Development of assessment instruments for students' graphical representation abilities in the topic of dynamic electricity

Rasna Solehayati, Ade Yeti Nuryantini*, Ida Farida Chaidir

Master's Programme in Science Education, Postgraduate,UIN Sunan Gunung Djati Bandung, A.H. Nasution St No. 105, Bandung, 40296, Indonesia

*Corresponding author, email: ade.yeti@uinsgd.ac.id

Article History
Received: 17 January 2026
Revised: 12 June 2026
Accepted: 22 July 2026

Keywords
Assessment
Dynamic electricity
Graphical representation
Instruments
Physics test

Abstract
This study focuses on the development of physics test instruments for high school students on the topic of Dynamic Electricity. This study aims to analyze the content quality of the test instrument context, evaluate the the feasibility of the test instrument, and the student's graphical representation abilities on the topic of Dynamic Electricity. This research and development studiy implemented the R&D method with the 4D model (Define, Design, Development, and Disseminate). The test instrument was trialed in class XII IPA at Karya Pembangunan 1 Ciparay High School with a total of 50 students as research subjects. Item analysis was conducted using Classical Test Teory (CTT) using SPSS 29.0 and Rasch Modeling using MINISTEP software to evaluate validity, reliability, difficulty level, and anad student response patterns for 12 item (6 items type A and 6 items type B). The results demonstrated that the test instrument possesses good content validity with overall adequate psychometric estimates. Out of 12 items tested, 7 items proved feasible and were retained, while 5 items required revision or elimination due to significant misfit, overfit, and cognitive redundancy. Students' graphical representation ability was predominantly in the fair category (46%), with 72% of students falling into the high-ability category, indicating a ceiling effect. These findings highlight the necessity of refining instrument design by incorporating non-linear graphs and higher difficulty levels to effectively measure higher-order thinking skills and support diagnostic assessment in physics learning.

Solehayati, R., Nuryantini, A, Y., Chaidir, I, F. (2026). Development of assessment instruments for students' graphical representation abilities in the topic of dynamic electricity. *Momentum: Physics Education Journal*, *10*(1), 214-229. https://doi.org/10.21067/mpej.v10i1.13759

1. Introduction

Education in the twenty-first century has driven shifts in learning paradigms due to globalization. The demand for the reconstruction of educational objectives and learning outcomes in line with 21st century competency requirements (Isabirye et al., 2025). Education in this era is directed at preparing students to face complex challenges and be able to adapt to various dynamics that occur (González-Pérez & Ramírez-Montoya, 2022). Therefore, there is a need for comprehensive competency-based skills that are student-cantered (Lu & Lin, 2025).

One of the essential competencies that should be developed in 21st century science education is is the ability to contruct and interpret multiple representations, particularly graphical representations (Casey et al., 2025). Graphical representation ability enables students to externalize their conceptual understanding through various forms, such as verbal, mathematical, visual, and graphical, thereby supporting a more in-depth scientific reasoning and problem-solving process (Masrurin et al., 2025; Surahmi, 2016). In physics learning, graphs serve not merely as visual illustration but as analytical tools that communicate relationships among physics variables reveal patterns, support predictions, and facilitate evidence-based reasoning (Munfaridah et al., 2021). Consequently, students' ability to interpret, contract, and evaluate graphs has become as important indicator of conceptual understanding and higher order thinking skills in physics learning (Kusumaningtyas et al., 2024). Students who process well developed graphical representation baillis are generally more capable of connecting mathematical relationships with underlying physical phenomena, whereas limited graphical representation skills often hider contextual reasoning and scientific argumentation (Fisher et al., 2026; Niyomufasha et al., 2024).

doi: 10.21067/mpej.v10i1.13759
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2548-9135

Particularly in the topic of dynamic electricity, graphical representation plays a strategic fundamental role because many electrical concepts cannot be directly observed but must instead be inferred through relationship between physical variables. Unlike many secondary physics topics in which phenomena can be visualized through direct motion or observable changes, understanding dynamic electricity requires students to interpret the relationships between voltage, current, resistance, and electrical power through graphical representations (Hermawati, 2022). Students are expected to interpret current voltage graphs, compare slopes, distinguish ohmic and nonohmic behaviour, analyse equivalent resistance, and infer electrical characteristics from graphical information. These cognitive processes require students not only to read graphs but also to integrate mathematical, conceptual, and visual representations simultaneously. This aligns with (Tesfaye, 2025) view graphical representation ability becomes a critical competency for understanding dynamic electricity and is more intensively required than in many other secondary physics topics where graphical interpretation plays a less central role.

In physics learning, graphical representation serves as a fundamental tool for understanding complex phenomena, particularly in topics such as dynamic electricity. However, research findings consistently indicate that students still face difficulties in reading, interpreting, and constructing physical graphs correctly. These persistent challenges do not merely affect their ability to complete procedural tasks but also significantly impacts their deeper conceptual understanding and the development of their scientific reasoning skills (Lichtenberger et al., 2025; Sharma, 2025). Consequently, students often resort to routine mathematical calculations while failing to grasp the underlying physical meanings and relationships embedded within the graphical data, which ultimately hinders their capacity to apply physics concepts to novel or real-world situations.

Considerable research efforts have been invested in developing assessment instruments for evaluating student's representational abilities in physics. Existing tools have largely focused on topics such as sound wave (Pradana, 2025; Rangkuti & Karam, 2022), mechanics (Istiyono & Fenditasari, 2021; Khotimah & Fariyani, 2025), heat and temperature (Aprilia & Dwandaru, 2024; Kim & Nam, 2021), or on general multiple-representation competencies, and have demonstrated satisfactory validity and reliability within those contexts. Despite these advances, limited attention has been given to developing an easement instrument specifically designed to measure graphical representation skills in dynamic electricity, a topic that requires students to interpret complex relationships among electrical variables through graphical representations. Furthermore, previous studies have predominantly reported conventional psychometric evidence, such as validity and reliability, with relatively limited emphasis on developing context-specific instruments capable of providing comprehensive information about students' graphical representation performance. This limitation restricts teachers' ability to identify specific learning difficulties and design targeted instructional support. Therefore, there remains a clear need for a context sensitive and psychometrically robust assessment instrument that can accurately assess students' graphical representation abilities, in dynamic electricity provide meaning evidence to support classroom assessment and instructional decision making.

The novelty of this study lies in the development of a context-specific assessment instrument designed to measure six dimensions of graphical representation skills in the context of dynamic electricity, a physics topic characterized by unique conceptual demands for interpreting relationships among current, voltage, resistance, and electrical power. The instrument is developed based on the six indicators of graphical representation proposed by Aprilia & Dwandaru (2024), which are adapted into authentic, domain-specific assessment tasks that reflect the conceptual characteristics of dynamic electricity. By focusing on a topic that has received limited attention in previous instrument development studies, this research extends the application of graphical representation assessment beyond existing contexts and provides a more focused framework for evaluating students' graphical representation competencies. The resulting instrument is expected to assist physics teachers in identifying students' strengths and learning difficulties in graphical representation, thereby supporting evidence-based instructional planning and targeted learning interventions.

