Indonesian Society for Science Educators
Journal of Science Learning
journal homepage: ejournal.upi.edu/index.php/jslearning

INFORMI: A Rasch-Validated Four-Tier Instrument to Diagnose Students' Conceptions of Momentum and Impulse

Shita Ayu Amalia[1], Dwi Sulisworo[1], Nuzulira Janeusse Fratiwi[2]*, Nurdini[2], Hera Novia[2], Achmad Samsudin[2], Hadi Nasbey[3], Firmanul Catur Wibowo[3], Mustafa Sozbilir[4]

[1]Physics Education Department, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Department of Physics Education Program, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3]Department of Physics Education Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, Jakarta, Indonesia
[4]Department of Mathematics and Science Education, Faculty of Mathematics and Science Education, Ataturk University, Erzurum, Turkey

*Corresponding author: nuzulira.janeusse.fratiwi@upi.edu

ABSTRACT This study aims to develop and validate the Instruments of Four-tier Multi-representational on Momentum and Impulse (INFORMI) as a diagnostic instrument to analyze students' conceptions in physics. The instrument was developed using the ADDIE model, which involves the stages of analysis, design, development, implementation, and evaluation. INFORMI consists of 19 items with four structured tiers: conceptual knowledge, confidence level, reasoning, and reasoning confidence. The instrument was administered to 37 eleventh-grade students at one of the Senior High Schools in Indramayu, Indonesia. Rasch analysis using MINISTEP software was conducted to evaluate the instrument's psychometric properties. The results indicated that the instrument fulfilled the requirement for construct validity (30%) and that all items met the fit criteria, with Outfit MNSQ values ranging from 0.60 to 1.46. The item reliability reached 0.86, and the Cronbach's alpha was 0.62, suggesting adequate internal consistency. The distribution of item difficulty ranged from very easy to very difficult, and all discrimination indices were positive, indicating that each item effectively distinguished students' conceptual understanding. Analysis of student responses revealed that most students fell into the Understands Concept (UC) and partially understands (PU) categories. The findings demonstrate that INFORMI is a valid, reliable, and effective tool for identifying students' misconceptions in the topic of momentum and impulse. It is recommended that teachers utilize INFORMI as a diagnostic instrument to inform and improve instructional strategies in physics learning.

Keywords: Four-tier test, INFORMI, Momentum and impulse, Rasch model, Students' conceptions

1. INTRODUCTION

Conceptual understanding is essential for achieving meaningful learning outcomes in physics. However, students often struggle with misconceptions, preconceptions that deviate from scientifically accepted concepts. These misconceptions, often formed through daily experiences and intuitive reasoning, can persist even after formal instruction, making them resistant to change. (Akcanca & Cerrah Özsevgeç, 2020; Kurniawan et al., 2020; Lassonde et al., 2016; Ozdemir, 2022). In particular, momentum and impulse are classical physics concepts that frequently give rise to misconceptions, as they involve abstract vector quantities and the force-time relationship, which are often counterintuitive. (Adianto & Rusli, 2021; Adimayuda et al., 2020; Amalia et al., 2019; Sakinah et al., 2023; Setiani et al., 2023).

Students often misconceive the distinction between impulsive force and impulse. They tend to assume that a collision involving a large force necessarily results in a large impulse, overlooking the importance of the contact time duration (Adianto & Rusli, 2021; Fang & Guo, 2023). A classic demonstration of misconceptions in momentum and impulse arises when students are shown an illustration of two spherical objects, a rubber ball and a plasticine ball,

Received: 15 July 2025
Revised: 27 August 2025
Published: 19 March 2026

thrown at the same velocity toward a rigid wall (Figure 1). The rubber ball bounces back after impact, while the plasticine ball sticks to the wall. Despite the plasticine ball experiencing a more minor change in momentum, many students claimed it received a larger impulse, often citing its more severe deformation. This misconception matters because it illustrates how students often rely on perceptual cues and everyday intuition rather than the impulse–momentum principle, which can hinder their ability to analyze force-corrected time interactions in more complex physics contexts.

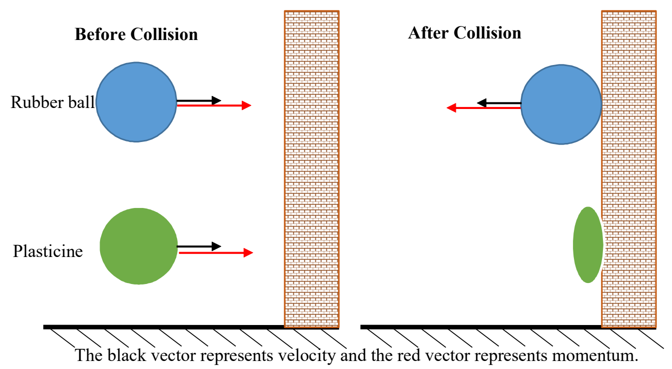

Figure 1 A rubber ball and plasticine are thrown with the same velocity toward a wall.

This suggests that students often estimate the magnitude of impulse based on perceived force or visible deformation, rather than analyzing the product of force and time or the change in momentum. The rubber ball, which actually undergoes a greater change in momentum (since it reverses direction), is often incorrectly assumed to experience less impulse.

Given the persistence and complexity of misconceptions, especially in topics like momentum and impulse, it becomes essential to adopt diagnostic assessments that move beyond merely evaluating the correctness of answers. Practical assessments must uncover the underlying reasoning and confidence levels that inform students' choices. (Grazziotin-Soares et al., 2021). Traditionally, educators have used multiple-choice tests (MCTs) due to their ease of use and objective scoring. However, MCTs fall short of revealing students' conceptual understanding, as they do not capture students' reasoning nor distinguish between correct answers resulting from understanding and those resulting from guessing. (Crogman et al., 2018; Gurel et al., 2015; Kaniawati et al., 2019).

To address this limitation, the two-tier diagnostic test was introduced, consisting of a content question followed by a reasoning question. While this format offers greater insight into student thinking, it cannot capture students' confidence in their responses. (Banawi et al., 2022; Gurel et al., 2015; Saat et al., 2016). Building upon this, the three-tier diagnostic test adds a confidence-rating component, typically for the selected answer. Nevertheless, it typically omits confidence ratings for the reasoning component, thereby limiting diagnostic depth. (Banawi et al., 2022; Kaniawati et al., 2019; Yang, 2022).

Recognizing these gaps, researchers have turned to the four-tier diagnostic test, which integrates both reasoning and confidence levels for each tier. This test format enables educators not only to detect the presence of misconceptions but also to distinguish between firmly held misconceptions and mere lack of knowledge. Previous research has demonstrated the effectiveness of such assessments in diagnosing alternative conceptions rooted in science learning. (Amalia et al., 2019; Kaltakci-Gurel et al., 2017; Kaniawati et al., 2020; Nurdini et al., 2020; Samsudin, Azizah, et al., 2020; Triyani et al., 2019). Istiyono, 2022 highlighted that the four-tier structure, comprising answer selection, confidence in answer, reasoning, and confidence in reasoning, enhances the precision of diagnostic feedback.

Moreover, recent empirical studies underscore several advantages of the four-tier format in physics education. Setiani et al., (2023) applied it to college students' understanding of momentum and impulse, demonstrating that the instrument could differentiate cognitive states such as scientific conception, lack of knowledge, and misconceptions, thereby revealing nuanced reasoning patterns. A systematic review by (Çelikkanlı & Kızılcık, 2022) confirmed that four-tier tests provide richer diagnostic validity compared to traditional multiple-choice formats, effectively distinguishing levels of certainty in both answers and justifications. Additionally, (Wu, Tian, Sun, Feng, & Luo, 2025) validated a four-tier instrument specifically for impulse concepts, exhibiting good item quality despite modest reliability, which indicates its capability to uncover persistent misconceptions even in conceptually abstract topics.

Building upon these prior developments, this study introduces the Instruments Four-tier Multi-representational on Momentum and Impulse (INFORMI), a diagnostic tool specifically designed to evaluate students' conceptual understanding and identify misconceptions in the topic of momentum and impulse. INFORMI employs multi-representational items, in which each concept is assessed through both verbal and mathematical formats. This approach not only enriches the diagnostic scope by capturing different dimensions of student understanding. (Fratiwi et al., 2020; Samsudin, Fratiwi, et al., 2020), but also enables variation between pre-test and post-test. By presenting the same underlying concepts through different representations, the instrument reduces the likelihood that students recognize repeated items, thereby increasing the accuracy with which conceptual change is detected.

