



# Development and Content Validation of a HOTS-Oriented Performance Assessment for Physics Practicum in Higher Education to Support SDG 4

Emiliannur<sup>1\*</sup>, Fuja Novitra<sup>1</sup>, Rahma Doni<sup>2</sup>, Metta Liana<sup>3</sup>, Nanang Winarno<sup>4</sup>, Lukman Hakim<sup>5</sup>

<sup>1</sup> Physics Education Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Indonesia

<sup>2</sup> Automotive Engineering Department, Faculty of Engineering, Universitas Negeri Medan, Indonesia

<sup>3</sup> Faculty of Teacher Training and Education, Universitas Maritim Raja Ali Haji, Tanjungpinang, Indonesia

<sup>4</sup> Science Education Department, Faculty of Mathematics and Natural Sciences, Universitas Pendidikan Indonesia, Indonesia

<sup>5</sup> Physics Education Department, Faculty of Mathematics and Natural Sciences, Universitas PGRI Palembang, Indonesia

Received: January 01, 2026

Revised: February 13, 2026

Accepted: March 25, 2026

Published: March 31, 2026

Corresponding Author:

Emiliannur

[emiliannur@fmipa.unp.ac.id](mailto:emiliannur@fmipa.unp.ac.id)

DOI: [10.29303/jppipa.v12i3.14471](https://doi.org/10.29303/jppipa.v12i3.14471)

 Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)



**Abstract:** Higher education is expected to equip students with Higher Order Thinking Skills (HOTS); however, assessment practices in physics laboratory learning still often emphasize procedural performance. This study aims to analyze assessment needs and to develop and conduct content validation of a HOTS-oriented performance assessment for physics laboratory learning in higher education to support Sustainable Development Goal 4 (Quality Education). This study employed a research and development (R&D) approach using the ADDIE model, limited to the Analyze, Design, and Develop stages. The participants were 49 undergraduate physics students selected using purposive sampling and laboratory assistants involved in practicum activities. Data were collected through essay tests, practicum module analysis, interviews, and observations, and analyzed using descriptive quantitative and qualitative methods. Content validity was evaluated through expert judgment using Aiken's V index. The results showed that existing assessments had not optimally measured HOTS, particularly evaluating (C5) and creating (C6). Therefore, a HOTS-oriented performance assessment integrating C4, C5, and C6 was developed, consisting of performance tasks, scoring rubrics, observation sheets, and self- and peer-assessment instruments. The content validation results indicated high validity with an Aiken's V value of 0.89. These findings provide a validated conceptual foundation for further implementation and evaluation to support SDG 4 in higher education.

**Keywords:** Development and Content Validation; Higher Education; Higher Order Thinking Skill; Performance Assessment; Physics Practicum

## Introduction

Higher education plays an important role in developing scientific knowledge and equipping students with advanced skills to tackle increasingly complex and dynamic global challenges (Dusdal, 2019; Khadka et al., 2025). In line with the objectives of Sustainable Development Goal 4 (SDG 4), higher education is expected to ensure the quality of learning

that fosters Higher Order Thinking Skills (HOTS), problem-solving, and lifelong learning capabilities. In the context of physics education, HOTS are particularly important because students are required to analyze experimental data, evaluate sources of error, and design alternative experimental procedures or solutions to scientific problems. Therefore, HOTS represent one of the key learning outcomes in physics practicum courses at the undergraduate and postgraduate levels.

### How to Cite:

Emiliannur, Novitra, F., Doni, R., Liana, M., Winarno, N., & Hakim, L. (2026). Development and Content Validation of a HOTS-Oriented Performance Assessment for Physics Practicum in Higher Education to Support SDG 4. *Jurnal Penelitian Pendidikan IPA*, 12(3), 681-692. <https://doi.org/10.29303/jppipa.v12i3.14471>

Assessment in higher education, particularly in physics practicum learning, still faces a number of significant challenges. These challenges are evident in learning environments that require authentic and contextual forms of evaluation, such as laboratory practical work. In practice, assessment is still often dominated by traditional methods that place greater emphasis on final outcomes than on students' cognitive engagement during the learning process (Braun, 2021). Performance tasks designed to reveal higher-order thinking skills have not yet been utilised to their full potential.

Preliminary observations and needs analysis conducted in physics practicum courses at a university indicate that assessment practices are still largely focused on the completion of experimental procedures and final reports. Analysis of practicum modules shows that most tasks are limited to analytical activities (C4), while tasks that require students to evaluate (C5) and create (C6) are still limited. In addition, structured performance rubrics, observation sheets, and self and peer assessment instruments are not yet systematically implemented. These findings indicate that assessment practices have not fully facilitated the development and measurement of students' higher-order thinking skills in practicum learning.

In recent years, assessment has increasingly been viewed as an integral part of the learning process, rather than merely a tool for measuring learning outcomes (Appiah & van Tonder, 2018; Calma & Davies, 2025; Zlatkin-Troitschanskaia & Shavelson, 2019). A number of studies emphasise the importance of assessment practices capable of revealing complex cognitive processes to support meaningful learning experiences (Shavelson et al., 2019). However, in the context of higher education, assessment still tends to focus on written tests and the completion of procedural tasks. Assessment approaches that require students to demonstrate HOTS through authentic performance remain limited, even though such approaches are crucial for preparing students to meet the demands of future learning and the professional workplace (Hyytinen & Toom, 2019; Vlachopoulos & Makri, 2024).

