

20-04-2026

Analysis of the difficulty level of physics cognitive test items using the Rasch approach: Implementation of analysis with R Software

Sri Agustini, Kaharuddin Arafah, Dirgah Kaso Sanusi, Salmah, Rahmat Faisal, Nuzul Fitria

To cite this article: Agustini, S., Arafah, K., Sanusi, D. K., Salmah, S., Faisal, R., & Fitria, N. (2026). Analysis of the difficulty level of physics cognitive test items using the Rasch approach: Implementation of analysis with R Software. *Priviet Social Sciences Journal*, 6(4), 433-444. <https://doi.org/10.55942/pssj.v6i4.1036>

To link to this article: <https://doi.org/10.55942/pssj.v6i4.1036>



Follow this and additional works at: <https://journal.privietlab.org/index.php/PSSJ>
Priviet Social Sciences Journal is licensed under a Creative Commons Attribution 4.0 International License.

This PSSJ: Original Article is brought to you for free and open access by Privietlab. It has been accepted for inclusion in Priviet Social Sciences Journal by an authorized editor of Privietlab Journals

Full Terms & Conditions of access and use are available at: <https://journal.privietlab.org/index.php/PSSJ/about>



Analysis of the difficulty level of physics cognitive test items using the Rasch approach: Implementation of analysis with R Software

Sri Agustini^{*}, Kaharuddin Arafah, Dirgah Kaso Sanusi, Salmah, Rahmat Faisal, Nuzul Fitria

Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar,
1Makassar, 90222, Indonesia

*e-mail: sri.agustini@unm.ac.id

Received 13 November 2025

Revised 0 February 2026

Accepted 20 April 2026

ABSTRACT

This study aims to analyze the difficulty level of Physics cognitive test items using the Rasch approach and the implementation of analysis through the application of R. The instrument consisted of 10 dichotomous items with three cognitive levels: C2 (comprehension), C3 (application), and C4 (analysis). A total of 100 students from SMAN 9 Gowa participated as respondents. The analysis was carried out using the Rasch 1PL model to obtain difficulty parameters (b), fit items and test information. The results showed that the difficulty of the item varied widely in the range of 0.12 to 2.22 logits, so that it was able to distinguish students' abilities, especially at the intermediate level, as evidenced by the peak of the Test Information Function at $\theta \approx 1$. However, some C2 items appear more difficult than the theory, which indicates the need for revisions to the redaction and effectiveness of the distractors. Items C3 and C4 were relatively consistent, although some items were classified as very difficult. The ICC analysis also showed differences in characteristics between items according to the principles of the Rasch model. Overall, the use of Rasch through R has proven to be effective in providing objective information about the quality of the items so that it can be used as a basis for improving the Physics cognitive assessment instrument to be more valid and in accordance with students' abilities.

Keywords: physics cognitive test; Rasch model; software R

priviet lab.
RESEARCH & PUBLISHING



1. INTRODUCTION

Technological assessment of students' cognitive performance in physics plays a central role in both instructional processes and educational evaluation. Physics cognitive tests are commonly employed to measure multiple dimensions of learning, including conceptual understanding, application of principles, scientific reasoning, and problem-solving ability. To produce valid and reliable measurement outcomes, the test items must comply with established psychometric standards. Among these standards, item difficulty is a key parameter, as it determines the capacity of an item to distinguish meaningfully among learners with different levels of proficiency. In educational settings, accurate information regarding item difficulty is essential for enhancing assessment quality and instructional decision-making. Knowledge of item difficulty levels allows educators and researchers to align test items with students' ability profiles, identify items that function inadequately, and refine instructional strategies based on the assessment results. However, in many contexts, analyses of physics test items remain limited in scope, with difficulty indices frequently derived from basic descriptive statistics that offer limited diagnostic value for teachers.

The evaluation of physics cognitive assessments has been largely grounded in Classical Test Theory (CTT). Although CTT provides a straightforward framework for item analysis, it is constrained by its dependence on the specific sample used for the estimation. Consequently, item difficulty parameters obtained through CTT are not invariant across populations, limiting their usefulness in comparative analyses, longitudinal evaluations, and systematic item development. These shortcomings underscore the need for more sophisticated measurement approaches in physical assessments. The Rasch model, a widely recognized model within Item Response Theory, offers a rigorous alternative for addressing these limitations. By estimating item difficulty and examinee ability on a common linear scale, the Rasch model supports invariant measurement and facilitates a clearer interpretation of assessment results. In addition, the model yields detailed diagnostic information, such as item fit statistics and person-item distribution maps, which provide deeper insights into the quality and functioning of test items beyond what is achievable using classical methods alone.

