Development of Integrated Competency-Based Physics Academic Test for Teacher Professional Education Program Students

Sujito[1]*, Ratna Ekawati[1], Hestiningtyas Yuli Pratiwi[2]

[1]Universitas Negeri Malang, Semarang St. No. 5, Malang, East Java, 65145, Indonesia
[2]Universitas PGRI Kanjuruhan Malang, S. Supriadi St. No. 48, Malang, East Java, 65148, Indonesia
*Corresponding author, email: sujito.fmipa@um.ac.id

Article History
Received: 10 September 2025
Revised: 10 October 2025
Accepted: 12 October 2025
Published: 14 October 2025

Keywords
Evaluation
Instrument
Physics
Professional
Teacher

Abstract
The teacher professional education program requires prospective physics teachers to demonstrate integrated academic, pedagogical, social, and personal competencies. However, current evaluation practices remain dominated by traditional multiple-choice tests that primarily assess lower-order cognitive skills and fail to capture critical thinking, problem-solving, and reflective abilities. This study aims to develop a problem-based professional competency test designed to predict the academic success of the teacher professional education program students in physics. The research employed a Research and Development (R&D) approach with a modified Borg & Gall model, covering needs analysis, design, expert validation, pilot testing, revisions, and field trials. The participants were 57 physics students of the teacher professional education program on one of the University in Malang, Indonesia at 2025. The initial instrument consisted of 35 items, which were refined to 25 valid items through rigorous analysis of validity, reliability, item difficulty, discrimination power, and distractor effectiveness. Findings revealed that all 25 items were valid, with an instrument reliability coefficient of 0.788. The distribution of item difficulty was dominated by easy to moderate levels, ensuring a balanced measurement range. These results highlight that a problem-based professional competency test can serve as a valid, reliable, and contextually relevant evaluation tool. It not only aligns with the demands of higher-order thinking but also reflects the pedagogical readiness required of future physics teachers.

How to cite: Sujito, S., Ekawati, R. & Pratiwi H. Y. (2026). Development of Integrated Competency-Based Physics Academic Test for Teacher Professional Education Program Students. *Teaching, Learning, and Development*, *4*(1). 49–58. doi: 10.62672/telad.v4i1.108

1. Introduction

The teacher professional education program represents a strategic initiative of the Indonesian government to ensure the preparation of qualified educators who demonstrate not only mastery of subject matter but also pedagogical, professional, social, and personal competencies. Within the field of physics education, the teacher professional education program students are expected to translate abstract concepts into meaningful learning experiences that foster conceptual understanding among learners (Sujito et al., 2024). The program also emphasizes the development of academic test instruments that holistically capture the multifaceted competencies required of professional teachers. In line with the Merdeka Curriculum, the teacher professional education program graduates are expected to design and implement teaching modules that reflect learner-centered principles and promote adaptive instructional practices. Importantly, the teacher professional education program is primarily intended for prospective non-civil servant teachers without formal teacher education backgrounds, thereby addressing the demand for a broader pool of qualified educators. By equipping participants with relevant knowledge, skills, and attitudes, the program seeks to foster the development of reflective and professional teachers. Furthermore, the teacher professional education program is envisioned as a means of enhancing teaching practices in a sustainable manner, providing long-term educational benefits even though the changes achieved are typically incremental in nature (Salsabila & Wahyudin, 2024).

However, student learning success depends not only on mastery of course content but also on the ability to think critically and apply knowledge to problem-solving. Although teachers are encouraged to adopt more effective instructional approaches, implementation is often constrained by competing priorities or limited professional capacity. A central challenge is the so-called “double discontinuity” in teacher education: the disconnection between school-based learning and university instruction, coupled with the limited transfer of university coursework into classroom teaching practice (Sunardi et al., 2022). Recognizing this gap, recent scholarship in teacher education highlights the importance of situating teacher preparation within authentic

doi: 10.62672/telad.v4i1.108

ISSN: 2988-2834

professional contexts to ensure that future teachers, particularly in physics, are adequately prepared for the complex demands of the profession.

As the demand for higher quality education continues to increase, there is a growing need for instruments that not only assess mastery of course content but also provide a more comprehensive prediction of student learning success. Such instruments should be designed to reveal the potential of higher-order thinking skills, which form a critical foundation for academic achievement (Yang et al., 2018). Physics students of the teacher professional education program often face difficulties when teaching abstract concepts such as mechanics, electromagnetism, and thermodynamics. This highlights the need for evaluation tools capable of measuring critical and analytical thinking skills rather than merely memorization. Problem-based tests offer one promising solution, as they encourage students to connect theoretical knowledge with contextual phenomena. For instance, Newton's laws can be better understood through everyday illustrations, while the concept of renewable energy can be explored through environmentally based learning. These instruments play an essential role in training the teacher professional education program students to integrate physics knowledge into real-life teaching contexts (Ma'ruf et al., 2020).

