1038/s41405-021-00067-4

Gustina, E., Viyanti, V., & Herlina, K. (2025). A need analysis for developing multi-representation assessment instruments assisted by thunkable to measure argumentation skills in fluid materials. *Indonesian Journal of Science and Mathematics Education*, *8*(2), 502–518. https://doi.org/10.24042/bb6k5224

Horcajo, J., Santos, D., & Higuero, G. (2022). The effects of self-efficacy on physical and cognitive performance: An analysis of meta-certainty. *Psychology of Sport and Exercise*, *58*(1), 1-9. https://doi.org/10.1016/j.psychsport.2021.102063

Irianti, E. (2021). Identifikasi miskonsepsi siswa pada materi cahaya menggunakan four tier diagnostics test. *Jurnal Pendidikan Fisika Undiksha*, *11*(2), 1–10. https://doi.org/10.26740/bioedu.v11n1.p60-67

Istiyono, E., Dwandaru, W. S. B., Fenditasari, K., Ayub, M. R. S. S. N., & Saepuzaman, D. (2023). The development of a four-tier diagnostic test based on modern test theory in physics education.

European Journal of Educational Research, *12*(1), 1-15. https://doi.org/10.12973/eu-jer.12.1.371
Janah, A., Mindyarto, B., & Ellianawati. (2021). Developing four-tier diagnostic test to measure students' misconceptions on simple harmonic motion material. *Journal of Physics: Conference Series, 1918*(5), 1-7. https://doi.org/10.1088/1742-6596/1918/5/052050
Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *ReseaRch in Science & Technological educaTion*, *35*(2), 238–260. https://doi.org/10.1080/02635143.2017.1310094
Kuo, B. C., Chen, C. H., & de la Torre, J. (2018). A cognitive diagnosis model for identifying coexisting skills and misconceptions. *Applied Psychological Measurement*, *42*(3), 179–191. https://doi.org/10.1177/0146621617722791
Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12
Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., Zhang, W., & Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, *72*(1), 19–26. https://doi.org/10.1016/j.pedn.2023.05.016
Linacre, J. (1994). Sample size and item calibration stability. *Rasch Mes Trans.*, *7*(1), 1-6.
Linn, R. L. (2008). *Measurement and assessment in teaching*. Pearson Education India.
Liu, G., & Fang, N. (2016). Student misconceptions about force and acceleration in physics and engineering mechanics education. *International Journal of Engineering Education*, *32*(1), 19–29.
Lubis, M. S. A., Fatmawati, E., Pratiwi, E. Y. R., Sabtohadi, J., & Damayanto, A. (2022). Understanding curriculum transformation towards educational innovation in the era of all-digital technology. *Nazhruna: Jurnal Pendidikan Islam, 5*(2), 526–542. https://doi.org/10.31538/nzh.v5i2.2110
Ma, H., Yang, H., Li, C., Ma, S., & Li, G. (2025). The effectiveness and sustainability of tier diagnostic technologies for misconception detection in science education: A systematic review. *Sustainability*, *17*(7), 1-28. https://doi.org/10.3390/su17073145
Maharani, L., Rahayu, D., Amaliah, E., Rahayu, R., & Saregar, A. (2019). Diagnostic test with four-tier in physics learning: Case of misconception in Newton's Law material. *Journal of Physics: Conference series, 1155*(1), 1-9. https://doi.org/10.1088/1742-6596/1155/1/012022
Mahmudi, I., Ismail, M., Rizki Sista, T., Saifulloh, A., & Syarifah. (2026). ACTFL-based computerised adaptive testing for quality management in Arabic instruction at an Islamic university. *Munaddhomah: Jurnal Manajemen Pendidikan Islam,* *7*(3), 604–620. https://doi.org/10.31538/munaddhomah.v7i3.2908
Maison, Asrial, Susanti, N., Effrita, A., & Tanti. (2021). Identification of students' misconception about light using a four-tier instrument. *Journal of Physics: Conference Series, 1876*(1), 1-7. https://doi.org/10.1088/1742-6596/1876/1/012063
Maiyanti, A. A., Huda, S., Anggraini, A., Laili, U. F., Muniroh, L., & Umam, R. (2025). Transformation of physics learning: Integrating virtual laboratories to improve students' scientific literacy skills. *Online Learning in Educational Research,* *5*(1), 155–172. https://doi.org/10.58524/oler.v5i1.661
Malaterre, C., Javaux, E. J., & López-García, P. (2023). Misconceptions in Science. *Perspectives on Science*, *31*(6), 717–743. https://doi.org/10.1162/posc_a_00590
Mardian, V., Samsudin, A., Utama, J. A., Suwarma, I. R., & Coştu, B. (2023). Validity and reliability of elasticity multiple-choice items (EMCI) using Rasch model. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni, 12*(2), 265–276. https://doi.org/10.24042/jipfalbiruni.v12i2.17037
Marinda, L. (2020). Teori perkembangan kognitif jean piaget dan problematika pada anak usia sekolah dasar. *Jurnal Kajian Perempuan Dan Keislaman*, *13*(1), 116–152. https://doi.org/10.35719/annisa.v13i1.26
Maulina, D., Wijaya, A. P., Viyanti, V., Rakhmawati, I., Ibrahim, N. H., & Meriza, N. (2026). The relationship between self-efficacy and problem-solving ability in biology learning at the senior

