Jurnal Pendidikan Progresif
e-ISSN: 2550-1313 | p-ISSN: 2087-9849
URL: https://jpp.fkip.unila.ac.id/index.php/jpp
Jounal's DOI: http://doi.org/10.23960/jpp

Vol. 16, No. 2, 2026
Article ID: 1990
Pages: 1299-1324

Assessing Students' Problem-Solving Ability on Renewable Energy Topics in Physics: A Rasch Analysis Based on Heller's Framework

Welli Andriani, Ida Kaniawati*, & Irma Rahma Suwarma
Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

*Corresponding email: kaniawati@upi.edu

Received: 15 May 2026 *Accepted: 18 June 2026* *Published: 03 July 2026*

Abstract: This study examined students' problem-solving ability in physics on renewable energy topics by evaluating a Heller-based assessment instrument using the Rasch model. The study focused on determining whether the instrument functioned adequately for profiling students' staged problem-solving performance and identifying how item difficulty was distributed across the Heller framework. A quantitative descriptive design was employed involving 35 students who completed five renewable-energy physics problems. The instrument consisted of 25 scored indicators representing five Heller stages: Focus the Problem, Describe the Physics, Plan the Solution, Execute the Plan, and Evaluate the Answer. Students' written responses were assessed using a 0–4 polytomous rubric. The data were analyzed using MINISTEP VERSION 5.11.2 to examine person and item reliability, separation indices, item fit, person fit, unidimensionality, item difficulty, and rating-scale functioning. The instrument demonstrated adequate psychometric quality for exploratory use. Person reliability was 0.89, item reliability was 0.86, person separation was 2.83, and item separation was 2.49. All 25 indicators met acceptable Rasch fit criteria, while 32 of 35 students showed response patterns consistent with model expectations. The unidimensionality analysis supported the presence of a dominant primary dimension, and the 0–4 rating categories functioned progressively. The item difficulty distribution showed that students' performance was influenced not only by the formal Heller stage but also by the conceptual and contextual demands of each renewable-energy problem. The Heller-based instrument was suitable for preliminary profiling of students' problem-solving ability in renewable-energy physics. However, further validation with larger and more diverse samples is recommended to strengthen the instrument's reliability, fairness, and instructional usefulness.

Keywords: renewable energy, physics problem solving, rasch analysis, heller's framework, cational assessment.

Article's DOI: https://doi.org/10.23960/jpp.v16i2.pp1299-1324

INTRODUCTION

Renewable energy has become a strategically important topic in science education because the global energy transition now requires students not only to understand scientific concepts, but also to reason about technological, environmental, and societal challenges (Attanayake et al., 2024; J.-L. Hu & Yang, 2024; Rehman et al., 2026). Recent educational research shows that renewable-energy learning is increasingly linked to energy literacy, sustainability awareness, and students' preparedness to engage with real-world socio-scientific issues (Majid et al., 2025; Martínez-Borreguero et al., 2024). In Indonesia, for example, awareness of renewable energy has been shown to vary across educational levels, indicating that students' understanding of this topic remains uneven and still requires stronger curricular attention (Agung et al., 2024; Hernandez & Prakoso, 2021). At the same time, recent work on energy literacy has emphasized that the transition from fossil-based systems to sustainable energy sources should be treated as a current socio-scientific issue, not merely as factual content to be memorized (Agung et al., 2024; Kellberg, Keller, Nordine, Moser, Keller, et al., 2025).

Problem-solving ability is, therefore, central to physics education. Students are expected not only to recall formulas, but also to interpret situations, represent relevant variables, select appropriate principles, execute solution procedures, and evaluate the reasonableness of their answers (Ario et al., 2025; Kuo et al., 2013). This emphasis is consistent with current international assessment discourse, which highlights scientific problem solving as a high-level educational outcome requiring idea generation, evaluation, and application in unfamiliar contexts (Nilimaa, 2023). Recent studies further show that many secondary students still struggle to coordinate the multiple subprocesses needed for successful physics problem solving. In contrast, new assessment work in energy-related physics contexts has begun to separate problem-solving into more specific components, such as representation, planning, execution, and evaluation (Evendi et al., 2024; Qotrunnada, 2022; Tong et al., 2023; Worku et al., 2025). These developments suggest that problem-solving should be assessed a structured, multidimensional competence rather than a single end-product score (Meyer & Friege, 2026; Musengimana, Yadav, Uwamahoro, & Nizeyiman, 2025).

A relevant framework for analyzing such competence is the problem-solving model developed by K. Heller & Heller (2010), which organizes students' performance into sequential and cognitively meaningful stages. In physics learning, this framework is especially useful because it makes visible how students move from identifying the problem to describing the underlying physics, planning a solution, carrying out the procedure, and finally evaluating the answer (Adams & Wieman, 2015; Musengimana, Yadav, Uwamahoro, & Nizeyimana, 2025; Subramaniam et al., 2024). Rather than treating success as simply correct or incorrect, the framework captures the quality of students' reasoning across the solution process (Looi et al., 2026; Mohammad Edy Nurtamam et al., 2025). For studies that aim to examine students' problem-solving ability in a diagnostic and instructional way, this process-oriented structure is more informative than using only total scores (P. Heller et al., 1992).

However, strong conceptual frameworks alone are not sufficient; they must be supported by sound measurement. In educational assessment, Rasch measurement has become increasingly important because it enables researchers to examine item functioning, person ability, rating-scale performance, dimensionality, and potential bias within a single, coherent model (Adu et al., 2017; Davis & Boone, 2021). Recent methodological work has reaffirmed that Rasch analysis is a robust approach for instrument validation and for improving the precision of educational measurement, including in science and STEM-related contexts (Boone, 2016; Qudratuddarsi et al., 2022). This is particularly important when the instrument uses partial-credit scoring, as in problem-solving assessments, because Rasch analysis can test whether categories function properly and whether items truly align with the intended construct. Thus, a Rasch approach is well-suited to evaluating a Heller-based instrument designed to measure students' problem-solving ability in physics (Sha et al., 2026; Shahat et al., 2024).

Despite the growing relevance of renewable-energy education and the continued importance of physics problem-solving, a significant gap persists in the literature. Although students may show varying levels of awareness of renewable energy, such awareness does not necessarily indicate that they are able to solve physics problems related to renewable-energy systems. This concern is supported by findings showing that students' average problem-solving ability on renewable-energy material reached only 45.00, which falls into the low category (Irfani & Simanjuntak, 2026). This suggests that students still face difficulties in identifying relevant physics quantities, connecting concepts to real energy situations, selecting appropriate principles, and evaluating whether their solutions are scientifically reasonable (Hasyim & Sujiono, 2023). Recent studies on renewable energy have more often focused on awareness, literacy, or broad instructional interventions, whereas studies on physics problem solving have focused on general baseline ability or on developing digital assessments for specific subprocesses (Rahayu et al., 2025; Widya et al., 2026). Taken together, these lines of work indicate progress, but they also suggest that empirical evidence remains limited on how students' problem-solving ability can be profiled specifically on renewable energy topics in physics using Heller's indicators and a Rasch measurement approach. This gap is important because renewable energy is both conceptually demanding and socially relevant, making it an especially appropriate domain for examining whether students can apply physics knowledge in meaningful contexts rather than only reproduce routine procedures (Agung et al., 2024; Kellberg, Keller, Nordine, Moser, & Lewalter, 2025; Meyer & Friege, 2026; Musengimana, Yadav, Uwamahoro, & Nizeyimana, 2025).

This study addresses that gap by assessing students' problem-solving ability in physics on renewable energy topics using a Heller-based instrument analyzed with the Rasch model. The novelty of the study lies in three main aspects. First, it situates the assessment of physics problem-solving within the highly relevant context of renewable energy, thereby linking cognitive performance to a current sustainability-related topic. Second, it operationalizes students' responses using Heller's five indicators, allowing problem-solving to be examined as a staged reasoning process rather than a single aggregate score. Third, it applies Rasch measurement to evaluate instrument quality and to map student ability, item difficulty, and rating-scale functioning. In this way, the study contributes not only to physics assessment but also to the broader effort to develop more valid and instructionally useful measures for contemporary science learning.

The study is expected to provide a clearer profile of students' strengths and weaknesses in solving renewable-energy physics problems. Such information is useful for teachers because it can help identify whether students need support in understanding the problem, representing the situation, selecting principles, executing procedures, or evaluating answers. It is also useful for researchers because it demonstrates how a conceptual problem-solving framework can be combined with Rasch measurement to produce a more rigorous assessment of students' competence. Accordingly, this study is guided by the following research questions:

RQ1. How well does the Heller-based problem-solving instrument on renewable energy topics fit the Rasch model?

RQ2. How are students' problem-solving abilities and item difficulties distributed across the five Heller indicators?

RQ3. Does the scoring scale function properly?

These questions are intended to clarify both the psychometric quality of the instrument and the substantive profile of students' problem-solving ability in renewable-energy physics contexts. Overall, this study contributes to the literature by providing a psychometrically grounded profile of students' problem-solving ability in physics on renewable energy topics. It also provides empirical evidence that may support the refinement of problem-solving instruments and the design of more targeted instruction in physics learning on sustainable energy issues.