Recent advances in statistical computing have expanded access to Rasch model applications through open-source platforms such as R. A range of R packages, including *eRm*, *TAM*, *ltm*, and *psychotools*, enables comprehensive Rasch analyses, encompassing parameter estimation, model fit evaluation, visualization of item-person relationships, and examination of unidimensionality assumptions in a transparent and reproducible manner (Desjardins & Bulut, 2018). The flexibility of R further supports automated analysis workflows, particularly for large-scale and repeated assessments. Despite these methodological advancements, the adoption of Rasch-based analysis using R software in physics education remains limited to date. Many practitioners continue to rely on conventional analytic techniques, often because of limited exposure to Rasch modeling and statistical programming tools. Moreover, prior research applying the Rasch model in physics contexts has predominantly emphasized overall instrument validity, with relatively little attention devoted to the detailed empirical characterization of item difficulty or to the provision of replicable analytical procedures that can be readily implemented by educators.

To address these gaps, the present study examines the difficulty levels of physics cognitive test items using the Rasch modeling framework and demonstrates a systematic implementation of the analysis using the R software. This study contributes to the existing literature by providing empirical evidence on the difficulty characteristics of physics assessment items and offering practical methodological guidance to support the development of more valid, reliable, and equitable measurement instruments in physics education.

2. LITERATURE REVIEW

2.1. Rasch Model as a Measurement Framework

The Rasch model is one of the basic forms of Item Response Theory, which models the probability of correct answers as a function of the difference between the participant's ability and the difficulty of the item on the logit scale (Rasch, 1960). The advantage of the Rasch model is its ability to separate the participant parameter (ability) and the item parameter (difficulty) on a single interval scale, resulting in a more metrically meaningful measure than classical statistics. This model allows for the examination of the suitability of each item to the model (item fit) and the creation of a person-item map (Wright map) that visualizes the distribution of abilities and the location of the item at the same scale. The item fit model allows for the identification and correction of items that do not match the model assumptions (Tennant et al., 2023). Meanwhile, the Wright map can provide an understanding of the participants' ability to the difficulty of the item on a continuum of the same scale and assess whether the instrument covers a well-measured range of abilities (Wright & Masters, 1982).

Research in the field of science education, including physics, has utilized the Rasch model to (a) determine the difficulty level of concept items, (b) identify misfitting items, and (c) map the applicability of instruments to student populations (Boone et al., 2014; Bond & Fox, 2015; Sumintono & Widhiarso, 2014). International studies have shown that items containing contextual context often have a higher level of difficulty than items that are purely mathematically inherent. Planinic et al. (2019) found that context-rich problems in physics assessments require the integration of conceptual knowledge and complex problem-solving skills; therefore, the level of difficulty increases significantly. Rasch's analysis in these studies helped test developers identify imbalances between content and context, as well as adjust the level of difficulty so that the instrument was able to optimally measure the range of students' abilities. Rasch analysis using software such as Winsteps, Ministep, or R eRm can provide in-depth diagnostic information, such as item difficulty distribution, person reliability, item fit index, and unidimensionality check. These findings confirm that the Rasch model plays an important role in improving the quality of instruments for measuring physics concepts while ensuring that the instruments developed are in line with the cognitive characteristics of Indonesian students (Sumintono, 2021).

2.2. Cognitive Ability Physics

Cognitive ability in the context of physics learning refers to a student's capacity to understand physics concepts, process scientific information, apply principles to new situations, analyze phenomena, solve problems, and rationally assess evidence or arguments. This aspect includes understanding scientific principles, logical and analytical thinking skills, and the application of knowledge to solve problems and evaluate scientific arguments (Bybee, 2013; Treagust et al., 2017). Cognitive frameworks often use cognitive taxonomy (for example, Bloom/Revised Bloom) or a science-specific cognitive processing framework to describe levels from basic knowledge to high-level thinking (analysis, synthesis, evaluation). The emphasis on cognitive abilities in physics is important because the nature of the subject is conceptual and abstract and often requires the transformation of representations (symbolic, mathematical, and graphic). In the context of physics, various studies have adapted this framework to assess scientific thinking skills, especially at the medium to high level, involving the analysis of relationships between physical variables, the formulation of mathematical models, and evidence-based reasoning.