high school. *Biosfer: Jurnal Pendidikan Biologi*, *19*(1), 78–88. https://doi.org/10.21009/biosferjpb.58149
Morton, J. P., Doran, D. A., & MacLaren, D. P. (2008). Common student misconceptions in exercise physiology and biochemistry. *Advances in Physiology Education*, *32*(2), 142–146. https://doi.org/10.1152/advan.00095.2007
Mustofa, H. A., Sandanadas, E., Mohtar, L. E., & Faresta, R. A. (2024). Identifying student prior-knowledge and conceptual changes profile on Newton law using by using POE (predict, observe, and explain) as probing understanding strategy. *Journal of Science and Mathematics Letters*, *12*(1), 80–91. https://doi.org/10.37134/jsml.vol12.1.10.2024
Nahadi, N., Firman, H., & Kurniadi, H. (2018). Development and validation of chemistry virtual test based multiple representations. *Journal of Education and Learning (EduLearn)*, *12*(1), 44–51. https://doi.org/10.11591/edulearn.v12i1.6963
Nahak, R. L. (2024). Pengembangan asesmen diagnostik kognitif berbasis kearifan lokal masyarakat ende lio. *Edukatif : Jurnal Ilmu Pendidikan*, *6*(5), 5825–5835. https://doi.org/10.31004/edukatif.v6i5.7562
Nasution, R. H., Wijaya, T. T., Adi Putra, M. J., & Hermita, N. (2021). Analisis miskonsepsi siswa SD pada materi gaya dan gerak. *Journal of Natural Science and Integration*, *4*(1), 11-21. https://doi.org/10.24014/jnsi.v4i1.10851
Novela, E., Viyanti, V., & Rohman, F. (2025). Adaptive assessment tool: Analysis of digital instrument development needs to measure critical thinking skills in static fluid materials. *Indonesian Journal of Science and Mathematics Education*, *8*(3), 820–829. https://doi.org/10.24042/ijsme.v8i3.29563
Önder Çelikkanlı, N., & Kızılcık, H. (2022). A review of studies about four-tier diagnostic tests in physics education. *Journal of Turkish Science Education*, *19*(4), 1-21. https://doi.org/10.36681/tused.2022.175
Patricia, A., & Sabani, S. (2025). Analisis miskonsepsi siswa sma pada hukum newton menggunakan four-tier test berbasis CRI. *Relativitas: Jurnal Riset Inovasi Pembelajaran Fisika*, *8*(2), 160–168. https://doi.org/10.29103/relativitas.v8i2.22835
Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). Development and validation of a critical thinking assessment-scale short form. *Education Sciences*, *12*(12), 1-19. https://doi.org/10.3390/educsci12120938
Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, *18*(1), 25–34. https://doi.org/10.1016/j.jval.2014.10.005
Putranta, H., & Afifah, F. (2025). Development of the four-tier diagnostic test to identify student misconceptions in the static fluids chapter. *Journal on Efficiency and Responsibility in Education and Science*, *18*(4), 268–281. https://doi.org/10.7160/eriesj.2025.180403
Rasikah, S. 2024. Profil kognitif peserta didik berdasarkan taksonomi bloom revisi menggunakan asesmen mathsci konteks keislaman jenjang kelas XII SMA/MA (Skripsi). Universitas Islam Negeri Syarif Hidayatullah.
Resbiantoro, G., & Setiani, R. (2022). A review of misconception in physics: The diagnosis, causes, and remediation. *Journal of Turkish Science Education*, *19*(2), 403–427.
Richey, R. C., & Klein, J. D. (2014). *Design and development research.* Springer. https://doi.org/10.1007/978-1-4614-3185-5_12
Romine, W. L., Schaffer, D. L., & Barrow, L. (2015). Development and application of a novel Rasch-based methodology for evaluating multi-tiered assessment instruments: Validation and utilization of an undergraduate diagnostic test of the water cycle. *International Journal of Science Education*, *37*(16), 2740–2768. https://doi.org/10.1080/09500693.2015.1105398
Sagatbek, A., Oni, T. K., Miller, E. A., Gabdullina, G., & Balta, N. (2024). Do High school students learn more or shift their beliefs and attitudes toward learning physics with the social constructivism of problem-based learning? *Education Sciences*, *14*(12), 1-26. https://doi.org/10.3390/educsci14121280
Santosa, S., & Jazuli, M. F. (2022). The digital madrasah as an idea of IT-based Islamic education. *Nazhruna: Jurnal Pendidikan Islam, 5*(2), 379–391. https://doi.org/10.31538/nzh.v5i2.2121

