Berkala Ilmiah Pendidikan
Volume 6, Number 2, July 2026

scidac plus

Development of Teaching Modules Lesson Study Based Integrated Technology Generative AI in Indonesian Language Learning to Develop Critical Thinking Skills of Elementary School Students in Bungo Regency

Feerlie Moonthana Indhra[1*], Sonia Yulia Friska[2], M. Muzakki[1]
[1]Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, Indonesia
[2]Universitas Islam Yasni Bungo Jambi, Indonesia
monthana71@gmail.com*

Abstract

The low critical thinking skills of elementary school students in Indonesian language learning, exacerbated by the lack of adaptive teaching tools for technological disruption, poses a pressing pedagogical challenge in developing regions. This research aims to develop valid, practical, and effective teaching modules through the integration of a collaborative approach. *Lesson Study* and technology *Generative Artificial Intelligence* (GenAI). The research method used is *Research and Development* (R&D) by adopting the 4D model (*Define, Design, Develop, Disseminate*). The trial subjects involved 34 fifth-grade students at SD Negeri 1 Muara Bungo, Bungo Regency, Jambi. Data collection was carried out through expert validation sheets, response questionnaires, and observations. *Lesson Study*, and critical thinking essay test. The results of the study showed: (1) The teaching module obtained very good validity scores from material experts (93.2%), media (88.5%), and language (91.0%) with an Aiken's V value of > 0.85; (2) The level of practicality based on teacher and student responses reached 89.4% (very practical category); (3) The effectiveness test showed a significant increase in students' critical thinking skills with a value of *N-Gain* of 0.66 (medium-high category), where the analysis and evaluation indicators experienced the sharpest increase. GenAI integration in the cycle *Plan-Do-See* on *Lesson Study* It has been proven to facilitate teachers' pedagogical reflection and challenge students to go beyond simply accepting machine-generated information to logically critique technological output. This module is worthy of adoption as a supplement to Indonesian language learning in elementary schools, particularly in areas transitioning to digital literacy.

Keywords: Critical Thinking, Generative AI, Lesson Study, Teaching Modules, Elementary School

INTRODUCTION

The paradigm shift in education in the 21st century demands a fundamental transformation in how students interact with information. The Independent Curriculum emphasizes the development of essential competencies, one of which is critical thinking skills (*critical thinking*) (BSKAP, 2022). In the context of Indonesian language learning in elementary schools (SD), critical thinking is no longer simply defined as the ability to literally understand text, but rather as a high-level cognitive competency for analyzing, evaluating, and concluding information from a massive and unfiltered array of digital sources. However, this pedagogical urgency confronts the reality of technological disruption that threatens students' critical thinking, necessitating adaptive and collaborative teaching tool interventions.

The presence of technology *Generative Artificial Intelligence* (GenAI) has fundamentally changed the cognitive landscape of students. Based on a recent literature review, GenAI has triggered a phenomenon*cognitive offloading*, where students tend to delegate higher-order thinking processes to machines, leading to stagnation in analytical and evaluation skills (Kasneci et al., 2023; Su & Yang, 2023). On the other hand, GenAI often produces "hallucinations" or information bias because it is trained using global data that is not always relevant to local contexts (Mollick & Mollick, 2023). If not guided with appropriate pedagogy, GenAI will stifle critical thinking. Conversely, if strategically integrated, GenAI can transform from a mere "answering tool" into a "critical audit object" that challenges students to verify, debate, and correct machine output (Williamson, 2023). Therefore, a teaching module is needed that not only teaches language content but also frames GenAI as a stimulus for practicing critical evaluation skills.

Innovation in teaching materials alone is not enough without increasing teachers' capacity to implement them. This is where the approach comes in. *Lesson Study* become a key variable that bridges the gap between curriculum demands and classroom practice. *Lesson Study*with cycle*Plan-Do-See*offers a collaborative space for teachers to design, test, and reflect on pedagogical strategies that critically integrate technology (Lewis & Hurd, 2011). The interrelationships between these variables are logically constructed: GenAI provides authentic problems (information that needs to be verified), the Teaching Module becomes a structured activity platform, Lesson Study becomes a mechanism for developing teachers' capacity to facilitate critical discussions, and Critical Thinking is a*output*cognitive targets of the evaluation process of the technology.

