



The Role of Teachers' Pedagogical Competence in Designing Deep Learning With Hots Assessment on Students' Critical Thinking Skills in Elementary School

Endang Setiyowati¹, Ari Metalin Ika Puspita¹, Neni Mariana^{1*}

¹ Faculty of Education, State University of Surabaya, Surabaya, Indonesia

25010855150@mhs.unesa.ac.id, marethadellarosa@unesa.ac.id, aripuspita@unesa.ac.id

corresponding author: 25010855150@mhs.unesa.ac.id

ABSTRACT

This study seeks to describe the role of teachers' pedagogical competence in designing deep learning that integrates Higher Order Thinking Skills (HOTS) assessment on critical thinking skills of elementary students. The research uses an exploratory sequential mixed-methods approach, with three teachers in grades 4, 5, and 6 at UPT SDN 232 Gresik as research subjects, and involved 78 students. Data collection was conducted through learning observations, in-depth interviews, analysis of Deep Learning Plan (RPM) documents, and HOTS assessments. The results reveal that teachers' pedagogical competence in designing deep learning was good, with variation across classes: Grade 4 and Grade 5 teachers demonstrated strong competence with consistent integration of HOTS, whereas the Grade 6 teacher indicated a significant gap between conceptual planning and classroom implementation. Teachers' ability to compile HOTS assessments also varied: the Grade 5 teacher produced the most balanced instrument, with 10 of 15 items at the C4-C6 level, while the Grade 6 teacher designed all 9 items at C4-C6 but with poor alignment between the teaching content and the assessment difficulty. Students' critical thinking skills showed significant variation across classes, with an overall average score of 75.72 (Good category): Grade 5 achieved the highest average of 86.86 (Very Good), Grade 4 achieved 75.74 (Good), and Grade 6 achieved the lowest at 64.58 (Adequate). This research recommends continuous training for teachers in designing authentic HOTS assessments and in ensuring alignment between lesson planning, classroom scaffolding, and assessment demands.

Keywords: critical thinking skills, deep learning, elementary school, HOTS assessment, pedagogical competence

Submitted	Accepted	Published
20 May 2026	20 July 2026	30 July 2026

Citation	:	Setiyowati, E., Puspita, A.M.I., & Mariana, N. (2026). The Role Of Teachers' Pedagogical Competence In Designing Deep Learning With Hots Assessment On Students' Critical Thinking Skills In Elementary School. <i>Jurnal PAJAR (Pendidikan dan Pengajaran)</i> , 10(4), 683-692. DOI: http://dx.doi.org/10.33578/pjr.v10i4.1561 .
----------	---	--

INTRODUCTION

Critical thinking skills are among the essential competencies that must be developed from the elementary school level to prepare students for the challenges of the 21st century. Facione (1990) defines critical thinking as a deliberate, self-regulatory process of judgment that involves interpretation, analysis, evaluation, and inference, accompanied by a clear explanation of the evidence, concepts, methods, and criteria that inform such judgments. Results from the Program for International Student Assessment (PISA) 2022 reveal that Indonesian students' average mathematics score was 366, far below the OECD average of 472, with approximately 82 percent of students performing below Level 2. This reflects a broader weakness in critical-thinking capacities, particularly in analysis, evaluation, and inference skills that should have been developed since elementary education (OECD, 2023).

Local research elucidates similar patterns. Studies on elementary school students (aged 7-12 years) in Java and Kalimantan reported average critical thinking scores ranging from 45 to 62.5, placing them in the low-to-medium category. A study at SDN 12 Engkurai (Melawi Regency, West Kalimantan) reported a mean

critical thinking score of only 45 for Grade 5 students in mathematics, with the inference aspect scoring lowest at 30 to 35 and evaluation at 34 to 40, indicating persistent difficulty with logical reasoning and evidence assessment (Charles et al., 2024). Empirical data from the 2025 National Assessment at UPT SDN 232 Gresik also showed that while 93.75 percent of students met the minimum numeracy competency, the average numeracy score was 68.2, with reasoning skills at 65.2 and critical reasoning skills at 57.14, indicating that higher-order analytical capabilities remain underdeveloped in most students.

