

Development of a Testlet Based Two-Tier Diagnostic Test for Identifying Students' Problem Solving Difficulties in Newton's Laws

Salwa Atiqah Huriyah*, Haratua Tiur Maria Silitonga, Naim Sulaiman

Physics Education, Faculty of Teacher Training and Education, Tanjungpura University, Pontianak, Indonesia

*Correspondence to: satiqahhuriyah@gmail.com

Abstract: This study aimed to develop a testlet-based two-tier diagnostic test to identify students' problem-solving difficulties in Newton's Laws and to describe the profile of difficulties experienced by students based on Heller's problem-solving stages. The study employed a Research and Development (R&D) method using the 4D model, consisting of the Define, Design, Develop, and Disseminate stages. The developed instrument was a two-tier diagnostic test comprising 20 items covering Newton's First Law, Newton's Second Law, Newton's Third Law, and applications of Newton's Laws. The instrument was administered to senior high school students, and the data were analyzed using the Rasch Model. The results indicated that the developed instrument met the validity and reliability criteria, with a person reliability value of 0.71, an item reliability value of 0.95, and a Cronbach's Alpha value of 0.73. The item difficulty analysis showed that 30% of the items were categorized as very easy, 30% as easy, 15% as difficult, and 25% as very difficult. The analysis of students' problem-solving difficulties revealed that the highest difficulty occurred at the check and evaluate stage (75%), followed by the plan a solution stage (60%), the visualize the problem stage (58%), the execute the plan stage (56%), and the describe the problem in physics terms stage (36%). These findings indicate that students still experience considerable difficulties in evaluating problem solutions and planning appropriate solution strategies. Therefore, the developed two-tier diagnostic test is suitable for identifying students' problem-solving difficulties in Newton's Laws.

Keywords: Learning Difficulties, Newton's Laws, Problem Solving, Rasch Model, Two-Tier Diagnostic Test

Introduction

Problem-solving ability is one of the essential competencies that students must possess in 21st-century physics education. This ability is not only related to conceptual understanding but also encompasses the ability to identify problems, plan solution strategies, apply physics concepts, and evaluate the obtained results. In physics learning, problem-solving serves as an important indicator of learning success because physics requires not only theoretical understanding but also the ability to apply concepts to solve various contextual problems (Docktor & Mestre, 2014; Hutapea et al., 2020). However, numerous studies have reported that students' problem-solving abilities remain relatively low, particularly in Newton's Laws, which constitute one of the fundamental topics in mechanics (Amelia, 2024; Azizah & Yuliati, 2025; Ulviani et al., 2025). This limitation is reflected in students' difficulties in understanding problem information, relating concepts to physical phenomena, determining appropriate solution strategies, and evaluating their answers. These findings indicate that students still encounter obstacles in their thinking processes when solving physics problems.

Previous studies have shown that students' problem-solving difficulties are often caused by conceptual misconceptions, procedural errors, and weak metacognitive abilities in evaluating obtained solutions (Rosengrant et al., 2009; Setyono et al., 2016). To accurately identify these difficulties, assessment instruments are needed that not only measure students' final answers but also reveal the underlying reasoning processes. Conventional one-tier multiple-choice tests commonly used by teachers provide information only about whether an answer is correct or incorrect without explaining the reasoning behind it (Gurel et al., 2015). Consequently, teachers face difficulties in identifying students' specific thinking errors.

One alternative assessment instrument that can address this issue is the two-tier diagnostic test (Gurel et al., 2015; Tuysuz, 2009). A two-tier diagnostic test consists of two levels of questions: the first tier

requires students to answer a given problem, while the second tier requires them to provide the reasoning underlying their answer. This format enables teachers to obtain deeper information regarding students' conceptual understanding while simultaneously identifying difficulties occurring at each stage of problem solving. Several studies have reported that two-tier diagnostic tests are effective for identifying misconceptions and students' reasoning patterns in various physics topics (Auliya et al., 2025; Dirman et al., 2022). Nevertheless, the development of two-tier instruments specifically designed to identify students' problem-solving difficulties in Newton's Laws remains limited (Istiyono et al., 2023).