One assessment approach that has the potential to address these limitations is performance assessment. Performance assessment is a multi-stage assessment method that requires students to demonstrate their knowledge and skills through observable performance or meaningful products (Bland & Gareis, 2018; Braun et al., 2020; Pine & Baxter, 1991). This approach is closely aligned with the revised Bloom's taxonomy, which formulates cognitive processes in the form of action verbs such as analyzing, evaluating, and creating rather than static knowledge categories. This shift from nouns

to verbs is important because performance assessment measures students' observable actions and cognitive processes during practicum activities, not merely their final answers. Therefore, the revised Bloom's taxonomy provides a more appropriate framework for developing performance assessment instruments that aim to measure higher-order thinking skills.

In addition to serving as a measurement tool, performance assessment also has an instructional function. Through the use of clear criteria and formative feedback, this form of assessment can facilitate the development of students' skills during the learning process (Zlatkin-Troitschanskaia & Shavelson, 2019). These characteristics make performance assessment highly relevant for physics laboratory practicals, as students are expected to integrate conceptual understanding, procedural skills, and reflective thinking in completing experimental tasks (Braun, 2021; Hartina et al., 2019; Komisia et al., 2025).

In this study, performance assessment is situated within the perspectives of assessment for learning and assessment as learning. In this perspective, assessment functions as a learning resource that supports students' reflection, self-regulation, and continuous improvement through structured tasks and transparent assessment rubrics (Ismailia, 2021; Olson & Krysiak, 2021; Tan & Leong, 2015). However, based on the preliminary findings, HOTS based tasks, process assessment, and clearly formulated assessment criteria have not been systematically integrated into physics practicum assessment. This situation limits the potential of practicum learning to develop HOTS in a planned and sustainable manner.

Therefore, this study aims to conduct a needs analysis and develop a HOTS-oriented performance assessment for physics practicum learning in higher education up to the 'Develop' stage of the ADDIE model. The novelty of this research lies in the systematic integration of HOTS from the revised Bloom's taxonomy into physics practicum performance tasks, observation sheet, self and peer assessment, and structured scoring rubrics within a single assessment framework. This study focuses on establishing the content validity of the developed assessment framework as a foundation for future implementation and evaluation studies. This research is important because it provides a validated assessment framework that can support the development and measurement of HOTS in physics practicum learning in higher education.

## Method

This study represents the initial phase of a Research and Development (R&D) project employing the ADDIE

instructional design model, which consists of five stages: Analyze, Design, Develop, Implement, and Evaluate as proposed. However, the scope of this study was limited to the first three stages, namely Analyze, Design, and Develop. This limitation was applied because the primary objective of the study was to identify assessment needs in higher education practicum learning, formulate a preliminary design of a HOTS oriented performance assessment, and examine the content validity of the developed assessment instruments.

At the early stage of instrument development, conceptual clarity and content validity are essential prerequisites before proceeding to the implementation and evaluation stages. Implementing an assessment without a validated framework may result in inaccurate and less meaningful data. Therefore, this study focused on developing the assessment design and establishing content validity as a foundation for further research.

*Research Procedure*

*Analysis stage*

The analysis stage aimed to validate the performance gap and identify the instructional needs for developing a HOTS-oriented performance assessment in physics practicum learning. This stage consisted of five main activities: (1) validating the performance gap, (2) determining instructional goals, (3) confirming the intended audience, (4) identifying required resources, and (5) determining potential delivery systems.

The performance gap was validated through needs analysis, learner analysis, context analysis, task analysis, and assessment analysis. Empirical data were collected from students' essay test scores (midterm and final tests) to examine students' learning achievement and initial HOTS profile. In addition, practicum module analysis, interviews with practicum assistants, and observations of practicum implementation were conducted to identify current assessment practices, scoring procedures, and challenges encountered during practicum activities.

Based on the validated performance gap, the instructional goal of this study was determined, which was to develop a HOTS-oriented performance assessment instrument for physics practicum learning in higher education, focusing on analyzing (C4), evaluating (C5), and creating (C6) skills. The intended audience of this study was undergraduate students enrolled in physics practicum courses. The required resources included practicum modules, essay tests, observation sheets, interview guidelines, scoring rubrics, self-assessment and peer-assessment instruments, and expert validators. The potential delivery system of the assessment was designed to be integrated into practicum learning through performance tasks, observation during

practicum activities, scoring rubrics, self-assessment, and peer-assessment. The results of the analysis stage served as the foundation for the design stage.

*Design Stage*

The design stage aimed to design a HOTS-oriented performance assessment framework based on the results of the analysis stage. This stage consisted of four main activities: (1) design based on the analysis stage, (2) performance assessment framework design, (3) blueprint design of HOTS-oriented performance assessment, and (4) theoretical basis of the design.

At this stage, HOTS indicators were formulated based on the revised Bloom's taxonomy, particularly at the levels of analyzing (C4), evaluating (C5), and creating (C6). Furthermore, HOTS-oriented practicum tasks, scoring rubrics, observation sheets, self-assessment instruments, and peer-assessment instruments were designed. An assessment blueprint was developed to map the relationship between practicum tasks and HOTS indicators to ensure that the integration of HOTS in practicum learning was carried out systematically and consistently. The design was also developed based on the theoretical framework of performance assessment and HOTS, which emphasizes authentic assessment through real tasks, process-oriented assessment, and reflective assessment in practicum learning.