METHOD

Research Design

This study employed a quantitative descriptive design with a Rasch measurement approach to evaluate the psychometric quality of a problem-solving instrument and to profile students' problem-solving ability on renewable energy topics in physics (Barker et al., 2017; Qudratuddarsi et al., 2022). The design was appropriate because the purpose of the study was not to compare treatment effects but to examine whether the instrument functioned properly and whether students' responses could be reliably interpreted in terms of ability, item difficulty, category functioning, and possible item bias. In line with Rasch-based educational measurement, the analysis focused on the interaction between person ability and item difficulty on a common logit scale, thereby enabling evaluation beyond raw-score summaries alone (Adu et al., 2026; Khalaf & Omara, 2022; Shahat et al., 2024).

More specifically, the study addressed three analytical aims: first, to examine the fit of the Heller-based problem-solving instrument to the Rasch model; second, to map the distribution of student ability and item difficulty across the five problem-solving indicators; and third, to evaluate the functioning of the 0–4 rating scale. This design is consistent with current Rasch applications in educational assessment, where instruments are evaluated not only for reliability, but also for dimensionality, response consistency, scale structure, and fairness across groups.

Participants and Research Context

The study involved 35 students who completed a renewable-energy problem-solving test as part of the physics curriculum. In the context of Rasch analysis, both rating-scale and polytomous models are influenced by sample size, which affects the stability of item parameter estimates and the precision of fit statistics. Therefore, the present study was positioned as an initial instrument evaluation rather than a broad generalization study. Classical Rasch literature indicates that although larger samples yield more precise estimates, Rasch models can still be analyzed with relatively small samples, and approximately 30 participants are often cited as a practical minimum for initial item calibration to maintain reasonable measurement stability and detect potential misfit items (Linacre, 1994). or polytomous data, infit and outfit mean square statistics may remain stable even with small sample sizes, allowing misfitting items to be identified. However, some statistics remain sensitive to sample size and therefore need cautious interpretation (Smith et al., 2008).

Beyond sample size, literature also shows that test length, targeting between participant ability and item difficulty, and response structure influence Rasch reliability and interpretation. Recent simulation studies indicate that as sample size and test length increase, separation reliability improves; however, samples in the tens can still provide initial insights into item quality and response patterns (Bintang & Suprananto, 2024). Therefore, the use of 35 participants is considered adequate for initial instrument evaluation, particularly for analyzing category functioning, item fit, and preliminary estimates of item parameters and participant ability, which can serve as a foundation for larger-scale validation studies. Participants were selected through purposive sampling, with the criteria that they were actively enrolled in physics courses focused on renewable energy, had completed all test items, and provided analyzable responses. The sample consisted of 35 students, aged 16–17 years, all enrolled in the 11th grade on the science track at a science-focused high school.

The study was conducted within a physics-learning context in which renewable energy served as the substantive content domain for the problem-solving tasks. This context was chosen because renewable energy is both scientifically relevant and socially meaningful, making it suitable for assessing students' ability to interpret, represent, and solve contextualized physics problems rather than merely recalling formulas. The single-cohort design allowed for initial psychometric profiling of the instrument before implementation in larger, more diverse samples.

To clarify the characteristics of the sample, demographic and academic background information were recorded. Since this study used a single-cohort descriptive design and did not compare treatment groups, this information was used to describe the homogeneity of students' prior learning background rather than to establish equivalence between groups. The participant profile is presented in Table 1.

Table 1. Demographic and academic background of participants

Variable	Category	Frequency	Percentage
Age	16 years	11	31.43%
	17 years	24	68.57%
Grade level	11th grade	35	100%
Academic track	Science track	35	100%
Active semester	Second semester, 2026	35	100%
Previous exposure to energy-related physics topics	Yes	35	100%
Previous exposure to renewable-energy learning	Yes	35	100%
Research data status	Preliminary trial data	35	100%

Before data collection, permission was obtained from the school authority. Because the participants were minors, informed consent was obtained from parents or guardians, and student assent was obtained before the test was administered. Participation was voluntary, and students were informed that their responses would be used only for research purposes. All student identities were anonymized during data processing, analysis, and reporting.

The study was conducted within a physics-learning context in which renewable energy served as the substantive content domain for the problem-solving tasks. This context was chosen because renewable energy is both scientifically relevant and socially meaningful, making it suitable for assessing students' ability to interpret, represent, and solve contextualized physics problems rather than merely recalling formulas. The single-cohort design allowed for initial psychometric profiling of the instrument before implementation in larger, more diverse samples.

Instrument Development and Structure

The instrument was a written problem-solving test consisting of five renewable-energy problems. Each problem was constructed to elicit responses across the five problem-solving stages proposed by K. Heller & Heller (2010): Focus the Problem, Describe the Physics, Plan the Solution, Execute the Plan, and Evaluate the Answer. As a result, the analytical unit comprised 25 scored item indicators (5 problems × 5 indicators). This structure was adopted because the Heller framework is specifically designed to capture the sequential reasoning processes involved in physics problem-solving, rather than treating performance as a single, undifferentiated score (K. Heller, 2013).

Before administration, the instrument was reviewed by physics education experts to evaluate content relevance, conceptual accuracy, alignment between each item and Heller's five indicators, clarity of wording, and suitability for 11th-grade students. The expert review was used to strengthen content validity before the response data were analyzed using the Rasch model. Suggestions from the expert review were used to revise unclear wording, improve the contextual accuracy of the renewable-energy problems, and ensure that each item could elicit evidence of the intended problem-solving indicator. This validation step was necessary because Rasch analysis evaluates the empirical functioning of items, whereas content validation ensures that the items are theoretically and substantively aligned with the intended construct. This structure was adopted because the Heller framework is specifically designed to capture the sequential reasoning processes involved in physics problem-solving, rather than treating performance as a single, undifferentiated score. To improve clarity for readers and editors, the instrument blueprint can be presented in Table 2.

This blueprint makes it explicit that the instrument was designed to assess process-based problem-solving rather than only answer correctness. Such an approach is particularly relevant in physics education, where students' difficulties often arise not only at the final calculation stage but also in problem representation, principle selection, and result evaluation. The following describes the problem-solving skills used.

Table 2. Blueprint of the problem-solving instrument based on Heller's framework and cognitive taxonomy

Heller's problem-solving indicator	Operational description	Item codes	Cognitive taxonomy
Focus the Problem	Identifying the main problem, known information, unknown variables, and the objective of the solution	1FP, 2FP, 3FP, 4FP, 5FP	Understand–Analyze (C2–C4)
Describe the Physics	Representing the problem using relevant physical concepts, principles, equations, symbols, or diagrams	1DP, 2DP, 3DP, 4DP, 5DP	Understand–Apply (C2–C3)
Plan the Solution	Determining an appropriate strategy, sequence of steps, and formula selection	1PS, 2PS, 3PS, 4PS, 5PS	Analyze–Evaluate (C4–C5)
Execute the Plan	Carrying out the solution systematically through valid reasoning or calculation procedures	1EP, 2EP, 3EP, 4EP, 5EP	Apply–Analyze (C3–C4)
Evaluate the Answer	Checking the final result in terms of correctness, units, consistency, and reasonableness	1EA, 2EA, 3EA, 4EA, 5EA	Evaluate (C5)

Note. The cognitive taxonomy was assigned based on the dominant cognitive demand of each Heller problem-solving indicator, following the Revised Bloom's Taxonomy. The classification represents an a priori cognitive specification of the indicators, not an empirical estimate of Rasch item difficulty.

Although the example item in Figure 1 is presented in English for publication purposes, the actual test was administered to students in Indonesian. This was done to ensure that students' responses reflected their physics problem-solving ability rather than their English-language proficiency. The local context of the Cirata Floating Solar Power Plant was retained because it was familiar and relevant to Indonesian students. Before administration, the wording of the items was reviewed to ensure that the physics concepts, problem context, and response prompts were clear and appropriate for Grade 11 students.

In the Cirata Reservoir area, the Cirata Floating Solar Power Plant (PLTS Terapung Cirata) has an installed capacity of 192 MWp. This power plant helps supply electricity to West Java and the surrounding areas. However, during the rainy season, several days have cloudy skies, resulting in only about 3.5 hours of effective sunlight per day, which is lower than the average of 8 hours.

Meanwhile, several regencies surrounding the reservoir plan to fully utilize electricity from the PLTS, which serves 50,000 households, with an average daily electricity consumption of 4 kWh per household.

The local government is concerned about whether the Cirata PLTS will continue to meet this demand during the rainy season.

(Sumber: https://kumparan.com/)

Imagine yourself as an energy analyst working with the local government. Your task is to assess whether the Cirata Floating Solar Power Plant (PLTS Terapung Cirata) can still meet household electricity needs during the rainy season. Use relevant physics concepts to analyze the situation, then answer the following questions.

a. What is the main problem above?
b. State the physical principles/concepts related to the problem above! Include the physical quantities that must be known to solve the problem!
c. What is the proposed solution to this problem?
d. Use the proposed solution to solve the problem!
e. Review whether the solution is appropriate for the problem! Explain!

Figure 1. Problem-solving skills questions

Scoring Procedure

Each indicator was scored using an ordered polytomous scale from 0 to 4. A score of 0 indicated that the response was absent, irrelevant, or failed to show the target indicator. In contrast, a score of 4 indicated that the response was complete, accurate, systematic, and fully consistent with the intended indicator. Because the instrument used ordered response categories rather than dichotomous scoring, Rasch analysis was appropriate for evaluating whether the scoring categories functioned progressively and meaningfully across the latent trait. Rating-scale functioning is a critical component of Rasch analysis because the usefulness of an instrument depends not only on the quality of its items but also on whether the response categories function in the intended order. A concise scoring rubric is illustrated in Table 3.