There are five main components of cognitive ability in physics. First, concept understanding. In-depth understanding of physics concepts (e.g., force, energy, field) allows students to connect theoretical models with real phenomena (Redish, 2003). The literature shows that conceptual understanding is different from the ability to solve mathematical problems, students can solve procedural problems without actually mastering basic concepts. Second, problem solving. Problem solving in Physics involves the process of problem recognition, modeling, selection of mathematical strategies, and verification of results (Docktor & Mestre, 2019). Classical and modern research emphasizes the knowledge structures and metacognitive strategies that distinguish expert and novice students. Third, Higher-Order Thinking

Skills (HOTS). Analysis, evaluation, and the ability to synthesize scientific information are categorized as HOTS (King et al., 2020). In physics, HOTS is related to the ability to design experiments, evaluate model assumptions and predict the consequences of changing conditions. Systematic studies have found increased attention to the development of HOTS through active learning models. Fourth, metacognition and self-regulation. Awareness and regulation of thinking processes (e.g., strategy planning, monitoring, solution evaluation) have been proven to improve physics learning outcomes and problem-solving quality. Interventions that foster metacognition (prompt reflection, think-aloud, self-assessment rubrics) improve knowledge retention and transfer (Zohar & Barzilai, 2013). Fifth, scientific argumentation and reasoning skills. The skill of structuring, evaluating, and defending evidence-based arguments is an important cognitive element, especially in the field of argumentation.

3. RESEARCH METHOD

This study used a descriptive quantitative design focusing on the psychometric analysis of items using the Rasch model. The respondents were 100 students from SMAN 9 Gowa who were selected purposively. This number is considered adequate for the stable estimation of the dichotomous Rasch model, as recommended by Linacre (2023) and Boone et al. (2014), who suggested a sample size of 30–200 respondents for the basic Rasch model. The research instrument in the form of a physics cognitive test consisted of 10 dichotomous items (correct answer = 1, false = 0) that were distributed according to the cognitive level: three items at the C2 level (comprehension), four items at the C3 level (application), and three items at the C4 level (analysis). The items were arranged based on a grid that listed competency indicators, learning objectives, and cognitive levels according to the taxonomy used. Prior to administration, the items were tested for language feasibility and processing time.

Data collection was carried out in classroom situations with a uniform duration for each participant, and instructions were given on a standard basis. The collected response data were compiled in a matrix format (row = participant; column = item; value 0/1). The main analysis uses a dichotomous Rasch model estimated in the environment R. The recommended package plan is eRm for the Rasch model fitting RM function. The analysis steps include: (1) fitting the Rasch model to the response matrix to obtain an estimate of the item parameter (difficulty) and the participant parameter (ability) at the logit scale; (2) checking fit items using infit and outfit statistics; (3) the creation of a person-point map (Wright map) to visualize the distribution of participants' abilities to the difficulty position of the item so that it can be seen whether the distribution of items is in accordance with the range of students' abilities; (4) estimation of the reliability of instruments in the Rasch framework to assess the ability of the test to differentiate the level of participants

The results of the analysis will be reported in the form of a table of item parameters (logit difficulty value, standard error), fit statistics for each item, Rasch reliability value, Wright map, and supporting graphs. Interpretations include identification of items that are too easy or too difficult, items that are misfit, and recommendations for improvement (language, context, or content), as well as suggestions for rearranging the grid so that the distribution of difficulty covers the student's range of abilities proportionally (Planinic et al., 2019).

4. RESULTS AND DISCUSSION

The results of the analysis of the Rasch Model Item Parameter Analysis (1PL) for 10 items are shown in Table 1.

Table 1. Results of Parameter Analysis of Rasch Model Item (1PL)

Item	a_true	b_true	a_est	b_est
Item 1	1	-2.000000	1	0.121943
Item 2	1	-1.555556	1	0.244749
Item 3	1	-1.111111	1	0.583708
Item 4	1	-0.666667	1	2.015205
Item 5	1	-0.222222	1	2.221225
Item 6	1	0.222222	1	0.764502
Item 7	1	0.666667	1	0.539965
Item 8	1	1.111111	1	0.583708
Item 9	1	1.555556	1	1.404221
Item 10	1	2.000000	1	2.114128

Table 1 shows the results of the estimation of the item parameters using the Rasch Model (1 Logistic Parameter/1PL). This model assumes that each item has the same level of differentiating power (a), whereas the difficulty level (b) can differ for each item. Rasch analysis is increasingly emphasized in physics education research because it provides invariant measurements and strengthens the interpretability of item difficulty beyond classical indices (Planinic et al., 2019). In addition, Rasch modeling provides deeper diagnostic evidence, such as item fit statistics, person reliability, and Wright maps, that support more robust instrument refinement (Juandi et al., 2024). Therefore, all a_true and a_est values were set to 1, indicating that all items had the same discriminating ability according to the basic assumptions of the Rasch model. The b_true value is the difficulty parameter of the item determined based on the initial data, while b_est is the result of the model's estimation based on the participants' responses. The b_est value reflects the empirical difficulty of each question item after the calibration process is conducted (Sumintono & Widhiarso, 2014).