Figure 1. Research flowchart

*Develop stage*

The development stage aimed to refine the assessment instruments and examine their content validity. This stage consisted of three main activities: (1) expert validation, (2) content validity analysis, and (3) revision and prototype development. Expert validation was conducted by experts in pedagogy, educational evaluation, and physics education. The validation covered performance assessment tasks, scoring rubrics, observation sheets, self-assessment instruments, peer-assessment instruments, and HOTS indicators. The validation data were analyzed using Aiken’s V index to determine the content validity of each assessment component. The flowchart of this research as Figure 1.

*Research Subjects*

The subjects of this study were 49 undergraduate students enrolled in a physics education program who participated in practicum courses at a higher education institution. Students were selected as research subjects because they directly experienced the practicum learning process and the assessment system implemented, making them a relevant source of information regarding learning achievement, HOTS abilities, and assessment needs.

In addition to students, three practicum assistants were involved as interview informants. Practicum assistants were selected because they were directly involved in practicum implementation, assessment processes, and technical challenges encountered during practicum assessment.

*Research Instruments*

The instruments used in this study included essay tests to measure student learning outcomes and initial HOTS profiles, practicum module analysis sheets, interview guidelines for practicum assistants, observation sheets for practicum implementation, and expert validation sheets to examine the content validity of the HOTS-oriented performance assessment instruments.

*Data Analysis Techniques*

This study employed both quantitative and qualitative descriptive data analysis. Quantitative data were obtained from students’ essay test scores (midterm and final tests) and expert validation results. Essay test scores were analyzed descriptively to determine students’ learning achievement and to serve as the basis for needs analysis. Expert validation data were analyzed using Aiken’s V index to determine the level of content validity of the assessment instruments. The Aiken’s V formula used in this study is expressed as:

$$V = \frac{s}{[n(c-1)]} \tag{1}$$

Explanation:

S:  $r-l_0$  (the score given by the validator minus the lowest score)

r: evaluator score (scale 1-5)

$l_0$ : lowest score

n: number of validators

c: highest validity assessment score

Expert assessment data is calculated using Aiken’s V. The criteria for assessing the content validity correlation coefficient are presented in Table 1.

**Table 1.** Validity Test Correlation Coefficient (Retnawati, 2016)

| Interval Aiken’s V | Validity Category |
|--------------------|-------------------|
| $V \leq 0.4$       | Low Validity      |
| $0.4 < V \leq 0.8$ | Moderate Validity |
| $V > 0.8$          | High Validity     |

Qualitative data were obtained from practicum module analysis, interviews, and observations. These data were analyzed using thematic analysis to identify patterns in practicum assessment practices, the alignment between practicum tasks and HOTS indicators, and challenges in practicum assessment implementation. The results of the qualitative analysis were used as the foundation for developing the HOTS-oriented performance assessment.

**Result and Discussion**

This study reports the findings from the needs analysis, preliminary design, and initial development stages of a research and development process based on the ADDIE model. The research was conducted in the context of higher education practicum courses, where assessment plays a critical role in supporting students’ cognitive and practical skill development. The presentation of results follows the sequence of stages implemented in this study, namely needs analysis, design, and content validation (Develop stage).

*Analysis Stage*

*Validate the Performance Gap*

The performance gap was validated through needs analysis, learner analysis, context analysis, task analysis, and instructional and assessment analysis.

First, analysis stage, empirical data were collected to obtain an initial overview of students’ learning achievement in the physics practicum course. Students’ performance data were obtained from midterm and final essay examinations administered by the course

instructor as part of routine assessment. The results showed that the average midterm score was 61.2, while the average final test score was 58.5. These results indicate that students' learning achievement tends to be at a moderate level and even shows a slight decrease in the final test. This condition suggests that although students were able to meet the minimum course requirements, their learning outcomes have not yet reached the expected level, particularly in terms of developing Higher Order Thinking Skills (HOTS). The relatively moderate achievement indicates that existing practicum learning and assessment practices may not have fully facilitated students in developing higher-level cognitive skills such as analyzing, evaluating, and creating. This finding is consistent with (Abrahams & Millar, 2008), who argue that laboratory activities do not automatically promote higher-order thinking unless they are supported by appropriate instructional and assessment designs. Therefore, a gap exists between the expected learning outcomes of practicum learning and the actual student performance, indicating the need for developing a more comprehensive HOTS-oriented performance assessment framework.

Second, the learner analysis indicates that most students are capable of following practicum procedures and understanding basic concepts. However, they encounter difficulties when engaging in higher-level cognitive processes such as analyzing data, evaluating experimental procedures, and proposing alternative solutions. This finding aligns with Krathwohl (2002), who emphasizes that HOTS involve complex cognitive processes that require structured and continuous practice. Without appropriate assessment strategies, students tend to remain at lower cognitive levels and do not fully develop higher-order thinking skills.

Third, observations and interviews with lecturers and practicum assistants revealed that practicum learning is conducted using structured modules. However, assessment practices are still predominantly focused on final outcomes rather than the learning process. In addition, formative feedback during practicum activities is limited. This condition is in line with findings from (Zlatkin-Troitschanskaia & Shavelson, 2019), who report that assessment in higher education often prioritizes outcomes over processes. Effective assessment, however, should function as an integral part of learning to support students' cognitive development.

Fourth, the analysis of the practicum module shows that assessment tasks consist of preliminary and final tasks, both in the form of essay questions. Preliminary tasks primarily measure analytical skills at the analyzing level (C4) of the revised Bloom's taxonomy, while higher-level cognitive processes such as evaluating (C5)

and creating (C6) are not yet systematically included. In contrast, the final tasks already incorporate all three HOTS levels (C4-C6). These findings indicate that although HOTS elements are present, they are not consistently integrated throughout the practicum sequence. This lack of cognitive limits students' opportunities to gradually develop higher-order thinking skills. Similar findings were reported by (Indrasari et al., 2022), who emphasize that HOTS development requires continuous and structured integration across learning activities.