Data Analysis

The response data were analyzed using MINISTEP VERSION 5.11.2 with a Rasch polytomous model. The analysis focused on several aspects of instrument quality and student performance: person reliability, item reliability,

Table 3. General scoring rubric for each problem-solving indicator

Score	Operational criterion
0	None of the required components of the intended indicator is present. The response is blank, irrelevant, or does not provide evidence related to the indicator.
1	Only 1 of the 4 required components is present, or the student mentions one relevant element but does not connect it correctly to the problem. The response cannot be used to support the next step in problem-solving.
2	2 of the 4 required components are present. The response relates to the indicator, but one or more core components are missing and/or there is a major conceptual or procedural error that affects the solution's correctness.
3	3 of the 4 required components are present. The response is mostly correct and relevant, but one component is incomplete, imprecise, or contains a minor error, such as missing units, rounding inconsistency, or an unclear explanation.
4	All 4 required components are present. The response is complete, accurate, systematic, and fully aligned with the intended indicator. No conceptual or procedural error is found.

item fit, person fit, item difficulty, Wright Map distribution, rating-scale functioning, and unidimensionality. In Rasch-based assessment, these outputs allow the researcher to determine whether the instrument measures a coherent construct, whether items behave consistently with model expectations, whether respondents show interpretable response patterns, and whether the instrument is fair across groups (Alnahdi et al., 2025; Shindi et al., 2025).

The data analysis was conducted in several steps. First, students' written responses were scored using the 0–4 rubric for each problem-solving indicator. Second, the scores were arranged into a response matrix consisting of students as respondents and problem-solving indicators as items. Third, the response matrix was entered into MINISTEP VERSION 5.11.2 and analyzed using the Rasch polytomous model. Fourth, person and item reliability were examined to assess the consistency of students' response patterns and the stability of the item difficulty hierarchy. Fifth, item fit and person fit were analyzed to identify whether the items and respondents behaved consistently with Rasch model expectations.

In addition, an additional rescoring analysis was conducted to examine the stability of the polytomous rating scale. The original five-category scale (0–4) was compared with a collapsed four-category scale (0–3) using MINISTEP VERSION 5.11.2. This comparison was performed to determine whether reducing the number of scoring categories improved measurement quality, particularly with respect to person reliability, item reliability, person separation, item separation, and mean infit and outfit MNSQ values. Based on this comparison, the scoring scale that demonstrated better measurement quality was retained for the final interpretation.

Item fit was evaluated using Infit Mean Square (MNSQ), Outfit MNSQ, standardized fit statistics (ZSTD), and Point-Measure Correlation (PTMEA Corr). Following widely used Rasch guidelines, items were considered acceptable when Infit and Outfit MNSQ values fell within approximately 0.50–1.50, ZSTD values were within approximately "2.0 to +2.0, and PTMEA Corr was positive (Monigir et al., 2024; Stolt et al., 2021). Values above the acceptable range indicate noise or unpredictability, whereas values far below it may indicate overfitting or redundancy.

Sixth, the Wright Map was used to examine the relative distribution of person ability and item difficulty on the same logit scale (Stolt et al., 2021). This map was used to determine whether the instrument adequately targeted the sample and whether the item set covered the range of students' problem-solving ability. Seventh, rating-scale analysis was conducted to determine whether all response categories were used and whether the average measures increased monotonically across categories. Eighth, unidimensionality was examined to determine whether the instrument primarily measured one coherent construct of problem-solving ability. Through these steps, the analysis provided evidence of the instrument's quality and the profile of students' problem-solving ability in renewable-energy physics topics.

Analytical Focus in Relation to the Research Questions

The analytical procedure was aligned directly with the study's research questions. RQ1 was addressed through reliability, item fit, person fit, and unidimensionality analyses to determine how well the Heller-based instrument fit the Rasch model. RQ2 was addressed through item-difficulty estimates and Wright Map interpretation to examine the distribution of student ability and item difficulty across the five indicators. RQ3 was addressed through rating-scale analysis. In this way, the method was designed not merely to describe students' scores, but to generate a psychometrically grounded profile of problem-solving performance in physics on renewable-energy topics.

RESULT AND DISCUSSION

RQ1. How well does the Heller-based problem-solving instrument on renewable energy topics fit the Rasch model?

Overview of the dataset

The analysis was conducted on response data from 35 students and 25 scored indicators derived from five renewable-energy problems, each operationalized through the five Heller problem-solving indicators. Responses were scored on an ordered 0–4 scale and analyzed using MINISTEP VERSION 5.11.2 under a Rasch polytomous model. This dataset provided the basis for evaluating instrument fit, response consistency, dimensionality, item difficulty, person ability, and category functioning.

The dataset structure was appropriate for addressing the three research questions. RQ1 focused on the psychometric fit of the Heller-based instrument, RQ2 on the distribution of student ability and item difficulty, and RQ3 on rating-scale functioning.

Table 4. Dataset profile and analytical scope

Component	Description
Participants	35 students
Problems	5 renewable-energy problems
Heller indicators	5 indicators
Analytical items	25 scored indicators
Score categories	0–4
Software	MINISTEP VERSION 5.11.2
Measurement model	Rasch polytomous model

Reliability and overall measurement quality

The first step in the Rasch analysis was to assess the instrument's overall reliability. In Rasch measurement, reliability is assessed not only through internal consistency but also through person reliability, item reliability, and separation indices, because these statistics indicate how well the instrument distinguishes among students' abilities and calibrates item difficulty.

Table 5. Summary of reliability indices

Statistic	Value	Interpretation
Person Reliability	0.89	Good
Item Reliability	0.86	Good
Person Separation	2.83	Good
Item Separation	2.49	Good
Cronbach's Alpha	0.90	Excellent

The reliability results indicate that the instrument demonstrated a good, stable measurement pattern for both students and items. The person's reliability value of 0.89 shows that students' response patterns were sufficiently consistent to support the interpretation of their problem-solving ability. This means that the instrument did not produce random response patterns, but provided meaningful information about differences in students' performance. The Cronbach's Alpha value of 0.90 further supports the internal consistency of the scored responses, indicating that the instrument had strong overall reliability for measuring students' problem-solving ability on renewable-energy topics.

The item reliability value of 0.86 indicates that the 25 scored indicators formed a reasonably stable hierarchy of item difficulty. This result suggests that the indicators did not function as equally easy or equally difficult tasks, but represented different levels of challenge across the problem-solving process. This is important because a diagnostic problem-solving instrument should be able to identify which stages of problem solving are relatively easier or more difficult for students.

The separation indices provide additional evidence of the instrument's diagnostic quality. The person separation value of 2.83 indicates that the instrument was able to distinguish students into several levels of problem-solving ability. This is useful for instructional interpretation because students can be identified based on the different levels of support they may need. Meanwhile, the item separation value of 2.49 indicates that the scored indicators spanned more than one level of difficulty, rather than clustering within a narrow range of difficulty. Taken together, the reliability and separation results indicate that the Heller-based instrument was sufficiently consistent, informative, and diagnostically useful for profiling students' problem-solving ability on renewable-energy physics topics.

Based on Table 6, the Rasch summary statistics indicate that the students' response data showed a stable measurement pattern. For the person component, the mean Infit MNSQ of 1.01 and the mean Outfit MNSQ of 1.00 indicate that the students' response patterns were very close to the Rasch model's expectations. This means that the observed responses did not deviate substantially from the predicted pattern, so the students' answers can be considered consistent and interpretable as evidence of their problem-solving ability. The mean Infit ZSTD value of 0.08 and the mean Outfit ZSTD value of 0.05 further support the conclusion that the student responses did not show any serious misfit. In addition, the person reliability value of 0.89 and the separation index of 2.83 indicate that the instrument was able to distinguish students into several different levels of ability, making the measurement sufficiently

Table 6. Essential rasch summary statistics from ministep version 5.11.2 output

Component	N	Mean total score	SD measure	Mean Infit MNSQ	Mean Infit ZSTD	Mean Outfit MNSQ	Mean Outfit ZSTD	Real RMSE
Person	35	51.90	0.61	1.01	0.08	1.00	0.05	0.20
Item	25	72.70	0.45	1.02	0.12	1.00	0.06	0.17

informative for profiling variation in students' problem-solving performance on renewable-energy topics.

For the item component, the mean Infit MNSQ value of 1.02 and the mean Outfit MNSQ value of 1.00 suggest that the 25 indicators scored in accordance with the model's expectations. This means that each Heller-based item consistently contributed to measuring the same construct, without exhibiting response patterns that were too random or too uniform. The mean Infit ZSTD value of 0.12 and the mean Outfit ZSTD value of 0.06 reinforce that the item fit remained within an acceptable range. The item reliability value of 0.86 and the item separation index of 2.49 show that the item set formed a fairly stable hierarchy of difficulty and represented more than one level of challenge in the problem-solving process. Overall, the results in Table 5 indicate that the Heller-based instrument was reliable, consistent, and suitable for profiling students' problem-solving ability on renewable-energy topics.

To further examine the stability of the rating scale, an additional rescoring analysis was conducted by comparing the collapsed 0–3 scale with the original 0–4 scale. The comparison was intended to determine whether reducing the number of categories improved the overall measurement quality.