Based on the results of the estimation of the difficulty parameter of the item (b_est), a range of values between 0.121943 and 2.221225 was obtained. This range shows that the difficulty level of the question items in the test varied widely, ranging from relatively easy to very difficult. Item 1 (b_est = 0.121943) is the easiest item in the test because it has the lowest difficulty value, while Item 5 (b_est = 2.221225) is the hardest item with the highest difficulty compared to the other items. Some items, such as Item 3 (b_est = 0.583708) and Item 7 (b_est = 0.539965), were at a moderate difficulty level, so that they were able to optimally distinguish participants with intermediate abilities. Meanwhile, Item 9 (b_est = 1.404221) and Item 10 (b_est = 2.114128) fall into the difficult-to-very-difficult category, suggesting that the items can only be answered correctly by participants with high abilities (Linacre, 2023).

When compared to the cognitive level of each question item, the results of the analysis show that there is a mismatch between the empirical difficulty level generated through the Rasch model and the theoretical difficulty level based on Bloom's taxonomy. Empirical Rasch difficulty does not always align with Bloom's taxonomy expectations, as comprehension-level items may become difficult because of wording and contextual demands (Planinic et al., 2019). Moreover, context-rich science items demand deeper integration of reasoning, leading to higher empirical difficulty levels when calibrated using the many-facet Rasch measurement (Chi et al., 2021). In the C2 level item (comprehension) consisting of Items 1, 2, and 7, these items should be relatively easy because they only require the ability of participants to understand basic concepts. However, the results of the Rasch analysis showed that the three items had a positive b_est value (0.12–0.54), which indicates that these questions are quite difficult for students. This indicates that the C2 grain was not functioning as expected. Possible causes include the redaction of questions that are too long or complicated, contexts or keywords that are unfamiliar to students, or answer options that are too similar to be confusing. Thus, the C2 items need to be revised to measure comprehension skills, not higher reasoning or analytical skills. In this regard, metacognitive scaffolding has been shown to improve students' learning performance and problem-solving outcomes, thereby supporting higher order cognitive test achievement (Stanton et al., 2021).

Meanwhile, items with a C3 level (implementation), including Items 3, 8, 9, and 10, showed results that were more consistent with the theory. Conceptually, questions at the C3 level require a medium to

high ability to apply the concepts that have been learned. The results of the Rasch analysis showed that the b_{est} values for these items ranged from 0.58 to 2.11, indicating a logical and graded difficulty gradation from medium to difficult. This indicates that items at the C3 level function well in distinguishing participants based on their ability level, according to the construct validity principle (Boone et al., 2014). Only Item 8 seems a little easier than expected, likely due to an overly simple problem structure or the existence of a less effective distractor.

As for the items at the C4 level (analysis), consisting of items 4, 5, and 6, in theory, they are high cognitive levels, so they are expected to have a relatively high level of difficulty. The results of the Rasch analysis showed that the b_{est} values for the three items were in the range of 0.76 to 2.22, in accordance with theoretical expectations. However, Items 4 and 5 had very high b -values (>2), indicating the potential for an overdemanding cognitive load. This indicates that the two items tend to be too complex or disproportionate to the students' level of ability. However, in general, the items at the C4 level have worked in accordance with the theory, although some of them still need to be reviewed so that the level of difficulty is not too extreme and remains in accordance with the purpose of measuring the participants' analytical abilities.