Furthermore, previous research has generally focused on identifying conceptual misconceptions without systematically relating them to problem-solving stages (Meiliyadi et al., 2025). According to Heller's problem-solving model, the process of solving physics problems involves five main stages: visualizing the problem, describing the problem in physics terms, planning a solution, executing the plan, and checking and evaluating the results (Heller et al., 1992). Information regarding students' difficulties at each of these stages is crucial for helping teachers design more targeted instructional strategies. Therefore, a diagnostic instrument capable of mapping students' problem-solving difficulties based on these thinking stages is required.

Based on the foregoing discussion, there remains a research gap concerning the limited availability of diagnostic instruments capable of comprehensively identifying students' problem-solving difficulties in Newton's Laws. The novelty of this study lies in the development of a testlet-based two-tier diagnostic test constructed according to Heller's problem-solving stages to identify students' difficulties in solving physics problems. The developed instrument not only measures the accuracy of students' answers but also investigates the conceptual reasoning underlying those answers, thereby providing deeper information regarding students' profiles of problem-solving difficulties.

Therefore, this study aims to develop and analyze the characteristics of a testlet-based two-tier diagnostic test for identifying students' problem-solving difficulties in Newton's Laws and to describe students' profiles of problem-solving difficulties based on the results obtained from the developed instrument.

Methods

This study employed a Research and Development (R&D) approach using the 4D development model, which consists of four stages: Define, Design, Develop, and Disseminate (Okpatrioka, 2023; Thiagarajan et al., 1974). This model was selected because it provides a systematic procedure for developing assessment instruments, from needs analysis to the dissemination of the final product.

During the define stage, a needs analysis was conducted through curriculum review, literature study, and identification of students' difficulties in Newton's Laws. This stage resulted in indicators of problem-solving difficulties based on Heller's five problem-solving stages: visualize the problem, describe the problem in physics terms, plan a solution, execute the plan, and check and evaluate.

During the design stage, an instrument blueprint was developed, the format of the testlet-based two-tier diagnostic test was designed, and test items were constructed based on the established indicators. The developed instrument consisted of 20 items covering Newton's First Law, Newton's Second Law, Newton's Third Law, and applications of Newton's Laws in everyday life.

The develop stage included expert validation, small-scale testing, instrument revision, and large-scale field testing. Validation was conducted by experts in physics content, educational evaluation, and language (Arikunto, 2021; Azwar, 2022). The small-scale trial involved 12 students to examine the readability and preliminary quality of the instrument, whereas the large-scale trial involved 188 senior high school students in Pontianak City. The collected data were analyzed using the Rasch Model to determine validity, reliability, item difficulty, discrimination power, and distractor effectiveness (Bond & Fox, 2013; Boone et al.,

2014). Data analysis was performed using Winsteps software (Linacre, 2012), and the interpretation of results followed the Rasch Measurement approach (Sumintono & Widhiarso, 2015).

The disseminate stage was conducted through the publication of research findings, enabling the developed instrument, which met the feasibility criteria, to be utilized in physics learning and future research.

Results and Discussion

Characteristics of the Two-Tier Diagnostic Test

This study produced a testlet-based two-tier diagnostic test designed to identify students' problem-solving difficulties in Newton's Laws. The developed instrument consisted of 20 items covering Newton's First Law, Newton's Second Law, Newton's Third Law, and applications of Newton's Laws in everyday life. Each item was constructed based on Heller's problem-solving stages, enabling the instrument to identify students' difficulties more specifically at each stage of the problem-solving process.

