Edu Komputika 13 (1) (2026)

Edu Komputika Journal

https://journal.unnes.ac.id/journals/edukom/

Improving Higher Order Thinking Skills through Project-Based Learning with AI-Enhanced Formative Assessment

Tara Lestari Safitri[1)], Nuur Wachid Abdulmajid[1)]✉, and Md Baharuddin Haji Abdul Rahman[2)]

[1]Department of Technology and Information Systems Education, Kampus Purwakarta, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi No. 229, Bandung, 40154, Indonesia

[2]School of Educational Studies, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia

Article Info

Article History:
Received: 16 May 2026
Revised: 29 August 2026
Accepted: 31 August 2026

Keywords:
AI-Enhanced Formative Assessment, Computer Science, Higher-Order Thinking Skills, Mixed-Methods Design, Project-based Learning.

Abstract

The low level of students' Higher Order Thinking Skills (HOTS) in computer science presents a challenge in the implementation of the Merdeka Curriculum. This study examines the extent to which a Project-Based Learning (PBL) approach combined with AI-enhanced formative assessment can improve students' HOTS in computer networking, particularly their ability to analyze computer network concepts and apply them in real-life situations. This study contributes by presenting an innovative learning model that integrates PBL and AI-enhanced formative assessment as a practical alternative to improve HOTS in Indonesian high schools. The study used a classroom action research (CAR) design as the overarching framework, implemented as a single CAR cycle of planning, action, observation, and reflection; within this cycle, a mixed-methods approach combining a quasi-experimental one-group pretest-posttest method with a descriptive qualitative method was used to collect and analyze data, and the study involved 33 eleventh-grade students at SMA N 3 Purwakarta. The intervention integrated a digital poster project with layered formative assessment using ChatGPT and Padlet. Data were analyzed via paired t-tests and thematic analysis using NVivo. Results showed a significant cognitive improvement, with the average score rising from 74.55 to 83.64 ($p = 0.026$). Qualitatively, 58% of students were able to evaluate connectivity options based on real-world needs, and each group conducted an average of 3.2 rounds of revisions based on AI feedback. It was concluded that the integration of PBL with AI-enhanced formative assessment effectively fosters students' higher-order thinking skills (HOTS) at the school studied, though the single-school sample means the findings are best generalized at the regional rather than national level.

✉ Correspondence Address:
Jl. Dr. Setiabudhi No. 229, Bandung, 40154, Indonesia
E-mail: nuurwachid@upi.edu

ISSN 2252-6811
E-ISSN 2599-297X

INTRODUCTION

The Merdeka Curriculum marks a fundamental shift in the focus of learning from mere content mastery to the development of thinking skills. One of the key areas emphasized is *higher-order thinking skills* (HOTS), which refer to students' ability to analyze, evaluate, and create, rather than simply memorize or understand the information they receive. In Computer Science, this demand feels even more urgent because the field is evolving too rapidly to be taught solely through rote memorization; what is far more important is students' ability to reason and adapt to ever-changing technological challenges.

The problem is that classroom learning conditions have not yet shifted significantly in that direction. Findings from observations and literature reviews indicate that learning activities in secondary schools are still dominated by approaches that target lower cognitive levels, such as understanding and recall, while analytical and evaluative skills receive insufficient attention (Aydin & Birgili, 2023; Muhibbuddin et al., 2023). This pattern is also reflected in assessment designs that test mastery of facts more than students' critical thinking skills.

Project-Based Learning (PBL) is recognized as one of the approaches capable of creating a more conducive learning environment for Higher-Order Thinking Skills (HOTS). The logic is simple: when students are asked to work on real-world projects based on authentic problems, they are encouraged to think analytically, collaborate, and produce outcomes that reflect their comprehensive understanding, rather than merely reproducing material from textbooks (Chang et al., 2022; Manuaba et al., 2024). A number of recent studies confirm the effectiveness of PBL in improving students' HOTS, including in technology-based subjects (Affandy et al., 2024; Wijayanto et al., 2023).

However, PBL has an often-overlooked weakness: the quality of feedback during the process. In large classes, teachers are not always able to provide prompt and personalized *feedback* to every group. In fact, timely and specific feedback is one of the most powerful factors in promoting deep learning (Hepburn et al., 2022; Rodríguez et al., 2022). This gap in timely, individualized feedback is a central obstacle to realizing HOTS through PBL, because higher-order thinking develops through cycles of trying, receiving specific feedback, and revising, a cycle that stalls when feedback is delayed or generic. This is where the integration of AI is beginning to receive serious attention in educational research. Several studies indicate that AI can serve as a *feedback engine* capable of providing instant, repetitive, and adaptive responses tailored to students' needs a capability that is difficult for teachers to fulfill manually in a regular classroom setting (Steiss et al., 2024; Zhai & Nehm, 2023a; Zhao et al., 2025).