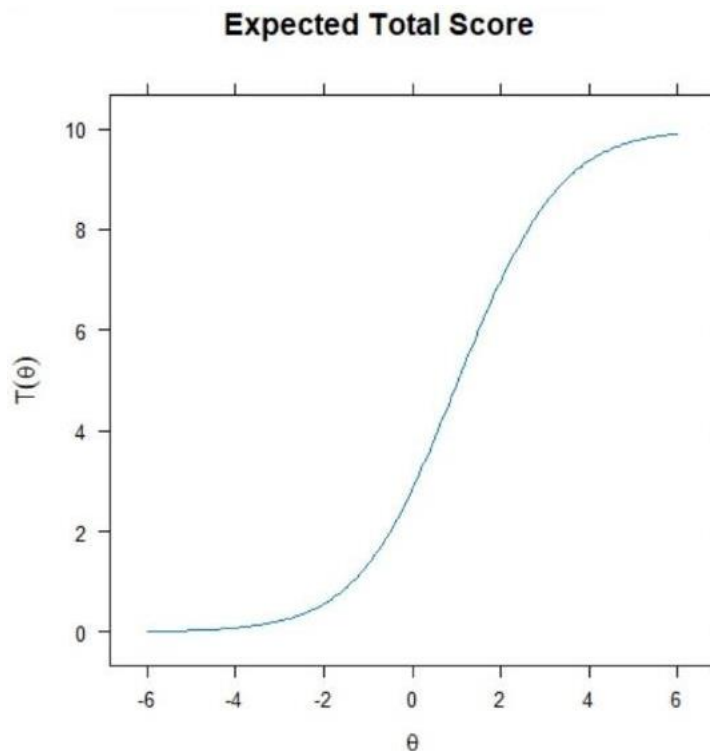


Figure 1. Expected Total Score

The graph above shows the relationship between the participant's ability (θ) and the expected total score ($T(\theta)$) based on the results of the Rasch model analysis of the 10 questions. The "S" curve indicates that the higher the participant's ability, the higher their expected total score. Participants with low abilities receive low scores, while participants with high abilities receive scores close to the maximum. The proportional shape of the curve around $\theta = 0$ indicates that the test has a good differentiating power in moderately capable participants.

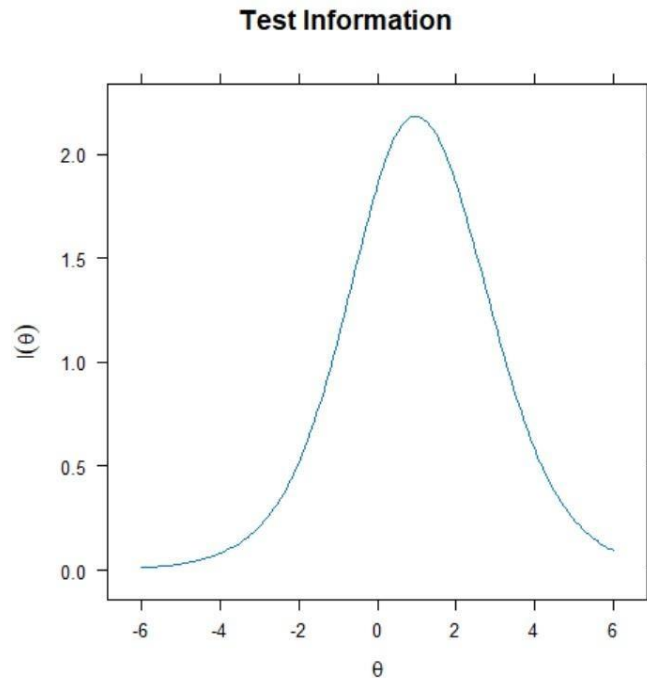


Figure 2. Test Information Function

The graph above shows the Test Information Function results of the Rasch model analysis of the 10 questions. The peak of the curve is around $\theta = 1$, which means that the test provides the highest or most accurate information in measuring the ability of participants at moderate to somewhat high levels. The decreased information value on both sides indicates that the measurement accuracy is reduced for participants with very low and very high abilities. Thus, the test is most effective for measuring participants with intermediate ability.