Teachers' suboptimal pedagogical competence has been identified as a primary factor behind this condition. Harun et al. (2022) found that elementary teachers tend to focus on procedural content mastery and factual tasks at cognitive levels C1 to C3, with limited holistic integration of HOTS assessment. Daga et al. (2022) similarly noted that teachers frequently rely on conventional approaches such as lectures and memorization, which prioritize procedural knowledge over analysis, evaluation, and creation at levels C4 to C6. Febriana et al. (2020) found that although teachers had produced HOTS-based lesson plans, classroom practice remained dominated by lecture methods and recall questions, making it difficult for students to respond to tasks requiring analytical and evaluative thinking.

Deep Learning has emerged as a promising approach to addressing low levels of critical thinking. Fullan et al. (2018) define deep learning as a process that develops six global competencies (6C): character, citizenship, collaboration, communication, creativity, and critical thinking. Deep learning is grounded in three mutually reinforcing principles: mindful, meaningful, and joyful. These three principles must be present simultaneously so that learning not only enhances high cognitive abilities but also shapes character and provides a pleasurable and relevant learning experience for students (Kemdikbudristek, 2025).

HOTS assessment plays a strategic role as the primary means of training and measuring students' critical thinking skills. According to Anderson and Krathwohl (2001), HOTS encompasses analyzing, evaluating, and creating, which correspond to Facione's critical thinking dimensions of analysis, evaluation, and inference. This study focuses specifically on these three critical thinking indicators, as they represent the primary intersection of HOTS and critical thinking and are measurable and relevant to the content of the area and perimeter of plane figures. Through authentic tasks that require contextual problem-solving, evidence assessment, and the development of innovative solutions, HOTS assessment serves not only as a measurement tool but also as a learning process that encourages students to think critically from an early age. Yusella et al. (2023) developed a HOTS-based assessment instrument proven to be valid, practical, reliable, and effective in improving students' critical thinking scores.

The urgency of this research lies in the limited literature at the elementary level that comprehensively examines the role of teachers' pedagogical competence in designing deep learning that integrates HOTS assessment to measure critical thinking skills. Most previous research has focused on secondary levels or above, or has studied these three elements (pedagogical competence, deep learning design, HOTS assessment) separately rather than as an integrated system. This study fills that gap by simultaneously examining all three elements within a single holistic framework at the elementary school level.

Based on the above background, this study aims to describe teachers' pedagogical competence in designing deep learning integrating HOTS assessment, to analyze teachers' ability to compile HOTS assessments, and to measure elementary students' critical thinking skills based on the results of teacher-designed HOTS assessments implemented in deep learning on the topic of area and perimeter of plane figures.

METHOD

This study employs an exploratory sequential mixed methods approach that combines qualitative and quantitative methods sequentially (Creswell and Plano Clark, 2018). The first phase is a qualitative case study exploring teachers' pedagogical competence in designing deep-learning activities and in developing HOTS assessments. The second phase was a descriptive quantitative approach to measure students' critical thinking skills based on HOTS assessment results. The third phase integrated qualitative and quantitative findings to

generate a holistic understanding of the role of teachers' pedagogical competence in developing students' critical thinking skills.

The study was conducted at UPT SDN 232 Gresik during the even semester of the 2025/2026 academic year, with three meetings per class. Research subjects were three classroom teachers in grades 4, 5, and 6, all holding S1 qualifications in Elementary School Teacher Education, along with 78 students, comprising 34 in Grade 4, 28 in Grade 5, and 16 in Grade 6. The school was selected because it has implemented the Merdeka Curriculum, has an active principal and teacher support, and allows intensive classroom observation.

Qualitative data collection was carried out through four techniques: learning observation, in-depth interviews, a pedagogical competence questionnaire, and document analysis. Learning observations were conducted across three meetings per teacher, using a structured observation instrument that covered 11 aspects. In-depth interviews were conducted after each learning session to explore the process of designing deep learning, challenges encountered, and teachers' reflections on their own pedagogical competence. The pedagogical competence questionnaire contained 18 open questions measuring three core competencies: organizing meaningful learning, utilizing ICT for learning, and conducting learning assessment. Document analysis was conducted on the Deep Learning Plans (RPM), Student Worksheets (LKPD), and HOTS assessment instruments designed by the teachers.