However, evidence for AI's role in feedback is still fragmented across different contexts. Comparing ChatGPT-generated feedback to trained human raters on secondary students' essays (Farrokhnia et al., 2026; Steiss et al., 2024) found that human feedback was still rated higher on most quality criteria, including accuracy and supportive tone with ChatGPT only matching humans on aligning feedback to the rubric; moreover, this comparison was made on single-shot writing correction rather than on an iterative, project-based cycle of revision (Acan & Kaban, 2026). Meanwhile, (Zhao et al., 2025) meta-analysis of 29 experimental and quasi-experimental studies confirms that generative AI has a moderate positive effect on higher-order thinking, strongest for problem-solving and weakest for creativity, yet the included studies largely measure AI's effect on isolated tasks rather than a sustained PBL process in which students must integrate feedback into an evolving product. This distinction matters because (Jose et al., 2025) caution that AI's effect on cognition is not uniformly positive: depending on how it is used, AI can function as either a cognitive enhancer or a source of cognitive offloading that erodes independent reasoning. This tension suggests that the instructional design surrounding AI feedback, not merely its availability, determines whether it supports or undermines higher-order thinking.

To date, no study has examined how AI-based formative feedback should be structured within a PBL cycle so that it strengthens, rather than substitutes for, students' own analytical and evaluative work. Existing PBL-HOTS studies (Affandy et al., 2024; Manuaba et al., 2024; Wijayanto et al., 2023) do not incorporate AI feedback, while existing AI-feedback studies (Steiss et al., 2024; Zhao et al., 2025) do not situate that feedback within a structured PBL design targeting HOTS specifically, nor do they address the enhancement-versus-offloading tension raised by Jose et al. (2025). This study addresses that gap by designing AI (ChatGPT) feedback as an *iterative scaffold* within a PBL poster project students received AI feedback multiple times and were required to analyze, evaluate, and revise their own content in response, rather than accepting AI output directly combined with peer and teacher assessment via Padlet, within the specific context of secondary

Computer Science education in Indonesia, where empirical evidence on this combination remains limited.

This study employed a mixed-methods design in the Computer Science course at SMA N 3, focusing on the topic of Computer Networks and the Internet. Through the implementation of PBL combined with AI-assisted (ChatGPT) formative assessment and the Padlet digital collaboration platform, this study aims to examine the extent to which this approach can improve students' higher-order thinking skills (HOTS), particularly their ability to analyze computer network concepts and apply them in real-life situations.

RESEARCH METHODS

This study employed a classroom action research (CAR) design as its overarching framework, following the cyclical planning-action-observation-reflection model of Kemmis and McTaggart, and implemented here as a single action research cycle appropriate to the two-session scope of the intervention. Within this single CAR cycle, data were collected and analyzed using a mixed-methods approach that combines two specific methods: a quasi-experimental one-group pretest-posttest method, used to measure the statistical significance of quantitative gains on the cognitive test, and a descriptive qualitative method, used to explore the depth of students' HOTS achievement through thematic analysis of their reflections and through classroom observation. In other words, the CAR cycle constitutes the overall research design, while the quasi-experimental and descriptive qualitative methods together form the mixed-methods approach applied within each stage of that cycle. This combination was chosen because it allows the statistical significance of gains on the cognitive test to be interpreted alongside the qualitative depth of students' reflections and classroom observations, providing a fuller picture of HOTS development than either data source alone, while the CAR cycle also allows the researcher, in collaboration with the subject teacher, to plan, implement, observe, and reflect on the intervention within the authentic classroom setting. The study was conducted over two face-to-face sessions in the Computer Science course at SMA N 3, focusing on Computer Networks and the Internet.

The study participants were 33 eleventh-grade students in the Computer Science class. The class was selected through *purposive sampling* based on the availability of technology and the teacher's willingness to collaborate in implementing the innovative learning model. The subject teacher served as a collaborator in designing and implementing the learning intervention.

The planning phase of the classroom action research (CAR) cycle involves designing a PBL scenario on the topic of Computer Networks, focused on developing students' higher-order thinking skills (HOTS). In this stage, the researcher prepares an AI-based formative assessment system (ChatGPT) and a digital collaboration platform (Padlet) as a repository for collecting assignments, as well as developing pre-tests, post-tests, assessment rubrics, and reflection questionnaires designed to measure students' higher-order thinking skills.

The action phase of the CAR cycle was conducted over two sessions. The first session began with a pre-test to assess students' initial understanding. Next, the teacher presented a contextual problem related to computer networks, after which students worked in groups to create digital posters. During this process, students utilized feedback from AI (ChatGPT) to iteratively improve the accuracy of their content and discussed their work via the Padlet platform. This iterative poster-making process is designed to encourage students to think analytically and evaluatively, which are indicators of higher-order thinking skills (HOTS). The second session focused on each group presenting their poster, followed by *peer assessment* via Google Forms and teacher assessment using a rubric. After the activity was completed, students took a post-test and filled out an online reflection questionnaire.

The observation phase of the CAR cycle was conducted on 20 November 2024 and involved collecting qualitative and quantitative data through an observation checklist rating students' focus/engagement and competitive orientation during discussions and presentations (on a 1–5 scale), Padlet activity logs, pre-test and post-test results, and responses to open-ended reflection questionnaires.

The reflection phase of the CAR cycle followed implementation: all collected data were analyzed to assess the intervention's effectiveness in improving students' higher-order thinking skills (HOTS), particularly their ability to analyze computer network concepts and apply them to real-life situations. This reflection, conducted jointly by the researcher and the collaborating subject teacher, completes the single action research cycle carried out in this study and informs the interpretation of findings presented in the Results and Discussion section.