Based on the identified research gap and the proposed contribution, this study aims develop and validate a context specific assessment instrument for measuring senior high school student's graphical representation ability in topic of dynamic electricity. The instrument was developed

through a systematic research and development process comprising needs analysis, design, development, and psychometric evaluation to ensure that it meets educational measurement standards and aligns with the competency demands of 21st learning. Ultimately, the developed instrument is expected to provide comprehensive diagnostic information about students' graphical representation skills, enabling physics teachers to identify students' strengths and learning difficulties and to design more effective, evidence-based instruction and assessment.

2. Method

2.1. Research Design

This study employed a Research and Development (R&D) approach to develop and validate a graphical representation assessment instrument for senior high school in the subject of Dynamic Electricity. The development process followed the Four-D (4D) model of learning device development according to Thiangarajan, Semmel and Semmel (Figure 1). Define, Design, Develop, and Disseminate phases. The selection of this model because it provides based on systematic framework of needs analysis, design, expert validation, empirical testing, and refinement before dissemination (Nuraini et al., 2025).

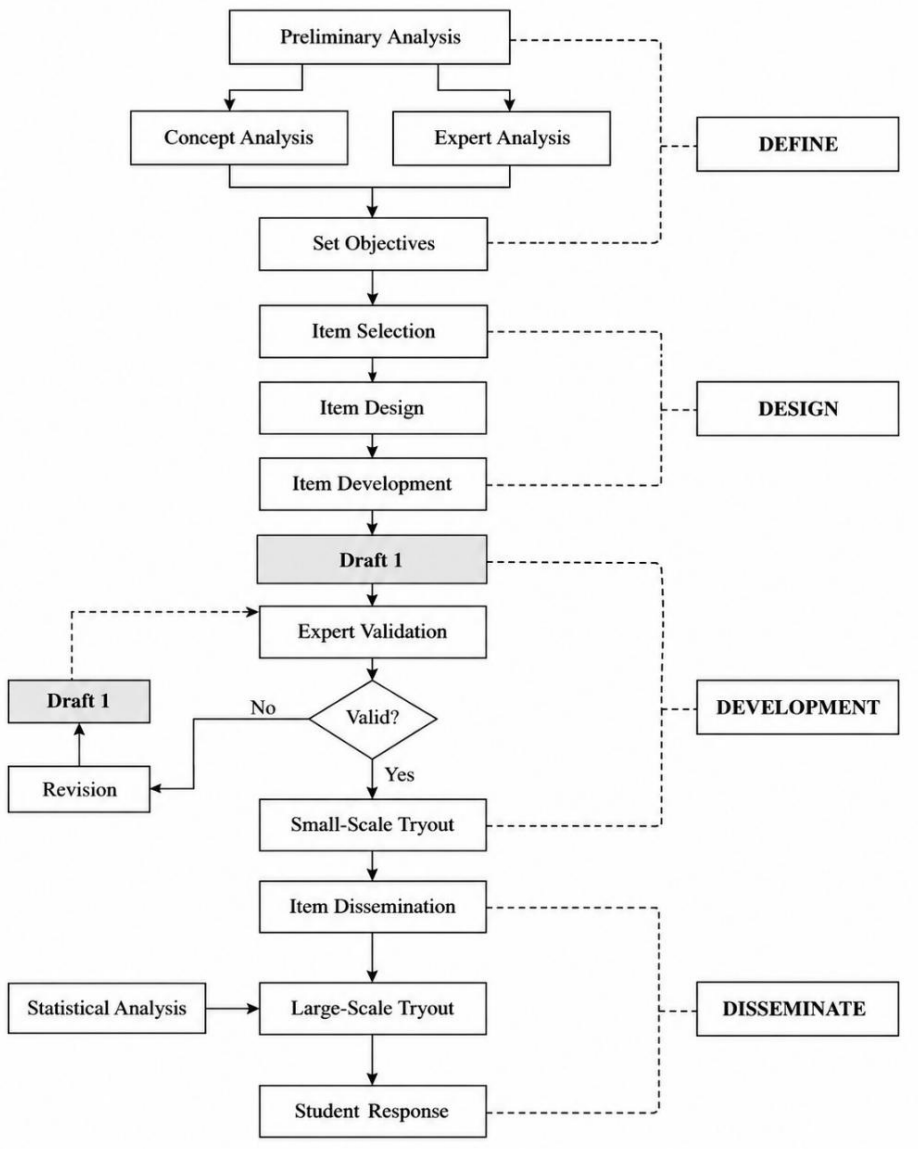


Figure 1. Research Flow Based on the 4D Development Model

Figure 1 illustrates the overall development procedure of the graphical representation assessment instrument based on the Four-D model. Each phase comprises specific activities to ensure the instrument's content validity, construct quality, and psychometric soundness before classroom implementation

2.2. Participants

The participants in this study consisted of expert validators and senior high school students. The expert validation involved two validators selected through purposive sampling based on their academic qualifications and professional expertise in physics education. The panel comprised two physics education lectures. The experts evaluated the instrument in terms of content relevance, contract appropriateness, language clarity, and its suitability for measuring student's graphical representation's ability in topic of dynamic electricity.

The empirical trial was conducted at Karya Pembangunan 1 Ciparay Senior High School, Grade XII (aged 16-18 years), involving 50 students. All participants had completed instruction on dynamic electricity prior to the study. The students were selected using purposive sampling because they inclusion criteria of being enrolled in Grade XII and having studied the topic of dynamic electricity. The sample was considered appropriate for the plenary psychometric evaluation of the developed assessment instrument.

2.3. Research Instrument

The object of this study was a graphical representation assessment instrument developed to measure senior high school students' graphical representation skills in the topic of dynamic electricity. The instrument was designed based on six indicators of graphical representation ability adapted from Aprilia and Dwandaru (2024), namely collecting data, graph interpretation, linking variables, comparing graphs, graph analysis, and making conclusion. These six indicators originally formulated by Aprilia and Dwandaru (2024) in the context of the topic of Temperature and Heat were adapted and applied to the topic of Dynamic Electricity in accordance with its conceptual characteristic and the intended learning outcomes. The instrument was designed to assess student's abilities through open ended descriptive questions. It consisted of 12-items essay, including six Type A and six Type B.

2.4. Research Procedure

The development procedure followed the four phases of the 4D model. During the Define phase, a needs analysis was conducted through curriculum analysis, literature review, and identification of students' difficulties in graphical representation on dynamic electricity. This stage established the conceptual framework and assessment objectives.