Table 7. Comparison of measurement quality between the 0–3 and 0–4 rating scales

Indicator	0–3 Scale	0–4 Scale
Person reliability	0.86	0.89
Item reliability	0.84	0.86
Person separation	2.51	2.83
Item separation	2.32	2.49
Cronbach's alpha	0.89	0.90
Mean person Infit MNSQ	1.02	1.01
Mean person Outfit MNSQ	0.99	1.00
Mean item Infit MNSQ	1.02	1.02
Mean item Outfit MNSQ	0.99	1.00

Table 7 shows that the original 0–4 rating scale produced slightly better measurement quality than the collapsed 0–3 scale. The 0–4 scale had higher person reliability, item reliability, person separation, item separation, and Cronbach's alpha, indicating that it provided more stable measurement and better distinguished students' ability levels and item difficulty. In addition, the mean Infit and Outfit MNSQ values for both persons and items were close to the expected value of 1.00 in both scales, showing acceptable fit to the Rasch model. Therefore, the original 0–4 scale was retained for the final analysis because it demonstrated better overall measurement quality and was more consistent with the staged scoring structure of Heller's problem-solving framework.

Item Fit to the Rasch Model

To answer RQ1, item fit was examined using Infit MNSQ, Outfit MNSQ, ZSTD, and PTMEA Corr. Based on predefined criteria, an item was considered acceptable when its Infit and Outfit MNSQ values fell within 0.50–1.50, its ZSTD values were within "2.0 to +2.0, and its PTMEA Corr was positive.

The analysis showed that the 25 scored indicators generally demonstrated acceptable fit to the Rasch model. This indicates that most items functioned as intended and contributed

Table 8. Summary of item-fit results

Category	Items	Interpretation
Acceptable fit	1EA, 1DP, 1PS, 2DP, 2EP, 2FP, 2PS, 3EA, 3PS, 4EA, 4PS, 5EA, 5EP, 5FP, 5PS, SDP, 2DREP	These items had infit and outfit MNSQ values within the acceptable range and showed response patterns consistent with the Rasch model.
Borderline but acceptable	1FP, 3EP, 4DP, 3DP, 3FP, 1EP, 2EA, 4FP	These items remained interpretable but were closer to the lower or upper boundaries of the fit range and should be reviewed substantively.
Items closest to the ideal model fit	4PS, 5EA, 2EP, 5FP, 5PS	These items had infit and outfit values closest to the ideal MNSQ value of 1.00. This category refers to statistical fit, not item difficulty or proportion correct.

meaningfully to measuring students' problem-solving ability. The positive PTMEA Corr values across the items further suggest that each indicator was positively aligned with the overall latent trait. However, several indicators require substantive review. Items 4FP, 5DP, 4EP, and 2DP showed relatively higher infit and/or outfit MNSQ values than the other items, suggesting a greater degree of unexpected response variation, although remained acceptable for interpretation.

By contrast, Items 1EP and 2EA were closer to the lower boundary of the fit range, indicating more predictable response patterns than the model would predict. These results do not justify removing the items. However, they suggest that the indicators should be reviewed in future refinements, particularly regarding item wording, scoring rubric clarity, scoring consistency, and alignment between task demand and indicator intent. Item 5EP was interpreted as a difficult but acceptable-fitting item. Although many students made procedural errors, its fit statistics remained within the acceptable range, indicating that the high error rate reflected substantive task difficulty rather than erratic or model-inconsistent response behavior.

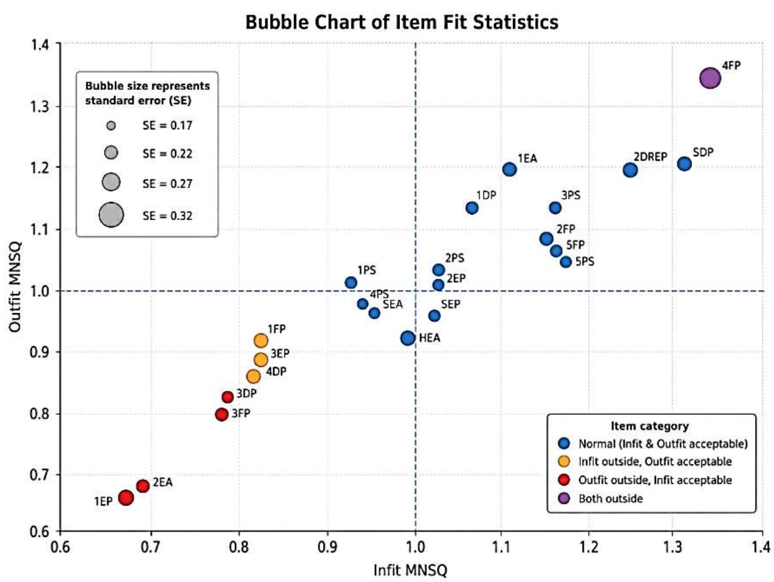

Figure 2. Item-fit Scatter plot

Figure 2 supports the pattern shown in Table 6 by demonstrating that the items were generally clustered within the acceptable Rasch fit boundaries. Overall, the item-fit evidence indicates that the Heller-based instrument was sufficiently aligned with Rasch model expectations, thereby supporting its use for further ability profiling.

Person fit and response consistency

Person fit was analyzed to determine whether students exhibited response patterns consistent with model expectations. In Rasch analysis, person misfit does not automatically invalidate the dataset, but it signals that some response patterns may be unusually predictable, erratic, or inconsistent relative to item difficulty. To provide a visual overview of these response patterns, the distribution of person-fit values for Infit MNSQ and Outfit MNSQ is shown in Figure 3.

Figure 3 shows that most students were distributed within the acceptable person-fit range, indicating that their response patterns were

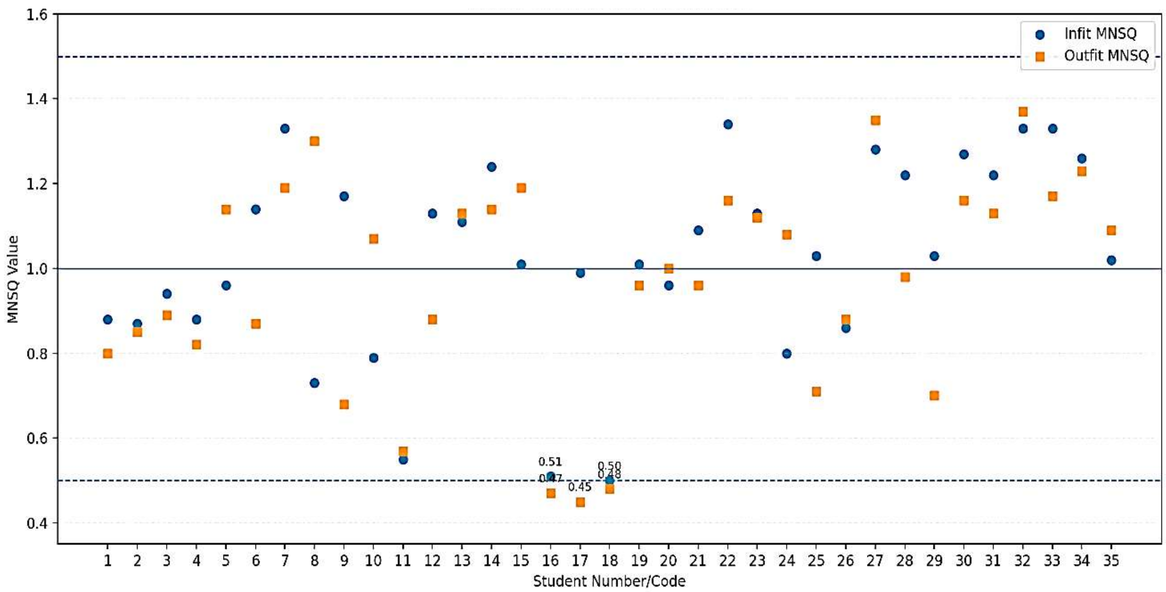

Figure 3. Person-fit plot based on Infit MNSQ and Outfit MNSQ

broadly consistent with Rasch model expectations. Nevertheless, a small number of students appeared near or below the lower fit boundary, which may indicate overly predictable or insufficiently varied response behavior. This visual evidence supports the more detailed person-fit classification presented in Table 9.

Table 9. Frequency distribution of person-fit statistics based on infit and outfit MNSQ ranges

Infit MNSQ range	Outfit MNSQ < 0.50	Outfit MNSQ 0.50–1.50	Outfit MNSQ > 1.50	Total
Infit MNSQ < 0.50	1 (2.86%)	0 (0.00%)	0 (0.00%)	1 (2.86%)
Infit MNSQ 0.50–1.50	2 (5.71%)	32 (91.43%)	0 (0.00%)	34 (97.14%)
Infit MNSQ > 1.50	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Total	3 (8.57%)	32 (91.43%)	0 (0.00%)	35 (100.00%)

Table 9 was revised into a frequency distribution table to address the reviewer's readability comment. Instead of listing all student codes in a single cell, the revised table cross-classifies students according to their infit and outfit MNSQ ranges. The results indicate that 32 students, or 91.43%, had both infit and outfit values within the acceptable range of 0.50–1.50. Three students, or 8.57%, showed person-fit values below the lower acceptable threshold and were therefore flagged for further review. No students showed infit or outfit values above 1.50.