To date, however, there have been no systematic efforts to develop such instruments specifically for physics education programs. A professional aptitude test can be designed to uncover students' academic potential by drawing upon tasks typically performed by both lecturers and students. These tasks are closely related to learning activities and reasoning processes that support graduate competencies in problem-solving, rather than simply mastery of subject matter. Once identified, academic potential can be used to predict how students' abilities will develop in completing academic tasks (Han, 2017; Simbolon & Silalahi, 2023). Tests of this type possess predictive characteristics, measuring fundamental abilities that can forecast academic success among physics education students. Academic achievement is determined not only by content mastery but also by general abilities combined with learning environments and reasoning skills for problem-solving. Such competencies are particularly relevant to higher education graduates, where intellectual development is a central focus. Test items are not tied to specific course content but instead address general problem contexts accessible to students from diverse backgrounds. The Academic Aptitude Test, therefore, is designed to predict learning achievement within a given educational framework (Benignus et al., 2023). In the context of the teacher professional education in physics education, professional competency tests aim to measure critical, creative, logical, numerical, analytical, and problem-solving skills.

Furthermore, the development of problem-based test instruments is of particular importance in learning, as they provide the means to evaluate higher-order thinking skills (Boa et al., 2018), problem-solving capacity (Bani-Hamad & Abdullah, 2019), and the integration of physics theory with teaching practice (Belland et al., 2019). Such instruments enable the teacher professional education program providers to map student readiness for the challenges of teaching while simultaneously serving as a selection tool to ensure that graduates of the Physics teacher professional education program are competent, innovative, and prepared to meet the demands of education. The primary issue at present is the dominance of traditional multiple-choice tests, which inadequately measure critical thinking, problem-solving, and student reflection. This creates a gap between expected learning outcomes and the actual competencies required of future teachers (Delisle, 1997). As a solution, a problem-based academic competency test has been developed. The instrument is designed to present authentic physics problems while also assessing pedagogical strategies that may be applied. A Research and Development approach was employed to ensure that the instrument is not only theoretically valid but also practically applicable within the teacher professional education program.

This study aims to develop and validate a problem-based academic competency test specifically designed for the teacher professional education program of physics students. Unlike conventional content-based evaluations, the instrument is intended to measure higher-order thinking, problem-solving, and pedagogical reasoning skills through authentic physics contexts. Employing a Research and Development (R&D) approach, the test was constructed to ensure validity, reliability, and practical applicability. Aligned with 21st-century skills, critical thinking, problem-solving, collaboration, and creativity (Boa et al., 2018). The professional ability potential test is expected to reveal student potential, predict academic success, and support the preparation of adaptive teachers capable of responding to curriculum reforms, technological advances, and diverse learner needs.

2. Method

This study employed a Research and Development (R&D) model with modified stages (Adri et al., 2020; Creswell, 2017), aiming to produce a specific product. The procedures were simplified into six steps: 1) Needs Analysis, in which weaknesses of existing Physics the teacher professional education program evaluation instruments were identified; 2) Initial Product Design, where a problem-based test blueprint was developed in alignment with the graduate learning outcomes and course learning objectives of the teacher professional education program; 3) Expert Validation, involving physics education specialists and the teacher professional education program practitioners; 4) Limited Trial, conducted with a small group of students to assess clarity

and feasibility; 5) Product Revision, carried out based on expert feedback and trial results; and 6) Field Testing, in which the instrument was applied in classes and analyzed for validity, reliability, difficulty level, and discrimination index.

The trial results were analyzed using both quantitative and qualitative approaches. Classical test analysis was employed to examine item difficulty, discrimination power, distractor effectiveness, validity, reliability, and overall item analysis. In parallel, qualitative data were collected through feedback from students and lecturers. The research subjects were 57 Physics students from Wave 1 at 2025, in one of University in Malang, Indonesia.

3. Results and Discussion

3.1. Instrument Development

The analysis of existing evaluation instruments in the Physics the teacher professional education program identified two primary needs: 1) the assessment of academic and pedagogical competencies, and 2) the implementation of authentic evaluation (Benignus et al., 2023; Boa et al., 2018). The teacher professional education program students are required not only to master physics concepts (cognitive domain) but also to demonstrate teaching skills (pedagogical domain) and professional attitudes. However, most current instruments focus predominantly on cognitive aspects through standard multiple-choice questions, while authentic evaluation remains urgent and underdeveloped. There is a pressing need for instruments capable of assessing problem-solving, critical thinking, and scientific reasoning within real teaching contexts. A review of prior studies and monitoring surveys revealed that evaluation weaknesses persist. Most instruments remain conventional multiple-choice tests measuring recall and comprehension, with limited capacity to assess problem-solving or the application of concepts to real-life contexts. Empirical findings show that 72% of Physics the teacher professional education program test items are still at Bloom's levels C1–C2, with only 28% reaching higher-order levels (C3 and above) (Salsabila & Wahyudin, 2024). This indicates the necessity of improving test quality to measure higher-order thinking skills such as analysis and synthesis, which are essential for developing physics teacher competencies.