The Design phase involved preparing the assessment blueprint, selecting the graphical representation indicators, constructing the scoring rubric, and developing the initial draft of the assessment items according to the intended learning outcomes.

During the Develop phase, the instrument underwent expert validation involving physics education lecturers and an experienced physics teacher. Feedback obtained from the validators was used to revise the assessment items before conducting a small-scale review and a large-scale empirical trial. The empirical data obtained from the trial were subsequently analysed to evaluate the psychometric quality of the instrument.

Finally, during the Disseminate phase, the validated assessment instrument was finalized and prepared for classroom implementation. Due to time and research scope limitations, the dissemination process was confined to the participating school and did not include wider implementation or cross-regional validation. Consequently, future studies are encouraged to administer the instrument in different schools and educational settings to further examine its psychometric properties and generalizability.

2.5. Data Analysis Techniques

Quantitative data obtained from the empirical trial were analysed using both Classical Test Theory (CTT) and Rasch Measurement Theory (RMT). The combination of these two analytical

approaches provides complementary evidence regarding conventional item quality and modern psychometric characteristics, thereby strengthening the overall validation of the developed assessment instrument.

CTT analysis was performed using IBM SPSS Statistics version 29 to examine item validity, internal consistency reliability, item difficulty, and discrimination index. Item validity was evaluated using Pearson's product–moment correlation, while reliability was determined through Cronbach's alpha coefficient (DeVon et al., 2007). Item difficulty and discrimination indices were subsequently calculated to assess each item's quality based on established educational measurement criteria.

For a more comprehensive psychometric evaluation, Rasch analysis was conducted using MINISTEP Software version 5.11.2.0 to examine item and person fit, reliability, separation indices, item difficulty, and ability estimates (JM, 2013). Following (Boone et al., 2014; Boone & Staver., 2020), acceptable item fit was defined by Infit Mean Square (INFIT MNSQ) and Outfit Mean Square (Outfit MNSQ) values between 0.50 to 1.50, Standardized Fit Statistics (ZSTD) values within −2.0 to +2.0, and Point Measure Correlation (Pt-Measure Corr) values ranging from 0.40 to 0.85 to evaluate item fit.

Instrument reliability was further evaluated using both Cronbach's alpha and Rasch reliability indices (person reliability, item reliability, and separation indices) (Noperi et al., 2025). These separation values indicated the instrument's capacity to distinguish varying levels of students' graphical representation skills and item difficulty. Additionally, a Wright Map was employed to compare the distribution of student ability and item difficulty on the same logit scale, enabling a comprehensive assessment of the instrument's targeting and diagnostic capability (Noperi et al., 2025; Nurussaniah et al., 2025). Item difficulty parameters were interpreted on the logit scale, where higher positive values denote more difficult items and lower negative values denote easier items. To ensure a comprehensive evaluation of the developed assessment instrument, both conventional psychometric indices and Rasch measurement parameters were employed. The interpretation criteria adopted for each analysis are presented in Table 1.

Table 1. Level for difficulty values (Hambleton et al., 1991; Rizbudiani et al., 2021)

Item Difficulty Level	Measure Value
Very Difficult	$b > 2$
Difficult	$1 < b \leq 2$
Moderate	$-1 < b \leq 1$
Easy	$-2 < b \leq -1$
Very Easy	$b \leq -2$

2.6. Level of Graphic Representation Ability of Students

The percentage of each student's graph representation indicator can be determined using the Equation (1)

$$NP = \frac{S}{SM}\ x100\% \quad (1)$$

Explanation:
NP = percentage value of graph representation
S = score obtained by the student
SM = maximum test score

Next, to calculate the average score across all aspects of graphic representation, the following Equation (2) and level of graph representation ability can be seen in Table 2.

$$\underline{X} = \frac{\Sigma\ students'\ total\ scores}{number\ of\ students} \quad (2)$$

Table 2. Level of Graph Representation ability Purwanto (2019)

Percentage	Category
$85.0\% \leq NP \leq 100\%$	Very Good
$75.0\% \leq NP \leq 85.0\%$	Good
$59.0\% \leq NP \leq 75.0\%$	Fair
$53.0\% \leq NP \leq 59.0\%$	Poor
$NP < 53.0\%$	Very Poor

The output results of the MINISTEP software can display the level of student's abilities in answering graphical representation questions. The level of student's abilities is categories into three, namely high, moderate, or low. The analysis results of student's abilities levels can be observed in Table 3.

Table 3. Estimate student's ability (Apriliani et al., 2026)

Estimated Value	Ability Level
>1.00	High
-1.00 - +1.00	Moderate
< -1.00	Low

3. Results and Discussion

The results obtained in this study are described in accordance with the 4D development stages, namely Define, Design, Development, and Disseminate.

3.1. Initial Analysis (Define)

In the initial analysis stage, researchers reviewed articles related to graphical representation instruments in physics learning and studied Learning Outcomes regarding "dynamic electricity" topic at the high school level. The concepts that are important for students are related to Ohm's law, wire resistance, electrical power, and series-parallel resistance circuits. The results of the analysis show the need for evaluation instruments that not only measure conceptual understanding but also students' ability to interpret and interpret graphs as a form of 21st-century representation.

3.2. Design Stages

The design stage is the second stage in this study, which is carried out to compile test specifications. This stage consists of four steps, namely determining the test objectives, determining the form and length of the test, determining the test grid, and writing the test. The final result of this stage is a test consisting of 12 items representing each indicator of graphic representation, with a test duration of 90 minutes. Graphical representation's indicator adapted from (Aprilia & Dwandaru., 2024). The indicators for the question items can be seen in Table 4.

Table 4. Indicators for graphical representation question items

Question item	Graphical representation indicators	Description
1A, 1B	Collecting Data	The ability to collect and examine data presented on a graph.
2A, 2B	Graph interpretation	The ability to interpret graphs related to relationships among variables.
3A, 3B	Linking variables	The ability to connect variables on a graph.
4A. 4B	Comparing graphs	The ability to compare graphs
5A, 5B	Graph analysis	The ability to analyses and transform data into a graph according to the concept given.
6A, 6B	Making a conclusion	The ability to create and review results based on data or information provided.

3.3. Characteristics of the Pilot Study

3.3.1. Empirical Validation

In developing test instruments, it is necessary to assess the test items to measure their suitability and quality based on empirical validation. In addition, through this empirical validation, researchers can analyse the percentage of students' ability to represent graphs. Question instruments based on graph representation indicators are first validated by subject matter experts

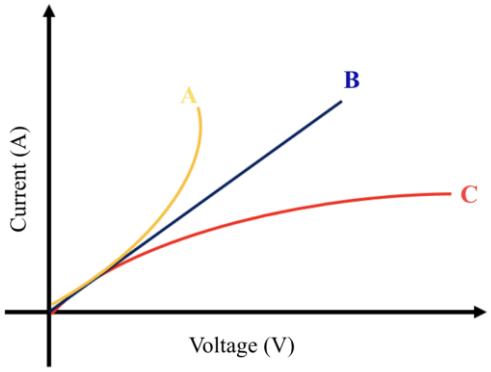
to determine the suitability of the questions. Both question types A and B are given to experts in the field of educational evaluation to examine the suitability of the questions with the graph representation indicators and dynamic electricity material to be measured.