Sari, A. N., Viyanti, V., & Ertikanto, C. (2023). Teachers' perceptions of physics scientific argumentation test instruments based on modern test theory using question modeling through e-learning edpuzzle LMS. *Indonesian Journal of Science and Mathematics Education*, *6*(3), 405–415. https://doi.org/10.24042/ijsme.v6i3.18991

Segado, M. A. 2025. Intuitive but wrong: Uncovering student misconceptions about force and motion with bayesian item-response methods (Doctoral dissertation). Massachusetts Institute of Technology.

Septiyani, E., & Nanto, D. (2021). Four-tier diagnostic test assisted website for identifies misconceptions heat and temperature. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *7*(1), 35–42. https://doi.org/10.21009/1.07104

Soeharto, S., Csapó, B., Sarimanah, E., Dewi, F., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649

Suhandi, A., Samsudin, A., Fratiwi, N. J., Nurdini, N., Feranie, S., Purwanto, M. G., Linuwih, S., & Coştu, B. (2025). Altering misconceptions: How e-rebuttal texts on Newton's laws reconstructs students' mental models. *Frontiers in Education,* *10*(4), 1-15. https://doi.org/10.3389/feduc.2025.1472385

Sumintono, B. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim Komunikata.

Supurwoko, S., Gunawan, K. D. H., & Hudha, M. N. (2025). Validity analysis of VR-based particle dynamics module development using the rasch model. *Online Learning In Educational Research (OLER)*, *5*(2), 441–455. https://doi.org/10.58524/oler.v5i2.803

Sutopo. (2016). Pemahaman mahasiswa tentang konsep-konsep dasar gelombang mekanik. *Jurnal Pendidikan Fisika Indonesia 12*, *12*(5), 41–53.

Sutrisno, S., Maison, M., Marzal, J., & Kurniawan, W. Rapid artificial diagnostic test in intelligent teacher assistant system (ITAS) to identify misconceptions on kinematics. In: M. L. Firdaus & A. Defianti, Ed., *Mathematics and Science Education International Seminar 2021 (meseis 2021)*, Atlantis Press, Bengkulu, 2023, pp. 316-324. https://doi.org/10.2991/978-2-38476-012-1_39

Suwasono, P., Koes H. S., Adi P. N., & Saniso, E. (2025). A significant reducing misconception on newton's law under purposive scaffolding and problem-based misconception supported modeling instruction. *Open Education Studies*, *7*(1), 1-26. https://doi.org/10.1515/edu-2025-0091

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tan, A. L., Ong, Y. S., Ng, Y. S., & Tan, J. H. J. (2023). STEM problem solving: inquiry, concepts, and reasoning. *Science & Education*, *32*(2), 381–397. https://doi.org/10.1007/s11191-021-00310-2