The instruments used in this study consist of three types. First, cognitive tests in the form of pre-tests and post-tests comprising multiple-choice questions that measure understanding of computer network concepts at the analysis and application levels according to Bloom's Revised Taxonomy. Second, a poster product assessment rubric covering four aspects: conceptual understanding (40%), visual creativity (25%), practical examples (20%), and completeness (15%). The conceptual understanding and practical examples aspects were specifically designed to measure students' ability to analyze and apply the material, serving as indicators of Higher-Order Thinking Skills (HOTS). Third, an open-ended reflection questionnaire consisting of two short essay questions: "What have you learned about networks and the internet?" and "How would you choose the right connectivity for your needs?" The second question is specifically designed to measure students' ability to transfer knowledge into real-life situations, which serves as an indicator of HOTS at the application and evaluation levels. Content validity of the cognitive test and the poster assessment rubric was established through expert judgment by the study's academic advisor prior to implementation. Reliability of the cognitive test was estimated using the Kuder–Richardson formula 20 (KR-20), yielding coefficients of 0.41 for the pre-test and 0.42 for the post-test (both n = 33), indicating moderate and comparable internal consistency across the two administrations. Reliability of the poster assessment rubric was examined through the correlation between peer- and teacher-assigned scores for the same six posters (r = 0.10, reported in the Results and Discussion section), which indicated low inter-rater consistency.

Consistent with the CAR design and its embedded quasi-experimental and descriptive qualitative methods named above, data collected during the intervention were analyzed using a *mixed-methods* approach. Quantitative data from the pre-test and post-test were analyzed using descriptive statistics (mean, standard deviation) and paired t-tests with SPSS to test the significance of improvements in students' cognitive understanding ($\alpha = 0.05$). Qualitative data from the reflection questionnaire were analyzed using NVivo software through thematic *coding* to identify the main themes reflecting students' HOTS achievements, particularly in conceptual understanding, differentiation skills, and the transfer of knowledge to real-life situations. Data triangulation was performed by comparing the results of observations, cognitive assessment scores, teacher rubric evaluations, and student reflections to validate findings related to improvements in HOTS.

RESULT AND DISCUSSION

Consistent with the mixed-methods design and the classroom action research (CAR) cycle described in the Research Methods section, the following results are presented in two integrated strands: quantitative results from the pretest-posttest cognitive test, and descriptive qualitative results from classroom observation, the poster assessment rubric, and thematic analysis of student reflections. These two strands are discussed together within each subsection below to reflect the reflection stage of the CAR cycle, in which quantitative and qualitative evidence were jointly interpreted to evaluate the effectiveness of the intervention.

A. Improvement in Students' Cognitive Understanding

The results of the paired t-test analysis using SPSS indicate a significant improvement in students' cognitive understanding following the implementation of PBL supported by *AI-enhanced formative assessment*. The average pre-test score of 74.55 increased to 83.64 on the post-test, representing an increase of 9.09 points. As summarized in Table 1, this difference between the pre-test and post-test means is statistically significant. The corresponding effect size (Cohen's d_z) was 0.41, indicating a small-to-medium practical improvement according to Cohen's (1988) benchmarks, in addition to the statistical significance reported above. This means that the PBL intervention, combined with layered *feedback* from AI, peers, and teachers, has a measurable positive impact on mastery of Computer Networks material. Figure 1 illustrates the pre-test to post-test gain for each of the 33 students.

Table 1. Summary of Paired-Samples t-Test Results (Pre-Test vs. Post-Test)

Measure	Pre-Test	Post-Test	Mean Difference	t (df=32)
Mean	74.55	83.64	9.09	
Std. Deviation	21.95	19.66	22.41	
Sig. (2-tailed)			p = 0.026	t = –2.367

n = 33; the post-test mean was significantly higher than the pre-test mean, t(32) = –2.367, p = 0.026 (p < 0.05).

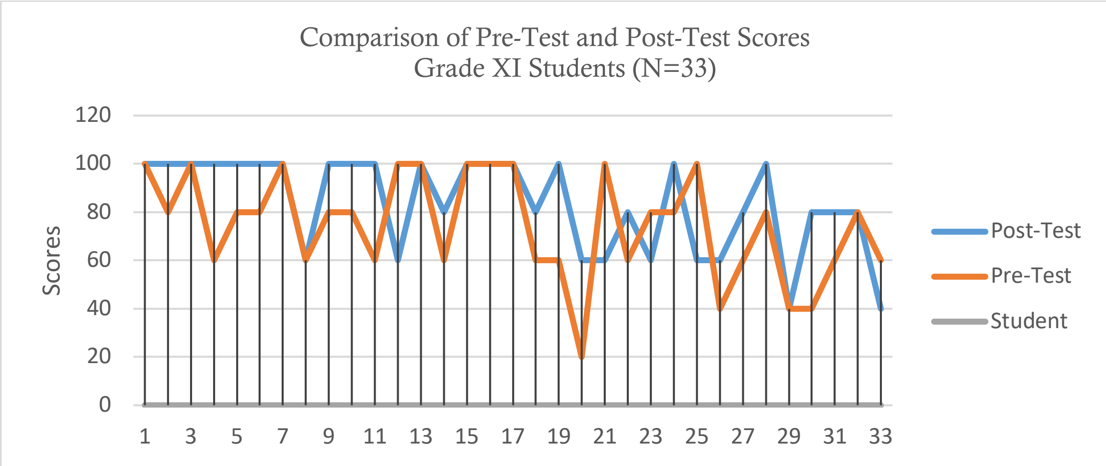

Figure 1. Graph comparing students' pre-test and post-test scores

This improvement aligns with the theory that continuous formative *feedback* promotes retention and deeper understanding. In terms of HOTS, this increase in cognitive scores does not merely reflect memorizing facts but rather reflects students' ability to understand and analyze computer network concepts in a more structured manner. This is supported by the post-test question design, which was crafted at the analysis and application levels of the Revised Bloom's Taxonomy. These findings align with the results of the study Zhao et al. (2025), which demonstrated that AI-assisted learning approaches can significantly enhance student learning outcomes, with the additional observation that the integration of PBL reinforces higher-order thinking skills through contextual problem-solving.