Suggestion for Question Number 5

The reason is that the graph presented only includes 3 curves, so it does not fully describe the diversity of interpretations by students. Additional graph variations could include non-linear curves, lines with different gradients, saturated graphs, or representations resembling the characteristics of certain electrical components.

Before:

A graph shows the relationship between electric voltage (V) and electric current (I) for conductors A, B, and C. Each graph illustrates the nature of the relationship between these two variables. Observe the following graph!



Based on the graph,

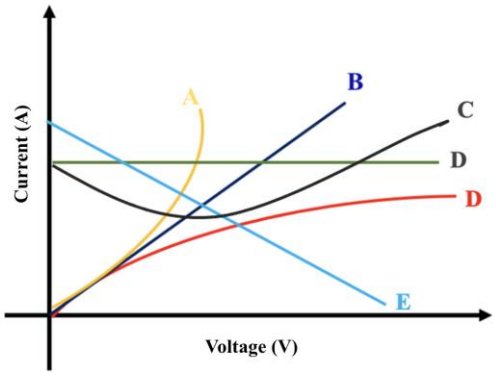
Explain the relationship between voltage (V) and electric current (I) for each type of conductor!

Determine which conductors show ohmic and non-ohmic properties? Explain your reasoning!

Compare the resistance of the three conductors. Which conductor has the greatest resistance? Explain.

After:

A graph shows the relationship between electric voltage (V) and electric current (I) for conductors A, B, and C. Each graph illustrates the nature of the relationship between these two variables. Observe the following graph



Based on the graph,

Explain the relationship between voltage (V) and electric current (I) for each type of conductor!

Determine which conductors show ohmic and non-ohmic properties? Explain your reasoning!

Compare the resistance of the three conductors. Which conductor has the greatest resistance? Explain.

3.4. Psychometric Characteristics of the Item

This study developed a test instrument to assess graphical representation skills in the topic of Dynamic Electricity, comprising two parallel forms are Type A and Type B with identical indicators, each consisting of six items. Raw scores for each item with maximum scores varying between 8, 12, and 16 were converted into an ordinal categorical scale prior to analysis using the Rasch model. Two response categorization schemes were tested and compared to determine the one best suited to the data characteristics. Based on the comparison results, student responses were modeled using a three-category polytomous scale 0, 1, 2 within the framework of the Partial Credit Model (PCM).

3.4.1. Instrument Pilot Testing and Internal Consistency Analysis

The instrument was pilot tested on 50 students using a split form design, with 25 students completing Form A and 25 completing Form B. Analyses using Classical Test Theory (CTT) and the Rasch Model indicated that the instrument's psychometric quality was generally adequate, though there were variations in quality at both the form and item levels that warrant attention. A comparison of the internal reliability of the two forms is presented in Table 5.

Table 5. Internal Consistency Reliability Items of Type A and B

Form	Number of Items	Number of Students	Cronbach's Alpha (α)	Interpretation
Type A	6	25	0,797	Good
Type B	6	25	0,675	Questionable

The results show a noticeable difference in internal consistency between the two parallel forms. Form Type A demonstrated a higher Cronbach's Alpha than form Type B, indicating that the response patterns across the six items were more internally consistent in Form A. However, the lower coefficient obtained for Form B should not be interpreted solely as evidence of inadequate item quality, because both forms consisted of only six items. Reliability coefficients are sensitive to test length, and shorter instruments tend to produce lower estimates of internal consistency (Busnawir et al., 2023; Shang et al., 2024).

To examine the potential influence of the test length on the observed reliability of Form type B, the Spearman-Brown prophecy formula was applied (Equation 3). Assuming that additional items would have characteristics comparable to those of the existing items, doubling the number of items from six to twelve yields the following projected reliability (Ellis, 2013; Zhao & Aryadoust, 2025).

$$\rho_n = \frac{n\rho_1}{1+(n-1)\rho_1} \tag{3}$$

$$\rho_n = \frac{2(0.675)}{1+0.675} = 0.801 \approx 0.81$$

The projection suggests that extending Form B from six to twelve items could increase its estimated reliability to approximately 0.81. This finding indicates that test length may partly account for the relatively lower internal consistency observed in Form B. Nevertheless, the value of 0.81 represents a projected rather than empirically observed reliability coefficient and therefore should not be interpreted as evidence that the 12 item instrument has actually achieved this level of reliability. Rather, this suggest that increasing the number of of well functioning items may improve measurement consistency.

3.4.2. Analysis of Item Validity and Fit Using the Rasch Model

However, internal consistency alone is insufficient to determine whether the instrument functions effectively as a diagnostic assessment. While Cronbach's Alpha provides information on overall score consistency, it does not reveal which individual items contribute to measurement quality, whether the items adequately represent varying levels of student ability, or whether specific items exhibit unexpected response patterns. Therefore, the Partial Credit Rasch Model with *MINISTEP* software to evaluate item characteristics, such as item difficulty, fit statistics, person item targeting, and student ability distribution (Table 6).

Table 6. Characteristic validity of the question item instrument

Item	Measure (Logit)	Difficulty Level	Outfit MNSQ	ZSTD	Pt-Measure Corr	Status
A1	-2.53	Very Easy	1.19	0.52	0.66	Valid
A2	0.52	Moderate	1.10	0.39	0.76	Valid
A3	0.52	Moderate	0.75	-0.65	0.61	Valid
A4	0.52	Moderate	1.51	1.34	0.54	Misfit
A5	0.75	Moderate	0.47	-1.81	0.74	Overfit
A6	1.86	Difficult	1.52	2.02	0.81	Misfit
B1	-1.97	Easy	2.02	1.51	0.48	Misfit
B2	-0.26	Moderate	0.79	-0.53	0.64	Valid
B3	-1.20	Easy	0.74	0.55	0.66	Valid
B4	-0.26	Moderate	1.04	0.23	0.51	Valid
B5	1.15	Difficult	1.24	0.67	0.46	Valid
B6	0.91	Moderate	0.57	-1.10	0.60	Overfit

Table 6 presents the item validity based on the Partial Credit Rasch Model. According to (Linacre, 2002), a good item fit requires an Outfit MNSQ between 0.5 and 1.5. Based on this criterion, seven items (A1, A2, A3, B2, B3, B4, and B5) are classified as Valid, indicating they function consistently with the model. Conversely, three items are classified as Misfit (B1, A4, and A6) with MNSQ values exceeding 1.5, suggesting their response patterns are highly unpredictable and produce statistical noise. Meanwhile, two items are categorized as Overfit (A5 and B6) with MNSQ values below 0.5, indicating overly deterministic responses that fail to effectively differentiate student abilities. In terms of difficulty, a notable redundancy is observed: items A2, A3, and A4 share an identical difficulty level of 0.52 logits, while B5 and B6 are closely positioned. Although all items exhibit positive point-measure correlations, the statistical anomalies and evident redundancies necessitate a strategic refinement of the instrument before proceeding to the main data collection.