The analysis further revealed minimal integration of pedagogical contexts. Test items rarely connected physics concepts with teaching strategies, assessment, or instructional practice, leaving students' competencies as "teachers of physics" insufficiently assessed (Tiruneh et al., 2017). Findings also showed that the teacher professional education program of Physics students performed 15% lower on pedagogically oriented items compared to content-based items, suggesting that existing instruments fail to measure content–pedagogy integration fairly (Sutaphan & Yuenyong, 2023). Moreover, authentic evaluations are largely absent, with minimal use of portfolios, projects, or performance assessments. Yet, physics teachers are expected to design instructional media, develop worksheets, and manage problem-based learning. The main weaknesses identified include: 1) an emphasis on rote memorization rather than problem-solving, 2) a lack of pedagogical contextualization, making items less relevant to classroom practice, and 3) insufficient authentic assessments, with little integration of portfolios, peer assessment, or field-based practice (Varas et al., 2023).

The development of the academic ability test instrument was carried out with a focus on improving the evaluation approach by designing problem-based tests, performance assessments, and validated reflective rubrics. The current evaluation instruments used in the teacher professional education program of physics remain largely content-based, with an emphasis on lower- to mid-level cognitive skills. In contrast, professional physics teachers are required to integrate content knowledge (CK), pedagogical knowledge (PK), and technological knowledge (TK) within the technological, pedagogical, content, and knowledge (TPACK) framework (Jena, 2016). Therefore, this study developed instruments aimed at preparing teacher professional education students of physics, as prospective physics teachers, to become critical, creative, and reflective educators. The developed test instrument consisted of 35 items. In addition to problem-based items, performance-based assessments and reflective rubrics were also designed. The results of this development are presented in Figure 1.

The difficulty level of test items is intended to determine whether an item is classified as easy, moderate, or difficult for test participants. Analysis of the 35 items administered to 57 students showed that 19 items were categorized as easy, 10 items fell into the moderate or ideal category, and 6 items were considered difficult. An item is classified as difficult if fewer than 30% of respondents answer it correctly, indicating that only a small proportion of students are able to solve it. Items answered correctly by 30% to 69.9% of respondents are categorized as moderate or ideal in difficulty. Meanwhile, items answered correctly by more than 70% of respondents are classified as easy, meaning that the majority of students were able to answer them correctly.

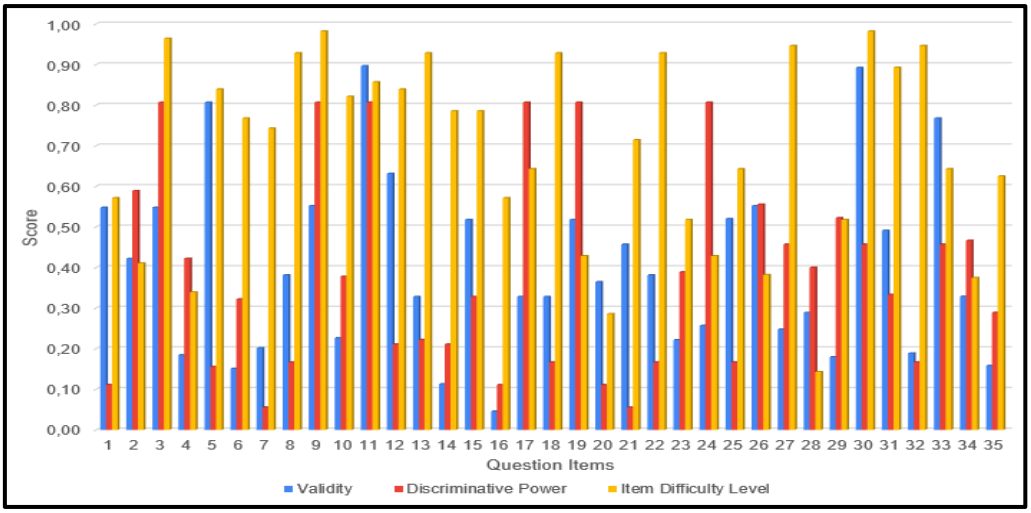

Figure 1. Characteristics of the Question Items from Initial Development Results

An item is categorized as having low discrimination power when the difference in the proportion of correct answers between the upper and lower groups is less than 30%. This indicates that the item can be answered correctly by both groups, with less than a 30% difference in correct responses. If the proportion of correct answers between the two groups is identical, the difference equals zero (0), meaning the item fails entirely to distinguish between high- and low-ability groups. When the difference falls within the range of 30% to 69.9%, the item is classified as having moderate or ideal discrimination power. Conversely, if the difference exceeds 70%, the item is considered to have high discrimination power, as it can effectively differentiate between more than 70% of respondents in the lower group who fail to answer correctly. Out of the 35 items tested, 18 were categorized as having ideal discrimination power, 17 fell into the low category, and none reached the high category. Discrimination analysis is essential to ensure that items are not only valid in terms of content but also functional as measurement tools. Items with low or negative discrimination power must be revised (e.g., by clarifying the wording or answer options) or eliminated altogether.