B. The Quality of PBL Products as an Indicator of Higher-Order Thinking Skills

Teachers' assessment of poster products using a rubric with four aspects showed varying results across groups. The aspects of conceptual understanding (40% weight) and practical examples (20% weight) specifically reflect students' HOTS abilities, namely the ability to analyze material and apply it in the real world. Figure 2 and Figure 3 present, respectively, the peer- and teacher-assessed scores for each group's final poster.

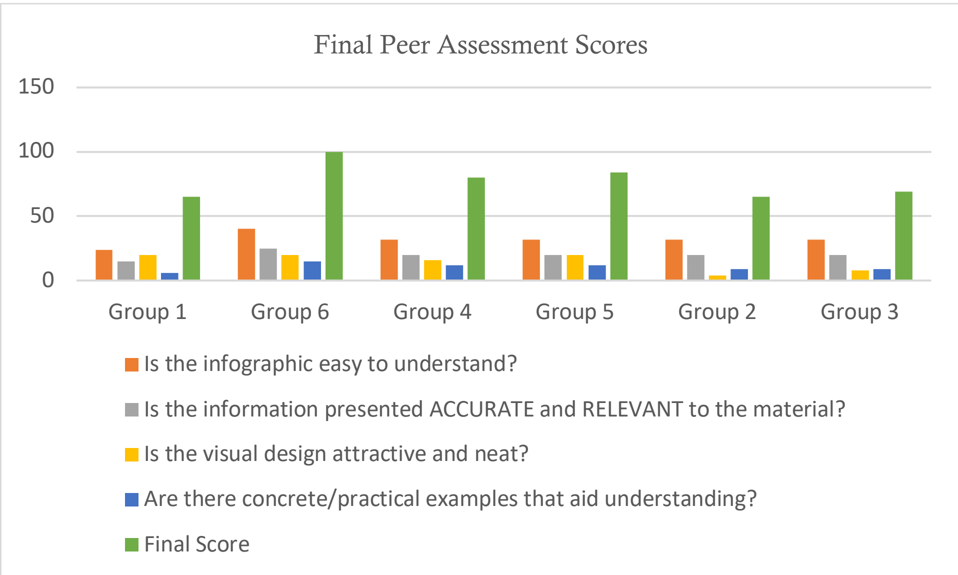

Figure 2. Final peer assessment score graph

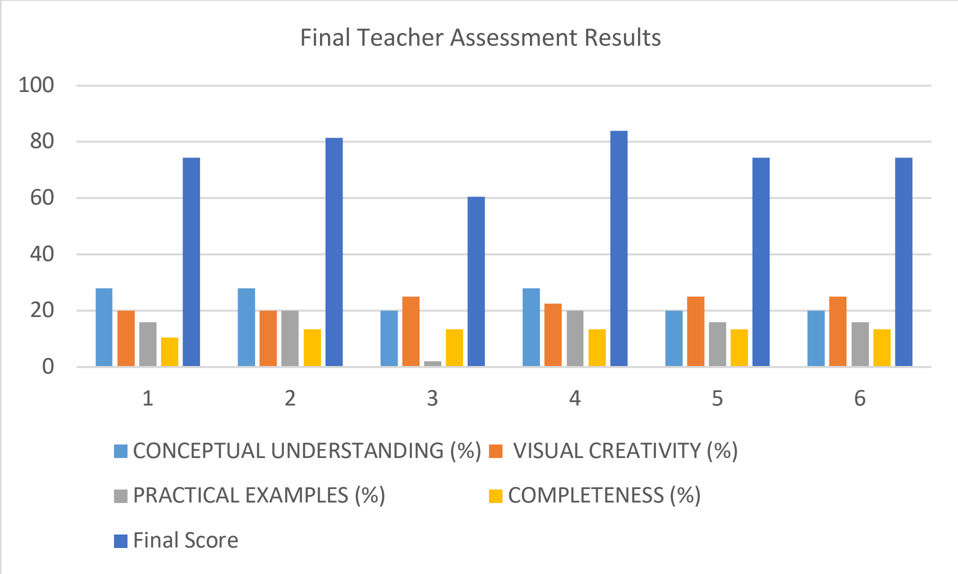

Figure 3. Graph of final teacher assessment scores by group

Figure 2 presents the peer-assessment scores awarded by classmates for each group's final poster, while Figure 3 presents the teacher's final rubric-based assessment of the same posters; together, the two figures allow the peer and teacher evaluations of HOTS-related product quality to be compared side by side. Based on the teachers' assessments, it appears that most groups were able to present content with adequate conceptual understanding, although the depth of the practical examples still varied. Groups that actively utilized AI *feedback* in an iterative manner tended to produce posters with greater depth of content. Padlet log data shows that, on average, each group performed 3.2 rounds of revisions based on AI and peer *feedback*, and all groups (100%) successfully uploaded both their initial drafts and final versions of their posters.