Quantitative data were collected through HOTS assessments administered to all students after completing learning sessions on the area and perimeter of plane figures. HOTS assessment instruments were designed by each teacher with reference to Facione's (2015) critical thinking indicators covering analysis, evaluation, and inference at cognitive levels C4-C6 of the Revised Bloom's Taxonomy. Each item was designed around a contextual stimulus relevant to students' lives and was accompanied by an analytic scoring rubric on a 0-4 scale.

Qualitative data analysis used thematic analysis with an inductive approach, adopting Miles, Huberman, and Saldana's (2014) interactive model through four components: data collection, data reduction, data presentation in thematic matrix form, and conclusion drawing with verification. The validity of findings was maintained through source triangulation, method triangulation, time triangulation, member checking, and audit trail. Quantitative data analysis was conducted descriptively to assess the quality of HOTS assessment items based on content validity, difficulty level, discrimination index, and scoring rubric quality. Students' critical thinking skills were categorized as follows: Very Good (86-100), Good (76-85), Adequate (60-75), Poor (55-59), and Very Poor (below 54).

RESULTS AND DISCUSSION

Teachers' Pedagogical Competence in Designing Deep Learning

Analysis of teachers' pedagogical competence in designing deep learning reveals variation in quality across the three classes, with an overall good category. Findings were obtained through data triangulation from RPM observation, learning observation, in-depth interviews, and the pedagogical competence questionnaire completed by the three teachers in Grades 4, 5, and 6. Analysis was conducted on 11 aspects of deep learning covering learner identification, material identification, graduate profile dimensions, learning achievement and learning objectives, learning topic, pedagogical practice, learning partnerships, learning environment, utilization of digital technology, learning steps, and assessment instrument appendices.

The following table presents the results of the RPM quality analysis compiled by the three teachers based on the deep learning framework. The table illustrates that the Grade 4 teacher (Teacher A) and the Grade 5 teacher (Teacher B) demonstrate strong and consistent deep learning planning, while the Grade 6 teacher (Teacher C) shows notable weaknesses, particularly in learner identification, learning environment, and digital technology utilization.

Table 1. Quality Analysis of Deep Learning Plans (RPM)

Assessment Aspects	Teacher A (Grade 4)	Teacher B (Grade 5)	Teacher C (Grade 6)	Average
Learner Identification	4	4	2	3.3
Material Identification	4	3	3	3.3
Graduate Profile Dimensions	4	4	3	3.7
Learning Achievement and Objectives	4	4	3	3.7
Learning Topic	4	3	3	3.3
Pedagogical Practice	4	3	3	3.3
Learning Partnership	2	2	2	2.0
Learning Environment	3	3	2	2.7
Digital Technology Utilization	3	3	2	2.7
Learning Steps	4	4	3	3.7
Appendix (LKPD and Assessment)	4	3	2	3.0

Note: Assessment scale 1-4 (1=Poor, 2=Adequate, 3=Good, 4=Very Good) Source: Research Data (2026)

The table above illustrates that Teacher A and Teacher B demonstrate strong competence across most aspects, particularly in formulating learning objectives, identifying learner characteristics, and designing structured learning steps consistent with the principles of the Merdeka Curriculum. Both teachers explicitly conduct diagnostic assessments at the start of learning to map students' prior knowledge, learning styles, and individual needs. Teacher B states: "I design learning objectives in three interconnected layers: first, so that students truly understand the concept through simulation and discussion; second, therefore, they can apply the calculation in real life, for example, calculating the fencing needs of a field or carpet for a room; and third, hence, students perform a brief reflection at the end of learning." This approach reflects the deep learning principle of mindfulness, in which students are encouraged to be conscious of their own learning process (Kemdikbudristek, 2025).