The validity of each test item was analyzed by correlating the item score with the total score obtained by respondents. Of the 35 items tested, 14 met the validity criteria, while 21 were classified as invalid or had low validity because their correlation values were below 0.30. The score of each item represented the measurement of a specific indicator, whereas the total score reflected the overall measurement of the intended construct. A correlation value of 0.30 was used as the benchmark for validity (Peris-ortiz, n.d.). If an item's correlation value was below 0.30, it was considered to contribute less than 30% to the measurement of the construct, and thus categorized as invalid. Based on the 35 items tested, 27 still contained distractor options that were not selected at all by respondents, indicating the need for revision of those distractors. Meanwhile, the remaining 8 items already had functioning distractors and therefore required no modification. Each item consisted of five alternative answers (A, B, C, D, and E), designed to serve as distractors. An option is considered a functioning distractor if it successfully attracts respondents' attention and is selected as an answer. Conversely, if an option is not chosen by any respondent, it is regarded as non-functioning, ineffective as a distractor, and thus must be revised.

3.2. Revision Process

The revision process of test items in educational research is a critical stage that determines whether an assessment instrument truly measures what it is intended to measure. In this study, the revision process was carried out systematically to enhance both the psychometric quality and pedagogical relevance of the items developed for the physics teacher professional education program. The main aim of the revision was to ensure that the instrument reflected the intended competencies, including academic mastery, pedagogical reasoning, and higher-order thinking skills, while also meeting established standards of validity and reliability.

The initial trial of 35 items served as a diagnostic stage to uncover weaknesses in the instrument. Based on the results of the initial trial, several weaknesses were identified, particularly in terms of item validity, discrimination power, and the functioning of distractors. The product of analysis revealed several recurring issues, namely deficiencies in validity, discrimination power, and the functioning of distractors. These weaknesses are not unusual in early drafts of assessment instruments, as item development is inherently iterative. However, their identification provided a foundation for targeted revisions designed to improve both the statistical properties of the items and their alignment with the curricular goals of the physics teacher professional education program. Items with a validity coefficient below 0.30 were classified as weak and thus required modification, while items with low or negative discrimination indices were considered problematic. In

addition, distractors that were not selected at all by respondents were marked as ineffective and needed improvement. After conducting an initial analysis of the developed instrument, taking these considerations into account, the results were obtained as in Table 1.

Table 1. Summary of Item Revision Results

Criteria	Before	After	Note
Valid Items (r ≥ 0.30)	14	25	Weak items (< 0.30) were revised or eliminated
Difficulty Index	Easy: 19 Moderate: 10 Difficult: 6	Easy: 12 Moderate: 9 Difficult: 4	Distribution adjusted to achieve balance
Discrimination Power	High: 0 Moderate: 18 Low: 17	High: 0 Moderate: 20 Low: 5	Low/negative items revised or replaced
Reliability (Cronbach α)	0.721	0.788	Increased after revision

3.2.1. Validity Concerns and Revisions

Item validity is one of the cornerstones of test development. In this study, validity was analyzed through item-total correlations, where the score on each item was correlated with the overall test score. Items with correlation coefficients below 0.30 were deemed weak, as they contributed little to the measurement of the construct. Out of the 35 items, 21 were found to be invalid by this criterion. Items with low validity typically suffer from one of two problems: they may test trivial knowledge not central to the construct, or they may be ambiguously worded, leading to inconsistent responses. For example, some items asked students to recall isolated definitions without requiring application or analysis. While these items could be answered by rote memorization, they did not reflect the higher-order competencies emphasized in the physics teacher professional education program.

The revision strategy for validity involved several measures: **Rewriting item stems**, Ambiguities were removed, and stems were revised to be clear, concise, and directly aligned with the intended indicators of competence. **Embedding contextual problems**, Items that previously tested factual recall was redesigned into contextualized problem-solving scenarios. For instance, rather than asking students to define Newton's second law, items were reframed into situations requiring application of the law to real-life contexts, such as analyzing motion in transportation or sports. **Aligning with graduate learning outcomes and course learning objectives**, Each revised item was cross-checked with graduate learning outcomes of the physics teacher professional education and course learning objectives to ensure coherence between the test and curricular expectations.

3.2.2. Discrimination Power and Its Role

Another critical dimension revealed by the analysis was discrimination power. Discrimination indices reflect the ability of an item to distinguish between high- and low-performing students. Ideally, items should be more likely answered correctly by students who possess greater understanding, while students with weaker understanding should find them more challenging. In the initial trial, 17 items were found to have low or even negative discrimination indices. This meant that students across ability levels were performing similarly on these items, which undermines the test's ability to identify differences in competency. A lack of discrimination can result from items being too easy, too difficult, or misleading due to poor construction.