This iterative revision process is a key indicator of higher-order thinking skills (HOTS) engagement in learning. When students receive *feedback* from AI regarding content gaps, they are required to evaluate the available information, identify gaps in understanding, and reconstruct more accurate explanations. This process aligns with the evaluation and creation levels in the revised Bloom's Taxonomy. These findings reinforce the argument that AI-supported PBL creates learning conditions that naturally foster HOTS through a repeated cycle of thinking, revising, and improving (Chang et al., 2022).

To make the iterative revision process easier to interpret than the raw Padlet screenshots, Table 2 summarizes the content changes from the initial draft (Progress 1) to the final version (Progress 2) for the groups whose cards were fully visible on the Padlet board.

C. Classroom Observation: Focus and Competitive Orientation

The classroom observation was conducted on 20 November 2024, across the two implementation sessions, using a checklist that rated each student's focus/engagement and competitive orientation during discussions and presentations on a 1–5 scale, in addition to the Padlet activity logs described above.

The focus on the learning process did not improve significantly over the two-day implementation: the mean focus score remained stable at 3.5, while the median declined slightly from 4 to 3. The distribution also became more polarized, with about 30% of students showing an increase, 33% a decrease, and the remainder unchanged; the standard deviation roughly doubled (from 0.7 to 1.4), indicating that the intervention's effect on sustained focus was uneven across students rather than uniformly positive.

Students' competitive orientation decreased over the same period, with the mean score falling from 3.5 to 3.2 and the proportion of students in the high-competitiveness category dropping from 48.5% to 36.4%. This reduction is interpreted as a positive outcome for deep learning, since a less competitive classroom climate is generally more conducive to collaborative, reflective engagement with feedback.

Table 2. Summary of Poster Revisions from Draft to Final Version (Padlet)

Group	Draft (Progress 1)	Final (Progress 2)	Nature of Revision
Group 1	No draft (Progress 1) card was captured on the Padlet board for this group; the pre-revision content is not documented.	Full written explanation distinguishing LAN and Internet: LAN defined as a network connecting devices within a small area (home, school, office, library) that allows fast local sharing of data, files, and resources, with high speed, greater security, lower cost, and use of Ethernet cable or Wi-Fi; Internet defined as the world's largest network connecting millions of networks globally, enabling information access, messaging, gaming, shopping, and communication without distance or time limits, via cable, fiber optic, Wi-Fi, or cellular technology.	Cannot be determined: without a documented draft, the extent of revision between Progress 1 and Progress 2 is unknown for this group.
Group 2	Poster template and title only ("Jaringan Lokal dan Internet – Kelompok 2"); the group had agreed only on a blue color scheme and general concept, with no written content or explanation added yet.	Full written explanation comparing LAN and Internet ("Internet Vs Jaringan Lokal"), covering their definitions, functions, and coverage areas, with example use cases for each.	Content added from scratch: explanatory text distinguishing LAN and Internet, along with examples, was written between the draft and final stages.
Group 3	Only group members' names listed (Laras, Tania, Aldi, Alfajri, Fitri, Shadiq); poster visual partly uploaded, no written explanation yet.	Full written explanation comparing LAN and Internet, describing the poster's visual layout (an illustration of a person using a computer, with LAN explained as a small-area network for sharing data/devices, and Internet explained as a global network) plus a rationale for the color scheme.	Content added from scratch: the explanatory text and design rationale were written between the draft and final stages.
Group 4	Poster image uploaded, but only a placeholder caption ("poster photo") with no written description.	Full written explanation of what LAN and the Internet are, example devices, advantages/disadvantages, and a rationale for the visual choices (bright colors, character/robot illustrations, device icons, and light effects to make the poster look modern and easy to understand).	Content added from scratch, similar to Group 3.
Group 5	A full explanatory paragraph on the difference between a local network and the internet was already present, including the rationale for the blue/pink/purple color palette and the studying-student illustration.	The core explanation is retained essentially unchanged; group members' names are added.	Minor revision: content was essentially finalized already at the draft stage; the final version mainly adds authorship.
Group 6	Brief planning note only ("Day 1"): the group discussed the theme and topic, then built the poster's element structure and looked for content, no substantive poster content yet.	Full explanation ("Day 2"): rationale for the forest theme (symbolizing calm and balance, contrasted with a crowded, distraction-filled internet) and a structured comparison between LAN (Local Network) and the Internet (Global Network).	Content substantially expanded: from a one-line planning note to a complete, reasoned explanation.

Agreement between peer and teacher assessments of the posters was low: the Pearson correlation between peer and teacher scores was only r = 0.10, with an average gap of 12.25 points between the two sources. Peer assessors tended to rate posters more favorably, plausibly reflecting a halo effect driven by visual appeal, whereas teachers focused more closely on the substance and accuracy of the content. This gap suggests that peer assessment scores in this study should be interpreted with caution rather than treated as interchangeable with the teacher's rubric-based evaluation.