The INFORMI was validated through Rasch analysis to ensure its psychometric soundness (Anandi et al., 2024; Boone, 2020; Planinic et al., 2019). The Rasch model, grounded in probabilistic measurement theory, maps the likelihood of a correct response as a function of the difference between a student's ability and the item's difficulty on a standard logit scale. This allows both item difficulty and student ability to be measured independently and placed on the same linear continuum (Hayat et al., 2020). Recent studies emphasize that the Rasch model enhances the objectivity of test construction by enabling the identification of misfitting items, the evaluation of item-person interactions, and the detection of item bias through Differential Item Functioning (DIF) analysis (Wei et al., 2012). For teachers, these features provide clear evidence of which test items truly reflect students' conceptual understanding, helping them to refine assessments, reduce ambiguity, and design targeted interventions based on more valid and fair diagnostic information.

By integrating correctness, certainty, and rationale of student responses, INFORMI refines the existing four-tier diagnostic approach to provide a comprehensive view of students' conceptions. Supported by Rasch analysis, the instrument demonstrates both statistical validity and pedagogical usefulness for classroom application. Rather than introducing a new theoretical framework, INFORMI adapts the four-tier structure with multi-representations in the context of momentum and impulse, enabling teachers to identify students' misconceptions and design targeted instructional interventions more accurately.

2. METHOD

This research employed the ADDIE development model, shown in Figure 2, consisting of Analysis, Design, Development, Implementation, and Evaluation. (Jung et al., 2019; Salas-Pilco et al., 2022; Spatioti et al., 2022).

In the analysis stage, the researcher conducted a literature review to identify common misconceptions among students regarding momentum and impulse. The results of this review were used to guide the development of test items that specifically target these misconceptions. In the design stage, a two-tier diagnostic test was initially constructed. The first tier consisted of multiple-choice questions focused on core concepts, while the second tier required students to select a reason to support their answer. At this stage, reasoning options were not yet provided; instead, students were asked to write their own explanations in an open-ended format. This approach was used to collect authentic student reasoning that reflected their actual thought processes.

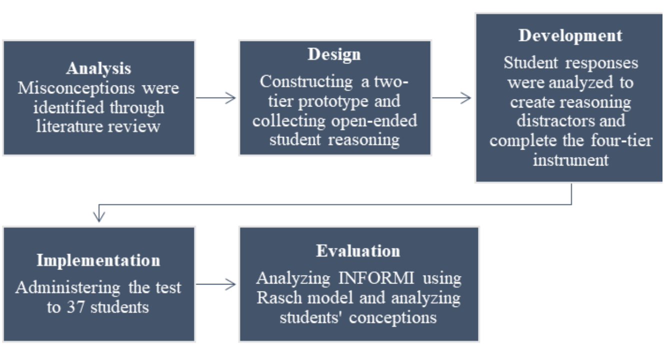

Figure 2 ADDIE model development flow

During the development stage, the open-ended responses gathered from the design stage were analyzed and categorized. The most frequently occurring and representative reasoning patterns were then used to construct distractors for the third tier (reasoning choices). Subsequently, the complete four-tier instrument was assembled, consisting of: (1) a content-based multiple-choice question, (2) a confidence rating for the selected answer, (3) a reasoning selection to justify the answer, and (4) a confidence rating for the reasoning.

The implementation stage involved administering the INFORMI to 37 eleventh-grade students selected through purposive sampling from a public Senior High School in Indramayu, Indonesia, accredited A with approximately 1.000 students and adequate classroom and laboratory facilities. The participants consisted of 26 female and 11 male students. These students, aged approximately 16 to 17, had completed the topic of momentum and impulse. As is typical for this educational level, they were assumed to have varying levels of conceptual understanding and confidence in learning physics. The test was administered during regular class hours under the researchers' supervision, with a 1-hour time allotment.

At the evaluation stage, data were collected and analyzed using Rasch modeling via MINISTEP (Student Version of Winsteps 5.8.5.0) software to examine unidimensionality (construct validity), item fit, reliability, difficulty levels, and discrimination indices. (Bond & Fox, 2015; Boone, 2020; Sumintono, 2018). The criteria for interpreting the Rasch analysis results are presented in Table 1 to provide a precise reference for evaluating the quality and performance of each test item.

While Table 1 provides the technical criteria for evaluating test item quality, Table 2 complements this by describing how students' responses were categorized into conceptual understanding levels. In other words, Table 1 addresses the psychometric properties of the instrument, whereas Table 2 focuses on how the instrument outcomes were interpreted in terms of students' conceptions.

Table 1 Aspect and interpretation criteria (Rasch Analysis)

Aspect	Criteria	Interpretation
Unidimensionality	< 20%	Fulfilled
	< 40%	Appropriate
	< 60%	Excellent
Fit Item Test	0.5 < MNSQ < 1.5; and -2.0 < ZSTD < +2.0	Acceptable fit (Valid)
Reliability (Person/Item)	0.94 ≤ value	Excellent
	0.90 ≤ value < 0.94	Very Good
	0.80 ≤ value < 0.90	Good
	0.67 ≤ value < 0.80	Fair
	value < 0.67	Poor
Cronbach Alpha (α)	0.8 ≤ α	Very Good
	0.7 ≤ α < 0.8	Good
	0.6 ≤ α < 0.7	Fair
	0.5 ≤ α < 0.6	Poor
	α < 0.5	Very Poor
Item Difficulty Level (Dif)	Dif > 0.86	Very Difficult
	0.00 < Dif ≤ 0.86	Difficult
	-0.86 < Dif ≤ 0.00	Easy
	Dif ≤ -0.86	Very Easy
Item Discrimination (PTMEASURE-AL CORR Score)	ID > 0.40	Very Good
	0.30 < ID ≤ 0.40	Good
	0.20 < ID ≤ 0.30	Fair
	ID ≤ 0.20	Poor

(Boone, 2020; Sumintono, 2018)

Table 2 Conceptual category

Category	Tier 1	Tier 2	Tier 3	Tier 4	Skor
Understand Concept (UC)	Correct	Confident	Correct	Confident	3
Partially Understand (PU)	Correct	Confident	Correct	Not Confident	2
	Correct	Not Confident	Correct	Confident	
	Correct	Not Confident	Correct	Not Confident	
	Correct	Confident	Incorrect	Confident	
	Correct	Confident	Incorrect	Not Confident	
	Correct	Not Confident	Incorrect	Confident	
	Correct	Not Confident	Incorrect	Not Confident	
	Incorrect	Confident	Correct	Confident	
	Incorrect	Confident	Correct	Not Confident	
	Incorrect	Not Confident	Correct	Confident	
	Incorrect	Not Confident	Correct	Not Confident	
Do Not Understand (NU)	Incorrect	Confident	Incorrect	Not Confident	1
	Incorrect	Not Confident	Incorrect	Confident	
	Incorrect	Not Confident	Incorrect	Not Confident	
Alternative Conceptions (AC)	Incorrect	Confident	Incorrect	Confident	0
No Coding (NC)	Responses do not match one or more tiers (incomplete)				

(Kaltakci Gurel et al., 2015; Kaniawati et al., 2020)

Then, students' conceptions are also evaluated in this stage, adopting the classification scheme proposed by (Gurel et al., 2015), which has been further employed in recent works (e.g.,(Diani et al., 2023)) to diagnose misconceptions in physics learning. Students' conceptual understanding was grouped into five categories: Understand Concept (UC), Partially Understand (PU), Alternative Conceptions (AC), Do Not Understand (NU), and No Coding (NC). The detailed tier-based criteria for classifying these categories are presented in Table 2, which outlines the combinations of answers, reasoning, and confidence levels used to determine students' conceptual understanding.

3. RESULT AND DISCUSSION

3.1 Analysis Stage

In the analysis stage, a comprehensive review of previous studies was conducted to identify common misconceptions related to momentum and impulse. To identify common misconceptions related to momentum and impulse, the relevant literature was searched across databases such as Scopus, ERIC, and Google Scholar using keywords "momentum misconceptions", "impulse misconceptions", and "student conceptions in physics". Priority was given to peer-reviewed studies that explicitly addressed students' alternative conceptions in momentum–impulse topics, both at the secondary and undergraduate levels. Studies outside the physics education domain or that did not directly examine misconceptions were excluded. This review identified recurrent misconceptions that served as the basis for developing the instrument and for the analysis in this study. One of the most recurring misconceptions identified in the literature is the confusion between force and impulse. Students tend to believe that a collision involving a large force must also produce a large impulse, regardless of the contact time (Adianto & Rusli, 2021; Fang & Guo, 2023). These misconceptions neglect the definition of impulse as the product of force and time. They are rooted in students' tendency to associate a greater impact with a larger change in momentum.