Teacher C, on the other hand, underscores a significant gap between conceptual understanding and planning practice. Learning objectives in the RPM were general, lacked the ABCD structure (Audience, Behavior, Condition, Degree), and were not consistently oriented toward high-order cognitive levels. Teacher C acknowledges: "In the RPM document I prepared, the objectives were formulated with AI assistance, and I thought the key points were met, but I forgot to write the ABCD elements completely; hence, the formulation was less systematic." Furthermore, although Teacher C actively uses the PhET Simulation in classroom sessions, this is not documented in the RPM, indicating a gap between planning and implementation.

Learning partnership receives the lowest average score of 2.0 across all three teachers because none of the teachers has optimally involved external partners, such as parents, community members, or professionals, in the learning process. Parents' involvement remains limited to routine communication, with no active role in supporting deep learning. This finding underscores the need to strengthen learning partnerships, as emphasized in the deep learning framework, where collaboration with diverse stakeholders can enrich students' learning experiences (Fullan et al., 2018).

Digital technology use varied across the three classes. Teacher B is the most systematic, actively utilizing YouTube videos and PhET Area Builder simulation not merely as presentation media but as tools for conceptual exploration and visualization. Teacher C also uses the PhET simulation during classroom sessions, but does not document it in the RPM. Teacher A is the most limited in digital technology use, relying primarily on conventional materials such as grid paper and LKPD. This finding aligns with Daga et al. (2022), who found that ICT integration in learning requires continuous teacher training to use technology strategically in support of deep learning.

Learning models chosen by each teacher support deep learning principles. Teacher A applies Discovery Learning with grid-paper manipulatives to help students concretely discover the concepts of area

and perimeter. Teacher B consistently applies Problem-Based Learning (PBL) with five structured stages that direct students to analyze problems, discuss, collaborate, and present solutions. Teacher C applies a simplified Deep Learning approach with three stages (observation, analysis, contextual application) without implementing the full PBL syntax, which contributes to lower student engagement during classroom sessions compared to Grades 4 and 5.

This finding aligns with Sa'diyah et al. (2024), who found that PBL has a significant positive effect on elementary students' critical thinking development, especially when designed with authentic and contextual problems. Harun et al. (2022) also found that teachers with good pedagogical competence in the design, implementation, and evaluation of HOTS-based learning contribute significantly to the development of higher-order thinking skills in elementary students.

Teachers' Ability to Compile HOTS Assessments

Analysis of teachers' ability to compile HOTS assessments reveals variation in quality across the three classes. All three teachers successfully designed items at cognitive levels C4 to C6 aligned with Facione's critical thinking indicators and the Merdeka Curriculum learning achievement. The following table presents the findings from the HOTS evaluation instrument analysis for each class.

Table 2. Summary of HOTS Assessment Findings by Class

Aspects	Teacher A (Grade 4)	Teacher B (Grade 5)	Teacher C (Grade 6)
Total items/sub-items	12 sub-items	15 sub-items	9 sub-items
Items at HOTS level (C4-C6)	7 out of 12 (58%)	10 out of 15 (67%)	9 out of 9 (100%)
Distribution of HOTS levels	C4-C5 dominant	C5-C6 strong with balanced distribution	C4-C6 balanced
Authentic contextual use	Yes	Yes	Yes
Consistency with CP and Graduate Profile	Good	Good	Good, but the gap between the taught content and the assessment
Scoring rubric quality	Consistent C4/C5/C6 criteria	Systematic 4-level rubric	Present but not for all items
Main strength	Scaffolded, step-by-step approach	Balanced, comprehensive, contextual HOTS	Strongest instrument design
Main challenge	Some C3 routine items remain	Some C3 items need strengthening	Misalignment between teaching content and evaluation

Source: Research Data (2026)

The authentic and contextual use of assessment is a primary strength across all three teachers. Teacher A designs items in the context of "a living room where Mrs. Rina would install tiles and carpet," Teacher B uses the context of "land to be planted with trees and Mr. Budi's garden with area comparisons," and Teacher C designs items around "a square park with a circular pool in the center." The use of authentic contexts aligns with the deep learning principle of meaningfulness, where students can perceive the relevance of mathematics to daily life (Kemdikbudristek, 2025).