Taken together, the observation data indicate that the AI-enhanced formative assessment intervention was feasible and had a discernible positive effect on cognitive understanding and on reducing competitive orientation, but had not yet produced a clear, uniform improvement in students' focus on the learning process a pattern most plausibly attributable to the short two-session implementation period, the absence of a control group, and the lack of standardized pre/post-test administration conditions, all of which are revisited in the Limitations subsection below.

D. Knowledge Transfer as Evidence of HOTS

This section presents the most crucial findings of this study. A qualitative analysis using NVivo of 33 student reflection responses identified two main categories of findings that reflect the achievement of higher-order thinking skills (HOTS): conceptual understanding and the application of knowledge to real-life situations.

1. Conceptual Understanding (Understanding & Analyzing Levels)

In response to the first reflective question regarding what had been learned, the NVivo analysis identified several dominant themes, as presented in the Table 3.

Table 3. Dominant Themes in Student Reflections on Question 1 Results of the NVivo Analysis

Theme	Example Answers	Number of Students	Percentage
Understanding the concept of the internet	"The Internet is a global network that connects computers around the world"	22	67%
Understanding networks and local area networks	"A network is a collection of connected devices; a local network is for small areas such as a home or school"	18	55%
Learning process and understanding of the material	"I have studied networks and the internet"	15	45%
Understanding the difference between the internet and networks	"The main difference is scope: the internet is global, while local networks are smaller"	10	30%
The global concept and wide reach of the internet	"The Internet is a global network connecting millions of devices"	9	27%
Understanding of computer concepts and devices	"Computers connected to share information"	5	15%

A total of 67% of students (n=22) were able to articulate the concept of the internet as a global network connecting devices worldwide. A total of 55% of students (n=18) understood the concept of a local area network (LAN) along with its coverage characteristics. More significantly from a HOTS perspective, 30% of students (n=10) were able to explicitly distinguish between the internet and a local network based on their scope and functions. This ability to distinguish reflects the analysis level in Bloom's Taxonomy, where students do not merely memorize definitions but are able to independently identify conceptual differences.

In addition, 45% of students (n=15) reflected on their awareness of the learning process they had undergone, indicating the emergence of metacognitive skills as a key component of HOTS. This finding is consistent with the assertion that meaningful learning occurs when students are able to connect new knowledge with their existing cognitive structures (Cottingham & Ausubel, 2023).

2. Transfer of Knowledge into Real Life (Application & Evaluation Levels)

The second reflective question, "How would you choose the right connectivity for your needs?", was specifically designed to measure students' ability to transfer knowledge to real-life situations. This ability represents the highest-order thinking skill (HOTS) indicator in this study because it requires students to apply and evaluate knowledge in contexts that have not been explicitly taught.

The analysis results in Table 4 show that 58% of students (n=19) were able to consider speed and activity requirements as the basis for selecting an internet connection, rather than simply choosing at random. A total of 42% of students (n=14) emphasized the importance of tailoring their choice of connection to their individual needs, reflecting their evaluative skills. A total of 24% of students (n=8) demonstrated more complex thinking by considering the appropriate type of network (local vs. internet) according to the scope of their needs.

Table 4. Dominant Themes in Students' Reflections on Question 2 Results of the Qualitative Analysis Using NVivo

Theme	Sample Answers	Number of Students	Percentage (%)
Considering personal needs	"I will choose the connectivity that suits my needs"	14	42%
Choosing based on speed and internet needs	"Choosing fast connectivity for online activities and learning"	19	58%
Considerations regarding network type (local vs. internet)	"Choosing the right network, whether local or internet"	8	24%
Focus on online learning	"Choosing connectivity that supports online learning"	8	24%
More in-depth consideration of connectivity	"Would prefer to make a more informed decision"	5	15%
Cost and efficiency considerations	"Choosing what fits the budget and isn't wasteful"	7	21%

In fact, 21% of students (n=7) had already considered cost efficiency as a variable in their decision-making. These four themes were coded from responses that could contain more than one idea, so the percentages are not mutually exclusive and do not sum to 100%; students whose answers did not fit any of the four dominant themes (for example, single-word or off-topic responses) were excluded from this thematic count.

These findings confirm the occurrence *of a transfer of learning* from conceptual mastery to practical decision-making within a relatively short period of time. In contrast to traditional learning, which often results *in "inert knowledge"* knowledge that cannot be applied outside the learning context (Don et al., 2023) the AI-assisted PBL in this study successfully encouraged students to apply their knowledge in a functional manner. This ability aligns with the 21st-century skills framework, which emphasizes *knowledge application over knowledge acquisition* (Chang et al., 2022), and serves as concrete evidence that students' higher-order thinking skills (HOTS) developed through the intervention designed in this study.

E. The Role of AI as a Scaffold for Higher-Order Thinking Skills in PBL

The use of AI (ChatGPT) as *a feedback engine* in this study has proven to serve as *scaffolding* that fosters students' higher-order thinking processes. Unlike conventional, one-way *feedback* from teachers, AI *feedback* is instant, iterative, and personalized. Students can immediately test their understanding by asking the AI questions, receiving conceptual explanations, and then revising their posters based on that feedback.

This iterative interaction between students and AI creates a thinking cycle that naturally requires analytical and evaluative skills. When the AI identifies flaws in the poster's content, students must analyze which parts are inaccurate, evaluate better alternative explanations, and create revised content. This cycle is repeated an average of 3.2 times per group, indicating a high level of cognitive engagement throughout the learning process.