Another standard error involves the conservation of momentum during collisions. For instance, in perfectly elastic collisions, students often fail to understand that total momentum is conserved even when the objects' velocities change. As shown in previous studies (Ben-Abu, 2019; Liu & Fang, 2016; Wong & Wong, 2022)Many students assume that if kinetic energy changes or an object's velocity differs after a collision, then momentum must not be conserved. This confusion between the conservation of momentum and the conservation of kinetic energy leads to incorrect assumptions.

A third misconception identified in the literature is the belief that an object's momentum is determined solely by its mass, neglecting the role of velocity. These misunderstandings lead students to assume that heavier objects always have greater momentum, regardless of their speed. Multiple studies have reported that students often state, "greater mass always means greater momentum," even when lighter objects travel at significantly higher velocities, an error that disregards the fundamental definition of momentum as the product of mass and velocity. (Amalia et al., 2019; Diyanahesa et al., 2017; Triyani et al., 2019).

These misconceptions tend to persist because they are strongly influenced by students' everyday experiences, intuitive reasoning, and oversimplified interpretations of formulas. For example, the tendency to associate a stronger collision with a greater impulse reflects students' reliance on perceptual cues rather than abstract definitions. At the same time, confusion between momentum and kinetic energy often stems from the simultaneous introduction of both concepts in physics instruction. Similarly, the belief that momentum depends only on mass arises from prior everyday observations that "heavier objects feel stronger," which reinforces incomplete scientific reasoning.

These findings are further detailed in Table 3, which maps each misconception identified in the literature to specific learning concepts within the topic of momentum and impulse. The table compiles common conceptual difficulties and links them to clearly defined conceptual indicators. This mapping provides a systematic foundation for designing assessment items that are conceptually focused and diagnostically rich.

Each misconception listed in Table 3 was then carefully transformed into two-tier items and developed into four-tier items in the INFORMI. The items were developed to probe not only the correctness of student answers but also their reasoning patterns and confidence levels, using both verbal and mathematical representations. By grounding the test items in validated misconceptions from the literature, the instrument ensures content validity. It targets the specific conceptual challenges that frequently hinder students' understanding of momentum and impulse.

3.2 Design Stage

Based on the analysis, a two-tier diagnostic test was first developed. The first tier consisted of a concept-based multiple-choice question, while the second tier asked students to provide their own reasoning in an open-ended format. This approach was adopted to gather authentic student reasoning. The inclusion of an open-ended reasoning tier enables researchers to access richer qualitative data and uncover subtle variations in student thinking that may not be visible through multiple-choice responses alone. Moreover, open-ended justifications promote student metacognition, as learners are required to explicitly articulate their thought processes, thereby enhancing the diagnostic function of the assessment. (Marthaliakirana et al., 2022). Several studies, including those by (Zarepour et al., 2024) have successfully used this format to reveal nuanced misconceptions in science education, particularly in topics involving abstract and context-dependent reasoning such as momentum and impulse.

In this study, the misconceptions identified in Table 3 were directly operationalized into two-tier items. For instance, the misconception that "momentum becomes zero at the peak of projectile motion" was transformed into an item asking students to determine the momentum at the highest point and justify their answer. In this way, each misconception served as the foundation for constructing items that probe both the answer choices and the underlying reasoning.

Therefore, adopting a two-tier format in the initial phase of instrument development is an essential step to ensure the instrument can uncover authentic, meaningful insights into students' conceptual understanding. An example of a two-tier diagnostic test is shown in Figure 3.

10.1 If two objects collide in a closed system (no external forces acting), then the total momentum of the system after the collision is ….
A. greater than the momentum before the collision
B. smaller than the momentum before the collision
C. the same as the momentum before the collision
D. depends on the type of collision
E. zero

10.2 Give a reason!
……………………………………………………………………………

Figure 3: Example of a two-tier diagnostic test

Table 3 Conceptual indicators and misconceptions in momentum and impulse

No.	Sub-topic	Common Misconception	Concept	Indicator
1	Fundamental Concepts of Momentum and Impulse	Momentum depends only on mass, not velocity.	Analyzing momentum changes in moving objects	Students are able to calculate and compare momentum of objects based on mass and velocity.
2		Objects with equal momentum always travel the same distance, regardless of mass or velocity.	Determining the range (displacement) of an object based on momentum	Students are able to compare the distance traveled by objects with the same or different momentum.
3		Momentum becomes zero at the peak because velocity is zero.	Determining the direction of momentum at the peak of vertical projectile motion	Students are able to determine the direction of momentum at the highest point of a projectile's motion.
4		Greater bounce means greater impulse, regardless of contact time or initial velocity.	Determining impulse magnitude when an object bounces off a wall	Students are able to determine impulse using momentum change before and after a bounce.
5		Both objects experience impulse in the same direction during a collision.	Describing impulse direction in two-object collisions	Students can describe the direction of impulse experienced by each object during a collision.
6	Impulse and Force-Time Relationship	Larger force always results in larger impulse, ignoring time duration.	Determining the effect of force applied to an object	Students can explain the relationship between force and impulse.
7		Time has no effect on impulse; only force matters.	Analyzing time interval in relation to momentum and impulse	Students can explain how increasing or decreasing time affects the impulse for a given force.
8		Heavier object always has greater momentum, even if the lighter one moves faster.	Analyzing momentum of two pushed objects	Students can compare momentum of objects with different masses and velocities that are pushed identically.
9		If kinetic energy changes, then momentum must also change.	Analyzing kinetic energy based on momentum and impulse	Students can relate kinetic energy changes to momentum and vice versa.
10	Conservation of Momentum in Collisions	If object velocities change after collision, momentum is not conserved.	Comparing total momentum before and after an elastic collision	Students can identify that total momentum is conserved in perfectly elastic collisions.
11		In inelastic collisions, both kinetic energy and momentum are lost.	Comparing total kinetic energy before and after an inelastic collision	Students can analyze changes in kinetic energy and recognize that momentum is still conserved, while part of the kinetic energy is transformed into other forms.

After analyzing student responses, a pool of reasoning choices was generated to represent both scientifically correct and common misconceptions. These choices later became part of the INFORMI design.

At this stage, an initial pool of 19 two-tier items was developed from the misconceptions listed in Table 3. These items were then pilot-tested with 36 students of the same grade level to evaluate item clarity, distractor plausibility, and the quality of students' reasoning. Although no expert review was conducted, the pilot provided essential feedback that informed the refinement of reasoning choices and ensured that the items functioned as intended. The revised items were subsequently integrated into the INFORMI structure.

3.3 Development Stage

During development, 19 items were initially constructed based on the misconceptions identified in Table 3. The reasoning options were not subjected to formal expert review; instead, they were validated and refined through analysis of authentic student responses during pilot testing, ensuring the inclusion of both scientifically correct explanations and common misconceptions. These two-tier diagnostic items were then expanded into a four-tier format to enhance the instrument's diagnostic power. The goal was not only to identify students' conceptions and misconceptions, but also to measure their confidence in both their answers and their reasoning. The instrument was then named the Instruments Four-tier Multi-representational on Momentum and Impulse (INFORMI), representing its specific focus on diagnosing misconceptions in the domain of momentum and impulse. The INFORMI example is shown in Figure 4.