Fifth, interviews with lecturers and practicum assistants revealed that formal scoring rubrics have not yet been developed. As a result, assessment decisions are often based on general impressions rather than explicit and transparent performance criteria, increasing the risk of subjectivity and inconsistency. This finding is supported by Reynders et al. (2020), who highlight the importance of clear rubrics in improving the reliability and quality of laboratory assessment. Furthermore, Olson & Krysiak (2021); Villarta et al. (2021); Wongvorachan et al. (2022) emphasize that self-assessment and peer-assessment can enhance students' reflective and metacognitive skills.

In addition, observational data indicate that process-oriented assessment is not systematically implemented during practicum activities. Students mainly follow procedural instructions, and opportunities for reflection and formative feedback are limited. According to Ahmed et al., (2011); Ismailia, (2021), observation-based assessment is essential for capturing important aspects of student performance.

Based on the needs analysis, learner characteristics, learning context, task analysis, and assessment practices, it can be concluded that physics practicum learning in higher education has not yet optimally supported the development of HOTS. This limitation is mainly due to the lack of systematic integration of higher-order tasks, the absence of structured assessment rubrics, and the limited implementation of process oriented and reflective assessment. These findings highlight the need to develop a HOTS-oriented performance assessment framework integrated into practicum learning.

#### *Determine Instructional Goals*

Based on the identified performance gap, the instructional goal of this study is to develop a HOTS-oriented performance assessment instrument for physics practicum learning in higher education. The assessment is designed to facilitate and measure students' higher-order thinking skills, particularly analyzing (C4), evaluating (C5), and creating (C6), through practicum tasks, observation sheets, scoring rubrics, self-assessment, and peer-assessment.

*Confirm the Intended Audience*

The intended audience of this study is undergraduate students enrolled in physics practicum courses in higher education. In this study, the subjects were 49 undergraduate physics education students who were taking practicum courses. In addition, practicum assistants were involved as interview informants because they are directly involved in practicum implementation and assessment practices.

*Identify Required Resources*

The resources required in this study include practicum modules, essay test instruments, observation sheets, interview guidelines, assessment rubrics, self-assessment and peer-assessment instruments, and expert validators. Expert validators are needed to assess the content validity of the developed assessment instruments.

*Determine Potential Delivery Systems*

The assessment instruments developed in this study are designed to be implemented in physics practicum learning. The delivery system of the assessment includes performance tasks, observation during practicum activities, scoring rubrics, self-assessment, and peer-assessment. The instruments are designed to be integrated into existing practicum activities so that they can be implemented without significant changes to the practicum structure.

*Design stage*

Based on the findings of the analysis stage, the main objective of this design stage was to develop a HOTS-oriented performance assessment framework that can support the development of higher-order thinking skills in higher education practicum learning. The assessment design was intended to measure not only students' conceptual understanding, but also HOTS. Therefore, the assessment framework was designed to align learning objectives, HOTS indicators, practicum activities, and assessment strategies within an integrated practicum learning system.

*Design Based on Analysis Stage*

Based on the needs analysis results, several recurring issues were identified in existing practicum assessment practices. These issues, along with the corresponding design recommendations, are summarized in Table 2.

Drawing on these findings, the assessment design was developed based on the principle that assessment should directly respond to identified learning and evaluation needs rather than merely replicate existing assessment formats. Therefore, limitations related to task cognitive levels, the absence of structured rubrics,

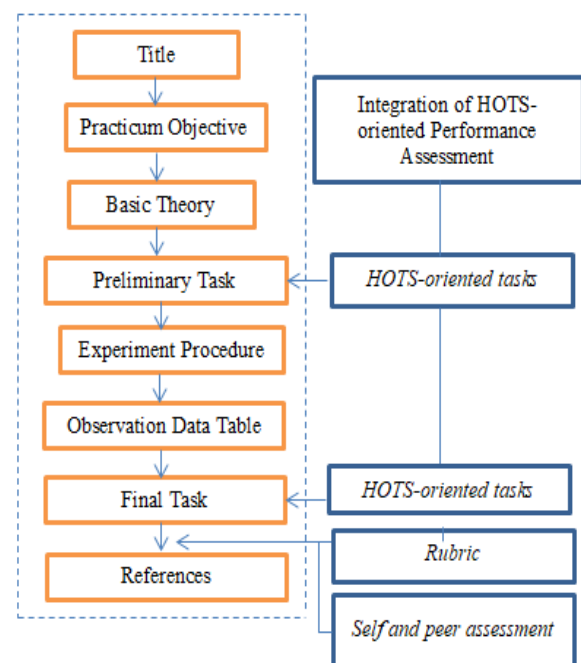
and the lack of process-based assessment were treated as key design priorities.