These findings align with the research by Steiss et al. (2024), which demonstrates that *AI-assisted formative assessment* can significantly improve the quality *of feedback*. Furthermore, in HOTS, AI functions not merely as a fact-checker but as a thinking partner that encourages students to continually question and deepen their understanding. This operationalizes the principle that technology in education is most effective when used as a tool to expand students' cognitive capacities, rather than simply replacing the teacher's role (Jose et al., 2025).

These findings can be situated within established learning theory. The gains in cognitive understanding, achieved as students iteratively revised their posters in response to AI feedback, align with a social-constructivist view of learning, in which knowledge is actively built through interaction rather than transmitted directly, a mechanism also demonstrated by (Trustee, 2024) in their study of Vygotsky-based project-based learning on secondary students' problem-solving. In this study, ChatGPT functioned less as an answer-provider and more as a *scaffold*: consistent with (Rodríguez et al., 2022) finding that scaffolded feedforward improves problem-based learning outcomes in large classes, AI feedback here supported students at the point of difficulty while leaving the analytical and evaluative work of identifying gaps, deciding what to revise, to the students themselves.

The intervention also illustrates how generative AI can extend the practical reach of *formative assessment*. (Zhai & Nehm, 2023b) argues that AI does not replace formative assessment but changes its scale and cadence, making the elicit–interpret–act cycle achievable far more frequently than a single teacher could sustain across many groups; the 3.2 average revision rounds observed per group in this study are consistent with that shift toward more frequent, lower-stakes feedback cycles. Finally, the fact that students had to actively decide how to respond to AI feedback, rather than simply accept it, reflects processes central to *self-regulated learning (SRL)*. (Zhai & Nehm, 2023b) similarly found that ChatGPT can support the forethought, performance, and self-reflection phases of SRL when learning activities are designed to require students' own judgment rather than passive acceptance of AI output; the design choice in this study to require students to analyze and justify their revisions (rather than copy AI suggestions directly) reflects this same principle, and offers one explanation for why cognitive gains were observed despite the AI enhancement-versus-erosion tension raised earlier in this paper.

F. Limitations of the Findings

Although the results of this study show positive indications of improvements in HOTS through AI-assisted PBL, several limitations must be acknowledged. First, the implementation period of only two sessions limited the depth of change that could be achieved, particularly for more complex aspects of HOTS such as creativity and deep evaluation. Second, the absence of a control group means that the causal relationship between the intervention and improvements in HOTS cannot yet be fully confirmed. Third, the measurement of HOTS in this study is still limited to analytical and application skills, while the aspects of deep evaluation and creativity have not been comprehensively measured. Further research with a longer duration, a control group, and more comprehensive HOTS instruments is needed to confirm and expand upon these findings. Fourth, the instruments used showed moderate-to-low reliability: the cognitive test's internal consistency was moderate (KR-20 = 0.41 pre-test; 0.42 post-test), and agreement between peer and teacher raters on the poster rubric was low (r = 0.10). These figures suggest that both instruments would benefit from further refinement for example, revising weak-performing distractors in the cognitive test and providing raters with more detailed scoring anchors and calibration training for the rubric in future iterations of this research.

CONCLUSION

This study demonstrates that the implementation of *Project-Based Learning* (PBL) combined with *AI-enhanced formative assessment* effectively improves students' *Higher Order Thinking Skills* (HOTS) in the Computer Science course at SMA N 3 Purwakarta. The improvement in cognitive understanding was statistically significant, marked by an increase in the average score from 74.55 on the pre-test to 83.64 on the post-test ($p = 0.026$). Beyond mere score improvements, qualitative analysis of student reflections revealed more substantive HOTS achievements: 30% of students were able to analyze the conceptual differences between the internet and local networks, while 58% successfully evaluated connectivity options based on their real-world needs concrete evidence of knowledge transfer into daily life. The role of AI (ChatGPT) as *a feedback engine* proved crucial in creating an iterative thinking cycle, where, on average, each group made 3.2 revisions based on AI feedback, a process that naturally demands analytical and evaluative skills as core indicators of HOTS.

These findings directly answer the research objective stated at the outset of this study: implemented through a single classroom action research cycle, the AI-enhanced PBL intervention was able to improve students' higher-order thinking skills, specifically their ability to analyze computer network concepts and apply them in real-life situations, as evidenced jointly by the statistically significant quantitative gains on the cognitive test and the qualitative evidence of conceptual analysis and knowledge transfer described above.

Nevertheless, this study has limitations that should be noted, namely the short duration of the intervention (only two sessions), the absence of a control group, and the scope of the HOTS assessment, which is still limited to the analysis and application levels. Therefore, further research is recommended to use an experimental design with a control group, a longer intervention duration, and a more comprehensive HOTS instrument that includes the aspects of evaluation and creation; because the sample was limited to 33 students at a single senior high school, the findings of this study should not be generalized to the national level; given the uneven distribution of technology access across Indonesia, the results are more appropriately generalized at the regional or provincial level, and this would be useful for confirming and expanding these findings regarding secondary education in Indonesia.

ACKNOWLEDGEMENTS

This research was conducted at SMA Negeri 3 Purwakarta, and the authors are deeply grateful to the school principal and institution for granting permission and providing a supportive environment for this research. Special appreciation is extended to the Informatics subject teacher who served as a collaborative partner in designing and implementing the Project-Based Learning intervention, whose cooperation and dedication were indispensable to the success of this study. The authors also sincerely thank the 33 eleventh-grade students who participated with enthusiasm and commitment throughout the learning process.