VERBAL	MATEMATICAL
8.1 Two objects, the second object having twice the mass of the first object, are initially at rest. Both objects are then pushed for the same amount of time and with the same force. Determine the ratio of their final momenta! A. The final momentum of the first object is four times the final momentum of the second object. B. The final momentum of the second object is two times the final momentum of the first object. C. The final momenta of both objects are the same. D. The final momentum of the first object is two times the final momentum of the second object. E. The final momentum of the first object is four times the final momentum of the second object. 8.2 Are you sure about your answer to question 8.1? A. Sure B. Not Sure 8.3 Reason for the answer to question 8.1: A. When the initial state is at rest, the magnitude of momentum is influenced by the time interval. B. When the initial state is at rest, the magnitude of momentum is influenced by force. C. When the initial state is at rest, the magnitude of momentum is influenced by mass. D. When the initial state is at rest, the magnitude of momentum is influenced by the object's velocity. E. When the initial state is at rest, the magnitude of momentum is influenced by force and time interval. 8.4 Are you sure about your answer to question 8.3? A. Sure B. Not sure	8.1 Two boxes of books have masses of 10 kg and 20 kg, initially at rest. Both boxes are then pushed for 3 seconds. Given the same force on each box, what is the ratio of their final momenta? A. 1:4 B. 1:2 C. 1:1 D. 2:1 E. 4:1 8.2 Are you sure about your answer to question 8.1? A. Sure B. Not Sure 8.3 Reason for the answer to question 8.1: A. $v_0 = 0; p$ is only influenced by F and Δt B. $v_0 = 0; p$ is only influenced by Δt C. $v_0 = 0; p$ can be influenced by m D. $v_0 = 0; p$ is not influenced by m E. $v_0 = 0; p$ is not influenced by v 8.4 Are you sure about your answer to question 8.3? A. Sure B. Not sure

Figure 4 Example of INFORMI

The INFORMI structure illustrated in Figure 4 was adapted from previous studies by (Kaltakci-Gurel et al., 2017). The diagnostic items were designed to assess students' conceptual understanding of momentum and impulse through both verbal and mathematical representations. Each item consists of four levels. The first tier presents a multiple-choice question that tests the core physics concept: the influence of force, time, and mass on the final momentum when two initially stationary objects are acted upon by equal forces for equal durations. The second tier asks students to indicate their confidence in the answer they selected in the first tier, thereby capturing their metacognitive awareness. The third tier provides a set of reasoning choices related to the concept being assessed. The fourth tier then asks students to indicate their confidence in the reasoning they selected in the third tier. These choices are constructed to reflect scientifically accurate reasoning as well as common misconceptions, such as the belief that final momentum depends solely on mass or is unrelated to time.

In this study, confidence was measured using a binary scale ("Sure" or "Not Sure"). Responses marked "Sure" were categorized as high confidence, whereas "Not Sure" responses were categorized as low confidence. This threshold was used in combination with the accuracy of the first- and third-tier responses. Each student's response was then coded into one of the five diagnostic categories (UC, PU, NU, AC, NC) by combining the accuracy of Tier 1 and Tier 3 with the confidence levels reported in Tier 2 and Tier 4. The scoring methodology is summarized in Table 2, which outlines the coding rules and corresponding scores for each category. Notably, responses that were blank or did not fit any of the coding rules across the four tiers were categorized as NC (No Coding). (Kaltakci Gurel et al., 2015).

Through this structure, the instrument categorizes student conceptions into five diagnostic categories: Understand Concept (UC), Partially Understand (PU), Alternative Conception (AC), Do Not Understand (NU), and No Coding (NC). For instance, students who select the correct answer with correct reasoning and high confidence are classified as UC. In contrast, students who select incorrect answers with high confidence despite flawed reasoning are categorized as having Alternative Conceptions (AC), indicating strong misconceptions. The use of both verbal and mathematical formats in the item pairs allows for the examination of how students apply their understanding across different representational contexts. This structure is especially valuable for diagnosing whether misconceptions are stable across contexts or vary depending on how a question is presented. However, creating parallel items across verbal and mathematical representations posed challenges. In particular, ensuring that both formats targeted the same underlying concept without making one representation easier than the other required careful item design and piloting. Minor wording or symbolic differences could unintentionally influence difficulty, so iterative expert validation and student tryouts were necessary to refine item equivalence.

3.4 Implementation Stage

In the implementation stage, the Instruments Four-tier Multi-representational on Momentum and Impulse (INFORMI) was administered digitally to a group of 37 eleventh-grade science students (MIPA) at a senior high school in Indramayu, Indonesia. The students, aged approximately 16 to 17 years old, consisted of 11 males and 26 females. The sample was selected using a purposive sampling technique, as this school provided access to students who had already studied the concepts of momentum and impulse. The instrument was delivered via a digital platform (Google Forms), enabling efficient distribution and automated recording of responses. Prior to the primary test, students were provided with short instructions and two practice items to familiarize them with the four-tier format and the use of confidence scales. Data collection was conducted in supervised classroom conditions, with students completing the test individually under teacher and researcher supervision to ensure independent work and minimize external assistance. The participants were already familiar with the Google Forms platform, as it was routinely used for school assessments. Students received clear instructions on how to select their answer, provide reasoning, and apply the confidence scales. All responses were automatically stored in the Google Forms database and subsequently exported into a spreadsheet format for further analysis. The implementation took approximately 1 hour and was conducted after students had completed instruction on momentum and impulse.

3.5 Evaluation Stage

Instrument Evaluation

The psychometric analysis of the Instruments Four-tier Multi-representational on Momentum and Impulse (INFORMI) was conducted using the Rasch model based on student responses from 37 participants. The evaluation focused on unidimensionality (construct validity), item fit, reliability, item difficulty, and discrimination.

Based on Figure 5, the raw variance explained by the measure's value of 30.0% indicates that the proportion of variance accounted for by the primary measured construct has met the criteria for construct validity. In the context of Rasch analysis, an instrument is considered to meet construct validity if the observed variance explained by measures exceeds 20% (Sumintono, 2018), suggesting that the data are consistently structured along a single expected latent dimension. In addition, the PCA of residuals was examined to confirm unidimensionality; the results showed that the first contrast had an eigenvalue of 3.54 (13.1%), which slightly exceeded the suggested threshold of 2.0. However, given that the variance explained by the Rasch measures was 30%, the INFORMI instrument can still be considered sufficiently unidimensional for diagnostic purposes. Therefore, achieving 30% provides strong evidence that the INFORMI instrument consistently measures the same conceptual construct: students' understanding of momentum and impulse. The item-level analysis further supports the construct validity evidence.

As shown in Figure 6, all 19 items demonstrated acceptable fit to the Rasch model. The Outfit Mean Square (MNSQ) values range from 0.59 (Item S3) to 1.59 (Item S2), and the standardized Z-scores (ZSTD) values range between –2.34 and 1.95. These values fall within the generally accepted fit range (MNSQ between 0.5 and 1.5, and ZSTD between -2 and +2), suggesting that all items operated as expected without significant misfit. It is worth noting that Item S2's Outfit MNSQ value (1.59) is near the

```
TABLE 23.0         .xlsx                        ZOU953WS.TXT   Jul 14 2025 19:48
INPUT: 37 PERSON  19 ITEM  REPORTED: 37 PERSON  19 ITEM  4 CATS MINISTEP 5.8.5.0
--------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units
                                          Eigenvalue   Observed   Expected
Total raw variance in observations     =      27.1245 100.0%         100.0%
  Raw variance explained by measures    =       8.1245  30.0%          32.5%
    Raw variance explained by persons   =       2.8829  10.6%          11.5%
    Raw Variance explained by items     =       5.2417  19.3%          20.9%
  Raw unexplained variance (total)      =      19.0000  70.0% 100.0%   67.5%
    Unexplned variance in 1st contrast =       3.5492  13.1%  18.7%
    Unexplned variance in 2nd contrast =       2.2337   8.2%  11.8%
    Unexplned variance in 3rd contrast =       2.2095   8.1%  11.6%
    Unexplned variance in 4th contrast =       1.8438   6.8%   9.7%
    Unexplned variance in 5th contrast =       1.5318   5.6%   8.1%
```

Figure 5 Result of unidimensionality (construct validity)