**Table 2.** Results of Needs Analysis and Design Recommendations

| Identified Issues                           | Recommendations   |
|---|---|
| Practicum tasks are not fully HOTS-oriented | Enrich preliminary and final tasks with HOTS-oriented items |
| Practicum module lacks scoring rubrics      | Develop clear scoring rubrics                               |
| No assessment of practicum process          | Develop observation rubrics for process assessment          |
| No self- and peer-assessment                | Develop self- and peer-assessment instruments               |

*Performance Assessment Framework Design*

The proposed performance assessment framework consists of several interrelated components: HOTS oriented task, analytic scoring rubrics, observation rubrics for assessing practicum processes, self and peer assessment instruments. These components were intentionally embedded within the practicum module so that assessment functions not only as a tool for measuring learning outcomes, but also as an integral part of the learning process. Through this integration, assessment is expected to provide continuous feedback, encourage student reflection, and facilitate the gradual development of higher-order thinking skills throughout practicum activities. The integration of the HOTS-oriented performance assessment within the practicum module is illustrated in Figure 2.



**Figure 2.** Integration design of a HOTS-oriented performance assessment within a higher education practicum module.

*Blueprint Design of HOTS-Oriented Performance Assessment*

To ensure systematic integration of higher-order thinking skills, a blueprint was developed to guide task formulation and assessment development. The blueprint ensures that each practicum component

measures HOTS levels, including analyzing (C4), evaluating (C5), and creating (C6). Table 4 is blueprint was used to guide task formulation and to ensure systematic integration of higher-order cognitive processes throughout practicum activities.

**Table 3.** Blueprint of HOTS-Oriented Practicum Tasks

| Practicum Task Component | Task Description  | HOTS Indicator  |
|--------------------------|---|-----------------|
| Preliminary Task         | Analyzing practicum objectives, identifying relevant variables, and explaining relationships between variables based on underlying theory | Analyzing (C4)  |
|                          | Evaluating the suitability of experimental procedures and predicting possible sources of error before conducting the experiment           | Evaluating (C5) |
|                          | Proposing modifications or alternative experimental designs to achieve the practicum objectives   | Creating (C6)   |
| Observed Performance     | Analyzing real-time experimental data and identifying inconsistencies or anomalies during data collection                                 | Analyzing (C4)  |
|                          | Evaluating the accuracy of measurements, use of instruments, and procedural decisions during the experiment                               | Evaluating (C5) |
|                          | Developing alternative strategies or solutions when encountering experimental difficulties  | Creating (C6)   |
| Final Task               | Analyzing experimental results and interpreting findings in relation to theoretical concepts  | Analyzing (C4)  |
|                          | Evaluating the validity of results, sources of error, and limitations of the experiment   | Evaluating (C5) |
|                          | Designing improved or extended experiments based on findings and reflections  | Creating (C6)   |
| Observation Rubric       | Analyzing students' ability to interpret procedures, data trends, and experimental phenomena during practicum                             | Analyzing (C4)  |
|                          | Evaluating students' decision-making, accuracy, and adherence to scientific procedures  | Evaluating (C5) |
|                          | Identifying creative problem-solving actions and adaptive strategies during practicum   | Creating (C6)   |
| Self-Assessment          | Analyzing personal understanding and performance throughout the practicum process   | Analyzing (C4)  |
|                          | Evaluating strengths, weaknesses, and learning progress based on rubric criteria  | Evaluating (C5) |
|                          | Formulating improvement plans and alternative learning strategies for future practicum activities   | Creating (C6)   |
| Peer-Assessment          | Analyzing peers' practicum performance using structured criteria  | Analyzing (C4)  |
|                          | Evaluating peers' experimental procedures, collaboration, and problem-solving approaches  | Evaluating (C5) |
|                          | Providing constructive suggestions and alternative solutions to improve peers' performance  | Creating (C6)   |

*Theoretical Basis of the Design*

From a theoretical perspective, the design of the HOTS-oriented performance assessment in this study is grounded in constructivist learning theory and authentic assessment theory, which emphasize that learning is an active process in which students construct knowledge through meaningful activities and reflection on their experiences. In this perspective, assessment should not only function as a tool to measure learning outcomes, but also as an integral part of the learning process that facilitates cognitive development, reflection, and skill improvement. Therefore, assessment

must be aligned with learning objectives, learning activities, and expected competencies (Brown & Mevs, 2012; Daly et al., 2010; Gale et al., 2016; Schildkamp et al., 2020; Tan & Leong, 2015; Villarta et al., 2021).

Performance assessment is a form of authentic assessment that requires students to demonstrate their knowledge and skills through real or simulated tasks that reflect actual practices in a discipline. Performance assessment is designed to measure complex competencies such as problem-solving, critical thinking, decision-making, and the ability to apply knowledge in practical contexts (Gale et al., 2016; Wested & Shavelson,

1997; Zlatkin-Troitschanskaia & Shavelson, 2019). This approach is particularly relevant in practicum learning because practicum activities involve procedural skills, conceptual understanding, data analysis, and problem-solving processes. Therefore, performance assessment is considered more appropriate than traditional paper-and-pencil tests for measuring learning outcomes in practicum-based courses.

In addition, the design of this assessment is based on the theoretical framework of Higher Order Thinking Skills (HOTS), which refers to higher-level cognitive processes in the revised Bloom's taxonomy, namely analyzing (C4), evaluating (C5), and creating (C6) (Krathwohl, 2002). These cognitive processes involve complex thinking skills such as reasoning, decision-making, problem-solving, and the ability to generate new ideas or solutions (Khadka et al., 2025; Shavelson et al., 2018; Ubaidillah et al., 2022). In higher education, HOTS are essential because students are expected not only to understand concepts but also to critically evaluate information, solve complex problems, and develop innovative solutions. In the context of physics practicum learning, HOTS are reflected in students' ability to analyze experimental data, evaluate experimental procedures and results, and design alternative experiments or solutions to experimental problems.