Despite the clear urgency, the reality on the ground reveals a wide gap. A preliminary study conducted in August 2025 at five public elementary schools in Bungo Regency, Jambi Province, identified concerning issues. Initial data were collected using critical thinking diagnostic test instrument (adapting the Facione (2015) rubric given to 120 fifth-grade students, as well as structured interviews with 10 Indonesian language teachers. The results showed that the average score of students' critical thinking skills (especially in the argument evaluation indicator) only reached 48.2 on a scale of 100. In addition, 85% of teachers admitted that they were not yet able to design activities that critique technology, and students' use of gadgets was still consumptive. This empirical data confirms the existence of pedagogical stagnation that requires immediate intervention.

The local context of Bungo Regency offers unique opportunities to address these issues. Bungo Regency boasts a wealth of ethnopedagogy and local wisdom, such as the ecosystem of Lake Kaco, the dynamics of the Tembesi River, and traditions *Palabe*. This regional characteristic has not been optimally utilized in the design of teaching modules. In fact, in the context of artificial intelligence, large language models (LLMs) often fail or experience hallucinations when confronted with hyper-local contexts. These local characteristics of Bungo will be designed as "verification baseline" in the teaching module. Students not only learn the language, but use their local knowledge of Bungo to "catch" logical or factual errors generated by GenAI. This ethnopedagogical approach makes the local context not just an add-on, but the core of students' validation and critical thinking processes (Indhra et al., 2025).

A review of previous research shows that research on GenAI in primary education is still very limited and tends to be normative. Pradana and Setyawan's (2024) study discusses the integration of AI for literacy, but its focus remains on*prompt engineering*without a critical evaluation framework. Susanti and Wibowo's (2025) research analyzed critical thinking through project-based learning, but ignored the disruption of GenAI. Meanwhile, research on*Lesson Study* (Widodo & Riandi, 2022) generally focuses on conventional pedagogy of the pre-AI era.

The research gap in this study lies in the absence of a teaching tool model that positions GenAI not as a "source of truth", but as an "object of criticism" facilitated through a collaborative cycle. *Lesson Study*, especially in areas that are transitioning towards digital literacy. The claim of novelty of this research includes three specific aspects: (1) Conceptual Novelty: Shifting the paradigm of AI integration from *answer generator* become *object of critical audit* through the "AI Fact Detective" activity; (2) Contextual Novelty: Utilization of local wisdom of Bungo Regency as *baseline* verification to train students to detect AI hallucinations; (3)Methodological Novelty: Embedding technology criticism into the reflection cycle *Lesson Study* (*Plan-Do-See*) to improve teacher professionalism in a sustainable manner.

Based on the urgency, research gaps, and novelty, this research was formulated to answer the question: (1) What are the characteristics of learning modules based on *Lesson Study* integrated GenAI with valid and practical Bungo ethnopedagogy content? (2) How effective is the teaching module in developing the critical thinking skills of elementary school students? This research is expected to provide theoretical contributions to the critical pedagogy model of the AI era, as well as practical contributions in the form of contextual teaching tools for elementary school teachers in developing areas.

METHOD

Research Design

This study employs a Research and Development (R&D) approach to design, develop, and empirically test an educational product. We adopted the 4D development model developed by Thiagarajan et al. (1974), which comprises four systematic and iterative stages: (1) Define, (2) Design, (3) Develop, and (4) Disseminate. This model provides an explicit structure that bridges needs analysis, expert validation, field trials, and product dissemination, making it highly suitable for iterative educational interventions.