3.4.3. Mapping of Item Difficulty and Students Ability

Figure 2 (wright map) below visulatized the aligments between the distribution of student ability on the left and item difficulty levels on the right.

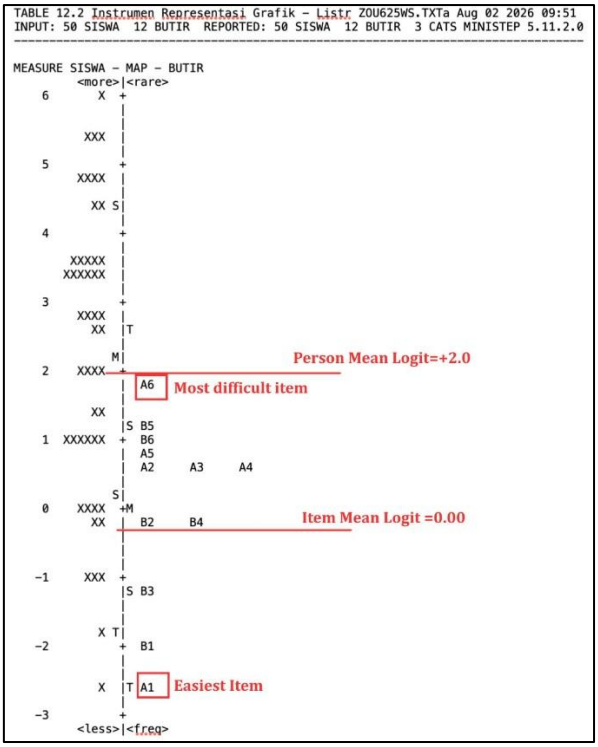


Figure 2. Wright Map of Person-Item in Graphical Rapresentation Abilities

After analyzing the validity and suitability of each test item in Table 6, the next step is to visualize the distribution of student ability versus item difficulty using the Person-Item map/Wright Map shown in Figure 2. The map reveals that the mean student ability marked with the letter M on the left side lies at a logit value significantly higher than the mean item difficulty marked with M on the right side. This indicates that, overall, the developed instrument tends to be too easy for the study subjects, resulting in a targeting imbalance. Furthermore, there are gaps at the high-ability (above +2 logits) and low-ability (below -2 logits) ends of the spectrum that are not captured by the existing items.

These visual findings are quantitatively supported by the summary statistics from the Rasch analysis. The average student ability (person measure) is recorded at 2.20 logits, wich is well above the average item difficulty level standardized at 0.00, comfirming the targeting imbalance (Tesio et al., 2024). Meanwhile, the Person Separation Reliability is recorded at 0.68, and the Item Separation Reliability at 0.82. An item separation value exceeding 0.80 indicates that the instrument demonstrates good consistency in its difficulty hierarchy (Bintang & Suprananto, 2024). To further examine potential issues in the item structure and inform the final instrument refinement, an in-depth analysis of item interdependency and misfit characteristics is presented in the following section

3.4.4. Synthesis of Instrument Quality and Recommendations

Following the visual and statistical examination of the item distribution in the previous section, a synthesis of the instrument's overall quality is necessary to guide the refinement process. A fundamental principle of the Rasch model is that, once students' latent ability is statistically controlled, the response to one item should neither influence nor be influenced by the response to another (Hrnjicic & Alihodžic, 2024). Based on the visualization provided by the Wright Map and a review of Table 6, there are strong indications of item redundancy (overlap). Items A3 and A4 exhibit identical difficulty levels, both positioned at a logit of 0.52. Meanwhile, items B5 and B6 also show very close difficulty levels at logits of 1.15 and 0.91, respectively. Coupled with the misfit and overfit statistics reported in Table 6, there is a strong indication that these item pairs potentially violate the local independence assumption. Specifically, item A4 exhibits misfit (Outfit MNSQ = 1.51), and item B6 exhibits overfit (Outfit MNSQ = 0.57), suggesting that student response patterns to these redundant items are overly uniform and dependent on one another.

To provide a comprehensive visual overview of the decisions derived from these findings, Figure 3 and **Error! Reference source not found.** present the classification of all 12 items into their respective final categories.

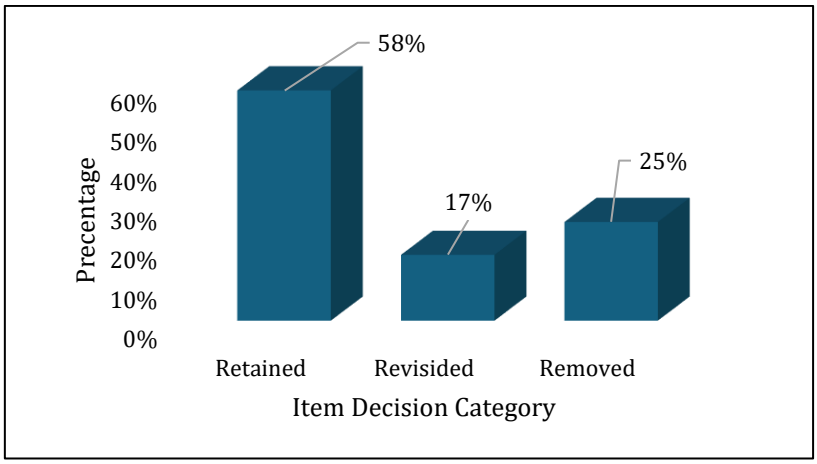


Figure 3. Percentage of item decision categories

As illustrated in Figure 3, 58% of the items are retained for the main study. Based on this classification, the following specific revisions are proposed to refine and finalize the instrument for the main data collection.

Based on the empirical findings, the instrument refinement is structured as follows. Items A1, A2, A3, B2, B3, B4, and B5 are retained, as they demonstrate satisfactory fit to the Rasch model and appropriate difficulty (Linacre, 2002). Items A5 and A6 require minor revisions to enhance response variance and model fit (Boone & Staver, 2020) . Conversely, items A4, B1, and B6 are removed due to significant misfit, overfit, or confirmed redundancy particularly the identical difficulty of A4 and A3, which violates the assumption of local independence (Hrnjicic & Alihodžic, 2024). Finally, since the mean student ability (2.20 logits) greatly exceeds the item difficulty, future development should include higher-level items (above +2 logits) to mitigate the ceiling effect and better differentiate top-performing students (Boone et al., 2014).