Revisions to improve discrimination involved: **Increasing cognitive demand**: Items classified as too easy were rewritten to require multi-step reasoning. For example, instead of asking students to directly calculate electrical resistance using Ohm's law, they were asked to interpret a circuit diagram and select the correct reasoning process to solve the problem. **Simplifying overly complex items**: Some items were found to be unnecessarily difficult due to convoluted wording or excessive information. These were revised to focus on the essential concept, reducing cognitive overload while maintaining appropriate challenge. **Ensuring construct-relevant difficulty**: Items were checked to ensure that difficulty arose from the targeted concept rather than from confusing phrasing or irrelevant details.

Distractor analysis provided further insights into item quality. Of the 35 items, 27 contained distractors that were not chosen by any respondents. This indicated that the distractors were implausible and thus ineffective. Effective distractors are critical in multiple-choice tests because they increase the diagnostic value of the item by reflecting common misconceptions or errors. Revisions to distractors followed three principles: **Plausibility**: Distractors were revised to appear credible to students with partial understanding. **Attractiveness**: Distractors were designed to compete with the correct answer, encouraging students to carefully apply reasoning rather than guessing. **Balance**: Care was taken to ensure that distractors were similar in length and format to the correct answer, reducing the risk of test-wise students identifying the answer through superficial cues.

3.2.3. Expert Validation and Pedagogical Relevance

Based on Table 1, it can be known that a central requirement of any educational measurement instrument is that item validity was analyzed by correlating the score of each item with the total test score. Items with correlation coefficients below 0.30 contributed less than 30% to the measurement of the intended construct and were therefore deemed invalid. The identification of 21 items in this category during the initial analysis highlighted the necessity of substantial revision. Following statistical revisions, the items were subjected to expert validation. Experts in physics education and experienced teacher professional education program practitioners evaluated the revised items for both content accuracy and pedagogical relevance. Their feedback confirmed that the revisions enhanced alignment with the competencies expected of future physics teachers. Experts particularly highlighted the integration of pedagogical elements into physics problems. For example, items were commended for linking physics concepts to classroom teaching strategies, such as designing experiments or interpreting student misconceptions. This integration ensured that the test assessed not only academic mastery but also pedagogical reasoning, reflecting the dual focus of the teacher professional education program.

3.2.4. Adjustments to Difficulty Levels

Item difficulty is another important consideration. The initial analysis revealed that 19 items were classified as easy, 10 as moderate, and 6 as difficult. This distribution was skewed toward the easy category, reducing the test's capacity to measure the full spectrum of student abilities. Revisions aimed at balancing the distribution were implemented. Easy items were revised by embedding them into real-world scenarios requiring higher-order thinking. For instance, rather than asking students to recall the definition of acceleration, items were reframed to analyze acceleration in contexts such as vehicles navigating slopes or athletes running races. Conversely, overly difficult items were simplified to focus on the targeted concept without extraneous barriers. After revision, the distribution shifted toward a more balanced set of easy, moderate, and difficult items. This balance is essential for producing a test that is both fair and discriminative across a range of ability levels.

3.2.5. Limited Trials and Student Feedback

A limited trial with a small group of teacher professional education students was conducted to evaluate the revised items. Feedback indicated that the items were clearer and more engaging than before. Students reported that the contextualized problems encouraged them to think critically rather than rely on rote memorization. Item analysis from the limited trial confirmed improvements in discrimination indices, distractor functioning, and overall reliability. The reliability coefficient (Cronbach's Alpha) increased from 0.721 in the initial trial to 0.788 after revision, indicating greater internal consistency. This improvement provided empirical evidence that the revisions had enhanced the overall quality of the test.

3.2.6. Implications for Physics Teacher Education

The significance of this revision process extends beyond psychometrics to the broader goals of physics teacher education. By embedding authentic contexts and higher-order thinking requirements, the revised items contribute to preparing the teacher professional education students for the complexities of teaching. The test not only measures knowledge but also encourages the reasoning and problem-solving skills that teachers must foster in their own classrooms. This aligns with global trends in teacher education, which emphasize authentic assessment as a means of preparing teachers to be adaptive, reflective, and capable of integrating content knowledge with pedagogy. The revised instrument therefore represents not just a technical improvement but a pedagogical innovation. The outcomes of this study are consistent with previous research in educational measurement. Studies have shown that poorly functioning distractors are common in early test drafts and that targeted revisions improve item discrimination and reliability. Similarly, research on teacher education assessments has highlighted the need for instruments that assess beyond content recall, focusing instead on critical thinking, problem-solving, and pedagogical integration. By addressing these issues, the present study contributes to filling a gap in physics teacher education, where assessment instruments often focus narrowly on content knowledge. The revised test integrates both content and pedagogy, providing a more holistic evaluation of student competence.

A systematic revision process was carried out on the developed test items, transforming the initial draft, which had substantial weaknesses, into a robust and pedagogically meaningful assessment instrument. This was achieved by addressing issues of validity, discrimination, distractor function, and difficulty level. The revised test achieved higher reliability and stronger alignment with required competencies. This process highlighted the importance of iterative development, where empirical evidence and expert judgment work together to refine educational assessment. Some aspects reviewed included technical improvements, the instrument's pedagogical relevance, and alignment with global trends in teacher education that prioritize authentic, problem-based assessment.