Teacher B produces the most balanced and comprehensive assessment instrument. Out of 15 sub-items, 10 are at level C4-C6, with Level C6 (Create) appearing strongly in five items and C5 (Evaluate) in four items. The LKPD is equipped with a systematic rubric that measures four aspects of performance: understanding the problem information, determining the correct formula, the calculation process, and the final answer. Teacher B explains: "I design items based on real contexts with stimuli such as texts, images, or tables that require students to think critically and analytically." This practice aligns with the findings of Yusella et al. (2023) that well-designed HOTS assessment instruments function not only as measurement tools but also as learning processes that encourage critical thinking.

Teacher C designs the assessment with the highest level of HOTS orientation, with all 9 sub-items at levels C4-C6, but reveals a critical misalignment: the teaching process across two prior sessions covers only the area and perimeter of simple composite figures (squares and rectangles), while the evaluation items include trapezoids, semicircles, and complex composite analysis. Teacher C explains this discrepancy by noting that the other shapes "had been previously introduced," but numerous students still face major difficulty on the evaluation. This finding supports Nitko and Brookhart (2011), who state that quality item analysis must consider content validity, the discrimination index, and the alignment between item indicators and expected competencies. Without ensuring such alignment, even a technically complex HOTS instrument can undermine rather than support critical thinking development.

Teacher A demonstrates a scaffolded approach that progresses from application (C3) toward analysis, evaluation, and creation (C4-C6). Of 12 sub-items, 7 reach the HOTS level, consistent with problem-based practice and a scoring rubric aligned with Merdeka Curriculum Phase B. The rubric contains four descriptors for C4 (Analysis), C5 (Evaluation), and C6 (Creation), each with clear performance criteria at four levels. This approach gradually enables more students to access higher-order thinking, supporting the scaffolding principle emphasized by Mariana (2017).

Regarding item discrimination, the HOTS assessments across the three classes collectively reveal strong differentiating power. Grade 6 achieves the highest average discrimination index at 0.64, followed by Grade 4 at 0.57 and Grade 5 at 0.49, with the majority of items in the "Very Good" category. The high discrimination index for Grade 6, combined with the lowest average student score, confirms the presence of the misalignment between instruction and assessment rather than poor instrument design.

Students' Critical Thinking Skills Based on HOTS Assessment Results

HOTS assessment results underscore that critical thinking skills varied significantly across classes. The overall average score was 75.72, placing students in the Good category. The following table presents the frequency distribution of students' critical thinking skills by class and category.

Table 3. Frequency Distribution of Critical Thinking Skills by Category

Categories	Score Ranges	Grade IV (n=34)	Grade V (n=28)	Grade VI (n=16)	Total	Percentage
Very Good	86-100	9	18	2	29	37.18%
Good	76-85	5	8	1	24	30.77%
Adequate	60-75	17	1	6	17	21.79%
Poor	55-59	1	0	1	2	2.56%
Very Poor	Below 54	2	1	6	9	11.54%
Mean Score		75.74	86.86	64.58	75.72	
Categories		Good	Very Good	Adequate	Good	

Source: Research Data (2026)

The table above illustrates meaningful differences in achievement in critical-thinking skills across the three classes. Grade 5 with Teacher B demonstrates the strongest performance, while Grade 6 with Teacher C indicates the lowest achievement with the widest score distribution. The following table presents the descriptive statistics of HOTS assessment results per class.

Table 4. Descriptive Statistics of HOTS Assessment Results Class

Classes	N	Min	Max	Mean	Mean Discrimination Index
Grade IV	34	52	100	75.74	0.57
Grade V	28	53	98	86.86	0.49
Grade VI	16	44	94	64.58	0.64
Total/Average	78	49.67	97.33	75.72	0.56

Source: Research Data (2026)

Grade 5 students achieve the highest mean of 86.86 with a Very Good category and relatively uniform distribution: 64.29 percent (18 students) are in the Very Good category (86-100), 28.57 percent (8 students) in the Good category (76-85), and only one student each in the Adequate and Very Poor categories. The standard deviation of 11.02 is the smallest among the three classes, indicating that the assessment successfully differentiates student abilities, with the majority remaining at a high level. This outstanding performance reflects Teacher B's successful alignment among the RPM design, classroom scaffolding activities (including PBL stages, YouTube videos, and PhET Area Builder simulation), and HOTS assessment demands.