Evidence of unidimensionality

A key requirement of Rasch measurement is that the instrument should reflect a dominant underlying construct. In this study, unidimensionality was evaluated using the standardized residual variance.

Table 10. Standardized residual variance in eigenvalue units

Component	Eigenvalue	Observed	Expected
Total raw variance in observations	40.94	100.00%	100.00%
Raw variance explained by measures	15.94	38.90%	40.10%
– Raw variance explained by persons	10.03	24.50%	25.30%
– Raw variance explained by items	5.91	14.40%	14.90%
Raw unexplained variance (total)	25	61.10%	59.90%
– Unexplained variance in 1st contrast	3.21	7.90%	12.90%
– Unexplained variance in 2nd contrast	2.70	6.60%	10.80%
– Unexplained variance in 3rd contrast	2.23	5.50%	8.90%
– Unexplained variance in 4th contrast	2.01	4.90%	8.00%
– Unexplained variance in 5th contrast	1.75	4.30%	7.00%
Essential unidimensionality	–	57.20%	–

The results in Table 10 indicate evidence of a dominant primary measurement dimension, although the instrument cannot be regarded as perfectly unidimensional. The variance explained by the Rasch measures was 38.90%, indicating that the intended construct accounted for a substantial portion of the response structure. However, the first residual contrast reached 3.21 eigenvalue units, indicating the presence of a secondary response pattern that should be interpreted in light of the content of the tested renewable energy physics items. Based on the content review, this secondary dimension appears to be related to differences among the problem-solving indicators. Items representing Focus the Problem and Describe the Physics tend to require students to identify relevant information, recognize the physical context, and connect renewable-energy phenomena to appropriate physics concepts.

In contrast, items representing Plan the Solution, Execute the Plan, and Evaluate the Answer require more procedural reasoning, formula selection, calculation, unit interpretation, and evaluation of the plausibility of the result obtained. Thus, the secondary structure is likely not caused by an entirely different construct, but by variation in cognitive demand between conceptual-contextual tasks and procedural-evaluative tasks within the renewable energy physics context. Therefore, the instrument may still be considered sufficiently unidimensional for exploratory profiling and initial educational measurement. However, the results suggest that future refinement should pay closer attention to balancing conceptual, procedural, and evaluative indicators across the items tested.

The inter-cluster residual correlation analysis was conducted to examine whether the secondary dimension indicated by the first

Table 11. Inter-Cluster residual correlations based on the first contrast

Cluster comparison	Pearson correlation	Disattenuated correlation	Cluster sizes	Interpretation
Cluster 1–2	0.5531	0.7093	10 and 6	Moderate to strong positive correlation
Cluster 1–3	0.4362	0.5412	10 and 9	Moderate positive correlation
Cluster 2–3	0.7090	0.9301	6 and 9	Strong positive correlation

contrast interfered with the main construct. As shown in Table 10, all residual cluster correlations were positive, with Pearson correlation values ranging from 0.4362 to 0.7090 and disattenuated correlations ranging from 0.5412 to 0.9301. Although the correlation between Cluster 1 and Cluster 3 was moderate, the positive direction of all correlations indicates that the residual clusters did not represent opposing constructs. In addition, each cluster showed a strong positive correlation with the overall item set, ranging from 0.8243 to 0.8410. Therefore, although the first contrast eigenvalue suggested a possible secondary residual pattern, this pattern did not substantially interfere with the main construct of physics problem-solving ability.

RQ2. How are students' problem-solving abilities and item difficulties distributed across the five Heller indicators?

Item difficulty distribution

To address RQ2, item difficulty was examined through Rasch measure values. Positive logits indicate more difficult items, whereas negative logits indicate easier items. Values near zero indicate items close to the average difficulty level of the instrument. For reporting efficiency, Table 8 presents the most difficult, easiest, and representative items within the average difficulty range. The full ordering of item measures was further interpreted using the Wright Map.

The results show that the instrument had a useful range of item difficulty, from "0.99 to 0.87

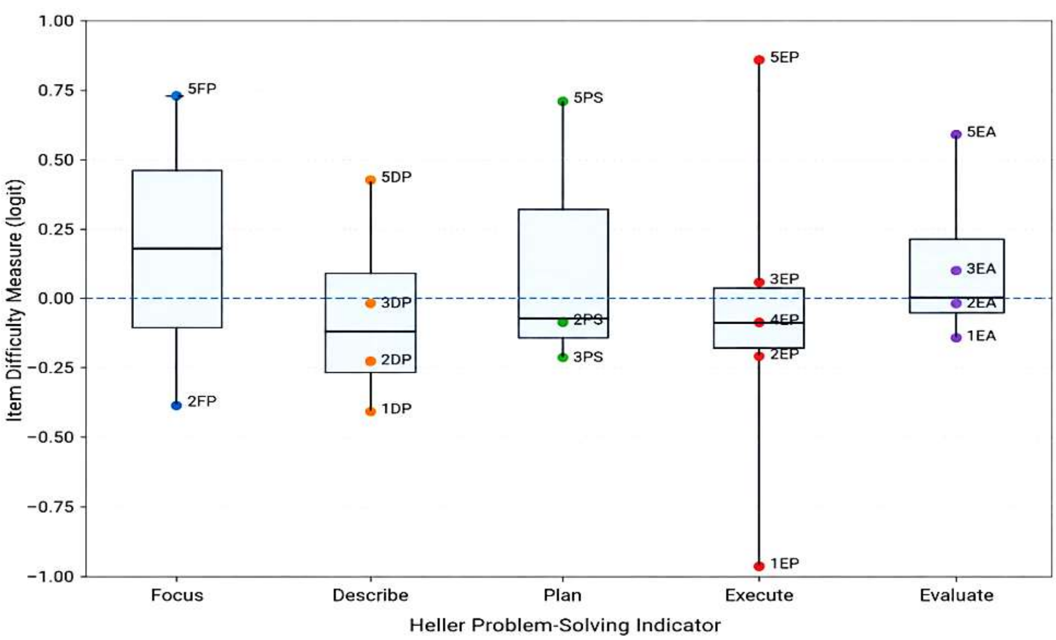

Figure 4 . Boxplot of item difficulty distribution by Heller problem-solving indicator

logits. As summarized in Table 8 and visualized more comprehensively in the Wright Map, the item set covered easy, moderate, and difficult levels rather than clustering at a single point on the continuum. The most difficult item was 5EP, whereas the easiest was 1EP. Interestingly, the difficulty pattern did not align with any single Heller indicator. For example, the Plan contained both the hardest item (5EP) and the easiest item (1EP). This suggests that item difficulty was shaped not only by the indicator type but also by the specific problem context in which the indicator was embedded.

A second notable pattern is that Problem 5 was the most demanding across several indicators (5FP, 5DP, 5PS, 5EP, 5EA), whereas Problem 1 contained several of the easiest indicators. This implies that the contextual and conceptual demands of the renewable-energy task substantially influenced student performance. Thus, the results indicate that the instrument captured meaningful differences in difficulty across tasks and not merely across scoring categories.

Wright map: alignment of student ability and item difficulty

The Wright Map was used to examine the distributions of student ability and item difficulty on the same logit scale.

The Wright Map shows that student ability ranged from “1.12 to 1.28 logits, while item difficulty ranged from “0.99 to 0.87 logits. The

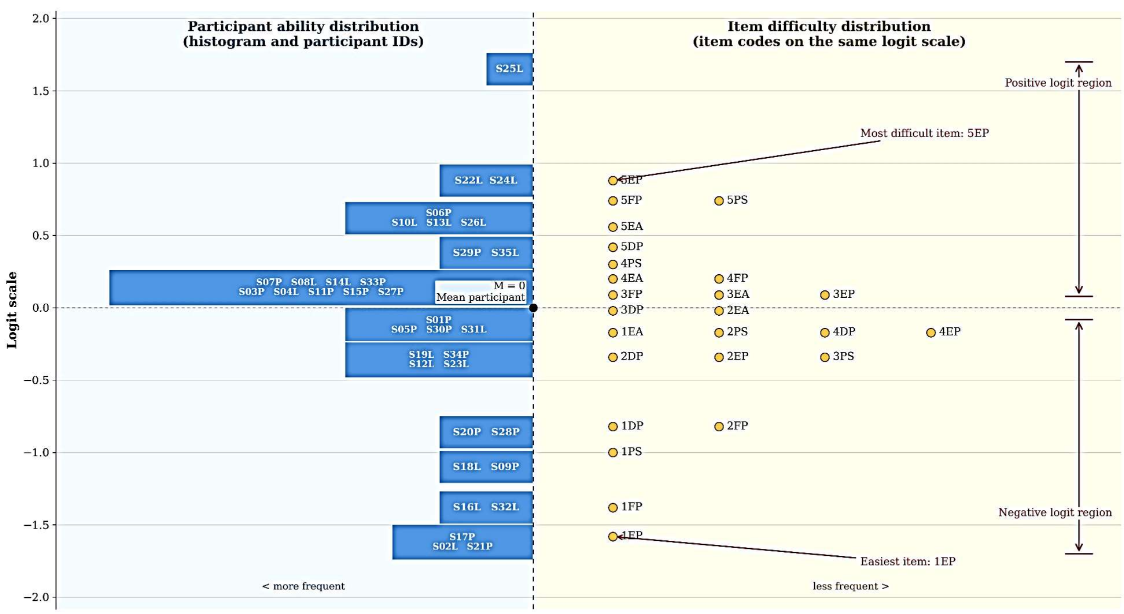

Figure 5. Wright’s map of person ability and item difficulty

student with the highest estimated ability was S25L (1.28 logits), while the lowest estimated ability was observed for S02L and S21P (“1.12 logits). On the item side, 5EP was the most difficult item and 1EP the easiest, consistent with the item-difficulty results reported in Table 8.