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
2	80	36	-.10	.21	1.78	2.37	1.59	1.78	A .33	.33	27.8	52.1	S2
8	76	36	.03	.19	1.52	1.88	1.35	1.22	B .52	.34	16.7	49.8	S8
1	37	37	1.22	.17	1.35	1.80	1.46	1.95	C .14	.49	16.2	27.5	S1
7	89	36	-.57	.25	1.44	1.22	1.36	1.06	D .33	.28	55.6	56.3	S7
11	96	36	-1.14	.32	1.34	.94	1.38	1.04	E .28	.23	75.0	68.4	S11
13	88	35	-.70	.27	1.34	.96	1.08	.33	F .46	.26	57.1	57.5	S13
16	59	36	.59	.17	1.24	1.33	1.29	1.36	G .20	.43	27.8	31.7	S16
5	67	36	.35	.18	1.06	.34	1.19	.83	H .14	.40	36.1	40.9	S5
6	46	36	.94	.17	1.05	.35	1.00	.09	I .55	.48	16.7	21.6	S6
12	89	35	-.78	.28	1.02	.16	.90	-.17	J .33	.24	57.1	58.4	S12
18	81	37	-.06	.20	.99	.06	.85	-.44	i .51	.34	54.1	52.1	S18
4	81	36	-.15	.21	.98	.03	.97	-.01	h .00	.33	52.8	52.9	S4
14	70	37	.32	.17	.82	-.87	.81	-.79	g .40	.39	37.8	41.3	S14
19	78	35	-.09	.21	.82	-.58	.70	-1.02	f .54	.33	62.9	52.2	S19
10	55	31	.45	.19	.81	-.91	.74	-1.09	e .62	.43	48.4	38.9	S10
9	96	36	-1.14	.32	.73	-.68	.70	-.79	d .36	.23	63.9	68.4	S9
15	80	36	-.11	.21	.67	-1.27	.71	-.98	c .45	.33	61.1	52.2	S15
17	72	37	.26	.18	.61	-2.10	.60	-1.82	b .24	.38	51.4	43.6	S17
3	56	36	.68	.17	.60	-2.74	.59	-2.34	a .46	.44	38.9	28.2	S3
MEAN	73.5	35.8	.00	.21	1.06	.12	1.01	.01			45.1	47.1	

Figure 6 Result of item fit and difficulty statistics

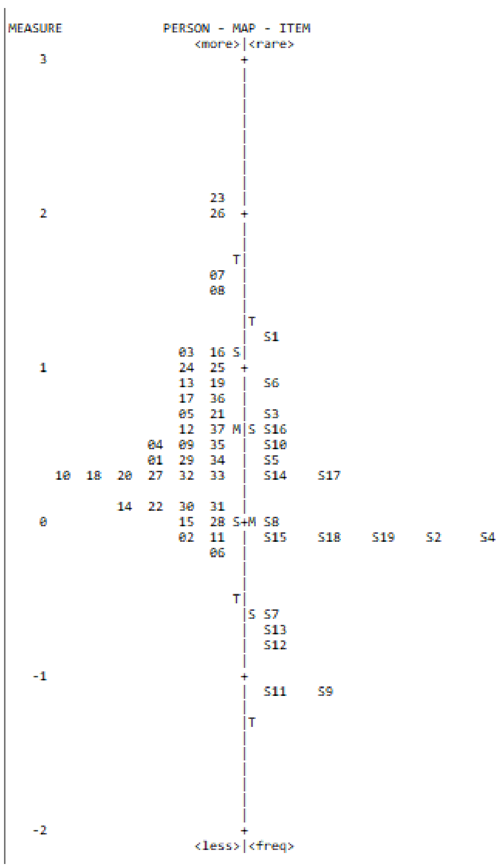

Figure 7 Wright map

upper limit, suggesting that some student responses were less predictable than expected by the Rasch model. This could be due to variation in students' conceptual understanding or occasional guessing on this particular item The item difficulty measures (JMLE) showed a range distribution spanning from –1.14 logits (Item S11 and S9, the easiest) to 1.22 logits (Item S1, the most difficult), indicating that the instrument successfully captures a broad spectrum of conceptual mastery levels related to momentum and impulse.

Regarding item discrimination, the Point-Measure Correlation (PTMEA Corr.) values are shown in Figure 6, ranging from 0.00 to 0.62. Most items fell into the "good" and "excellent" categories, showed satisfactory discrimination, with values above 0.30 considered acceptable. Items such as S1 (0.62), S4 (0.58), and S15 (0.46) effectively distinguished between high- and low-performing students, likely because these items target key conceptual aspects of momentum and impulse and include well-constructed distractors that challenge students' understanding. However, one item, S4, had a PTMEA Corr. of 0.00, indicating that it failed to correlate with the underlying construct and may not meaningfully differentiate students' levels of understanding. Such items often involve either ambiguous distractors or abstract concepts, such as the vector nature of momentum and the force–time interaction in impulses, which may not be directly observable by students and are therefore challenging to understand, a common issue in multi-tier diagnostic assessments. (Gurel et al., 2015). This particular item should be revised or replaced in future development stages to improve the instrument's overall precision.

To further illustrate the relationship between student ability and item difficulty, a Wright Map was constructed (Figure 7). The map displays both the distribution of person abilities (on the left) and item difficulties (on the right) along the same logit scale. The results indicate that the items are generally well-targeted to the participants' ability levels. Item S1 was identified as the most difficult, while Items S9 and S11 were the easiest. Although most items clustered around the average student ability, a small gap appeared at the higher end of the scale, suggesting that the instrument has limited capacity to differentiate among the most advanced students. This insight is valuable for future refinement, ensuring the instrument maintains adequate coverage across all ability levels.

The instrument's reliability statistics, shown in Figure 8, further reinforce its psychometric strength. The item reliability was 0.86, categorized as very good, indicating that the hierarchy of item difficulty is stable and replicable. The person's reliability, reported at 0.67, is moderate. Rasch analysis further indicated a person separation index of 1.43, suggesting that the instrument can adequately distinguish between students of different ability levels. As diagnostic instruments are designed to probe diverse reasoning processes rather than produce high-stakes scores, moderate person reliability is both expected and acceptable, especially in conceptual domains where student reasoning tends to be variable and context-sensitive. (Setiani et al., 2023; Sternberg, 2018).

Additionally, the instrument's internal consistency, as indicated by the KR-20 coefficient, was 0.62. While this figure is below the standard threshold for summative tests, it is adequate for formative diagnostic purposes. In line with previous studies (Istiyono, 2022; Samsudin, Fratiwi, et al., 2020; Uccio et al., 2019)This level of consistency suggests that the instrument is dependable for identifying patterns of understanding and misconception rather than assigning definitive scores. Moreover, several methodological references emphasize that coefficients between 0.6 and 0.7 remain acceptable for exploratory research and early-stage instrument development, particularly in educational contexts. (Hair, 2009; Taber, 2018)

Overall, these results indicate that the INFORMI instrument demonstrates strong psychometric properties. Most items are valid, reliable, and able to discriminate well across different levels of student understanding. The variation in item difficulty enhances the instrument's diagnostic power, while the multi-representational, four-tier structure allows for deeper insight into students' reasoning. Nevertheless, some refinement is warranted, particularly for Item S4, to strengthen further the instrument's capacity to detect and analyze conceptual difficulties in momentum and impulse topics.

Diagnosis of Students' Conceptions

The use of the Instruments Four-tier Multi-representational on Momentum and Impulse (INFORMI) provided a comprehensive framework for identifying students' conceptions of momentum and impulse. The instrument's four-tier structure enabled detailed classification of student responses into five conceptual categories: Understand Concept (UC), Partially Understand (PU), Alternative Conceptions (AC), Do Not Understand (NU), and No Coding (NC).

As shown in Figure 9, the most significant proportion of student responses fell into the Partially Understand (PU) category, with 43% for mathematical items and 34% for verbal items, indicating that many students had fragmented or incomplete conceptions across item formats. A similar pattern was observed in the Understand Concept (UC) category, where 42% of students correctly answered

```
TABLE 3.1 bismillah.xlsx                            ZOU156WS.TXT  Aug 21 2025 09:08
INPUT: 37 PERSON  19 ITEM  REPORTED: 37 PERSON  19 ITEM  4 CATS MINISTEP 5.8.5.0
--------------------------------------------------------------------------------

     SUMMARY OF 37 MEASURED PERSON
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN      37.7      18.4         .59     .29      1.01    .10   1.01    .09 |
|  SEM       1.0        .1         .09     .01       .06    .16    .07    .17 |
| P.SD       6.2        .7         .55     .06       .33    .97    .43   1.00 |
| S.SD       6.3        .7         .56     .06       .34    .99    .43   1.02 |
| MAX.      52.0      19.0        2.08     .49      1.86   2.34   2.42   2.69 |
| MIN.      28.0      16.0        -.16     .24       .41  -1.66    .45  -1.41 |
|------------------------------------------------------------------------------|
| REAL RMSE     .32 TRUE SD      .45  SEPARATION  1.43  PERSON RELIABILITY  .67 |
|MODEL RMSE     .30 TRUE SD      .46  SEPARATION  1.55  PERSON RELIABILITY  .71 |
| S.E. OF PERSON MEAN = .09                                                    |
-------------------------------------------------------------------------------
PERSON RAW SCORE-TO-MEASURE CORRELATION = .96 (approximate due to missing data)
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .62  SEM = 3.83
STANDARDIZED (50 ITEM) RELIABILITY = .86
```