3.4.5. The Empirical Linkage Between Item Analysis Metrics and Instrument Indicator Validity

Based on the estimated student abilities presented in Figure 3, the student who participated in the pilot study were classified according to the following level criteria. The result indicate that majority of students were classified as having high ability, as shown in the diagram in **Error! Reference source not found.**.

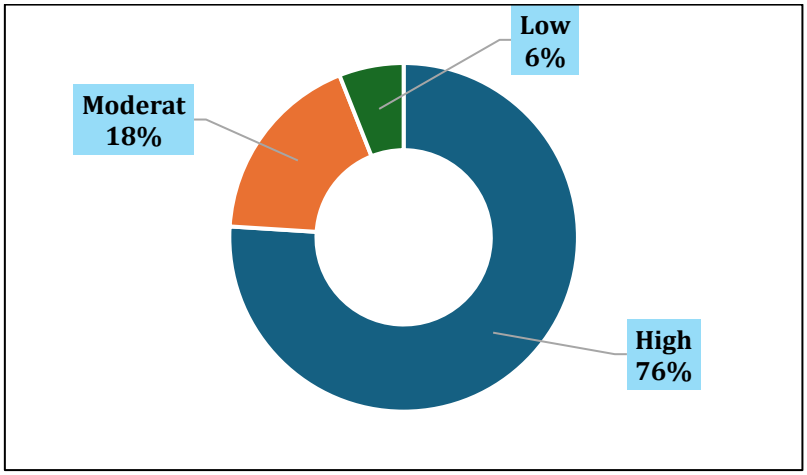


Figure 4. Distribution of student's ability levels

3.4.6. The Empirical Linkage Between Item Analysis Metrics and Instrument Indicator Validity

The summary of student ability estimates shows a person reliability value of 0.74, indicating a moderate level of overall consistency in student ability, although this falls short of the high category (> 0.80). While MINSTEP analyzes individual student ability levels via specific logit values, the person reliability summary assesses the overall ability level (mean of +2.29 logits). These observations suggest that, despite the high average student ability, the instrument does not optimally cover the full ability spectrum; a ceiling effects evident, indicating a lack of items with high difficulty levels (> +2.00 logits). Consequently, measuring students' graphical representation skills regarding Dynamic Electricity requires an instrument with a broader difficulty range, particularly to assess inferential analysis and graph extrapolation capabilities. The instrument development results demonstrate an interplay between validity, reliability, and the distribution of student abilities. Good validity specifically Point-Measure Correlation does not automatically guarantee the instrument's suitability for the sample if the ability distribution lacks diversity such as when it is dominated by a high-ability group (Halim et al., 2026; Wulandari et al., 2025). In other words, item validity alone is insufficient to ensure instrument quality without a balanced Wright Map and a more heterogeneous sample.

Following the statistical item analysis and the subsequent generation of the Wright Map, the inquiry proceeds with an investigation into the etiologies of statistical anomalies, namely item misfit, overfit, and redundancy. This analytical progression entails a critical re-evaluation of both the substantive item content and the graphical representation indicators embedded within the instrument. Designed to assess graphical representation competencies in topic of Dynamic Electricity, the instrument operationalizes six distinct indicators adapted from Aprilia et al (2024). This discourse seeks to explicate the differential statistical functionality of the items contrasting optimal performance with aberrant or invariant response patterns thereby establishing a rigorous theoretical framework to justify the item modification and deletion protocols articulated previously.

For the Collecting Data indicator presented in two Questions, items A1 and B1 were designed to test students' basic procedural ability to apply Ohm's Law equation ($I = V/R$) to complete data tables and plot linear $V - I$ graphs. Statistically, A1 was confirmed as the most difficult item while maintaining high validity (Outfit MNSQ = 1.19), and A2 also fell within the valid category. This indicates that students were highly capable of understanding the steps to calculate current and representing them in a linear graph. However, a significant anomaly occurred in item B1, which was categorized as very easy (-1.97 logits) yet exhibited the highest misfit value (Outfit MNSQ = 2.02). Conceptually, this anomaly can be attributed to potential visual ambiguities in the graph or table presented in Form B. Although the item was cognitively straightforward, the less intuitive visual presentation likely created confusion, leading even high-ability students to respond in ways that deviated from the Rasch model's expectations.

For the Graph Interpretation and Graph Comparison indicators found in Questions 3, 4, and 7, students were required to compare the resistance values of two or more conductors based on the slope (gradient) of the voltage–current relationship graph. Statistical findings revealed that items A2, A3, and A4 shared completely identical difficulty levels (0.52 logits). In the context of dynamic electricity concepts, this equivalence indicates a high degree of cognitive redundancy. The three items required an identical level of content analysis, differing only in numerical values or the colours of the graph lines. If students were able to compare resistance in item A2, they could easily apply the same skill to A3 and A4 without requiring any new conceptual understanding. Regrettably, item A4 was also flagged as misfitting (Outfit MNSQ = 1.51), which can be conceptually attributed to potential ambiguity in the wording or the graph scaling on Form B. This finding reinforces the argument that item A4 should be eliminated, as it is not only redundant but also suffers from a defective visual construct

For the Linking Variables indicator (Questions 5 and 6), students compared ohmic (linear) and non-ohmic (curved) graphs. Item A6 was the most difficult (1.86 logits) and misfitting (Outfit MNSQ = 1.52), reflecting its high cognitive complexity (C4-Analysis). High school students often struggle with non-ohmic concepts because they require understanding that resistance (R) changes dynamically with voltage (V). Poor visual quality in the non-linear graph also increased response inconsistency. Conversely, item A5 (0.75 logits) was overfitting (Outfit MNSQ = 0.47), indicating it

was too simple and allowed even low-ability students to guess correctly. Item A5 needs revision with sharper conceptual distractors.

For the Graph Analysis and Drawing Conclusions indicator (Questions 8, 9, and 10), students analyzed current I versus inverse total resistance (1/R) in parallel circuits and concluded the relationship between wire length (L) and resistance (R). Understanding that the slope of an I and 1/R graph represents voltage (V) is abstract and requires strong algebra skills. An anomaly appeared in item B6 (0.91 logits), which was overfitting (Outfit MNSQ = 0.57). This overfit suggests that the item was too predictable; the highly uniform response pattern failed to distinguish students with genuine mathematical understanding from those who simply memorized straight-line graph patterns. Therefore, graph analysis items on Form B should be redesigned with greater variety rather than relying on generic graphs

Overall, these findings confirm that item difficulty strongly correlates with the hierarchy of graphical representation indicators; items requiring procedural calculations (Indicator 1) proved to be the easiest and most valid, whereas those demanding non-linear curve analysis and graph slope interpretation (Indicators 3, 5, and 6) were the most difficult and prone to misfit. The misfit and overfit anomalies in items B1, A4, A6, and B6 are not merely mathematical; they reflect underlying content and visual issues. Therefore, the revision recommendations are not just based on blind statistical decisions, but on the need to refine non-ohmic graph design, curve slope comparison, and item wording. These adjustments ensure the instrument can accurately distinguish students who truly understand Direct Current principles from those who merely guess graph patterns.