Grade 4 students achieve a mean score of 75.74, placing them in the Good category. The distribution indicates 26.47 percent of students (9 students) in the Very Good category, 35.29 percent (12 students) in the Good category, and 32.35 percent (11 students) in the Adequate category. Teacher A's use of Discovery Learning, with concrete grid-paper manipulatives and consistent scaffolding during group work, supports students in progressively developing their critical-thinking skills. The result underscores that even without advanced digital technology, a structured and hands-on learning approach can effectively support students' analytical development.

Grade 6 students achieve the lowest mean (64.58) in the Adequate category and the widest score distribution, with a standard deviation of 15.36. Only 18.75 percent of students (3 out of 16) reach the Good to Very Good category, while 37.50 percent (6 students) are in the Adequate category and another 37.50 percent (6 students) are in the Very Poor category. The primary cause is the misalignment between teaching content, which focuses on simple composite figures of squares and rectangles, and the evaluation items, which involve trapezoids, semicircles, and complex composite analysis. This misalignment is compounded by limited scaffolding during classroom sessions and insufficient structured reflection activities. Teacher C's RPM also lacks explicit learner identification, and the actual use of PhET simulation is undocumented, indicating that the design-to-practice consistency required for effective deep learning was not achieved.

These differences align with the integration of qualitative and quantitative findings. Teacher B demonstrated the clearest alignment across all three components: RPM quality, classroom scaffolding, and HOTS assessment design, producing the highest and most equitable critical thinking outcomes. Teacher A demonstrates solid competence, with room for improvement in content depth and the enrichment of technology integration. Teacher C demonstrates strong assessment instrument design but lacks alignment between what is taught and what is assessed.

The finding that Grade 5 far outperformed the other two grades supports Harun et al.'s (2022) argument that teachers with strong pedagogical competence in HOTS-based learning design, implementation, and evaluation contribute significantly to the development of higher-order thinking skills. It also aligns with research by Aprilia et al. (2024), which concludes that HOTS-based learning effectively supports critical thinking skills in elementary mathematics, with overall achievement at the Good category and above indicating that deep learning is feasible for broad implementation, especially on content with high relevance to real life, such as area and perimeter of plane figures.

The data also confirm the findings of Susilowati and Suyatno (2021) that teachers' competence in implementing HOTS-oriented learning is closely related to students' critical thinking skills at the elementary school level. Quality planning and quality assessment constitute the two primary pillars that determine the success of deep learning. When these two pillars are well-aligned, as demonstrated in Grade 5, the result is optimal and equitable critical thinking skill development.

Integration of Qualitative and Quantitative Findings

The integration of qualitative and quantitative findings yielded a holistic understanding of how teachers' pedagogical competence influences students' critical thinking skills through deep learning and HOTS assessment. Five key integrated findings emerged from this analysis.

The first finding confirms a strong positive relationship between RPM quality and students' critical thinking outcomes. Grade 5 has the most structured learning design, supported by clear LKPD, strong scaffolding, and digital media, including PhET Simulation, leading to the highest student achievement. This

confirms that mature lesson planning aligned with HOTS assessment demands can improve students' abilities in analysis, evaluation, and creation.

The second finding reveals a significant gap in Grade 6. Although the HOTS instrument was of the highest technical quality with all 9 items at level C4-C6, the evaluation content was far more complex than what had been taught during learning sessions, involving trapezoids, semicircles, and advanced composite analysis. The absence of adequate scaffolding and minimal use of structured reflection further weakened the impact of the assessment on critical thinking development.

The third finding confirms that deep learning with the principles of mindful, meaningful, and joyful consistently supports students' critical thinking skills. Observational data reveal that all three teachers applied these principles across their sessions, and the overall average critical thinking score of 75.72, in the Good category, reflects the effectiveness of this approach, despite variations in implementation quality.

The fourth finding underscores that Problem-Based Learning (PBL) supports the development of critical thinking skills. All three teachers reported positive experiences with problem-based models and designed authentic and contextual problems. Observational data show students actively engaging in problem-solving, group discussion, and the explanation of their reasoning processes.