Based on the Rasch Model analysis, the developed instrument satisfied the validity and reliability criteria (Bond & Fox, 2013; Boone et al., 2014). All 20 items were found to be valid and suitable for measuring students' abilities. The reliability analysis showed a person reliability value of 0.71, which falls into the adequate category, while the item reliability value of 0.95 was categorized as excellent. Furthermore, the Cronbach's Alpha value of 0.73 indicated good internal consistency. A summary of the instrument characteristics is presented in Table 1.

Table 1. Characteristics of the Two-Tier Diagnostic Test

Aspect	Result	Category
Valid Items	20	Valid
Person Reliability	0.71	Adequate
Item Reliability	0.95	Excellent
Person Separation	1.55	Adequate
Item Separation	4.35	Excellent
Cronbach's Alpha	0.73	Good

These results indicate that the instrument possesses good measurement quality (Soeharto, 2021). The high item reliability value suggests that the item difficulty levels were well distributed, allowing the instrument to measure students across different ability levels. In addition, the obtained separation indices demonstrate that the instrument is capable of distinguishing students according to their ability levels.

The item difficulty analysis revealed a relatively balanced distribution of item difficulty levels. Most items were classified as easy to very easy, while several items were categorized as difficult and very difficult. Such a distribution is essential in a diagnostic instrument because it enables the identification of students' difficulties across a wide range of ability levels.

Table 2. Distribution of Item Difficulty Levels

Category	Number of Items	Percentage
Very Easy	6	30%
Easy	6	30%
Difficult	3	15%
Very Difficult	5	25%
Total	20	100%

The diverse distribution of item difficulty levels indicates that the instrument is capable of identifying not only students who experience difficulties with basic concepts but also those who struggle with items requiring higher-order analysis and application of concepts. Therefore, the developed instrument is considered suitable for identifying students' problem-solving difficulties in Newton's Laws.

Students' Problem-Solving Difficulties Based on Heller's Stages

The analysis of problem-solving difficulties was conducted by mapping each test item onto Heller's five problem-solving stages: visualize the problem, describe the problem in physics terms, plan a solution,

execute the plan, and check and evaluate. The level of difficulty was calculated based on the average percentage of students experiencing conceptual difficulties, procedural difficulties, and a lack of understanding at each stage.

The results revealed variations in students' difficulties across the different problem-solving stages. A summary of the findings is presented in Table 3.

Table 3. Analysis of Problem-Solving Difficulties Based on Heller's Stages

Problem-Solving Stage	Average Correct Response	Average Difficulty	Difficulty Level
Visualize the Problem	42%	58%	Moderate
Describe the Problem in Physics Terms	64%	36%	Low
Plan a Solution	40%	60%	High
Execute the Plan	44%	56%	Moderate
Check and Evaluate	25%	75%	Very High

As shown in Table 3, the highest level of difficulty was found in the check and evaluate stage, with a difficulty percentage of 75%. This finding indicates that most students are not accustomed to evaluating the solutions they obtain (Heller et al., 1992; Setyono et al., 2016). Students tend to accept calculation results without rechecking units, assessing the logical consistency of their answers, or verifying the appropriateness of the physics concepts applied.

A high level of difficulty was also observed in the plan a solution stage, with a percentage of 60%. This result suggests that students encounter obstacles in determining appropriate solution strategies and selecting relevant physics principles or laws (Reif & Heller, 1982). Many students are able to identify the information provided in a problem but are unable to connect that information to suitable solution procedures.

In the visualize the problem stage, the difficulty percentage reached 58%. This difficulty was reflected in students' inability to represent problems using diagrams (Rosengrant et al., 2009), identify relevant objects, and determine appropriate physical conditions. Poor visualization skills may consequently hinder performance in subsequent stages of problem solving.

The execute the plan stage showed a difficulty percentage of 56%. Errors at this stage generally involved incorrect substitution of values into equations, mathematical operation errors (Halliday et al., 2014; Serway et al., 2016), and inaccuracies during calculations. Nevertheless, the level of difficulty at this stage was lower than those observed in the planning and evaluation stages.