3.4.7. Pedagogical Implication for Physics Learning and Assessment

Based on the results of the trial test analysis, in this section the researcher describes the percentage of achievement of each indicator of students' representation abilities to see which aspects have been mastered well and which indicators still need strengthening (Figure 5).

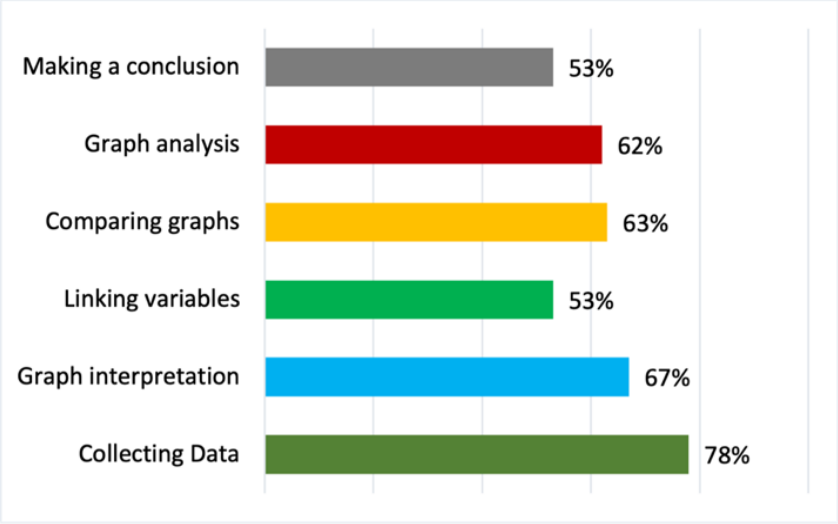


Figure 5. Percentage of each indicator graphical representation's students

These empirical findings have clear implications for physics teaching and assessment. The study shows that students' graphing skills vary greatly across cognitive levels. Specifically, students performed better on basic data collection (78%) than on higher-order tasks like drawing conclusions (53%). This trend proves that students read direct graph information more easily than they perform complex reasoning. This finding aligns with (Altindis et al., 2026), who noted that graphing skills develop gradually from visual recognition to conceptual inference. Similarly, (Termini et al., 2025) found that most students only recognize graph shapes without understanding the physical variables behind them. The statistical results where basic procedural items (A1, A2) were highly valid while complex items (A6, A4) showed misfit further support these studies, highlighting the need for physics instruction to move students from simple graph reading to deep conceptual understanding

The instrument's novelty offers a way to address this learning gap. Based on expert validator feedback, limiting graphs to three standard linear curves reduces the test's ability to measure deep understanding. In response, this study added varied gradients, non-linear curves, and graph shapes reflecting real electrical components. Using varied graph formats helps uncover student misconceptions more effectively (Ayeh, 2025; Royani et al., 2025). Applying graphical representation specifically to Dynamic Electricity is a novel approach rarely found in previous assessment tools. By incorporating curve of non-ohmic graphs and different slopes, this instrument prevents students from simply guessing based on basic visual patterns, thereby improving the test's ability to differentiate student ability levels

These improvements provide practical benefits for the classroom. Teachers can use this instrument as a diagnostic tool to assess how well students understand relationships between variables in current-voltage graphs. As suggested (Worku et al., 2026), low scores in graph analysis and drawing conclusions show that teaching strategies must integrate multiple representations. Furthermore, using varied graph formats improves assessment quality (Choi & Jung, 2025)and better tests higher-order thinking skills (Vrbanović Lisac et al., 2025). Ultimately, this study not only creates a valid physics test, but also offers a pedagogical approach that uses graphs as learning tools to build critical thinking and scientific literacy, rather than treating them as simple formula illustrations.

4. Conclusion

This study successfully developed and evaluated a graphical representation test for Dynamic Electricity using the 4D model, CTT, and Rasch analysis. While the instrument demonstrated good content validity, statistical analysis showed that seven items were feasible, whereas five required revision or removal due to misfit, overfit, and cognitive redundancy. Additionally, a strong ceiling effect occurred, as student ability (+2.20 logits) significantly exceeded item difficulty (0.00 logits). These results offer clear practical insights for physics education. First, the refined test serves as a diagnostic tool to identify misconceptions, particularly regarding ohmic and non-ohmic concepts. Second, while students excel at basic data reading, they struggle with higher-order graph analysis, highlighting the need for teaching strategies that emphasize multi-representation over simple formula calculation. Finally, future assessments must include non-linear graphs and higher difficulty levels (> +2 logits) to effectively evaluate Higher-Order Thinking Skills (HOTS). Ultimately, graphical representation should be taught as a core reasoning skill rather than a passive visual.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The datasets generated during and/ or analysed during the current study are available from the corresponding author on reasonable request.

Declaration on AI Use

The authors declare that no artificial intelligence (AI) or AI-assisted tools were used in the preparation of this manuscript.

References

Altindis, N., Bowe, K. A., Bauer, C. F., & Aikens, M. L. (2026). Graphical chunking strategies in students' interpretation of a logistic function titration graph. *The Journal of Mathematical Behavior*, *83*, 101332.

Aprilia, A., & Dwandaru, W. S. B. (2024). Empirical Analysis of Physics Test Instruments to Measure Graphical Representation Abilities in "Temperature and Heat" Topics. *Science Education International*, *35*(3), 240–249. https://doi.org/10.33828/sei.v35.i3.7

Apriliani, S. S., Saepuzaman, D., & Suwarma, I. R. (2026). Characteristic of critical thinking test instrument and analysis students' ability using Rasch model Article History. *Momentum: Physics Education Journal*, *10*(1), 121–134. https://doi.org/10.21067/mpej.v10i1.13455

Ayeh, I. G. (2025). Students' mathematics conceptual challenges: Exploring students' thinking, understanding, and misconceptions in functions and graphs. *European Journal of Science and Mathematics Education*, *13*(3), 191–206.

Bintang, R. S., & Suprananto. (2024). The Impact of Sample Size, Test Length, and Person-Item Targeting on The Separation Reliability in Rasch Model: A Simulation Study. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, *13*(1), 1–13. https://doi.org/10.15408/jp3i.v13i1.27975

Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analyses in the human sciences*. Springer. https://doi.org/https://doi.org/10.1007/978-3-030-43420-5

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences* (Vol. 10). Springer. https://doi.org/https://doi.org/10.1007/978-3-030-43420-5

Busnawir, B., Kodirun, K., Zamsir, Z., Samparadja, H., & Hasnawati, H. (2023). Trends Stability of Reliability Coefficient Based on Sample Size and Ability of Test-takers. *PYTHAGORAS Jurnal Pendidikan Matematika*, *18*(2), 97–111. https://doi.org/10.21831/pythagoras.v18i2.59392

Choi, S., & Jung, Y. (2025). Knowledge graph construction: Extraction, learning, and evaluation. *Applied Sciences*, *15*(7), 3727.