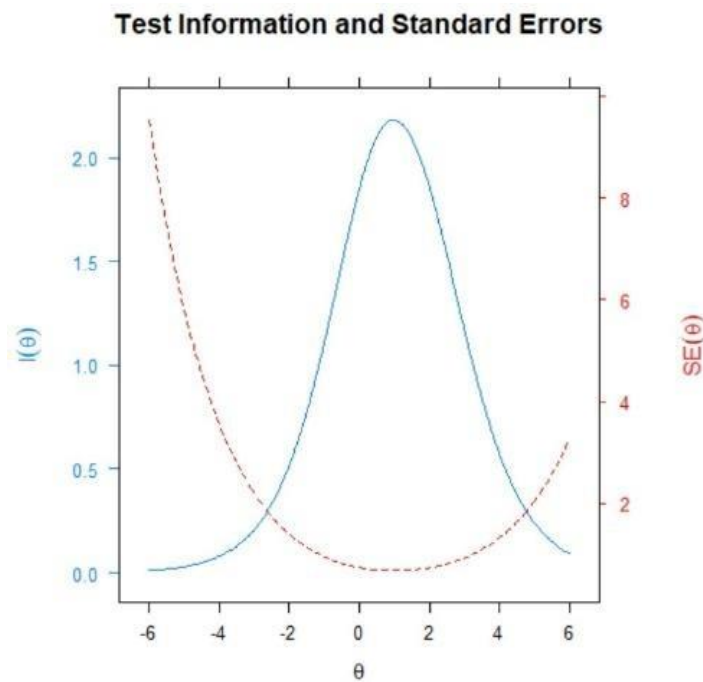


Figure 3. Test Information Function (blue line) and Standard Error of Measurement (dotted red line)

The graph above shows the relationship between the Test Information Function (blue line) and the Standard Error of Measurement (dotted red line) in the results of the Rasch model analysis for the 10 questions. The test information ($I(\theta)$) peaked at approximately $\theta = 1$, indicating that the test was most accurate in measuring participants with moderate to somewhat high ability. In contrast, the Standard Error ($SE(\theta)$) was at its lowest point in the same area, indicating the smallest measurement error rate. Error values increase at very low and very high ability levels; therefore, this test is most effectively used to measure participants with moderate ability.

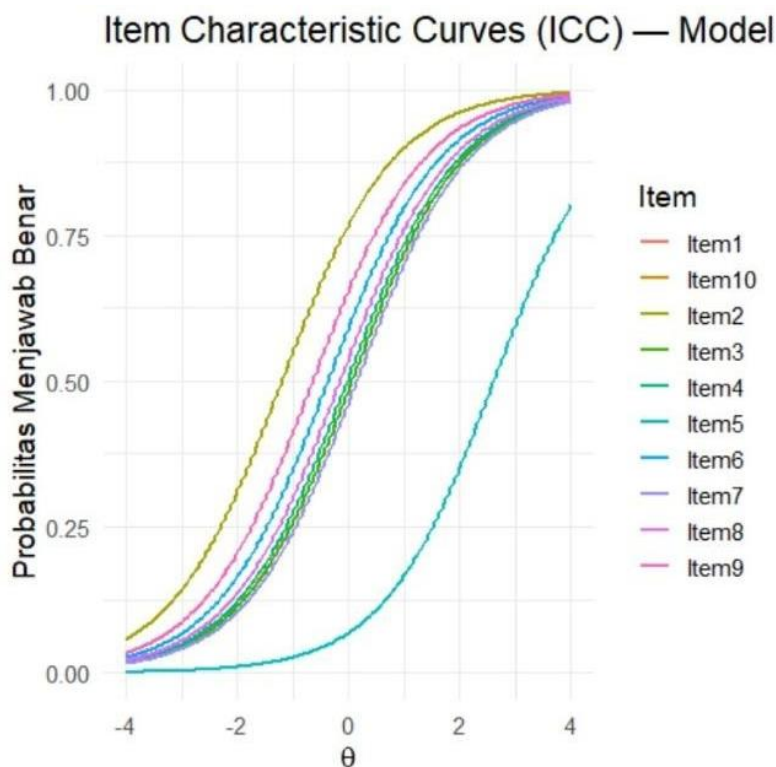


Figure 4. Item Characteristic Curves (ICC)

The graph above shows the Item Characteristic Curves (ICC) for the 10 questions based on the results of the Rasch model analysis. Each curve represents the relationship between the participant's ability (θ) and the probability of answering correctly for each item. The entire curve had a sigmoid shape (uphill from left to right), which indicates that the higher the ability of the participants, the greater the chance of answering correctly. The difference in the position of the curve along the θ -axis indicates the variation in difficulty between items: curves shifting to the right indicate more difficult items (such as Items 4, 5, and 10), while curves on the left side indicate easier items (such as Items 1 and 2). Overall, this graph shows that each item has different characteristics but is consistent with the principles of the Rasch model, namely, an increase in ability followed by an increase in the probability of answering correctly (Bond & Fox, 2015; Linacre, 2023).