To ensure consistency, we mapped each stage against specific operational outputs and strict exit criteria that had to be met before progressing to the subsequent phase:

1. Define: Produces a comprehensive needs analysis document.
2. Design: Generates the initial draft of the teaching modules.
3. Develop: Yields validated, practical, and field-tested instructional modules.
4. Disseminate: Results in effective, finalized modules implemented through Lesson Study cycles.

Participants and Setting

The study was conducted at SD Negeri 164/II Gapura Suci, Bungo Regency, Jambi Province, Indonesia, during the odd semester of the 2025/2026 academic year. We selected this location based on three criteria: (1) the school possesses basic technological infrastructure, including projectors and stable yet limited internet access; (2) it has not yet optimized technology integration for Higher-Order Thinking Skills (HOTS) instruction; and (3) it represents elementary schools in developing regions undergoing a transition toward digital literacy.

The research participants included:

1. Main Test Subjects: 34 fifth-grade students (aged 10–11 years) selected through purposive sampling. The selection was based on the Phase C Learning Outcomes (*Capaian Pembelajaran*) under the Indonesian Independent Curriculum (*Kurikulum Merdeka*), which require students to analyze explanatory and persuasive texts.
2. Collaborators: Fourth- and fifth-grade Indonesian language teachers who participated in the Lesson Study cycles.
3. Validators: Three independent experts specializing in Indonesian language materials, digital learning media, and educational evaluation.
4. Instrument Testing Group: 30 fifth-grade students from SD Negeri 2 Muara Bungo selected to test instrument reliability.

Development and Lesson Study Integration Procedures

The development procedure integrates the 4D model with Lesson Study cycles and Generative Artificial Intelligence (GenAI) tasks at every stage:

Define Stage (Definition)

This stage involved three core analyses:

1. Front-End Analysis: Mapping the pedagogical gap between the critical thinking demands of the Independent Curriculum and actual classroom realities.
2. Task Analysis: Identifying critical thinking indicators specifically interpretation, analysis, evaluation, and inference adapted from the framework by Facione (2015).
3. Concept Analysis: Mapping the Indonesian language curriculum materials (explanatory texts) and contextualizing them with the local wisdom of Bungo Regency, such as Kaco Lake, the Tembesi River, and traditional *Palabe* customs.

Design Stage (Planning)

We designed the draft teaching module to align with the Independent Curriculum structure, enriched by three components:

1. "AI Fact Detective" Activity: Students input specific prompts into a safe, content-screened educational version of GenAI to generate an explanatory text. Students then act as "Expert Editors" to critique the AI output, identifying hallucinations, cultural biases, and inconsistencies regarding the local Bungo context.
2. Lesson Study Scenarios: Detailed observation sheets and reflection guidelines structured around the *Plan-Do-See* cycle.
3. Contextual Worksheets: Student tasks based on Bungo ethnopedagogy used as a baseline to verify the accuracy of the AI-generated output.

Develop Stage (Development)

This stage focused on expert validation, iterative revisions, and field testing:

1. Evaluation of the module prototypes by the three experts using 1–4 Likert scale assessment sheets.
2. Product revisions based on qualitative and quantitative expert feedback.
3. A field trial with the 34 main subjects to evaluate practicality and preliminary effectiveness.
4. Execution of a full Lesson Study cycle (*Plan-Do-See*) to serve as a mechanism for collaborative pedagogical reflection.

Disseminate Stage (Dissemination)

The finalized module was implemented within a single-school context using a One-Group Pretest-Posttest design to assess its effectiveness. We recommend wider dissemination across regional school clusters for future research.

Research Instruments

We developed four instruments to ensure data validity and reliability:

Expert Validation Sheets

These sheets utilize a 1–4 Likert scale across 40 evaluation statements per aspect (content material, media design, and language readability). Content validity was calculated using Aiken's V formula:

$$V = \frac{\sum s}{n(c-1)}$$

Where $s = r - l$ (r represents the rating given by an evaluator, and l is the lowest possible score), n is the total number of evaluators, and c is the highest possible score. An instrument is considered valid if Aiken's $V > 0.80$.