The integration of performance assessment and HOTS is theoretically supported by the principle of constructive alignment, which states that learning objectives, learning activities, and assessment must be aligned to support meaningful learning (Indrasari et al., 2022). If practicum learning aims to develop higher-order thinking skills, then the assessment system must also be designed to measure and facilitate those skills. Therefore, the assessment tasks, scoring rubrics, observation sheets, self-assessment, and peer-assessment developed in this study were systematically designed to reflect HOTS indicators at each stage of practicum activities, including preliminary tasks, observed performance during practicum, and final tasks.

Furthermore, the use of analytic scoring rubrics, observation-based assessment, self-assessment, and peer-assessment is supported by assessment for learning theory, which emphasizes the role of formative assessment and feedback in improving student learning (Arnold, 2022; Black & Wiliam, 2009; Schildkamp et al., 2020). Observation rubrics allow lecturers and practicum assistants to assess students' performance

processes, while self-assessment and peer-assessment encourage reflection, metacognition, and students' responsibility for their own learning (Bhutta et al., 2019; Tan & Leong, 2015). Through this approach, assessment does not only measure learning but also supports learning and the development of higher-order thinking skills.

The HOTS-oriented performance assessment design developed in this study is based on several interrelated theoretical foundations: authentic assessment theory, performance assessment theory, Higher Order Thinking Skills theory, constructive alignment, and assessment for learning. The integration of these theoretical perspectives provides a strong conceptual foundation for developing a comprehensive assessment framework that is capable of measuring and facilitating higher-order thinking skills in higher education practicum learning.

#### *Development stage*

In the ADDIE model, the development stage focuses on producing and refining instructional products that have been designed in the previous stage. Development phase includes developing assessment instruments, conducting expert validation, revising the prototype, and preparing the product for implementation. Therefore, in this study, the development stage focused on refining the HOTS-oriented performance assessment and examining its content validity through expert judgment.

Content validation is a crucial step in development research because it ensures that the developed instruments are theoretically sound, pedagogically appropriate, and aligned with learning objectives before being implemented in real classroom settings. This step is widely recommended in educational measurement research to ensure that assessment instruments adequately represent the intended constructs (Reynders et al., 2020; Panadero et al., 2018).

#### *Content Validity Results*

The results of the content validity analysis for each assessment component are presented in Table 4.

As shown in Table 4, the performance assessment tasks obtained an Aiken's V value of 0.83, indicating a high level of content validity. This result suggests that the tasks were considered relevant to practicum learning objectives and sufficiently representative of higher-order cognitive processes, particularly analyzing, evaluating, and creating. The experts' evaluations indicate that the task demands are aligned with authentic laboratory activities, supporting their potential use as measures of students' performance in

practicum contexts. This finding is consistent with previous studies showing that performance-based tasks can effectively measure higher-order thinking and scientific reasoning in laboratory learning (Reynders et al., 2020).

**Table 4.** Content Validity Results of the HOTS-Oriented Performance Assessment

| Assessment Component         | Aiken's V | Validity Category |
|------------------------------|-----------|-------------------|
| Performance assessment tasks | 0.83      | High Validity     |
| Scoring rubric               | 0.94      | High Validity     |
| Self-assessment instrument   | 0.90      | High Validity     |
| Peer-assessment instrument   | 0.90      | High Validity     |
| HOTS indicators              | 0.88      | High Validity     |
| Overall mean                 | 0.89      | High Validity     |

Among all components, the scoring rubric achieved the highest validity index, with an Aiken's V value of 0.94, which falls within the very high validity category. This finding reflects strong agreement among experts regarding the clarity and consistency of the rubric criteria and performance descriptors. In practicum-based assessment, rubrics play an essential role in reducing subjectivity, supporting consistent scoring, and communicating performance expectations to students (Bhutta et al., 2019).

The self and peer-assessment instruments each obtained an Aiken's V value of 0.90, indicating very strong content validity. These results suggest that both instruments are appropriate for encouraging student reflection and metacognitive awareness. The inclusion of self- and peer-assessment supports the concept of assessment as learning, in which students are actively involved in monitoring and evaluating their own learning processes. Previous studies have shown that self- and peer-assessment can improve students' metacognitive skills, responsibility for learning, and critical thinking (Yan et al., 2023).

The HOTS indicators yielded an Aiken's V value of 0.88, indicating strong validity. This result confirms that the indicators are conceptually aligned with the higher cognitive levels of the revised Bloom's taxonomy, namely analyzing (C4), evaluating (C5), and creating (C6). Such alignment is important because these indicators serve as the conceptual foundation for task construction and rubric development within the HOTS-oriented performance assessment framework. Alignment between learning objectives, assessment tasks, and cognitive levels is a key principle in assessment design to ensure construct validity (Ichsan et al., 2020; Ramos et al., 2013).

### *Revision and the Prototype*

Based on expert suggestions and validation results, several revisions were made to improve the clarity of task instructions, rubric descriptors, and assessment criteria. This revision process is an essential part of the development stage because instructional products must be refined iteratively before implementation. The final output of this development stage is a validated prototype of a HOTS-oriented performance assessment consisting of: HOTS-oriented practicum tasks, analytic scoring rubrics, observation rubrics, self and peer assessment. This prototype is considered conceptually and pedagogically appropriate for practicum learning and is ready for the implementation stage.