Finally, the authors would like to express their deepest gratitude to someone who once became a place to ask, to share stories, and to find comfort in moments when words were difficult to carry alone. Although things are no longer the same, the kindness, conversations, and quiet presence once shared remain a meaningful part of the author's journey. To someone who was never quite given the chance to be called mine, thank you for having once been there, for the questions you answered, the stories you listened to, and for leaving behind lessons that the author will carry long after the paths have grown apart.

The authors also acknowledge the use of ChatGPT as an AI-assisted formative assessment tool, Padlet as a digital collaboration platform, and NVivo and SPSS software for qualitative and quantitative data analysis, all of which were essential in realizing the instructional model examined in this study. This research was self-funded by the authors and did not receive any grant or financial support from external funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Acan, Y. G., & Kaban, A. L. (2026). Human vs. Machine Feedback: Evaluating ChatGPT-4 in the Assessment of Secondary EFL Learners' Writing. *Asia Pacific Journal of Educators and Education*, *41*(1), 159–187. https://doi.org/10.21315/apjee2026.41.1.8

Affandy, H., Sunarno, W., Suryana, R., & Harjana. (2024). Integrating creative pedagogy into problem-based learning: The effects on higher order thinking skills in science education. *Thinking Skills and Creativity*, *53*, 101575. https://doi.org/10.1016/j.tsc.2024.101575

Aydin, U., & Birgili, B. (2023). Assessing Mathematical Higher-Order Thinking Skills: An Analysis of Turkish University Entrance Examinations. *Educational Assessment*, *28*(3), 190–209. https://doi.org/10.1080/10627197.2023.2202311

Chang, T.-S., Wang, H.-C., Haynes, A. M., Song, M.-M., Lai, S.-Y., & Hsieh, S.-H. (2022). Enhancing student creativity through an interdisciplinary, project-oriented problem-based learning undergraduate curriculum. *Thinking Skills and Creativity*, *46*, 101173. https://doi.org/10.1016/j.tsc.2022.101173

Don, H. J., Goldwater, M. B., & Livesey, E. J. (2023). Cognition of relational discovery: Why it matters for effective far transfer and effective education? *Frontiers in Psychology*, *14*, 957517. https://doi.org/10.3389/fpsyg.2023.957517

Farrokhnia, M., Latifi, S., Papadopoulos, P. M., Hogenkamp, L., Gijlers, H., Khosravi, H., & Noroozi, O. (2026). Generative AI offers more, but students revise less: Comparing the effects of teacher and AI feedback on student essay revisions. *International Journal of Educational Technology in Higher Education*, *23*(1), 6. https://doi.org/10.1186/s41239-026-00579-9

Hepburn, L.-A., Borthwick, M., Kerr, J., & Vasnev, A. (2022). A strategic framework for delivering ongoing feedback at scale. *Assessment & Evaluation in Higher Education*, *47*(5), 742–754. https://doi.org/10.1080/02602938.2021.1959517

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology*, *16*, 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

Manuaba, I. B. A. P., -No, Y., & Wu, C.-C. (2024). Correction: The effectiveness of problem based learning in improving critical thinking, problem-solving and self-directed learning in first-year medical students: A meta-analysis. *PLOS ONE*, *19*(5), e0303724.

https://doi.org/10.1371/journal.pone.0303724

Muhibbuddin, M., Artika, W., & Nurmaliah, C. (2023). Improving Critical Thinking Skills Through Higher Order Thinking Skills (HOTS)-Based Science. *International Journal of Instruction*, *16*(4), 283–296. https://doi.org/10.29333/iji.2023.16417a

Rodríguez, M. F., Nussbaum, M., Yunis, L., Reyes, T., Alvares, D., Joublan, J., & Navarrete, P. (2022). Using scaffolded feedforward and peer feedback to improve problem-based learning in large classes. *Computers & Education*, *182*, 104446. https://doi.org/10.1016/j.compedu.2022.104446

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, *91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Trustee, E. (2024). Secondary School Learners' Perceptions of the Impact of Project-Based Learning on Academic Performance. *ETDC: Indonesian Journal of Research and Educational Review*, *4*(1), 57–67. https://doi.org/10.51574/ijrer.v4i1.2689

Wijayanto, B., Sumarmi, S., Hari Utomo, D., Handoyo, B., & Aliman, M. (2023). Problem-based learning using e-module: Does it effect on student's high order thinking and learning interest in studying geography? *Journal of Technology and Science Education*, *13*(3), 613. https://doi.org/10.3926/jotse.1965

Zhai, X., & Nehm, R. H. (2023a). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, *60*(6), 1390–1398. https://doi.org/10.1002/tea.21885

Zhai, X., & Nehm, R. H. (2023b). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, *60*(6), 1390–1398. https://doi.org/10.1002/tea.21885

Zhao, Y., Yue, Y., Sun, Z., Jiang, Q., & Li, G. (2025). Does Generative Artificial Intelligence Improve Students' Higher-Order Thinking? A Meta-Analysis Based on 29 Experiments and Quasi-Experiments. *Journal of Intelligence*, *13*(12), 160. https://doi.org/10.3390/jintelligence13120160