Taslidere, E. (2016). Development and use of a three-tier diagnostic test to assess high school students' misconceptions about the photoelectric effect. *Research in Science & Technological Education*, *34*(2), 164–186. https://doi.org/10.1080/02635143.2015.1124409

Tsui, C., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, *32*(8), 1073–1098. https://doi.org/10.1080/09500690902951429

Uswatun, & Mubarak, S. (2024). Pengembangan instrumen tes diagnostik four-tier untuk mengidentifikasi miskonsepsi materi konsep mol dan stoikiometri. *Al Kawnu: Science and Local Wisdom Journal*, *4*(1), 34–49.

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences*, *39*(2), 163–173. https://doi.org/10.1080/01490400.2015.1127189

Vicovaro, M. (2023). Grounding intuitive physics in perceptual experience. *Journal of Intelligence*, *11*(10), 1-20. https://doi.org/10.3390/jintelligence11100187

Viyanti, Shintya, R. E., Rosidin, U., Suyatna, A., & Farawahidah, N. (2025). Analysis of project-based problem-solving skills of optical equipment materials with heller and heller steps. *Jurnal Pendidikan IPA Indonesia*, *14*(2), 241–250. https://doi.org/10.15294/jpii.v14i2.23810

Viyanti, Sunyono, Rosdiana, T., Anggreini, Jarwadi, & Priowicaksono, N. (2024). Developing electronic student worksheets based on ethnoscience to train smart risk-taking behaviour and growth mindset. *Jurnal Pendidikan MIPA*, *25*(1), 127–141. https://doi.org/10.23960/jpmipa/v25i1.pp127-141

Weber, J., & Wilhelm, T. (2024). Contributing factors to the improvement of conceptual understanding in a computer-based intervention in Newtonian dynamics. *Physical Review Physics Education Research*, *20*(2), 1-17. https://doi.org/10.1103/PhysRevPhysEducRes.20.020130

Wedana, I. W. W., Viyanti, & Permadi, D. (2025). Development of e-LKPD based on android-assisted problem-based learning to improve critical thinking ability of learners. *Journal of Physics: Conference Series*, *3132*(1), 1–11. https://doi.org/10.1088/1742-6596/3132/1/012028

Winarti, W., Pardiyanto, P., & Kurnianto, A. (2021). Senior high school students' understanding of Newton's laws in outer space: Identification of misconceptions. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni, 10*(2), 219–230. https://doi.org/10.24042/jipfalbiruni.v10i2.8997

Yang, D., & Sianturi, I. A. J. (2019). Assessing students' conceptual understanding using an online three-tier diagnostic test. *Journal of Computer Assisted Learning*, *35*(5), 678–689. https://doi.org/10.1111/jcal.12368

Yordan, A., & Fahyuni, E. F. (2021). Child-friendly IRE learning through digital storytelling in the COVID-19 pandemic. *Nazhruna: Jurnal Pendidikan Islam,* *4*(3), 590–605. https://doi.org/10.31538/nzh.v4i3.1682

Yulia, G., Liudmila, L., Anna, B., Valentina, L., & Sergei, S. (2022). The development of critical thinking skills in mobile learning: Fact-checking and getting rid of cognitive distortions. *International Journal of Cognitive Research in Science, Engineering and Education*, *10*(2), 51–68. https://doi.org/10.23947/2334-8496-2022-10-2-51-68

Zain, A. Z., Samsudin, A., & Hasanah, L. (2025). Diagnostic instrument test of renewable energy topics: Analysis of conceptual change students using the rasch approach. *Online Learning In Educational Research (OLER)*, *5*(2), 231–243. https://doi.org/10.58524/oler.v5i2.675

Zainuri, A., & Saepuloh, S. (2022). Electronic media pillar of learning evaluation in madrasah. *Nazhruna: Jurnal Pendidikan Islam,* *5*(3), 1216–1227. https://doi.org/10.31538/nzh.v5i3.2383

Zhai, X. (2022). Assessing high-school students' modeling performance on Newtonian mechanics. *Journal of Research in Science Teaching*, *59*(8), 1313–1353. https://doi.org/10.1002/tea.21758