Overall, the HOTS-oriented performance assessment achieved an average Aiken's V score of 0.89, which indicates a very good level of content validity. These findings demonstrate that the developed assessment instrument meets essential conceptual and pedagogical requirements and is suitable for progression to the implementation stage. However, this study is limited to expert-based content validation at the development stage. Therefore, further research is necessary to examine inter-rater reliability, practicality, and empirical effectiveness of the assessment in enhancing students' Higher Order Thinking Skills in higher education practicum learning.

### **Conclusion**

The results of this study indicate that practicum learning in higher education has not fully supported the development of Higher Order Thinking Skills (HOTS). Existing practicum modules and assessment practices still tend to focus primarily on analytical tasks and final outcome assessment, and have not systematically integrated evaluating (C5) and creating (C6) skills. In addition, process-oriented assessment and formative feedback have not been optimally implemented. Therefore, a more comprehensive assessment approach is needed to measure cognitive abilities, procedural skills, and reflective abilities in an integrated manner through performance assessment. Based on this need, this study developed a HOTS-oriented performance assessment framework integrated into practicum learning, which consists of HOTS-based practicum tasks, observation instruments, self-assessment and peer-assessment instruments, as well as structured scoring rubrics. The results of expert validation indicate that the developed instrument has high content validity and is conceptually and pedagogically appropriate for use in higher education practicum learning. This study contributes theoretically by providing a conceptual

framework for integrating HOTS into performance assessment in practicum learning. Practically, this study provides a validated assessment prototype that can be used by lecturers in physics practicum courses. However, this study is limited to content validation; therefore, future studies should investigate the reliability, practicality, and effectiveness of the assessment through field implementation.

#### Acknowledgments

Thank you to all those who have supported this research. Hopefully, this research will be useful.

#### Author Contributions

E., F.N., R.D.: preparation of the initial draft HOTS oriented performance assessment, results, discussion of methodology, analysis, conclusions; M.L., N.W., L.H: Participation in the validation process.

#### Funding

The author would like to thank the Institute Research and Community Services Universitas Negeri Padang for funding this work with a contract number: 1125/UN35.15/LT/2023.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### References

- Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1945–1969. <https://doi.org/10.1080/09500690701749305>
- Ahmed, K., Miskovic, D., Darzi, A., Athanasiou, T., & Hanna, G. (2011). Observational tools for assessment of procedural skills: A systematic review. *American Journal of Surgery*, 202, 469–480.e6. <https://doi.org/10.1016/j.amjsurg.2010.10.020>
- Appiah, M., & van Tonder, F. (2018). E-Assessment in Higher Education: A Review. *International Journal of Business Management and Economic Research*, 9(6), 1454–1460. Retrieved from <http://www.ijbmer.com/docs/volumes/vol9issu e6/ijbmer2018090601.pdf>
- Arnold, J. (2022). Prioritising students in Assessment for Learning: A scoping review of research on students' classroom experience. *Review of Education*, 10(3), 1–36. <https://doi.org/10.1002/rev3.3366>
- Bhutta, S. M., Chauhan, S., Ali, S. K., Gul, R., Cassum, S., & Khamis, T. (2019). Developing a rubric to assess critical thinking in a multidisciplinary context in higher education. *SOTL in the South*, 3(1), 6–25. <https://doi.org/10.36615/sotls.v3i1.69>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Bland, L. M., & Gareis, C. R. (2018). Performance Assessments: A Review of Definitions, Quality Characteristics, and Outcomes Associated with Their Use in K-12 Schools. *Teacher Educators' Journal*, 11, 52–69. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1174728.pdf>
- Braun, E. (2021). Performance-based assessment of students' communication skills. *International Journal of Chinese Education*, 10(1). <https://doi.org/10.1177/22125868211006202>
- Braun, H. I., Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Borowiec, K. (2020). Performance Assessment of Critical Thinking: Conceptualization, Design, and Implementation. *Frontiers in Education*, 5(September), 1–10. <https://doi.org/10.3389/feduc.2020.00156>
- Brown, C., & Mevs, P. (2012). *Quality Performance Assessment: Harnessing the Power of Teacher and Student Learning*. 1–6. Retrieved from <https://www.qualityperformanceassessment.org>
- Calma, A., & Davies, M. (2025). Studies in Higher Education Assessing students' critical thinking abilities via a systematic evaluation of essays. *Studies in Higher Education*, 1–16. <https://doi.org/10.1080/03075079.2025.2470969>
- Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010). Exploring formative e-assessment: Using case stories and design patterns. *Assessment and Evaluation in Higher Education*, 35(5), 619–636. <https://doi.org/10.1080/02602931003650052>
- Dusdal, J. (2019). Higher education in Germany – recent developments in an international perspective. In *Comparative Education*, 55(2). <https://doi.org/10.1080/03050068.2019.1575023>
- Gale, J., Wind, S., Koval, J., Dagosta, J., Ryan, M., & Usselman, M. (2016). Simulation-based performance assessment: an innovative approach to exploring understanding of physical science concepts. *International Journal of Science Education*, 38(14), 2284–2302. <https://doi.org/10.1080/09500693.2016.1236298>
- Hartina, L., Rosidin, U., & Suyatna, A. (2019). Pengaruh Penerapan Instrumen Performance Assessment pada Pembelajaran IPA Berbasis Laboratorium Real terhadap Hasil Belajar Siswa. *Jurnal Penelitian Pendidikan IPA*, 6(1), 25. <https://doi.org/10.29303/jppipa.v6i1.299>
- Hyytinen, H., & Toom, A. (2019). Developing a performance assessment task in the Finnish higher education context: Conceptual and empirical