Critical Thinking Essay Test

This test consists of five essay questions set within the local Bungo context, adapted from Facione's (2015) core indicators: interpretation, analysis, evaluation, and inference. Student responses were scored using a holistic 0–4 rubric per indicator.

Prior to field implementation, we piloted the test with 30 students outside the main sample. Reliability analysis yielded a Cronbach's Alpha (α) of 0.82, with an average item discrimination index of 0.41, confirming that the instrument is highly reliable and possesses good discriminatory power.

Teacher and Student Response Questionnaires

These questionnaires measure the practical utility of the module using a 1–4 Likert scale. Construct validity was verified via Corrected Item-Total Correlation (retaining items > 0.30). Internal consistency testing showed high reliability, with a Cronbach's Alpha of $\alpha = 0.87$ for the teacher questionnaire and $\alpha = 0.84$ for the student questionnaire.

Lesson Study Observation Sheets

Structured observation guidelines were used to document the students' cognitive dynamics when evaluating AI outputs and to record the teachers' pedagogical reflections. To maintain validity, we triangulated the observation data across three distinct sources: the participating teachers, the students, and independent observers.

Data Collection Techniques

Data were gathered using a method-triangulation approach across four primary techniques:

1. Expert Validation: Administered during the *Develop* stage to establish product validity.
2. Surveys: Teacher and student response questionnaires distributed immediately following the intervention to assess practicality.
3. Field Observations: Documented using Lesson Study observation sheets during the ongoing *Plan-Do-See* cycles.
4. Learning Outcomes Testing: The critical thinking essay test administered twice: as a pretest before the intervention and as a posttest after completing the Lesson Study cycles.

Research Design Justification

This study applied a One-Group Pretest-Posttest design (Creswell & Guetterman, 2019), systematically represented as follows:

We selected this design due to two primary research constraints:

- Population Limitations: The sample consisted of only 34 students within a single classroom. Dividing these students to establish a separate control group within the same school environment was deemed methodologically unrepresentative and ethically problematic.

- R&D Framework: The overarching objective of this study is to evaluate the developmental viability and relative effectiveness of the new educational product rather than to isolate pure experimental causality between isolated variables.

To mitigate threats to internal validity, we conducted treatment fidelity checks through Lesson Study observations to guarantee that the intervention strictly followed the planned instructional design. Potential history effects were minimized by keeping the intervention duration brief, spanning exactly three instructional sessions.

Data Analysis Techniques

Quantitative Analysis

1. Product Validity: We calculated the validity percentage using the following formula:
2. $\text{Validity Percentage} = \left(\frac{\text{Acquired Score}}{\text{Maximum Possible Score}}\right) \times 100\backslash\%$
3. Scores were categorized based on Akbar's (2013) criteria: $81\backslash\% - 100\backslash\%$ indicates very valid, $61\backslash\% - 80\backslash\% valid, and 41\backslash\% - 60\backslash\%$ moderately valid.
4. Practicality: Determined by computing the percentage of positive responses from the teacher and student questionnaires.
5. Effectiveness: Measured using the normalized Hake's Gain (N-Gain) formula:

$$g = \frac{\text{Posttest Score} - \text{Pretest Score}}{\text{Maximum Score} - \text{Pretest Score}}$$

The resulting g scores were classified as high ($g > 0.7$), moderate ($0.3 \leq g \leq 0.7$), or low ($g < 0.3$). We verified data normality using the Shapiro-Wilk test, followed by a paired-sample t-test at a significance level of $\alpha = 0.05$ to evaluate the statistical significance of the post-intervention improvements.

Qualitative Analysis

Qualitative data derived from Lesson Study reflections and field notes were analyzed using the interactive model by Miles, Huberman, and Saldaña (2014), which involves three concurrent flows of activity:

1. Data Condensation: Selecting, focusing, and transforming raw field notes to isolate students' cognitive struggles and teachers' pedagogical shifts.
2. Data Display: Organizing the condensed data into structured descriptive narratives and thematic matrices.
3. Conclusion Drawing and Verification: Identifying recurring patterns in teacher behavior and student critical thinking to formulate grounded conclusions.