The fifth finding highlights that appropriate use of digital technology supports deep learning and critical thinking development. Teacher B's use of PhET Area Builder and YouTube videos as conceptual exploration tools, rather than only as presentation aids, effectively increased student engagement and conceptual understanding. This supports Fullan et al. (2018), who emphasize the role of digital technology as a catalyst for interactive, contextual, and deep learning.

CONCLUSIONS AND RECOMMENDATION

This study underscores that teachers' pedagogical competence in designing deep learning integrating HOTS assessment varied across the three classes. Grade 4 and Grade 5 teachers demonstrated strong pedagogical competence through structured planning, consistent integration of HOTS from the planning stage, and problem-based LKPDs that promoted critical thinking. Both were consistent with the student-centered principles of the Merdeka Curriculum, resulting in more meaningful and enjoyable learning. The Grade 6 teacher showed a gap between conceptual understanding and real implementation, with an RPM that remained conventional in some respects, learning objectives that were not fully HOTS-oriented, and classroom practice focused on abstract content with limited student engagement.

Regarding the compilation of HOTS assessments, the Grade 4 teacher designed a scaffolded instrument with 7 of 12 items at the HOTS level and a consistent rubric. The Grade 5 teacher produced the most balanced assessment with 10 out of 15 items at C4-C6 levels and a comprehensive scoring rubric. The Grade 6 teacher designed the most technically advanced instrument, with all 9 items at C4-C6, but the assessment lacked alignment with the content taught in prior sessions, resulting in a major discrepancy between assessment expectations and students' preparation.

Students' critical thinking skills showed clear and significant variation across classes. Grade 5 achieved the highest mean of 86.86 (Very Good category), with 92.86 percent of students at the Good to Very Good level, demonstrating that Teacher B's alignment of RPM, scaffolding, and HOTS assessment resulted in optimal and equitable critical thinking development. Grade 4 achieved a mean of 75.74 (Good category), showing solid development with room for improvement. Grade 6 achieved the lowest mean of 64.58 (Adequate category), with only 18.75 percent of students reaching the Good to Very Good level, primarily due to the misalignment between instructional content and assessment demands. The overall average across all three classes was 75.72, with 55.12 percent of students (43 out of 78 students) achieving the Good to Very Good category.

These findings confirm that the alignment of lesson planning, classroom implementation, and assessment is the decisive factor in ensuring that HOTS assessment within deep learning effectively and equitably develops students' critical thinking skills. This research recommends continuous and structured training for teachers in designing authentic HOTS assessments with emphasis on ensuring alignment across

planning, scaffolding, and evaluation. Strengthening the deep learning principles of mindfulness, meaningfulness, and joyfulness through intensive mentoring and teacher learning communities remains essential. The use of digital technology should be elevated beyond presentation media to serve as a tool for exploration, simulation, and student collaboration. Development of learning partnerships with parents and local communities also needs strengthening so that learning becomes more meaningful and connected to students' real lives.

The limitation of this research lies in its focus on a single school with a limited number of research subjects. Future research may involve more schools with diverse characteristics to obtain a more comprehensive understanding. Additionally, this study only measured critical thinking skills in mathematics on the topic of area and perimeter of plane figures, and future research may explore the development of critical thinking skills in other mathematics topics or other subjects within the deep learning and HOTS assessment framework.

Credit authorship contribution statement

Endang Setiyowati: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. Ari Metalin Ika Puspita: Supervision, Validation, Methodology, Writing – review & editing, Project administration. Neni Mariana: Supervision, Formal analysis, Validation, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Declaration

All participants provided informed consent prior to their involvement in the study. They were informed about the study's purpose, procedures, and their right to withdraw at any time without consequence. Permission to conduct the study was obtained from the school authorities, and all data collected were kept confidential and used solely for research purposes.