The lowest level of difficulty was found in the describe the problem in physics terms stage, with a difficulty percentage of 36%. This finding indicates that most students were able to identify the relevant physics concepts associated with the given problems. However, some students still experienced difficulties in determining physical quantities and constructing accurate free-body diagrams.

Profile of Students' Problem-Solving Difficulties

Based on the analysis of each problem-solving stage, a profile of students' problem-solving difficulties was obtained, as presented in Table 4.

Table 3. Profile of Students' Problem-Solving Difficulties

Rank	Heller's Stage	Type of Difficulty	Percentage
1	Check and Evaluate	Evaluation Difficulty	75%
2	Plan a Solution	Planning Difficulty	60%
3	Visualize the Problem	Visualization Difficulty	58%
4	Execute the Plan	Execution Difficulty	56%
5	Describe the Problem in Physics Terms	Description Difficulty	36%

The findings indicate that the most dominant difficulty experienced by students was evaluation difficulty in the check and evaluate stage, with a percentage of 75%. This difficulty was reflected in students' limited ability to verify units, evaluate the reasonableness of solutions, and review the procedures used to solve problems. These findings suggest that students' metacognitive skills in physics learning still require improvement.

The second most prominent difficulty was planning difficulty in the plan a solution stage, with a percentage of 60%. Students experienced challenges in determining appropriate solution strategies and selecting the correct Newtonian principles to solve problems. This condition indicates that students have not yet fully integrated conceptual understanding with the problem-solving process.

In addition, students experienced visualization difficulties (58%) and execution difficulties (56%). These difficulties indicate that students still encounter obstacles in representing physics problems and applying mathematical procedures accurately. Meanwhile, description difficulty showed the lowest percentage (36%), suggesting that most students were relatively capable of identifying relevant physics concepts.

Overall, the findings demonstrate that students' problem-solving difficulties are not solely attributable to inadequate conceptual understanding but are also influenced by their ability to plan strategies, evaluate solutions, and reflect upon the problem-solving process (Dirman et al., 2022; Lengkong et al., 2021). Therefore, physics instruction should be designed to systematically develop all stages of problem solving so that students can solve physics problems more effectively.

Conclusion

This study successfully developed a testlet-based two-tier diagnostic test for identifying students' problem-solving difficulties in Newton's Laws. Based on the Rasch Model analysis, the developed instrument met the validity and reliability criteria and is therefore suitable for use as a diagnostic tool in physics education. The findings revealed that the most dominant problem-solving difficulty occurred at the check and evaluate stage (75%), followed by the plan a solution stage (60%), visualize the problem stage (58%), execute the plan stage (56%), and describe the problem in physics terms stage (36%).

These results indicate that students tend to experience difficulties in evaluating the reasonableness of their answers and determining appropriate problem-solving strategies. Therefore, physics instruction should be designed to provide systematic training in evaluation and solution-planning skills to enhance students' overall problem-solving abilities. Furthermore, the developed diagnostic instrument can serve as an effective tool for identifying students' difficulties and supporting the design of targeted instructional interventions in physics learning.

References

- Amelia, Y. S. (2024). *Pengembangan E-LKPD Berbasis Learning Cycle 7E Untuk Menstimulus Complex Problem Solving Dan Self Efficacy Siswa*. Universitas Lampung.
- Arikunto, S. (2021). *Dasar-dasar Evaluasi Pendidikan Edisi 3*. Bumi Aksara.
- Auliya, H. R., Christianti, M., & Hidayat, B. (2025). Development of Multi-Tier Diagnostic Tests in Primary Schools: A Systematic. *Jurnal Prima Edukasia*, 13(1), 59–74.
- Azizah, N., & Yuliati, L. (2025). Systematic Literature Review : Analysis of Students ' Reasoning Ability and Difficulties in Newton ' s Law Material. 11(5), 50–55. <https://doi.org/10.29303/jppipa.v11i5.10433>
- Azwar, S. (2022). *Reliabilitas dan Validitas Edisi 4*.
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*.
- Dirman, H. M., Mufit, F., & Festiyed, F. (2022). Review and comparison of four-tier multiple choice and five-tier multiple choice diagnostic tests to identify mastery of physics concepts. *Jurnal Penelitian Pendidikan IPA*, 8(1), 1–12.
- Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical*