Figure 8 Summary of Rasch statistics

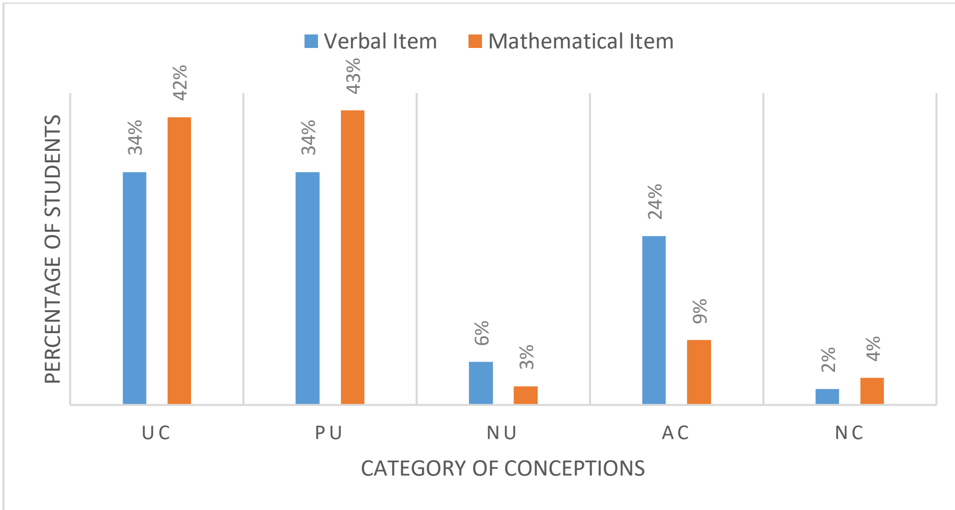

Figure 9 Distribution of student conceptions based on verbal and mathematical items

mathematical items, compared to 34% for verbal items. Interestingly, verbal items revealed more Alternative Conceptions (AC), with 24% of students demonstrating consistent misconceptions, compared to only 9% in mathematical items. This indicates that the most prevalent misconception occurred in verbal representations, suggesting that while students may find structured mathematical problems easier to navigate, verbal or conceptual formats are more likely to expose underlying misunderstandings. The relatively low percentages in the Do Not Understand (NU) and No Coding (NC) categories for both item types further support the instrument's diagnostic power in capturing meaningful cognitive patterns rather than random guessing.

This pattern provides empirical evidence of the instrument's effectiveness, as the use of multi-representational items (verbal and mathematical) revealed misconceptions that would have remained hidden if only one representation had been used. In particular, the consistently higher percentage of Alternative Conceptions in verbal items demonstrates the instrument's capacity to expose students' intuitive but incorrect reasoning, thereby validating its role as an effective diagnostic tool.

These findings align with prior studies emphasizing the importance of using multi-representational and multi-tier assessments to diagnose students' conceptions. For instance, (Bayuni, Sopandi, & Sujana, 2018; Milenković, Hrin, Segedinac, & Horvat, 2016) highlighted that tiered diagnostic tools can more effectively distinguish between misconceptions and a simple lack of knowledge. Similarly, (Munfaridah, Avraamidou, & Goedhart, 2021) found that students often exhibit inconsistent reasoning when switching between representational formats, particularly in abstract topics like force and momentum. (Samsudin, Fratiwi, et al., 2020) Also reported that verbal representations are more prone to elicit entrenched misconceptions, as they require interpretation beyond procedural calculations. Thus, the pattern observed in Figure 9 reinforces the notion that verbal and conceptual formats are essential for revealing deep-seated misunderstandings that might remain hidden in purely mathematical contexts. Incorporating both representations in diagnostic instruments such as INFORMI not only increases diagnostic validity but also provides richer insights for instructional decision-making and targeted remediation.

Building on this, a closer look at specific indicators further illustrates the persistence and distribution of these misconceptions. Although the instrument demonstrated strong psychometric properties, several students still exhibited misconceptions. The results revealed that misconceptions (AC) were prevalent across several indicators. The highest rate of misconceptions was identified in the indicator "analyzing the magnitude of momentum between two objects being pushed." In this indicator, 43% of students exhibited misconceptions when responding to the verbal item, and 24% did so on the mathematical item. This finding aligns with the study of (Gurel et al., 2015), who emphasized that even after instruction, students often retain misconceptions rooted in intuitive or everyday experiences. These misconceptions may persist because students rely on surface features of problems or on prior knowledge that conflicts with scientific principles.

To further illustrate the finding, the distribution of students' conceptions on this particular indicator is illustrated in Figure 10. The dominance of alternative conceptions demonstrates that the instrument is effective in detecting specific misconceptions rather than merely assessing correctness. This provides clear evidence of how the instrument correlates with students' conceptual understanding, as it not only identifies the prevalence of scientifically accurate conceptions but also captures the persistence of alternative conceptions. Such diagnostic insight highlights the instrument's effectiveness for guiding targeted conceptual clarification in future instruction.

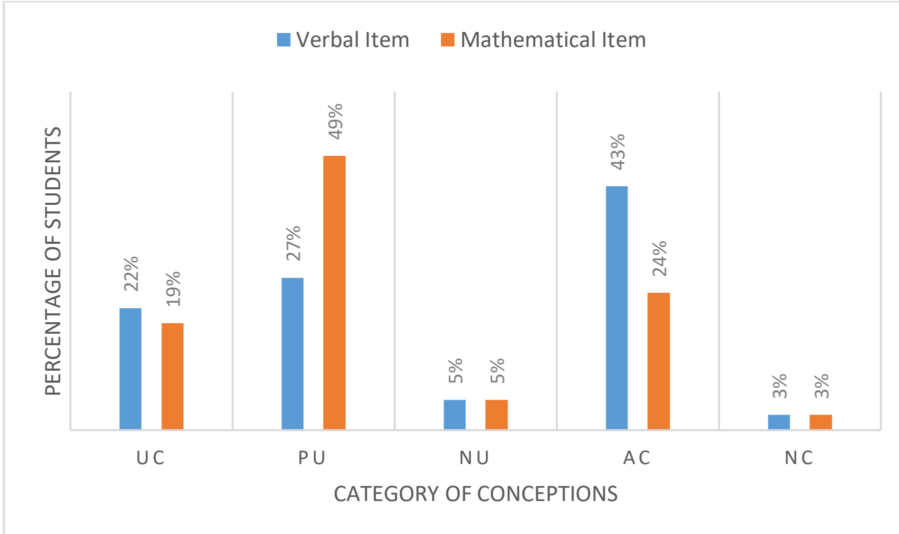

Figure 10 Distribution of student conceptions on the indicator "Analyzing the magnitude of momentum between two objects" based

The higher percentage of misconceptions found in verbal items compared to mathematical items can be attributed to the abstract and interpretive nature of verbal representations, which often rely heavily on students' prior knowledge and everyday experiences. Unlike mathematical items, which may offer structured cues through equations or numerical relationships, verbal questions demand deeper conceptual understanding and the ability to model physical phenomena mentally. As noted by (Setiani et al., 2023)Verbal representations tend to activate intuitive but scientifically inaccurate reasoning, particularly in topics such as momentum and impulse, where students often relate motion and force to real-life perceptions rather than formal physics principles. Similarly, (Gurel et al., 2015), emphasized that verbal diagnostic items are more effective in revealing deeply rooted misconceptions because they expose the learner's conceptual framework without procedural support. These findings suggest that while mathematical items may mask conceptual difficulties in procedural fluency, verbal items are more effective at uncovering inconsistencies between students' reasoning and scientific concepts, thereby providing clear evidence of INFORMI's diagnostic effectiveness. Moreover, the results highlight specific alternative conceptions, such as the belief that a greater mass automatically results in greater momentum regardless of velocity, or the confusion between force and impulse, which are rooted in everyday experiences and intuitive reasoning. By making these misconceptions visible, INFORMI provides teachers with concrete input for lesson planning, particularly in designing targeted instructional strategies to confront and remediate the most prevalent alternative conceptions. In practical terms, teachers can use INFORMI as a pre-test to identify prior knowledge, as a formative assessment to monitor conceptual change, and as a post-test to evaluate learning outcomes. The diagnostic results can directly inform lesson adaptations, for example, using vector diagrams, interactive simulations, or problem-based discussions when a large proportion of students hold misconceptions about the relationship between mass, velocity, and momentum. Furthermore, the instrument could be integrated into AI-assisted analysis, enabling teachers to obtain instant feedback on students' conceptual profiles and to make evidence-based instructional decisions.