REFERENCES

- Anderson, L. W., and Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Addison Wesley Longman, Inc.
- Aprilia, N., Wijaya, T. T., and Hidayat, W. (2024). HOTS-based learning and critical thinking in elementary mathematics: A classroom study. *Journal of Elementary Education Studies*, 7(1), 45-59.
- Arikunto, S. (2019). *Prosedur penelitian: Suatu pendekatan praktik* (Revised edition). Rineka Cipta.
- Charles, C., Mastiah, M., and Peterianus, S. (2024). Analisis kemampuan berpikir kritis siswa kelas V sekolah dasar pada pembelajaran matematika. *Jurnal Pendidikan Dasar*, 1(1), 138-147.
- Creswell, J. W., and Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Daga, A. T., Wahyudin, D., and Susilana, R. (2022). Implementation of teacher's pedagogic competence in the 2013 curriculum learning process in elementary school. *Eduvest - Journal of Universal Studies*, 2(5), 900-909.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The California Academic Press.
- Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Measured Reasons and The California Academic Press.
- Febriana, R., Kuntarto, E., and Alirmansyah, A. (2020). Kemampuan guru dalam merencanakan pembelajaran higher order thinking skills (HOTS) di sekolah dasar. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, 9(2), 153-166.

- Fitriyah, L., Amin, S. M., and Siswono, T. Y. E. (2020). Designing open-ended instruction for improving elementary school student's critical thinking ability. *International Journal of Innovative Science and Research Technology*, 5(8), 47-52.
- Fullan, M., Quinn, J., and McEachen, J. (2018). *Deep learning: Engage the world change the world*. Corwin.
- Harun, D., Hulukati, E. P., Hasan, A. M., and Ahmad, J. (2022). Profile of elementary school teachers' pedagogic competence in implementing scientific-based thematic learning and higher-order thinking skills. *Journal of Humanities and Social Sciences Studies*, 4(2), 217-222.
- Jayanto, I., Suharno, and Akhyar, M. (2024). The influence of problem-based learning on critical thinking skills and pedagogical competence at elementary schools. *Jurnal Pendidikan Dasar Nusantara*, 9(2), 112-128.
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2025). *Panduan pembelajaran mendalam*. Kemendikbudristek.
- Mariana, N. (2017). *Transforming mathematics problems in Indonesian primary schools by embedding Islamic and Indonesian contexts* [Doctoral dissertation, Murdoch University]. Murdoch University Research Repository.
- Mariana, N., and Sasmita, F. E. (2024). Designed activities on developing mental strategies using alms value-based learning trajectory in elementary school. *Journal on Mathematics Education*, 15(3), 859-882.
- Marzano, R. J. (2003). *Classroom management that works: Research-based strategies for every teacher*. ASCD.
- Miles, M. B., Huberman, A. M., and Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Nitko, A. J., and Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Pearson.
- OECD. (2023). *PISA 2022 results: The state of learning and equity in education*. OECD Publishing.
- Patricia, D., and Elfitra, E. (2024). Pengembangan instrumen asesmen HOTS untuk mengukur kemampuan berpikir kritis siswa SMP. *Jurnal Pendidikan Matematika*, 15(1), 23-37.
- Perdirjen GTK No. 2626/B/HK.04.01/2023 tentang Pedoman pengelolaan kinerja guru dan kepala sekolah.
- Permendiknas No. 16 Tahun 2007 tentang Standar Kualifikasi Akademik dan Kompetensi Guru.
- Sa'diyah, H., Umalihayati, Hidayah, R., Salimi, M., Fajari, L. E. W., Mashudi, and Aini, S. (2024). The effect of problem based learning model on critical thinking skills in elementary school: A meta analysis study. *Jurnal Iqra': Kajian Ilmu Pendidikan*, 9(1), 135-160.
- Sarwanto, S., Fajari, L. E. W., and Chumdari, C. (2021). Critical thinking skills and their impacts on elementary school students. *Malaysian Journal of Learning and Instruction*, 18(2), 133-161.
- Susilowati, D., and Suyatno, S. (2021). The relationship between teachers' pedagogical competence in HOTS-oriented learning and students' critical thinking skills at elementary school. *Jurnal Pendidikan Dasar*, 12(1), 78-91.
- Yusella, Y., Suciptaningsih, O. A., and Degeng, M. D. K. (2023). Development of HOTS-based assessment in the curriculum to increase critical thinking skills of elementary school students. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(7), 4991-4998.