The present study applied the Rasch model to examine the difficulty characteristics of physics cognitive test items and evaluate how well the instrument functions across different levels of student ability. The results indicated that the estimated item difficulty parameters ranged from 0.12 to 2.22 logits, demonstrating a relatively wide spread of difficulty levels. This variation suggests that, in general, the test can distinguish students with different levels of cognitive ability, particularly those within the intermediate ability range. The distribution of item difficulty is further clarified through the Expected Total Score curve (see Figure 1). The sigmoidal shape of the curve reflects the probabilistic nature of the Rasch model, showing that increases in student ability (θ) are associated with higher expected total scores. The steepest portion of the curve occurs around $\theta \approx 0$, indicating that the test has an optimal discriminative power for students with moderate ability. This finding is consistent with the fundamental

principle of Rasch measurement, which emphasizes maximum measurement precision, where item difficulty aligns closely with examinee ability.

Additional evidence regarding the effectiveness of the test is provided by the Test Information Function (see Figure 2). The peak of the information curve around $\theta \approx 1$ indicates that the instrument yields the most precise measurement for students with moderate to moderately high ability levels. This result implies that the test is particularly well-suited for diagnosing learning outcomes among students who have achieved a basic conceptual understanding but are still developing higher order reasoning skills. Conversely, the decline in test information at very low and very high ability levels suggests reduced measurement precision for students at the extremes, which is a common characteristic of instruments with a limited number of items. The relationship between test information and measurement error is more explicitly illustrated in Figure 3, which presents both the Test Information Function and the Standard Error of Measurement. As expected under Rasch theory, the lowest standard error coincides with the peak of the test information, again around $\theta \approx 1$. This inverse relationship confirms that the instrument provides its most reliable estimates of ability in the mid-range of the ability continuum, while the measurement error increases for students with very low or very high abilities. From a practical perspective, this finding suggests that the test is most appropriate for formative or diagnostic purposes among average-achieving students, while caution should be exercised when interpreting the results for students at the extremes of performance.

When item difficulty estimates were examined in relation to cognitive levels, several important patterns emerged. Items categorized at the C2 level (comprehension) were expected to be relatively easy. However, Rasch analysis revealed that these items exhibited positive difficulty values, indicating that they were more challenging for students than anticipated. This mismatch between the theoretical cognitive level and empirical difficulty suggests that these items may contain linguistic complexity, unfamiliar contexts, or distractors that inadvertently increase cognitive demand. These findings underscore the value of Rasch analysis in identifying items that do not function as intended and require revision to better align with their targeted cognitive constructs. In contrast, items at the C3 level (application) generally demonstrated difficulty levels consistent with theoretical expectations. The gradual increase in difficulty across these items indicates that they effectively differentiate students based on their ability to apply physics concepts in problem-solving situations. This alignment supports the construct validity of the C3 items and suggests that they are well designed to capture intermediate-to-higher-order cognitive processes.

Items at the C4 level (analysis) exhibited the highest difficulty estimates, as expected for higher order cognitive tasks. Nevertheless, the two items showed very high difficulty values, exceeding 2 logits. Such extreme difficulty may indicate an excessive cognitive load or overly complex problem structures that surpass the intended measurement purpose. While these items are effective in identifying high-ability students, they may provide limited information for the majority of test-takers. Consequently, revising these items to moderate their difficulty could improve the overall balance and measurement efficiency of the tests. The Item Characteristic Curves (see Figure 4) further support the internal consistency of the instrument. All items displayed the expected sigmoid shape, indicating that the probability of a correct response increased monotonically with student ability. Differences in the horizontal positioning of the curves reflect variations in item difficulty, with easier items located to the left and more difficult items shifted to the right along the ability scale. This pattern confirms that, despite differences in difficulty, the items conform to the assumptions of the Rasch model and function coherently within a unidimensional measurement framework.

Overall, the integration of graphical outputs and parameter estimates demonstrates that the Rasch-based analysis provides a comprehensive and theoretically grounded understanding of item functioning. By combining numerical difficulty indices with visual diagnostics such as information functions and item characteristic curves, this study highlights how Rasch modeling can inform both psychometric evaluation and pedagogical decision making. The findings emphasize the importance of aligning theoretical cognitive levels with empirical item difficulty and illustrate how Rasch analysis using R software can

support the development of physics assessment instruments that are more valid, reliable, and diagnostically useful.

5. CONCLUSION

Analysis of the 10 items of the physics cognitive test using the Rasch Model showed that the difficulty level of the items varied quite widely (0.12–2.22 logit), and the test most accurately measured students with moderate ability, marked by the peak of the Test Information Function at $\theta \approx 1$. However, some C2 level items appear more difficult than expected and required editorial or distractor revisions, while C3 and C4 level items were relatively consistent with theory, although some were too difficult. In general, the application of Rasch through R succeeds in providing an objective picture of the quality of the item and can be the basis for improving the Physics cognitive assessment instrument to be more valid and in accordance with students' abilities.