3.3. Instrument Revision

The next step taken by the researchers was to reduce the number of items from 35 to 25. This selection process was carried out by considering four main aspects: validity, difficulty level, discrimination power, and feedback from distractor analysis. Based on these criteria, each item was either retained without revision, retained with revision, or eliminated. The characteristics of the revised test items are presented in Figure 2.

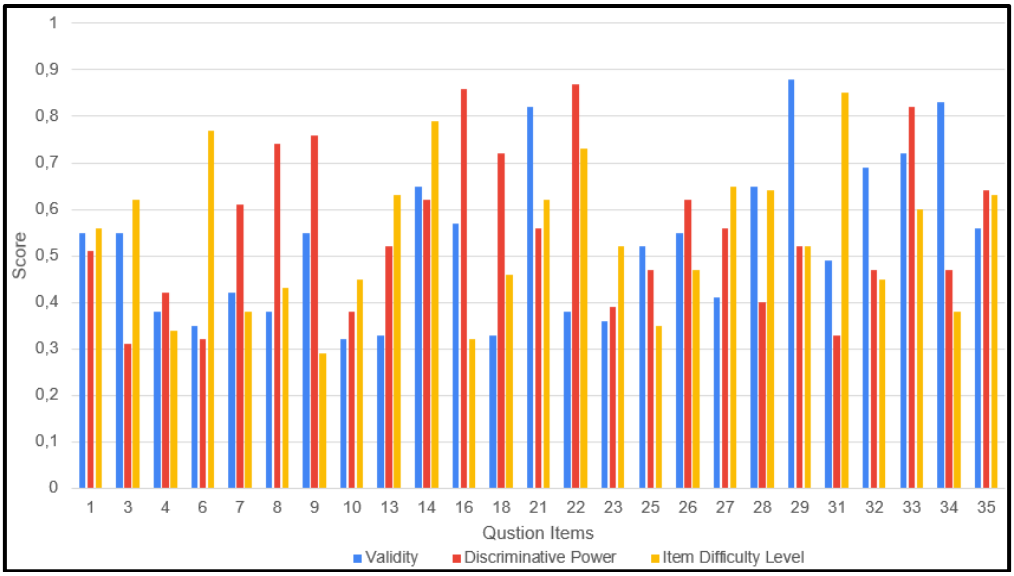

Figure 2. Item Characteristics After Revision

Based on Figure 2, it can be seen that after undergoing various revisions, only 25 out of the original 35 items were deemed suitable for assessing the pedagogical and academic abilities of the teacher professional education of Physics students. In terms of difficulty, the researchers upgraded several items to the moderate category. However, with respect to discrimination power, none of the items reached the high category across all aspects (zero realization). This suggests the need to further enhance items by increasing their difficulty level, thereby increasing their discriminatory power. The target that was nearly achieved was the proportion of items categorized as easy. Conversely, the target for items in the ideal (moderate) difficulty category was not met, as no items fell within this classification. On the other hand, targets related to ideal and low discrimination indices were mostly fulfilled. Therefore, revisions focused primarily on clarifying the wording of the item stems to eliminate ambiguity and ensure that each question measures the targeted indicator. Distractors were reconstructed to make them more plausible and engaging for respondents, thus increasing their effectiveness as choice alternatives. If the item's difficulty level was too low, it was adjusted by integrating it into a contextual problem or increasing its complexity. Conversely, items that were too difficult were simplified without compromising their conceptual depth. Following these modifications, the revised items were revalidated by physics education experts and the teacher professional education practitioners. Their feedback confirmed that the revisions had improved content validity and pedagogical relevance.

This iterative revision process demonstrates the importance of combining empirical evidence from item analysis with expert judgment to ensure the development of valid, reliable, and pedagogically meaningful assessment instruments. The revised instrument offers a valuable tool for assessing prospective physics teachers' readiness to meet the complex demands of their profession. Validity of each item was determined by correlating the item score with the total score obtained by respondents. A correlation value of 0.30 served as the benchmark for validity. Out of the 35 tested items, 14 were classified as valid, while 21 were considered invalid or had low validity (below 0.30). This result indicates that items with a correlation below 0.30 contributed less than 30% to the measurement of the intended construct. Conversely, items that achieved a validity score ≥ 0.30, with appropriate difficulty levels (easy, moderate, or difficult) and acceptable to high discrimination power, were retained without revision. Since the number of valid items did not yet meet the target across all measured aspects, the researchers considered including items with validity values approaching 0.30. Items with validity coefficients ranging from 0.24 to 0.29 were selected for revision in order to be utilized. The revisions were primarily carried out on items that represented certain indicators in the test blueprint but were still limited in number. Therefore, additional items were sought that could still be improved. Through this process, 11 items were revised, including an item with a validity value of 0.24 categorized as easy but with an ideal discrimination index, and another item with a validity value of 0.28 categorized as easy but with a low discrimination index.