- Review Special Topics-Physics Education Research, 10(2), 20119.
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science.
- Halliday, D., Resnick, R., & Walker, J. (2014). *Fundamentals of Physics* (10th ed.). Wiley.
- Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. In *American Journal of Physics* (Vol. 60, Issue 7, pp. 627–636). <https://doi.org/10.1119/1.17117>
- Hutapea, J., Sahyar, S., & Manurung, S. R. (2020). Analysis of Problem Solving Ability Tests on Dynamics Material in Basic Physics Courses BT - *Journal of Physics: Conference Series*. 1485(1), 12014.
- Istiyono, E., Sunu Brams Dwandaru, W., Fenditasari, K., Ayub, M. R. S. S. N., & Saepuzaman, D. (2023). The Development of a Four-Tier Diagnostic Test Based on Modern Test Theory in Physics Education. *European Journal of Educational Research*, 12(1), 371–385. <https://doi.org/10.12973/eu-jer.12.1.371>
- Lengkong, M., Istiyono, E., Rampean, B. A. O., Tumanggor, A. M. R., & Nirmala, M. F. T. (2021). Development of Two-Tier Test Instruments to Detect Student's Physics Misconception. *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, 528(Icriegms 2020), 561–566. <https://doi.org/10.2991/assehr.k.210305.082>
- Linacre, J. M. (2012). *Winsteps Rasch Measurement Computer Program*.
- Meiliyadi, L. A. D., Arizona, K., Wahyudi, M., & Damayanti, I. (2025). Research trend on Newton's law misconceptions from Scopus database using bibliometric. *Indonesian Journal of Educational Research and Technology*, 5(3), 275–290.
- Okpatrioka, O. (2023). Research and Development (R&D) Penelitian yang Inovatif dalam Pendidikan. *Dharma Acariya Nusantara: Jurnal Pendidikan, Bahasa Dan Budaya*, 1(1), 86–100.
- Reif, F., & Heller, J. I. (1982). Knowledge Structure and Problem Solving in Physics. *Educational Psychologist*, 17(2), 102–127. <https://doi.org/10.1080/00461528209529248>
- Rosengrant, D., Van Heuvelen, A., & Etkina, E. (2009). Do students use and understand free-body diagrams? *Physical Review Special Topics—Physics Education Research*, 5(1), 10108.
- Serway, R. A., Jewett, J. W., Wilson, K., & Rowlands, W. (2016). *Physics for Global Scientists and Engineers, Volume 2*. Cengage Learning Australia. <https://books.google.co.id/books?id=fVtnDwAAQBAJ>
- Setyono, A., Nugroho, S. E., & Yulianti, I. (2016). Analisis kesulitan siswa dalam memecahkan masalah fisika berbentuk grafik. *UPEJ Unnes Physics Education Journal*, 5(3), 32–39.
- Soeharto, S. (2021). Development of a Diagnostic Assessment Test to Evaluate Science Misconceptions in Terms of School Grades: A Rasch Measurement Approach. *Journal of Turkish Science Education*, 18(3), 351–370.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunika.
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana University.
- Tuysuz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, 4(6), 626–631.
- Ulviani, N., Maria, H. T., & Hidayatullah, M. M. S. (2025). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Peserta Didik Pada Hukum Newton Di SMA Negeri 1 Teluk Pakedai. *Jurnal Education And Development*, 13(1), 100–106.