Finally, we performed data triangulation by cross-examining the quantitative growth (N-Gain scores) against the qualitative observation metrics to ensure the validity of our findings.

Product Revision Process

Product modifications occurred iteratively across three developmental checkpoints:

1. Post-Expert Validation Revisions: Feedback from the three independent validators was logged into a specialized revision matrix and classified as either major (structural and conceptual modifications) or minor (editorial and typographic adjustments). Revised drafts were returned to the respective experts for final re-verification.
2. Post-Field Trial Revisions: Based on student and teacher feedback from the initial trial, we refined the module's instructional readability, adjusted activity durations, and optimized the level of difficulty of the GenAI prompts.

3. Post-Lesson Study Revisions: Teacher reflections during the *See* phase highlighted specific learning bottlenecks. In response, we integrated targeted scaffolding elements into the final module, including a logical conjunction guide and a localized Bungo context mini-dictionary, to assist lower-performing students.

All modifications were compiled into a transparent track-change matrix documenting: (a) the initial evaluator input, (b) the concrete revision made, (c) the pedagogical justification, and (d) the final expert sign-off.

Ethical Considerations

This study adhered to the educational research ethics guidelines approved by the institutional ethics commission of UIN Sulthan Thaha Saifuddin Jambi, structured around six core ethical principles:

1. Informed Consent: We secured formal, written consent from the school principal, participating teachers, and the students' parents or legal guardians. Students received an age-appropriate explanation regarding the study's purpose and were explicitly informed of their right to withdraw at any time without academic penalty.
2. Anonymity and Confidentiality: To protect participant privacy, all student identities were anonymized using alphanumeric codes (e.g., S1, S2). Research data were kept in encrypted digital storage accessible exclusively by the primary research team.
3. Beneficence: The intervention was designed to ensure no physical or psychological discomfort occurred. The "AI Fact Detective" task was presented as an engaging, low-stakes educational activity rather than a high-pressure examination.
4. Justice: All participants received equitable access to the technological tools and teacher guidance. Students facing literacy difficulties were provided with additional instructional scaffolding to ensure inclusive learning opportunities.
5. Ethical Use of GenAI: Students were explicitly taught digital ethics, emphasizing that GenAI text should not be plagiarized or presented as original work. Instructors highlighted technological limitations, and students interacted only with a secure, content-screened educational layout.
6. Academic Integrity: All research data, findings, and statistical outcomes are reported transparently, acc