Revisions focused on the wording of the item stems, both in terms of statements or questions and the answer options. For example, in item number 21, distractor analysis showed that all alternative answers were chosen by the 57 respondents, with the following distribution: option A = 5%, option B = 7%, option C = 75%, option D = 5%, option E = 4%, and no response = 4%. Based on these findings, revisions were directed at refining the item statement to ensure greater clarity and alignment. The overall development of this test was primarily intended to predict learning achievement in Physics, and at a later stage, it is projected to serve as a selection instrument for prospective the teacher professional education students. Therefore, aspects of validity, item difficulty, and discrimination power were considered as primary factors (Walsh et al., 2019; Yang et al., 2018). The 25 reconstructed items were each reanalyzed in terms of their difficulty level and discrimination index. The next step involved testing the reliability of these items using Cronbach's Alpha coefficient through SPSS version 21. The analysis yielded a reliability value of 0.79 for the overall set of 25 items, indicating that the assembled test instrument possessed a relatively high level of reliability (Bahri et al., 2021). These findings were then compared with the ultimate objective of the test instrument, namely, to produce items with strong discriminatory power between students predicted to succeed and those likely to encounter challenges in completing the teacher professional education program.

4. Conclusion

This study successfully developed a problem-based academic potential test instrument for students in the teacher professional education in Physics. The instrument was constructed through a Research and Development (R&D) approach, involving stages of needs analysis, design, expert validation, limited trials, revisions, and field testing. The findings indicate that the problem-based test effectively addresses the limitations of existing the teacher professional education evaluation tools, which have typically focused only on lower-order cognitive skills, lacked pedagogical integration, and offered limited authenticity. From an initial pool of 35 items, a selection process produced 25 items deemed suitable for use, characterized by varying levels of difficulty (easy, moderate, and difficult), strong discriminatory power in most items, and validity that met established standards. The overall reliability of the instrument was also high, with a Cronbach's Alpha coefficient of 0.788, confirming its consistency in measuring the intended construct. The developed instrument not only assesses content mastery but also reveals the teacher professional education Physics students' critical, creative, analytical, numerical, and problem-solving abilities. It can serve as both a predictor of academic success and a selection tool for competent the teacher professional education candidates. The problem-based test should be continuously applied in the teacher professional education of physics programs to evaluate both pedagogical and academic competencies of prospective teachers. Future development is recommended by enriching the item pool with questions that have high discrimination power and incorporating authentic assessment forms, such as portfolios and performance evaluations, to ensure more comprehensive results. Broader trials across multiple the teacher professional education institutions are also essential to strengthen external validity and confirm its wider applicability.

Author Contributions

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

Funding

This research is funded by the Teacher Professional Education Program of the Graduate School, Universitas Negeri Malang.

Declaration of Conflicting Interests

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

References

Adri, M., Rusdinal, Z., Zainul, R., Darni, Sriadhi, Wahyuningtyas, N., Khaerudin, Nasrun, Rahmulyani, Nuranjani, Nurmaniah, Wedi, A., Surahman, E., Aisyah, E. N., Oktaviani, H. I., Sri Martini Meilanie, R., Purnamawati, S. N., Hapidin, Listyasari, W. D., … Adnan, E. (2020). Development of content learning system in professional education subjects for educational institutions in Indonesia. *Journal of Physics: Conference Series, 1594*(1), 012022. doi:10.1088/1742-6596/1594/1/012022

Annisa, N., & Asrizal. (2022). Design and validity of STEM integrated physics electronic teaching materials to improve new literacy of class XI. *Jurnal Pendidikan Fisika, 10*(3), 177–192. doi:10.26618/jpf.v10i3.7900

Bahri, A., Jamaluddin, A. B., Muharni, A., Fikri, M. J. N., & Arifuddin, M. (2021). The need of science learning to empower high order thinking skills in 21st century. *Journal of Physics: Conference Series, 1899*(1), 012144. doi:10.1088/1742-6596/1899/1/012144

Bani-Hamad, A. M. H., & Abdullah, A. H. (2019). Developing female students' learning and innovation skills (4Cs) in physics through problem-based learning. *International Journal of Academic Research in Business and Social Sciences, 9*(12), 501–513. doi:10.6007/IJARBSS/v9-i12/6750

Belland, B. R., Weiss, D. M., Kim, N. J., Piland, J., & Gu, J. (2019). An examination of credit recovery students' use of computer-based scaffolding in a problem-based, scientific inquiry unit. *International Journal of Science and Mathematics Education, 17*(2), 273–293. doi:10.1007/s10763-017-9872-9

Benignus, C., Buschner, P., Meier, M. K., & Wilken, F. (2023). Patient-specific instruments and patient-individual implants: A narrative review. *Journal of Personalized Medicine, 13*(3), 426. Retrieved from https://www.mdpi.com/2075-4426/13/3/426

Boa, E. A., Wattanatorn, A., & Tagong, K. (2018). The development and validation of the blended Socratic method of teaching (BSMT): An instructional model to enhance critical thinking skills of undergraduate business students. *Kasetsart Journal of Social Sciences, 39*(1), 81–89. doi:10.1016/j.kjss.2018.01.001

Costa, A. L. (1991). *Developing minds: A resource book for teaching thinking* (Rev. ed.). Alexandria, VA: ASCD.