Overall, the map indicates that the instrument was reasonably well targeted to the sample, as the person and item distributions overlapped on a common scale. However, the map also shows a slight mismatch at both extremes. A few students were located above the most difficult item, suggesting that the instrument may not fully challenge the highest performers. Conversely, a few students fell below the easiest item, suggesting that it may still be too demanding for some lower-ability respondents. Thus, the Wright Map supports the instrument’s adequacy while suggesting that adding a few more difficult and a few more accessible items could improve targeting.

RQ3. Does The Scoring Scale Function Properly?

Rating-scale functioning

To answer the first part of RQ3, the ordered 0–4 scoring scale was examined. A well-functioning scale should show that all are used and that average measure values increase monotonically from lower to higher categories.

The results indicate that the rating scale functioned as intended. All five categories were used by respondents, indicating that none were redundant or empty. In addition, the average measure values increased in a monotonic pattern from "0.43 for category 0 to 0.50 for category 4. This progression indicates that higher assigned scores corresponded to higher estimated ability levels.

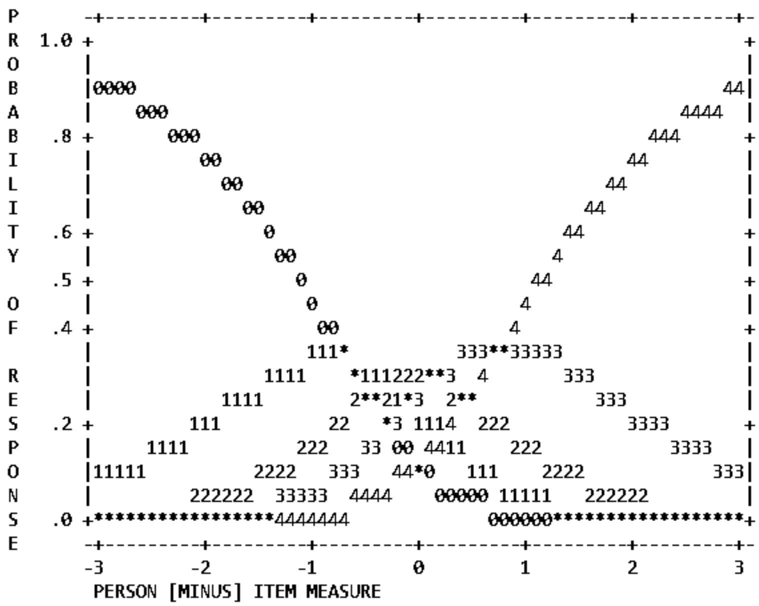

Figure 6. Rating-scale functioning

The most frequently used category was 3, followed by 2, while categories 0 and 4

were less frequent but still clearly represented. Taken together, these results suggest that the rubric distinguished progressive levels of problem-solving performance and that the ordered categories worked meaningfully in practice.

Summary of findings in relation to the research questions

Taken together, the results provide a coherent answer to the three research questions. In relation to RQ1, the Heller-based instrument showed good reliability, acceptable item fit, largely acceptable person fit, and reasonably adequate unidimensionality, indicating that it functioned sufficiently well under the Rasch model. In relation to RQ2, student ability and item difficulty were meaningfully distributed on the same logit scale, with the most difficult items concentrated mainly in Problem 5 and the easiest in Problem 1, suggesting that task context played an important role in shaping difficulty. Regarding RQ3, the 0–4 rating scale functioned properly.

Overall, these results indicate that the instrument can be used to generate a psychometrically grounded profile of students' problem-solving ability in physics on renewable-energy topics. However, several borderline items and the modest residual structure suggest that continued refinement would further strengthen measurement quality.

This study examined students' problem-solving ability in physics on renewable-energy topics using a Heller-based instrument analyzed with the Rasch model. Overall, the findings indicate that the instrument functioned adequately and that students' performance could be interpreted in a psychometrically meaningful way. The reliability results were good; the item-fit statistics were acceptable across all 25 scored indicators; most students showed response patterns consistent with model expectations; and the rating scale functioned progressively. Taken together, these findings suggest that the Heller-based instrument was suitable for profiling students' problem-solving ability in a renewable-energy context, while also revealing several areas for further refinement. In substantive terms, the results support the argument that renewable energy is not only a relevant content domain for physics learning but also a productive context for assessing how students interpret, represent, plan, execute, and evaluate solutions to complex and socially meaningful problems (Agung et al., 2024; Attanayake et al., 2024; Kellberg, Keller, Nordine, Moser, & Lewalter, 2025; Rehman et al., 2026).

The findings also reinforce previous work showing that problem-solving in physics should be treated as a structured, multidimensional competence rather than a single outcome score. This conclusion is supported by several parts of the Rasch results. First, the acceptable fit of all 25 indicators shows that the indicators derived from the five Heller stages worked coherently within the measurement model. This suggests that identifying the problem, describing the relevant physics, planning the solution, executing the procedure, and evaluating the answer were not isolated activities, but connected components of students' problem-solving performance. Second, the unidimensionality evidence indicates the presence of a dominant latent construct, although the result was not perfectly unidimensional. This pattern is pedagogically reasonable because physics problem solving requires students to coordinate several cognitive moves, including representation, conceptual reasoning, mathematical execution, and reflection. Therefore, the present findings do not merely repeat earlier scholarship but clarify how the Heller stages can operate as an integrated diagnostic structure in a renewable-energy context. The acceptable fit of all indicators suggests that the five Heller stages captured a coherent latent construct while still preserving meaningful variation in task demand. This is consistent with earlier scholarship that emphasizes the diagnostic value of process-based problem-solving frameworks in physics, especially those that distinguish between identifying the problem, describing the relevant physics, planning the solution, executing the procedure, and evaluating the answer (Adams & Wieman, 2015; K. Heller, 2013; K. Heller & Heller, 2010; P. Heller et al., 1992).

From a pedagogical perspective, the structured approach is useful because it helps teachers locate where students' reasoning begins to break down. For example, a student may understand the physical situation but fail to plan an appropriate solution, or may execute a calculation but fail to evaluate whether the answer is physically meaningful. In this sense, the Heller-based framework provides more instructional information than a single total score because it breaks the problem-solving process into stages that can be targeted directly through feedback and scaffolding. At the same time, the evidence for unidimensionality indicates that the instrument was not perfectly unidimensional, which is unsurprising, given that physics problem-solving often requires multiple interconnected subprocesses and cognitive moves. In that sense, the present results align with recent studies arguing that students' problem-solving performance in physics involves a combination of representation, reasoning, execution, and reflection that may cluster around a dominant construct while still retaining some internal heterogeneity (Meyer & Friege, 2026; Musengimana, Yadav, Uwamahoro, & Nizeyiman, 2025; Tong et al., 2023).

A particularly important result concerns the distribution of item difficulty. The analysis showed that difficulty did not map neatly onto a single Heller indicator. Instead, the hardest and easiest items could appear within the same indicator category, particularly within Execute the Plan. This suggests that students' performance was shaped not only by the formal problem-solving stage but also by the specific conceptual and contextual demands of each renewable-energy problem. This interpretation is supported by the pattern observed in Problem 5, especially in indicator 5EP, which became one of the most difficult indicators in the analysis. In students' written responses, difficulties with 5EP were not limited to the execution of mathematical procedures. Several responses showed that students struggled to translate the renewable-energy situation into the relevant physics quantities before completing the calculation. Some answers showed incomplete substitution of known variables, inconsistent use of units, or calculation steps that were not connected back to the physical meaning of the renewable-energy problem. This pattern indicates that the difficulty of 5EP was related to the interaction between conceptual understanding, contextual interpretation, and procedural execution. By contrast, Problem 1 contained several easier indicators because the required reasoning appeared to be more familiar and more directly connected to routine energy concepts. Therefore, the difference between Problem 5 and Problem 1 suggests that item difficulty was influenced by the task's conceptual and contextual features, not solely by the Heller stage itself. This finding extends earlier research on physics problem-solving, which has often shown that student difficulty arises from the interaction between content context and the demand for reasoning rather than from the procedural stage alone (Evendi et al., 2024; Kuo et al., 2013; Qotrunnada, 2022; Worku et al., 2025). It also resonates with research on renewable energy education, which shows that sustainability-related topics can be cognitively demanding because students must connect disciplinary knowledge to real-world technological and socio-scientific issues rather than merely applying routine formulas (L. Hu, 2024; Majid et al., 2025; Martínez-Borreguero et al., 2024).

Another meaningful contribution of the study lies in the evidence for scale functioning and fairness. The monotonic increase in the average measure across the 0-4 categories indicates that the scoring rubric worked as intended and successfully differentiated levels of student performance. This supports the use of ordered polytomous scoring for process-based physics problem solving, where partial understanding and partially correct reasoning are educationally important and should not be collapsed into dichotomous right-or-wrong judgments (Adu et al., 2026; Barker et al., 2017; Shahat et al., 2024). This finding is important because fairness is a central requirement of educational measurement, especially when the goal is to construct valid profiles of student competence rather than rank learners (Moradi et al., 2022; Wetzel & Böhnke, 2017).