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, *39*(2), 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.x

Ellis, J. L. (2013). A standard for test reliability in group research. *Behavior Research Methods*, *45*(1), 16–24. https://doi.org/10.3758/s13428-012-0223-z

Fisher, M., Volfson, A., Eshach, H., & Ben-Abu, Y. (2026). Visualizing Physics: Boosting Argumentation Skills Through Representation. *Journal of Science Education and Technology*, 1–20. https://doi.org/10.1007/s10956-025-10291-4

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Hermawati, T. (2022). Analysing science teachers' difficulties in teaching the concept of electricity in junior high school. *Research in Physics Education*, *1*(1), 33–44.

Hrnjicic, A., & Alihodžic, A. (2024). Measuring Students' Conceptual Understanding of Real Functions: A Rasch Model Analysis. *International Electronic Journal of Mathematics Education*, *19*(1).

Istiyono, E., & Fenditasari, K. (2021). Physics Graphical Representation Test of Straight Motion Kinematics Based on Boti Boat Local Wisdom: Development and Validity. *6th International Seminar on Science Education (ISSE 2020)*, 350–357. https://doi.org/10.2991/assehr.k.210326.050

JM, L. (2013). *A user's guide to WINSTEPS, MINISTEP Rasch-Model computer programs*.

Khotimah, K., & Fariyani, Q. (2025). Development of a Physics Assessment Instrument to Measure High School Students Multirepresentation Skills in Momentum and Impulse. *Impulse: Journal of Research and Innovation in Physics Education*, *5*(1), 49–61. https://doi.org/https://doi.org/10.14421/impulse.2025.51-04

Kim, J.-H., & Nam, J. (2021). Thermodynamic identities with sunray diagrams. *European Journal of Physics*, *42*(3), 035101. https://doi.org/https://doi.org/10.48550/arXiv.1912.11485

Kusumaningtyas, D. A., Manyunu, M., Kurniasari, E., Awalin, A. N., Rahmaniati, R., & Febriyanti, A. (2024). Enhancing Learning Outcomes: A Study on the Development of Higher Order Thinking Skills based Evaluation Instruments for Work and Energy in High School Physics. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 14–31. https://doi.org/10.23917/ijolae.v6i1.23125

Lichtenberger, A., Hofer, S. I., Stern, E., & Vaterlaus, A. (2025). Enhanced conceptual understanding through formative assessment: results of a randomized controlled intervention study in physics classes. *Educational Assessment, Evaluation and Accountability*, *37*(1), 5–33. https://doi.org/https://doi.org/10.1007/s11092-024-09445-6

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, *16*(2), 878.

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

Niyomufasha, T., Ntivuguruzwa, C., & Mugabo, L. R. (2024). The engineering students' use of multiple representations in mechanics problems solving at a selected public university in Rwanda. *Cogent Education*, *11*(1), 2372941. https://doi.org/10.1080/2331186X.2024.2372941

Noperi, H., Diani, R., Irwandani, I., Suryani, Y., Sodikin, S., & Saputro, T. (2025). Psychometric analysis of the critical thinking skills instrument in physics learning using the Rasch model. *Indonesian Journal of Science and Mathematics Education*, *8*(3), 627–641. https://doi.org/https://doi.org/10.24042/ijsme.v8i3.28770

Nurussaniah, N., Setyosari, P., Kuswandi, D., & Ulfa, S. (2025). Psychometric Validation of an Analytical Skills Test in Physics Using the Rasch Model. *Journal of Baltic Science Education*, *24*(3), 522–537. https://doi.org/10.33225/jbse/25.24.522

Pradana, P. W. (2025). Assessment Instrument of Graph Representations on Sound Wave Topic: Development and Measurement Implementation. *Knowledge Management & E-Learning*, *17*(2), 352–371. https://doi.org/https://doi.org/10.34105/j.kmel.2025.17.016

Rangkuti, M. A., & Karam, R. (2022). Conceptual challenges with the graphical representation of the propagation of a pulse in a string. *Physical Review Physics Education Research*, *18*(2), 020119. https://doi.org/https://doi.org/10.1103/PhysRevPhysEducRes.18.020119

Rizbudiani, A. D., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables (SLETV). *Al-Jabar: Jurnal Pendidikan Matematika*, *12*(2), 399–412.

Royani, S. N. M., Irvaniyah, B. N., & Kusairi, S. (2025). A Revealing Students' Struggles and Misconceptions in Kinematics through Multi-Representation Assignments. *JPI (Jurnal Pendidikan Indonesia)*, *14*(4), 810–821.

Shang, Y., Aryadoust, V., & Hou, Z. (2024). A Meta-Analysis of the Reliability of Second Language Listening Tests (1991–2022). In *Brain Sciences* (Vol. 14, Number 8). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/brainsci14080746

Sharma, A. K. (2025). Cognitive and psychological aspects of physics learning. *Physics Journal| Farhangian University*, *2*(1), 28–39. https://doi.org/10.48310/esip.2025.18642.1011

Termini, G., Battaglia, O. R., Grazia, I., & Fazio, C. (2025). Gaining insights through thematic analysis about learning dimensions in teaching learning sequences on surface phenomena in liquids. *Physical Review Physics Education Research*, *21*(2), 020143.

Tesfaye, D. (2025). *Enhancing Secondary School Students' Conceptual Understanding and Problem-Solving Performance in Electricity Using Multiple Representations-Based Instruction Integrated with Formative Assessment Practice*.

Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 1. The "most likely" measures coming from the model. *Disability and Rehabilitation*, *46*(3), 591–603. https://doi.org/10.1080/09638288.2023.2169771

Thiagarajan, S. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*.

Vrbanović Lisac, D., Radanović, I., & Šimić Šašić, S. (2025). Enhancing higher order thinking skills to improve understanding of biological content: A quasi-experimental study. *Natural Sciences Education*, *54*(2), e70023.

Worku, D. T., Ejigu, M. A., Gebremeskal, T. G., & Kassa Gogie, T. (2026). Assessing the impact of multiple representations based instruction integrated with formative assessment practice on secondary school students' problem-solving performance in Physics. *Research in Science & Technological Education*, *44*(1), 224–249.

Zhao, H., & Aryadoust, V. (2025). A meta-analysis of the reliability of second language reading comprehension assessment tools. *Studies in Second Language Acquisition*, *47*(1), 388–416. https://doi.org/DOI: 10.1017/S0272263124000627