Creswell, J. W. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: SAGE Publications.

Delisle, R. (1997). *How to use problem-based learning in the classroom*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Han, S. (2017). Korean students' attitudes toward STEM project-based learning and major selection. *Kuram ve Uygulamada Egitim Bilimleri (Educational Sciences: Theory & Practice), 17*(2), 529–548. doi:10.12738/estp.2017.2.0264

Jena, R. K. (2016). Learning styles and attitudes toward the use of wearable technology in higher education: A study among Indian students. *SMART Journal of Business Management Studies, 12*(1), [pages not provided]. Retrieved from https://www.indianjournals.com/ijor.aspx?target=ijor:sjbms&volume=12&issue=1&article=006

Lou, Y., Blanchard, P., & Kennedy, E. (2015). Development and validation of a science inquiry skills assessment. *Journal of Geoscience Education, 63*(1), 73–75. doi:10.5408/14-028.1

Ma'ruf, M., Setiawan, A., Suhandi, A., & Siahaan, P. (2020). Identification of the ability to solve the problem of contextual physics possessed by prospective physics teachers related to basic physics content. *Journal of Physics: Conference Series, 1521*(2), 022011. doi:10.1088/1742-6596/1521/2/022011

Oviawe, E. (2020). Cognitive task analysis: Guide to the development of an interprofessional simulation-based instruction. *The Journal of the American Osteopathic Association, 120*(4), [pages not provided]. doi:10.7556/jaoa.2020.035

Peris-Ortiz, M. (n.d.). *Sustainable learning in higher education*. [Publisher and location not provided].

Salsabila, S., & Wahyudin, D. (2024). Peran Program Profesi Guru Pra-Jabatan (PPG Prajab) terhadap kemampuan menentukan indikator pencapaian kompetensi (IPK) pada penyusunan modul ajar Kurikulum Merdeka. *Edukatif: Jurnal Ilmu Pendidikan, 6*(5), 5659–5670. doi:10.31004/edukatif.v6i5.7195

Simbolon, D. H., & Silalahi, E. K. (2023). Physics learning using guided inquiry models based on virtual laboratories and real laboratories to improve learning. *Journal for Lesson and Learning Studies, 6*(1), 55–62. doi:10.23887/jlls.v6i1.61000

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: The National Academies Press.

Stupple, E. J. N., Maratos, F. A., Elander, J., Hunt, T. E., Cheung, K. Y. F., & Aubeeluck, A. V. (2017). Development of the Critical Thinking Toolkit (CriTT): A measure of student attitudes and beliefs about critical thinking. *Thinking Skills and Creativity, 23*, 91–100. doi:10.1016/j.tsc.2016.11.007

Sujito, S., Liliasari, L., Suhandi, A., & Soewono, E. (2021). Description in course of mathematical methods for physics and possible development of course program. *Momentum: Physics Education Journal, 5*(1), 73–84. doi:10.21067/mpej.v5i1.5184

Sujito, S., Sulur, S., Hudha, M. N., Winarno, N., & Sunardi, S. (2024). Enhanced learning: Designing bifocal modeling practicum tools with ESP32 for exploring kinetic theory of gases. *Momentum: Physics Education Journal, 8*(2), 249–260. doi:10.21067/mpej.v8i2.10046

Sunardi, S., Suhandi, A., Darmawan, D., & Muslim, M. (2022). Investigation of student difficulties in physics learning and readiness to implement physics learning using bifocal modeling-based practicum in Indonesia. *[Journal name not provided], 11*(4), 1091–1102.

Sutaphan, S., & Yuenyong, C. (2023). Enhancing grade eight students' creative thinking in the water STEM education learning unit. *Cakrawala Pendidikan, 42*(1), 120–135. doi:10.21831/cp.v42i1.36621

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education, 15*(4), 663–682. doi:10.1007/s10763-016-9723-0

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: John Wiley & Sons.

Varas, D., Santana, M., Nussbaum, M., Claro, S., & Imbarack, P. (2023). Teachers' strategies and challenges in teaching 21st century skills: Little common understanding. *Thinking Skills and Creativity, 48*, 101289. doi:10.1016/j.tsc.2023.101289

Walsh, C., Quinn, K. N., Wieman, C., & Holmes, N. G. (2019). Quantifying critical thinking: Development and validation of the Physics Lab Inventory of Critical Thinking. *Physical Review Physics Education Research, 15*(1), 010135. doi:10.1103/PhysRevPhysEducRes.15.010135

Yang, Y., He, P., & Liu, X. (2018). Validation of an instrument for measuring students' understanding of interdisciplinary science in grades 4–8 over multiple semesters: A Rasch measurement study. *International Journal of Science and Mathematics Education, 16*(4), 639–654. doi:10.1007/s10763-017-9805-7