4. CONCLUSION

This study successfully developed and validated the Instruments Four-tier Multi-representational on Momentum and Impulse (INFORMI), a diagnostic tool designed to analyze students' conceptions of momentum and impulse. Rasch model analysis using MINISTEP software confirmed that all 19 items met the required psychometric standards, with Outfit MNSQ and ZSTD values falling within acceptable ranges. The instrument demonstrated strong construct validity, with a raw variance explained by measures reaching 30%, and high item reliability of 0.86, indicating a stable item hierarchy. Although person reliability was moderate (0.67), this is acceptable for diagnostic instruments targeting conceptual reasoning. The KR-20 value of 0.62 further supports INFORMI's internal consistency. The items spanned various difficulty levels and mainly exhibited good to excellent discrimination, confirming the instrument's sensitivity in identifying different levels of student understanding. Overall, INFORMI is a valid, reliable, and conceptually grounded four-tier diagnostic instrument capable of uncovering nuanced student misconceptions and supporting more informed instructional practices in physics education.

It is recommended that educators incorporate INFORMI into classroom assessment practices to identify students' conceptual gaps and to design more targeted, evidence-based remediation strategies. Teachers can administer INFORMI as a pre-instructional diagnostic tool or during instruction to monitor conceptual change. The results may guide grouping strategies, targeted scaffolding, and classroom discussions that directly address identified misconceptions. Given its multi-representational and four-tier format, INFORMI is particularly effective in diagnosing misconceptions that are not easily captured through traditional assessments. This is especially valuable in abstract physics topics such as momentum and impulse, where students often rely on intuitive reasoning.

In practice, INFORMI's diagnostic results can directly guide teachers in lesson planning. For example, common misconceptions, such as the belief that greater mass always means greater momentum, can be addressed through momentum-conservation experiments or interactive simulations. Similarly, confusion between force and impulse may be remediated by using motion sensors or PhET simulations to visualize force–time graphs. By linking diagnostic outcomes with targeted classroom strategies, INFORMI not only identifies misconceptions but also informs concrete instructional adaptations.

Future research may explore the adaptation of INFORMI to other domains in physics and investigate its effectiveness in improving conceptual change through instructional interventions. Beyond this, more specific directions include testing INFORMI with larger, more diverse samples to enhance generalizability, examining its applicability across cultural contexts, and developing digital versions with automated scoring and feedback. Further studies could also evaluate how integrating INFORMI into classroom interventions impacts long-term conceptual change, thereby strengthening the link between diagnostic assessment and instructional improvement.

Nevertheless, this study has some limitations. The sample size was relatively small (37 students), which restricts the generalizability of the findings. The instrument focused solely on momentum and impulse, so further studies are needed to test its applicability in other physics domains. In addition, the moderate person reliability (0.67) reflects the variability of students' reasoning and indicates that further refinement is needed. Finally, the verbal items, while effective in eliciting misconceptions, may disadvantage students with lower reading comprehension skills. Additionally, as this study was conducted in the Indonesian context, the potential influence of cultural and curricular factors on students' reasoning should be acknowledged, which may limit transferability to other educational settings. Moreover, while the four-tier format provides rich diagnostic information, it may also be cognitively demanding for some learners, suggesting that scaffolding or simplified versions could be considered in future adaptations. These limitations provide important directions for future research and refinement of the instrument.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to the Physics Education Master's Program of Universitas Ahmad Dahlan for academic assistance and to SMAN 1 Sliyeg, Indramayu, for the opportunity to collect research data. In addition, the authors extend their most profound appreciation to the Research and Community Service Forum of the Association of Indonesian Public Institutes for Teacher Education (FORUM LPPM ALPTKNI) for the research funding provided through the RKLI Program 2025, as stated in Decree No. 4/III/2025 on the Awardees of the 2025 RKLI Research Program..

REFERENCES

Adianto, T., & Rusli, M. A. (2021). Analysis of Student's Difficulties in Solving Physics Problem: Impulse and Momentum Topics. *Unnes Science Education Journal*, *10*(1), 24–33. https://doi.org/10.15294/usej.v10i1.41517

Adimayuda, R., Aminudin, A. H., Kaniawati, I., Suhendi, E., & Samsudin, A. (2020). A multitier open-ended momentum and impulse (MOMI) instrument: Developing and assessing quality of conception of 11th grade sundanese students with rasch analysis. *International Journal of Scientific and Technology Research*, *9*(2), 4799–4804.

Akcanca, N., & Cerrah Özsevgeç, L. (2020). Effect on Academic Achievement and Misconceptions of Pre-service Teachers through Combining Different Teaching Methods in a Preschool Science Course. *Journal of Science Learning*, *4*(1), 41–49. https://doi.org/10.17509/jsl.v4i1.24672

Amalia, S. A., Suhendi, E., Kaniawati, I., Samsudin, A., Fratiwi, N. J., Hidayat, S. R., Zulfikar, A., Sholihat, F. N., Jubaedah, D. S., Setyadin, A. H., Purwanto, M. G., Muhaimin, M. H., Bhakti, S. S., & Afif, N. F. (2019). Diagnosis of Student's Misconception on Momentum and Impulse Trough Inquiry Learning with Computer Simulation (ILCS). *Journal of Physics: Conference Series*, *1204*(1). https://doi.org/10.1088/1742-6596/1204/1/012073

Anandi, R. P., Sumintono, B., Zailaini, M. A., & Syafitri, R. (2024). *Using Rasch Model to Assess the Foreign Language Speaking Anxiety Scale (FLSAS) Among University Students in Salatiga, Indonesia* (Issue Proms 2023). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-494-5_6

Banawi, A., Sopandi, W., Kadarohman, A., & Solehuddin, M. (2022). Five-Tier Multiple-Choice Diagnostic Test Development: Empirical Evidences to Improve Students' Science Literacy. *Proceedings of the International Conference on Madrasah Reform 2021 (ICMR 2021)*, *633*. https://doi.org/10.2991/assehr.k.220104.020

Bayuni, T. C., Sopandi, W., & Sujana, A. (2018). Identification misconception of primary school teacher education students in changes of matters using a five-tier diagnostic test. *Journal of Physics: Conference Series*, *1013*(1). https://doi.org/10.1088/1742-6596/1013/1/012086

Ben-Abu, Y. (2019). The symmetrical energy and momentum principles in introductory university physics courses. *Physics Education*, *54*(4). https://doi.org/10.1088/1361-6552/ab0e3a

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences*. https://doi.org/10.1017/CBO9781107415324.004

Boone, W. J. (2020). Rasch basics for the novice. In *Rasch Measurement: Applications in Quantitative Educational Research*. https://doi.org/10.1007/978-981-15-1800-3_2

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, *19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175

Crogman, H., Peters, R., & Trebeaucrogman, M. (2018). Probing Students' Misconceptions: Results from a Concept Inventory and Their Understanding in Science Learning. *European J of Physics Education*, *9*(1), 1309–7202.

Diani, R., Anggraeni, Y. M., Fujiani, D., & Yuliani, E. (2023). *The identification of senior high school students' misconception on physics using the four-tier diagnostic test with certainty of response index*. 020016. https://doi.org/10.1063/5.0123837

Diyanahesa, N. E.-H., Kusairi, S., & Latifah, E. (2017). Development of Misconception Diagnostic Test in Momentum and Impulse Using Isomorphic Problem. *Journal of Physics: Theories and Applications*, *1*(2), 145. https://doi.org/10.20961/jphystheor-appl.v1i2.19314

Fang, N., & Guo, Y. (2023). Improving Student Learning of Impulse and Momentum in Particle Dynamics Through Computer Simulation and Animation. *Journal of Educational Computing Research*, *60*(8), 1969–1990. https://doi.org/10.1177/07356331221096979

Fratiwi, N. J., Samsudin, A., Ramalis, T. R., & Costu, B. (2020). Changing students' conceptions of Newton's second law through express-refute-investigate-clarify (ERIC) text. *Universal Journal of Educational Research*, *8*(6), 2701–2709. https://doi.org/10.13189/ujer.2020.080655

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). https://doi.org/10.1038/s41405-021-00067-4

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Hair, J. F. (2009). *Multivariate data analysis*.