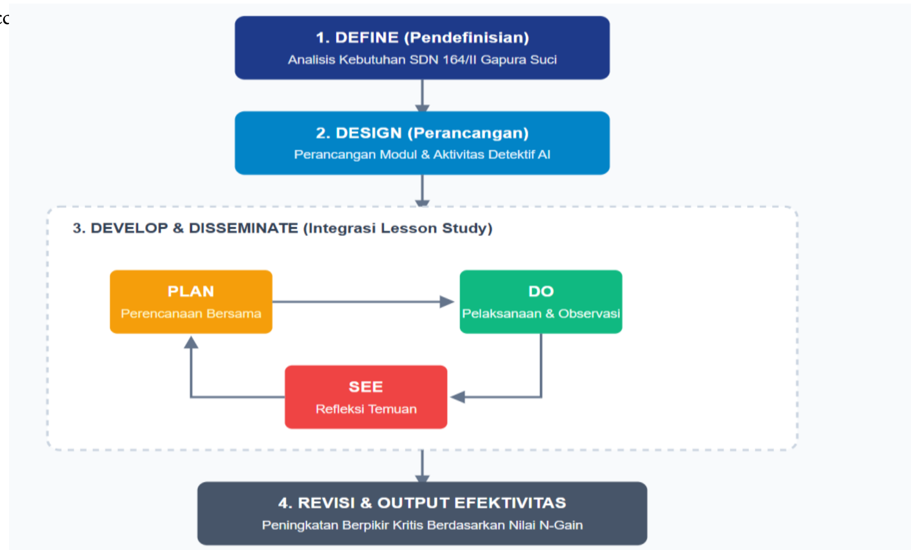

Figure 1. Lesson Study Integration Flowchart in the 4D Model

RESULTS AND DISCUSSION

Define Stage

The results of the needs analysis at SD NEGERI 164/II GAPURA SUCI revealed three crucial findings. First, 85% of teachers acknowledged the difficulty in designing learning activities that explicitly train critical thinking, beyond simply memorizing text structures. Second, the school's technological resources are adequate, but their use is still limited to displaying static presentation slides. Third, students demonstrate high enthusiasm for gadgets, but their digital literacy tends to be consumptive. They have not been trained to distinguish between facts, opinions, and algorithmic "hallucinations." These findings provide a strong foundation that the developed teaching modules must bridge the gap between technological availability and critical thinking pedagogy.

Design Stage (Planning)

The product designed is an Indonesian Language Teaching Module for the topic "Writing and Analyzing Simple Scientific Explanation Texts." The module structure refers to the Independent Curriculum format, but with the addition of special components that provide novelty value (*novelty*) this research:

Understanding Meaning

Emphasizing that "Good text is text whose logic can be accounted for, whether written by humans or machines."Core Activity (GenAI Integration): Students are divided into groups. Each group is given the same prompt to produce an explanatory text using the GenAI tool (e.g., an educationally safe version). The students' task is not to copy, but to become "Expert Editors." They must identify illogical sentences, biased facts, or inappropriate structures, and then revise them manually by referring to source books or local knowledge about the nature of Bungo Regency.

Lesson Study Worksheet: Includes an observation sheet for guest teachers (observers) who observe the group discussion, focusing on how students present arguments when rejecting or accepting AI output.

Develop Stage (Development and Validation)

After the draft was completed, it was validated by three experts. The assessment results are presented in Table 1.

Table 1. Expert Validation Results for the Teaching Module

Evaluation Aspects	Validator	Acquisition Score	Maximum Score	Percentage	Aiken's V	Category
Content/Material Suitability	Member 1	38	40	95,0%	0,93	Very Valid
Presentation & Media Suitability	Member 2	35	40	87,5%	0,88	Valid
Linguistic Eligibility	Member 3	36	40	90,0%	0,91	Very Valid
Overall Average		109	120	90,8%	0,90	Very Valid

Revision Note: The content expert suggested that the GenAI prompts provided in the module be more specific and contextual. For example, rather than asking the AI to explain "volcano" in general, the prompts should be directed to explain the geological phenomenon of "Lake Kaco" or the rubber plantation ecosystem in Bungo. This makes it easier for students to connect their prior knowledge to the machine's output, leading to a more rigorous evaluation process. These revisions were immediately implemented in the final draft.

Furthermore, a practicality test was conducted on two teachers and 34 students. The average teacher questionnaire response reached 91.5% (Very Practical), with comments that this module "helped teachers lead more lively discussions because students had concrete objects (AI texts) to critique, not just imagination." Student responses reached 87.3% (Very Practical), with indicators that students felt "more challenged and less bored" compared to conventional learning.

Disseminate Stage (Test of Lesson Study Effectiveness and Implementation)

The module implementation is carried out through a Lesson Study cycle for three meetings.

Plan: The teacher team collaborated to modify the "AI Fact Detective" activity to suit the dynamics of fifth grade, including preparing backup offline activities in case the internet connection is disrupted.

Do: The lesson was implemented. The researcher acted as an observer. It was observed that initially, students tended to simply accept the text from the AI because it was considered "smart." However, with the guidance of the teacher's provocative questions ("Does this sentence make sense if it happened in our village? Try checking this fact in a library book"), students began actively looking for anomalies. A healthy debate ensued among students regarding the accuracy of the data.