An important strength of the study is the internal triangulation of evidence across several Rasch outputs. The reliability indices indicated that the instrument consistently distinguished student ability and item difficulty; the item-fit and person-fit analyses showed that most indicators and respondents behaved in line with model expectations; the unidimensionality results supported the presence of a dominant common construct; the Wright Map demonstrated reasonable targeting between item difficulty and student ability; and the rating-scale analysis confirmed that the ordered categories worked progressively. Because these conclusions converge across multiple analytical layers, the study does not rely on a single statistic or one isolated output. Instead, the argument for instrument adequacy is supported by mutually reinforcing evidence, which strengthens the interpretive credibility of the findings (Boone, 2016; Davis & Boone, 2021; Qudratuddarsi et al., 2022; Stolt et al., 2021). This triangulated pattern is particularly valuable in physics assessment research, where the validity of an instrument depends not only on reliability, but also on construct coherence, scale functioning, and fairness.

From an instructional perspective, the findings suggest that renewable-energy topics can serve as effective contexts for diagnosing how students think through physics problems in a staged manner. The fact that item difficulty varied substantially across problems suggests that teachers should not assume that all energy-related problems require the same level of reasoning. Instead, instructional design may need to scaffold students differently depending on whether the task emphasizes identifying variables, representing the physics, planning a strategy, carrying out calculations, or evaluating the answer. This is consistent with recent calls in science and sustainability education to move beyond factual teaching and toward learning environments that require students to reason through authentic socio-scientific challenges (Agung et al., 2024; Kellberg, Keller, Nordine, Moser, & Lewalter, 2025; Martínez-Borreguero et al., 2024). In this sense, the present study contributes not only to measurement but also to pedagogy, as it shows how a Rasch-calibrated Heller-based instrument can help identify where students struggle within the broader problem-solving process in renewable-energy topics.

At the same time, the discussion should be interpreted with appropriate caution. The sample size was limited to a single cohort of 35 students, and several items remained borderline despite meeting the acceptable criteria. In addition, the residual structure suggests that the instrument, while sufficiently unidimensional for exploratory use, may still contain some heterogeneity linked to specific task features or indicator demands. These points do not weaken the main conclusion, but they do suggest that the present instrument should be viewed as a strong initial profiling tool rather than as a final, fully optimized measure. Future research should therefore examine the instrument using larger, more diverse samples, test its performance across different educational levels, and refine items that remain near the fit boundaries. It would also be valuable to compare student profiles across renewable-energy subtopics and to investigate whether particular Heller stages are more sensitive than others to conceptual complexity or contextual familiarity (Mayer, 2014; Musengimana, Yadav, Uwamahoro, & Nizeyiman, 2025; Widya et al., 2026).

LIMITATION AND OMPLICATIONS

This study has several limitations that should be considered when interpreting the findings. First, the analysis was based on a relatively small sample of 35 students from a single cohort and instructional context. Although this was sufficient for an initial Rasch-based profiling study, it limits the generalizability of the results to broader student populations, different school settings, and other curricular levels. Second, the instrument consisted of five renewable-energy problems, operationalized through 25 scored indicators, which provided useful initial coverage of Heller's framework but may not yet fully capture the conceptual and contextual complexity of renewable-energy physics learning. Third, although the instrument showed acceptable overall fit and reasonable unidimensionality, the residual structure and several borderline items indicate that some local heterogeneity persisted across problem contexts and indicator demands. Thus, the instrument can be regarded as psychometrically adequate for exploratory use, but not yet as a final optimized measure for large-scale assessment.

A further limitation concerns the scope of inference that the design supports. The study was descriptive and measurement-oriented rather than experimental, so it does not permit conclusions about the effectiveness of a particular instructional intervention for improving problem-solving ability. Likewise, the Wright Map generally indicated acceptable targeting, yet also showed slight mismatches at both ends of the ability continuum, suggesting that the instrument may not fully challenge the highest-performing students or sufficiently support the lowest-performing students. Future studies should therefore test the instrument on larger and more heterogeneous samples, include students from different educational levels and institutional contexts, and consider adding items at both higher and lower difficulty levels to improve scale targeting and construct coverage.

Despite these limitations, the study offers several important implications for both educational measurement and physics instruction. From a measurement perspective, the findings show that a Heller-based instrument, analyzed using the Rasch model, can provide a psychometrically grounded profile of students' problem-solving ability in renewable-energy topics. This is important because it demonstrates that problem-solving can be assessed as a staged reasoning process rather than as a single total score, while also allowing researchers to evaluate item functioning, person consistency, category performance, and fairness within a single coherent framework. From an instructional perspective, the results suggest that renewable-energy topics can serve as productive assessment contexts for diagnosing how students reason through contemporary socio-scientific physics problems. The variation in item difficulty across problems further indicates that teachers should pay closer attention to contextual demand, not only to the procedural stage, when designing learning tasks. More broadly, the study implies that problem-solving instruction in physics should be strengthened through context-sensitive scaffolding that helps students move from problem identification to representation, planning, execution, and evaluation in sustainability-related learning situations.

CONCLUSION

This study examined students' problem-solving ability in physics within renewable-energy contexts using a Heller-based assessment instrument analyzed through the Rasch model. The findings indicate that the instrument is suitable for providing an initial structured description of students' problem-solving performance in a classroom-based assessment context. The analysis also suggests that students' responses can be meaningfully interpreted through the five stages of problem solving. However, performance across these stages is influenced not only by procedural demands but also by the conceptual complexity and contextual familiarity of the renewable-energy tasks. Across the indicators, student performance varies, reflecting differences in how learners engage with problem representation, conceptual reasoning, planning, execution, and evaluation. This suggests that problem-solving ability in physics is not a single, unified skill but a multi-layered construct shaped by both cognitive processes and task characteristics.

The results further imply that the assessment of problem-solving should not rely solely on stage completion but should also consider the interaction between cognitive demand and the contextual features of physics problems. In this sense, renewable-energy contexts provide a meaningful setting for observing differences in how students construct and apply their understanding of physics. Overall, the study supports the use of structured problem-solving frameworks for diagnostic purposes in physics education. Future work should expand the scope of validation by involving larger, more diverse samples and refining item design to better capture variation across levels of cognitive demand and contextual complexity.

■ ACKNOWLEDGMENT

The first author (W.E) would like to express her gratitude to Education Fund Management Institution (*Lembaga Pengelola Dana Pendidikan*, LPDP), under the Ministry of Finance of the Republic of Indonesia, for supporting her studies and the writing of this review.

■ DECLARATION OF GENERATIVE AI USAGE IN THE WRITING PROCESS

During the writing process for this manuscript, the author used Grammarly for language editing, including improving sentence structure, grammar, spelling, and punctuation, as well as enhancing the text's readability. Following use of the tool, the author reviewed and revised the entire manuscript as necessary and assumes full responsibility for the final content of this article.

■ REFERENCES

Adams, W. K., & Wieman, C. E. (2015). Analyzing the many skills involved in solving complex physics problems. *American Journal of Physics*, *83*(5), 459–467. https://doi.org/10.1119/1.4913923

Adu, P., Bartholomew, E., & Medvedev, O. N. (2026). Rasch Methodology for Assessment in International Contexts. In *International Handbook of Behavioral Health Assessment* (pp. 1–30). Cham: Springer Nature Switzerland.

Agung, N., Riva, I., Tahta, F., & Zulfia, A. (2024). Renewable energy education and awareness among Indonesian students: Exploring challenges and opportunities for a sustainable future. *Sustainable Energy Technologies and Assessments*, *63*, 103631. https://doi.org/10.1016/j.seta.2024.103631

Alnahdi, A. H., Alsubiheen, A. M., & Aldaihan, M. M. (2025). Rasch measurement model supports the unidimensionality and internal structure of the arabic oswestry disability index. *Journal of Clinical Medicine*, *14*(4), 1–12. https://doi.org/10.3390/jcm14041259

Ario, M., Suhendra, Jupri, A., & Nurlaelah, E. (2025). Students' errors and learning obstacles in solving algebraic word problems: hermeneutic phenomenology. *Education Sciences*, *15*(12), 1–27. https://doi.org/10.3390/educsci15121674

Attanayake, K., Wickramage, I., Samarasinghe, U., Ranmini, Y., Ehalapitiya, S., Jayathilaka, R., & Yapa, S. (2024). Renewable energy as a solution to climate change: Insights from a comprehensive study across nations. *PLoS ONE*, 1–21. https://doi.org/10.1371/journal.pone.0299807

Barker, B. A., Donovan, N. J., Schubert, A. D., & Walker, E. A. (2017). Using rasch analysis to examine the item-level psychometrics of the infant-toddler meaningful auditory integration scales. *Speech Lang Hear*, *20*(3), 130–143. https://doi.org/10.1080/2050571X.2016.1243747

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

Davis, D. R., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, *2*(June), 100054. https://doi.org/10.1016/j.ijedro.2021.100054

Evendi, Saputri, M., Miza, H., & Hamid, A. (2024). Analysis of students' difficulties in solving physics problems using polya's theory. *Jurnal Penelitian Pendidikan IPA*, *10*(9), 6525–6532. https://doi.org/10.29303/jppipa.v10i9.7127

Hasyim, M., & Sujiono, E. H. (2023). The effect of applying the ethno-stem-project-based learning model on students ' higher-order thinking skill and misconception of physics topics related to Lake Tempe, Indonesia. *Jurnal Pendidikan IPA Indonesia, 12*(1), 1–13. https://doi.org/10.15294/jpii.v12i1.38703

Heller, K. (2013). Helping your students develop expertise in problem solving – while learning physics. *USA:University of Minnesota, AAPT New Faculty Workshop*.