Hayat, B., Putra, M. D. K., & Suryadi, B. (2020). Comparing item parameter estimates and fit statistics of the Rasch model from three different traditions. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *24*(1). https://doi.org/10.21831/pep.v24i1.29871

Istiyono, E. (2022). Diagnostic Tests as an Important Pillar in Today's Physics Learning: Four-tier Diagnostic Test a Comprehensive Diagnostic Test Solution. *Journal of Physics: Conference Series*, *2392*(1). https://doi.org/10.1088/1742-6596/2392/1/012001

Jung, H., Kim, Y. R., Lee, H., & Shin, Y. (2019). Advanced instructional design for successive E-learning: Based on the successive approximation model (SAM). *International Journal on E-Learning: Corporate, Government, Healthcare, and Higher Education*, *18*(2), 191–204.

Kaltakci Gurel, D., Eryilmaz, A., & McDermott, L. C. (2015). A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science. *EURASIA Journal of Mathematics, Science and Technology Education*, *11*(5). https://doi.org/10.12973/eurasia.2015.1369a

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 and Technological Education*, *35*(2), 238–260. https://doi.org/10.1080/02635143.2017.1310094

Kaniawati, I., Fratiwi, N. J., Danawan, A., Suyana, I., Samsudin, A., & Suhendi, E. (2019). Analyzing students' misconceptions about Newton's laws through four-tier Newtonian test (FTNT). *Journal of Turkish Science Education*, *16*(1), 110–122. https://doi.org/10.12973/tused.10269a

Kaniawati, I., Rahmadani, S., Fratiwi, N. J., Suyana, I., Danawan, A., Samsudin, A., & Suhendi, E. (2020). An analysis of students' misconceptions about the implementation of active learning of optics and photonics approach assisted by computer simulation. *International Journal of Emerging Technologies in Learning*, *15*(9), 76–93. https://doi.org/10.3991/ijet.v15i09.12217

Kurniawan, M. A., Rahayu, S., Fajaroh, F., & Almuntasheri, S. (2020). Effectiveness of Dual Situated Learning Model in Improving High School Students' Conceptions of Chemistry Equilibrium and Preventing Their misconceptions. *Journal of Science Learning*, *3*(2), 99–105. https://doi.org/10.17509/jsl.v3i2.22277

Lassonde, K. A., Kendeou, P., & O'Brien, E. J. (2016). Refutation texts: Overcoming psychology misconceptions that are resistant to change. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 62–74. https://doi.org/10.1037/stl0000054

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.

Marthaliakirana, A. D., Suwono, H., Saefi, M., & Gofur, A. (2022). Problem-based learning with metacognitive prompts for enhancing argumentation and critical thinking of secondary school students. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(9). https://doi.org/10.29333/ejmste/12304

Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Development of a Three-Tier Test as a Valid Diagnostic Tool for Identification of Misconceptions Related to Carbohydrates. *Journal of Chemical Education*, *93*(9), 1514–1520. https://doi.org/10.1021/acs.jchemed.6b00261

Munfaridah, N., Avraamidou, L., & Goedhart, M. (2021). The Use of Multiple Representations in Undergraduate Physics Education: What Do we Know and Where Do we Go from Here? *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(1), 1–19. https://doi.org/10.29333/ejmste/9577

Nurdini, N., Suhandi, A., Ramalis, T., Samsudin, A., Fratiwi, N. J., & Costu, B. (2020). Developing Multitier Instrument of Fluids Concepts (MIFO) to Measure Student's Conception: A Rasch Analysis Approach. *Journal of Advanced Research in Dynamical and Control Systems*, *12*(6), 3069–3083. https://doi.org/10.5373/JARDCS/V12I6/S20201273

Ozdemir, E. (2022). Animated Concept Cartoons as a Starter for Cognitive Conflict in Online Science Learning: A Case of Circular Motion. *Journal of Science Learning*, *5*(2), 242–249. https://doi.org/10.17509/jsl.v5i2.41191

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 20111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Saat, R. M., Fadzil, H. M., Aziz, N. A. A., Haron, K., Rashid, K. A., & Shamsuar, N. R. (2016). Development of an online three-tier diagnostic test to assess pre-university students' understanding of cellular respiration. *Journal of Baltic Science Education*, *15*(4), 532–546. https://doi.org/10.33225/jbse/16.15.532

Sakinah, N., Amalia, I. F., & Adimayuda, R. (2023). Identifying senior high school students' misconception on the momentum and impulse concepts uses three-tier diagnostic test. *Research in Physics Education*, *2*(1), 18–25. https://doi.org/10.31980/ripe.v2i1.2377

Salas-Pilco, S. Z., Yang, Y., & Zhang, Z. (2022). Student engagement in online learning in Latin American higher education during the COVID-19 pandemic: A systematic review. … *Journal of Educational* …. https://doi.org/10.1111/bjet.13190

Samsudin, A., Azizah, N., Sasmita, D., Rasmitadila, F., M. A., Supriyatman, & Wibowo, F. C. (2020). Analyzing the students' conceptual change on the kinetic theory of gases as a learning effect through a computer simulation-assisted conceptual change model. *Universal Journal of Educational Research*, *8*(2), 425–437. https://doi.org/10.13189/ujer.2020.080213

Samsudin, A., Fratiwi, N. J., Ramalis, T. R., Aminudin, A. H., Costu, B., & Nurtanto, M. (2020). Using Rasch analysis to develop a multi-representation of the tier instrument on Newton's law (motion). *International Journal of Psychosocial Rehabilitation*. https://doi.org/10.37200/IJPR/V24I6/PR260865

Setiani, A., Firmansyah, D., & Ramadhani, C. I. (2023). Analysis of Momentum and Impulse Misconceptions of College Students Using Four-Tier Diagnostic Tests. *International Journal of Education* *and Teaching Zone*, *2*(3), 349–359. https://doi.org/10.57092/ijetz.v2i3.146

Spatioti, A. G., Kazanidis, I., & Pange, J. (2022). A Comparative Study of the ADDIE Instructional Design Model in Distance Education. In *Information (Switzerland)* (Vol. 13, Issue 9). https://doi.org/10.3390/info13090402

Sternberg, R. J. (2018). Context-Sensitive Cognitive and Educational Testing. In *Educational Psychology Review* (Vol. 30, Issue 3, pp. 857–884). https://doi.org/10.1007/s10648-017-9428-0

Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assessment for Learning*. https://doi.org/10.2991/icei-17.2018.11

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Triyani, G., Danawan, A., Suyana, I., & Kaniawati, I. (2019). An investigation of students' misconceptions about momentum and impulse through interactive conceptual Instruction (ICI) with computer simulation. *Journal of Physics: Conference Series*, *1280*(5). https://doi.org/10.1088/1742-6596/1280/5/052008

Uccio, U. S. di, Colantonio, A., Galano, S., Marzoli, I., Trani, F., & Testa, I. (2019). Design and validation of a two-tier questionnaire on basic aspects in quantum mechanics. *Physical Review Physics Education Research*, *15*(1). https://doi.org/10.1103/physrevphyseducres.15.010137

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335–345. https://doi.org/10.1021/ed100852t

Wong, K. S., & Wong, H. (2022). Understanding the Law of Conservation of Momentum in One-Dimensional Collisions Between Two Objects Using the Velocity Space Approach. *The Physics Teacher*, *60*(2), 94–96. https://doi.org/10.1119/5.0024603

Wu, M., Tian, P., Sun, D., Feng, D., & Luo, M. (2025). Evaluating Students' Conceptual Understanding of Isomers Based on a Four-Tier Diagnostic Tool in Upper Secondary Schools. *International Journal of Science and Mathematics Education*, *23*, 907–947.

Yang, D. C. (2022). Investigating the differences between confidence ratings in the answer and reason tiers in fourth graders via online four-tier test. *Studies in Educational Evaluation*, *72*. https://doi.org/10.1016/j.stueduc.2022.101127

Zarepour, B., Soleimani, H., & Sanjari, M. (2024). Exploring Assessment Strategies Used to Promote Metacognitive Development in Science Education. *Assessment and Practice in Educational Sciences*, *2*(2), 1–9.