See: In the post-lesson reflection session, the teachers concluded that their role had shifted from "informers" to "facilitators of critical discussion." They also recognized the need to provide scaffolding in the form of a list of logical connectives to help students formulate rebuttals to the AI text. To measure effectiveness, a critical thinking test (pretest and posttest) was administered. Data were analyzed using descriptive statistics and the N-Gain test.

Table 2. Results of Pretest and Posttest of Critical Thinking Skills

Critical Thinking Indicators	Shoes Rata-rata Pretest	Average Posttest Score	N-Gain Value	Category
Interpretation	56,2	73,5	0,39	Currently
Analysis	49,0	79,4	0,59	Currently
Evaluation	43,5	82,0	0,68	Medium-High
Inference (Concluding)	51,1	76,2	0,51	Currently
Rate-Rate Total	50,0	77,7	0,66	Medium-High

Table 2 shows consistent improvement across all indicators. The most significant improvement occurred in the Evaluation indicator (N-Gain 0.68). This is logical, as the main activity in the module requires students to assess the accuracy and quality of machine-generated text. Students learn that authority does not always reside in technology, but rather in human reason, which is able to verify information.

In-depth Discussion

The findings of this study confirm that the integration of GenAI into teaching modules, when framed through a Lesson Study approach, does not pose a threat to the authenticity of students' work, but rather a catalyst for critical thinking. This aligns with the opinions of recent education experts (Kohnke et al., 2023; Pradana & Setyawan, 2024), who state that AI literacy in elementary schools should focus on critical evaluation rather than simply prompt engineering.

In the context of Bungo Regency, where exposure to advanced technology may not be as deep as in metropolitan cities, this approach has proven to be inclusive and adaptive. The module doesn't require each student to have a sophisticated device; a single device, operated by the teacher or in groups, is sufficient, while the cognitive focus remains on analyzing text on paper or a whiteboard. This addresses concerns about the digital divide, a frequent obstacle to technology R&D in the region.

The successful improvement in scores on evaluation and analysis indicators is inseparable from the Lesson Study mechanism. The Plan-Do-See cycle forces teachers to go beyond simply "completing the material" to truly observe how students think. During the See session, teachers realized that open-ended questions were more effective in sparking healthy debate among students about AI output than closed-ended questions. This collaborative reflection then refined the learning execution in subsequent sessions, creating a continuous spiral of pedagogical improvement (Lewis & Hurd, 2011).

However, this study also identified challenges that need to be addressed. Approximately 15% of students with low basic literacy skills experienced difficulty distinguishing between facts and opinions in complex AI texts. They tended to become frustrated when they couldn't find the absolute "right" answer. This suggests that this module needs to be supplemented with more intensive scaffolding, such as providing a mini-dictionary or a list of logical conjunctions, for students who require additional support. Furthermore, reliance on an internet connection, even if minimal, remains a risk factor that must be managed through a blended activity design.

Discussion

In-depth Analysis of Research Results: Shift from Consumer to Cognitive Auditor The results of this study show highly significant findings, particularly in the spike in the Evaluation indicator with the highest N-Gain value (0.68), compared to the Interpretation (0.39) and Inference (0.51) indicators. The researchers' analysis shows that the high evaluation scores are not coincidental, but rather a direct impact of the "AI Fact Detective" activity design.

Cognitively, when students are confronted with texts claimed to be "produced by intelligent machines," cognitive dissonance occurs when they discover a discrepancy between the AI's output and the empirical reality of their environment (Bungo Context). This process forces students to go beyond simply understanding the text (interpretation) to actively compare, dispute, and assess the credibility of the information (evaluation). These results demonstrate that pedagogical interventions that position AI as "wrong objects/need to be corrected" effectively break down the technology bias, where elementary school students generally tend to assume machines are always right.