Heller, K., & Heller, P. (2010). *Cooperative problem solving in physics: a user's manual*. University of Minnesota, 310.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physic*, *60*(7), 627–636. https://doi.org/10.1119/1.17117

Hernández, A. M., & Prakoso, Y. T. B. (2021). The learning activation approach—understanding Indonesia's energy transition by teaching it. *Energies*, *14*(17), 1–19. https://doi.org/10.3390/en14175224

Hu, J.-L., & Yang, P.-S. (2024). Interactive cycles between energy education and energy preferences: a literature review on empirical evidence. *Energies*, *17*, 1–31. https://doi.org/10.3390/en17205092

Hu, L. (2024). Programming and 21st century skill development in K-12 schools: A multidimensional meta-analysis. *Journal of Computer Assisted Learning*, *40*(2), 610–636. https://doi.org/10.1111/jcal.12904

Irfani, V. A., & Simanjuntak, M. P. (2026). *Pengaruh problem based learning berbantuan website html5 terhadap pemecahan masalah siswa pada materi energi terbarukan* [the effect of problem based learning assisted by html5 websites on students' problem solving in renewable energy material]. *Journal of Education Technology Information, Social Science and Health*, *5*(1), 71–82. https:// doi.org/10.57235/jetish.v5i1.8140

Kellberg, S., Keller, M., Nordine, J., Moser, S., Keller, M., & Nordine, J. (2025). Communication and public engagement energy literacy for all/ ? Exploring whether prior interest and energy knowledge mediate energy literacy development in a modern socio-scientific museum exhibition in a modern socio-scientific museum exhibition. *International Journal of Science Education, Part B*, *15*(2), 161–182. https://doi.org/10.1080/21548455.2024.2344129

Kellberg, S., Keller, M., Nordine, J., Moser, S.,

& Lewalter, D. (2025). Energy literacy for all? Exploring whether prior interest and energy knowledge mediate energy literacy development in a modern socio-scientific museum exhibition. *International Journal of Science Education, Part B: Communication and Public Engagement*, *15*(2), 161–182. https://doi.org/10.1080/21548455.2024.2344129

Khalaf, M. A., & Omara, E. M. N. (2022). Rasch analysis and differential item functioning of English language anxiety scale (ELAS) across sex in Egyptian context. *BMC Psychology*, *10*(1), 1–16. https://doi.org/10.1186/s40359-022-00955-w

Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education*, *97*(1), 32–57. https://doi.org/10.1002/sce.21043

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

Looi, C. K., Liu, Z., & Sun, D. (2026). Taming the black box: Design principles for rule-integrated LLM tutoring systems in primary school mathematical problem solving. *Computers and Education: Artificial Intelligence*, *10*(March), 100586. https://doi.org/10.1016/j.caeai.2026.100586

Majid, N. A. M. N. A., Osman, K., & Yee, T. S. (2025). Integrating energy literacy into science education: a comprehensive systematic review. *International Journal of Evaluation and Research in Education*, *14*(2), 1253–1263. https://doi.org/10.11591/ijere.v14i2.31873

Martínez-Borreguero, G., Maestre-Jiménez, J., Mateos-Núñez, M., & Naranjo-Correa, F. L. (2024). Integrating Energy and Sustainability into the Educational Curriculum: A Pathway to Achieving SDGs. *Sustainability (Switzerland)*, *16*(10). https://doi.org/10.3390/su16104100

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In *Cambridge University Press* (pp. 43–71). https://doi.org/10.1017/CBO9781139547369.005

Meyer, A., & Friege, G. (2026). An atomized approach to assessing energy problem solving in physics using multidimensional item response theory. *Frontiers in Education*, *11*, 1–12. https://doi.org/10.3389/feduc.2026.1759878

Mohammad Edy Nurtamam, I Ketut Budayasa, & Agung Lukito. (2025). Analysis of students' algebraic reasoning processes in expanding mathematical expressions in nonroutine problems. *International Journal of Educational Evaluation and Policy Analysis*, *2*(4), 116–130. https://doi.org/10.62951/ijeepa.v2i4.433

Monigir, N., Tumbol, Y. W., & Monolimay, N. V. (2024). Use of the rasch model for analysis of test instruments in class v science subjects, state 38 primary school, manado. *International Journal of Information Technology and Education*, *3*(2), 1–7. https://doi.org/10.62711/ijite.v3i2.171

Moradi, E., Ghabanchi, Z., & Pishghadam, R. (2022). Reading comprehension test fairness across gender and mode of learning: insights from IRT-based differential item functioning analysis. *Language Testing in Asia*, *12*(1), 1–18. https://doi.org/10.1186/s40468-022-00192-3

Musengimana, T., Yadav, L. L., Uwamahoro, J., & Nizeyiman, G. (2025). Effect of systematic physics problem-solving strategy on secondary school students' learning achievement. *Physics Education*, *60*(3), 035032. https://doi.org/10.1088/1361-6552/adc8ed

Musengimana, T., Yadav, L. L., Uwamahoro, J., & Nizeyimana, G. (2025). Assessing physics students' problem-solving skills: a baseline investigation. *Discover Education*, *4*(1), 1–15. https://doi.org/10.1007/s44217-025-00640-1

Nilimaa, J. (2023). *New examination approach for real-world creativity and problem-solving skills in mathematics*. 477–495.

Qotrunnada, N. A. (2022). Analysis of the difficulties of high school students in improving problem solving ability in physics learning. *International Journal of Current Educational Research*, *1*(1), 84–101. https://doi.org/10.53621/ijocer.v1i1.141

Qudratuddarsi, H., Hidayat, R., Raja Maamor Shah, R. L. Z. binti, Nasir, N., Imami, M. K. W., & Mat Nor, R. bin. (2022). Rasch validation of instrument measuring gen- z science, technology, engineering, and mathematics (stem) application in teaching during the pandemic. *International Journal of Learning, Teaching and Educational Research*, *21*(6), 104–121. https://doi.org/10.26803/ijlter.21.6.7

Rahayu, S., Abdurrahman, Herlina, K., Suyatna, A., & Ertikanto, C. (2025). Analysis of teachers' needs in renewable energy learning programs using ssi integrated with pjbl-stem to enhance collaborative problem-solving and entrepreneurial skills. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 774–782. https://doi.org/10.29303/jppipa.v11i1.9299

Rehman, A., Sanjari, M. J., Elavarasan, R. M., & Jamal, T. (2026). Sustainability-aligned pathways for energy transition: A review of low-carbon energy network solutions. *Renewable and Sustainable Energy Reviews*, *226*(PE), 116428. https://doi.org/10.1016/j.rser.2025.116428

Shahat, M. A., Boone, W. J., Al-Alawi, K. A., & Al-Balushi, S. M. (2024). Rasch analysis in developing a questionnaire to measure self-efficacy beliefs of Omani preservice science teachers for teaching through engineering design. *Humanities and Social Sciences Communications*, *11*(1), 1–17. https://doi.org/10.1057/s41599-024-04087-x

Shindi, Y. A., Alkharusi, H., Shebani, Z., Kazem, A., Alsoudi, S., Maqbali, A. Al, Al-Bahrani, M., Fahdi, R. Al, Al-Omari, A. A., & Alabri, K. (2025). Rasch analysis and differential item functioning of the time management questionnaire among university students. *Journal of Posthumanism*, *5*(6), 4072–4091. https://doi.org/10.63332/joph.v5i6.2581

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, *8*(1), 33.

Stolt, M., Kottorp, A., & Suhonen, R. (2021). A Rasch analysis of the self-administered Foot Health Assessment Instrument (S-FHAI). *BMC Nursing*, *20*(1), 1–9. https://doi.org/10.1186/s12912-021-00625-z

Subramaniam, R. C., Morphew, J. W., Rebello, C. M., & Rebello, N. S. (2024). Presenting a STEM ways of thinking framework for engineering design-based physics problems. *Physical Review Physics Education Research*, *21*(1), 10122. https://doi.org/10.1103/PhysRevPhysEducRes.21.010122

Tong, D., Liu, J., Sun, Y., Liu, Q., Zhang, X., Pan, S., & Bao, L. (2023). Assessment of student knowledge integration in learning work and mechanical energy. *Physical Review Physics Education Research*, *19*(1), 10127. https://doi.org/10.1103/PhysRevPhysEducRes.19.010127

Wetzel, E., & Böhnke, J. R. (2017). Differential item functioning. In *Encyclopedia of Personality and Individual Differences*. https://doi.org/10.1007/978-3-319-28099-8_1297-1

Widya, W., Andriani, R., Mujtahid, Z., Muttakin, M., & Nurpatri, Y. (2026). Enhancing students' problem-solving skills and energy-saving character through a physics e-module. *Indonesian Journal of Science and Mathematics Education*, *9*(1), 90–102. https://doi.org/10.24042/ijsme.v9i1.28694

Worku, D. T., Ejigu, M. A., Gebremeskal, T. G., & Kassa Gogie, T. (2025). 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 and Technological Education*, *00*(00), 1–26. https://doi.org/10.1080/02635143.2025.2469062