- insights. *British Journal of Educational Psychology*, 89(3), 551–563. <https://doi.org/10.1111/bjep.12283>
- Ichsan, I. Z., Hasanah, R., Ristanto, R. H., Rusdi, R., Cahapay, M. B., Widiyawati, Y., & Rahman, M. M. (2020). Designing an Innovative Assessment of HOTS in the Science Learning for the 21st Century. *Jurnal Penelitian Dan Pembelajaran IPA*, 6(2), 211. <https://doi.org/10.30870/jppi.v6i2.4765>
- Indrasari, W., Fitriani, W., & Budi, A. S. (2022). The Level of Higher-Order Thinking Skills in Basic Physics Practicum. *Journal of Physics: Conference Series*, 2377(1). <https://doi.org/10.1088/1742-6596/2377/1/012078>
- Ismailia, T. (2021). Performance assessment using rubric to improve students' speaking skill. *Journal of Applied Linguistics and Literacy*, 5(2), 65–82. <https://doi.org/10.25157/jall.v5i2.5786>
- Khadka, J., Dahal, N., Acharya, U., Puri, G., Subedi, N., & Hasan, K. (2025). Higher-order thinking skills in e-learning contexts in higher education: a phenomenological study. *Frontiers in Education*, 10, 1–13. <https://doi.org/10.3389/feduc.2025.1555541>
- Komisia, F., Tukan, M. B., Yanni, D., Fallo, A., Jenisa, G., Hale, P., Meta, G. B., Lanmai, P. O., Eflani, F., Aloisia, M., & Leba, U. (2025). *Validity Analysis of a Science Process Skills Assessment Instrument for High School Students in Chemistry*. 11(12), 185–193. <https://doi.org/10.29303/jppipa.v11i12.12860>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212–218. [https://doi.org/10.1207/s15430421tip4104\\_2](https://doi.org/10.1207/s15430421tip4104_2)
- Olson, J. M., & Krysiak, R. (2021). Rubrics as tools for effective assessment of student learning and program quality. *Curriculum Development and Online Instruction for the 21st Century, January 2021*, 173–200. <https://doi.org/10.4018/978-1-7998-7653-3.ch010>
- Pine, J., & Baxter, G. P. (1991). Performance Assessment in Science. *Applied Measurement in Education*, 4(4), 347–362. [https://doi.org/10.1207/s15324818ame0404\\_7](https://doi.org/10.1207/s15324818ame0404_7)
- Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2013). Higher Order Thinking Skills and Academic Performance in Physics of College Students: A Regression Analysis. *International Journal of Innovative Interdisciplinary Research*, 4, 48–60. <https://doi.org/ISSN 1839-9053>
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian*. Yogyakarta: Parama Publishing.
- Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education*, 7(1). <https://doi.org/10.1186/s40594-020-00208-5>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103(April), 101602. <https://doi.org/10.1016/j.ijer.2020.101602>
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., & Marino, J. P. (2019). Assessment of University Students' Critical Thinking: Next Generation Performance Assessment. *International Journal of Testing*, 19(4), 337–362. <https://doi.org/10.1080/15305058.2018.1543309>
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Mariño, J. P. (2018). International Performance Assessment of Learning in Higher Education (iPAL): Research and Development. *Methodology of Educational Measurement and Assessment, March*, 193–214. [https://doi.org/10.1007/978-3-319-74338-7\\_10](https://doi.org/10.1007/978-3-319-74338-7_10)
- Tan, K., & Leong, W. S. (2015). Assessment rubric for learning. *Assessment and Learning, August 2014*, 115–128. Retrieved from <https://www.researchgate.net/publication/270958483>
- Ubaidillah, M., Marwoto, P., & Subali, B. (2022). *Higher Order Thinking Laboratory (HOT Lab) -Based Physics Learning: A Systematic Literature Review*. 96–107. <https://proceeding.unnes.ac.id/ISET/article/view/1735>
- Villarta, Y. M., Atibula, L. T., & Gagani, F. S. (2021). Performance-based assessment: Self-efficacy, decision-making, and problem-solving skills in learning science. *Journal La Edusci*, 2(3), 1–9. <https://doi.org/10.37899/journallaedusci.v2i3.381>
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83(November), 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Wested, G. S. F., & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice*, 16(3), 16–24. <https://doi.org/10.1111/j.1745-3992.1997.tb00596.x>
- Wongvorachan, T., Bulut, O., Tsai, Y.-S., & Lindner, M. A. (2022). Improving Student Feedback Literacy in e-Assessments: A Framework for the Higher

- Education Context. *Trends in Higher Education*, 1(1), 16–29.  
<https://doi.org/10.3390/higheredu1010002>
- Yan, Z., Panadero, E., Wang, X., & Zhan, Y. (2023). A Systematic Review on Students' Perceptions of Self - Assessment: Usefulness and Factors Influencing. *Educational Psychology Review*, 35(3), 1–28. <https://doi.org/10.1007/s10648-023-09799-1>
- Zlatkin-Troitschanskaia, O., & Shavelson, R. J. (2019). Advantages and challenges of performance assessment of student learning in higher education. *British Journal of Educational Psychology*, 89(3), 413–415.  
<https://doi.org/10.1111/bjep.12314>
- Zlatkin-Troitschanskaia, O., Toepper, M., Pant, H. A., Lautenbach, C., & Kuhn, C. (Eds.). (2018). Assessment of learning outcomes in higher education: Cross-national comparisons and perspectives. Springer.  
<https://doi.org/10.1007/978-3-319-74338-7>