Ethical Approval

Not Applicable

Informed Consent Statement

Not Applicable

Contributions of Authors

SA conceptualized the study, designed the instrument, conducted the data collection, performed the Rasch analysis using R software, and prepared the original draft of the manuscript. KA, DKS, and S contributed to the development of the analytical framework, interpretation of the results, and validation of the findings related to item difficulty and fit statistics. RF and NF supported the literature review, data organization, and manuscript revision. All authors contributed to the discussion of the results, critically reviewed the manuscript, and approved the final version.

Disclosure Statement

The authors declare that there are no potential conflicts of interest related to this study.

Data Availability Statement

The data presented in this study are available on request from the corresponding author due to privacy reasons.

Funding

This study did not receive any external funding.

Notes on Contributors

Sri Agustini

<https://orcid.org/0009-0004-4481-8178>

Sri Agustini is affiliated with Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar.

Kaharuddin Arafah

<https://orcid.org/0000-0002-1847-981X>

Kaharuddin Arafah is affiliated with Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar.

Dirgah Kaso Sanusi

<https://orcid.org/0009-0002-4015-4373>

Dirgah Kaso Sanusi is affiliated with Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar.

Salmah

Salmah is affiliated with Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar.

Rahmat Faisal

Rahmat Faisal is affiliated with Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar.

Nuzul Fitria

Nuzul Fitria is affiliated with Department of Physics, Faculty of Mathematic and Natural Science, Universitas Negeri Makassar.

REFERENCES

- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315814698>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Bybee, R. W. (2013). *The case for STEM education: Challenges and opportunities*. NSTA Press. <https://books.google.com/books?id=gfn4AAAAQBAJ>
- Chi, S., Liu, X., & Wang, Z. (2021). Comparing student science performance between hands-on and traditional item types: A many-facet Rasch analysis. *Studies in Educational Evaluation*, 70, 100998. <https://doi.org/10.1016/j.stueduc.2021.100998>
- Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R*. CRC Press. <https://doi.org/10.1201/b20498>
- Docktor, J. L., & Mestre, J. P. (2019). *The Psychology of Learning and Teaching Physics*. Routledge.
- Juandi, T., Kaniawati, I., Samsudin, A., & Riza, L. S. (2024). The Application of Rasch Model to Analyse the Validity and Reliability of an Instrument for Reflective Thinking Skills on Topic of Wave-Particle Dualism. *Kappa Journal*, 8(2), 1-8. <https://doi.org/10.29408/kpj.v8i2.27049>
- King, F. J., Goodson, L., & Rohani, F. (2020). *Higher Order Thinking Skills: Definition, Teaching Strategies, Assessment*. Center for Advancement of Learning.
- Linacre, J. M. (2023). *A User's Guide to WINSTEPS/MINISTEP: Rasch-Model Computer Programs*. <https://www.winsteps.com/a/Winsteps-Manual.pdf>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 020111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut.
- Redish, E. F. (2003). *Teaching physics with the physics suite*. John Wiley & Sons. <https://www.wiley.com/en-us/Teaching+Physics+with+the+Physics+Suite-p-9780471393788>

- Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE—Life Sciences Education*, 20(2), fe3. <https://doi.org/10.1187/cbe.20-12-0289>
- Sumintono, B. (2021). *Model Rasch untuk Penelitian Pendidikan: Prinsip Dasar dan Aplikasinya dengan R dan Winsteps (The Rasch Model for Educational Research: Basic Principles and Applications with R and Winsteps)*. Universitas Negeri Jakarta Press.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (Application of the Rasch model for social science research)* (Edisi revisi [Revised ed.]). Trim Komunikata Publishing House. <https://eprints.um.edu.my/11413/>
- Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: Early developments, current practice, and future challenges. *Frontiers in Rehabilitation Sciences*, 4, 1208670. <https://doi.org/10.3389/fresc.2023.1208670>
- Treagust, D. F., Duit, R., & Fischer, H. E. (Eds.). (2017). *Multiple representations in physics education*. Springer. <https://doi.org/10.1007/978-3-319-58914-5>
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press. <https://eric.ed.gov/?id=ED436551>
- Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. *Studies in Science Education*, 49(2), 121–169. <https://doi.org/10.1080/03057267.2013.847261>