Dialogue with Theory: Beyond Cognitive Offloading through Epistemic Vigilance These findings directly align with Kasneci et al.'s (2023) theory of cognitive offloading, which warns that AI can lead to stagnation of reasoning as students delegate their thinking processes to machines. This research demonstrates that this danger can be mitigated through appropriate module design. Rather than allowing students to engage in cognitive offloading, the module forces "reverse cognitive loading" where the cognitive burden is returned to the students to audit the machine.

Furthermore, when compared with Facione's (2015) Critical Thinking framework, posttest results confirm that evaluation skills are at the top of the cognitive hierarchy in the AI era. Students not only use Facione's criteria (argument analysis), but also develop what Williamson (2023) calls epistemic vigilance the innate ability to filter out false information. This teaching module successfully translates the abstract concept of epistemic vigilance into concrete activities in fifth-grade elementary school through cross-verification with local knowledge.

Dialogue with Previous Research: Affirming Research Gaps and Novelty This research offers a fundamentally different contribution compared to previous studies. Pradana and Setyawan (2024) focused on integrating AI into literacy through prompt engineering (how students command AI). This study critiques this approach as insufficient; teaching students how to "ask" AI is useless if they lack the ability to "challenge" AI's answers. The novelty of this research lies in the shift in focus from prompting to auditing.

On the other hand, Susanti and Wibowo (2025) found that Project-Based Learning (PjBL) can enhance critical thinking in developing regions. However, this study shows that conventional PjBL does not specifically train students' resilience to algorithmic disinformation. The integration of GenAI in this study provides a measurable "hallucinatory" stimulus, which conventional PjBL cannot provide. Thus, this study fills a gap by providing a specific pedagogical model for the era of information disruption, which has previously been discussed more in higher education contexts (Kohnke et al., 2023), rather than in primary education.

The Crucial Role of Lesson Study and Hyper-Local Context The success of this critical thinking improvement is inseparable from the Lesson Study mechanism. The Plan-Do-See cycle acts as pedagogical scaffolding for teachers. In the See stage, teacher reflections found that students had difficulty formulating rebuttals due to limited logical vocabulary. This qualitative finding was immediately corrected in the next meeting by providing a "Logical Connective Word Dictionary" (e.g., however, contrary to the facts, logically unreasonable). This proves that Lesson Study is not merely an administrative procedure, but rather a driving force for real-time pedagogical adaptation (Lewis & Hurd, 2011).

Furthermore, the use of the local context of Bungo Regency (Lake Kaco, the Palabe tradition) serves as an epistemological anchor. Large language models (LLMs) are trained on global data and often experience hallucinations when confronted with hyper-local contexts. The AI's unfamiliarity with Bungo's local details actually creates a "gap" that allows elementary school students who have prior knowledge of their region to feel empowered and confident in critiquing the machine. Local wisdom transforms into a powerful tool for scientific validation.

Limitations and Implications (Critical Analysis) While effective, this study has limitations that must be honestly analyzed. Qualitative data indicates that 15% of students with low basic literacy experienced cognitive frustration when critiquing complex AI texts. This indicates that AI integration cannot replace a foundation of basic literacy. Critical thinking about AI requires students to have sufficient declarative knowledge (basic facts) to make comparisons. Therefore, this module is not recommended for students who have not reached the minimum literacy level, and requires modifications to the complexity of the AI texts.

The theoretical implication of this research is the emergence of an "AI-as-Auditor" model for primary education, shifting the "AI-as-Tutor" paradigm. For practitioners in developing regions, this research provides evidence that infrastructure limitations (e.g., only one device per class) are not a barrier, as long as the activity design focuses on collaborative discussion and argumentation, rather than individual access to technology.

CONCLUSION

Based on the results of the research and discussion, it can be concluded that: (1) The Indonesian language teaching module is based on *Lesson Study* integrated *Generative AI* The module developed has met the criteria of being very valid (90.8%) and very practical (89.4%) for use at elementary school level; (2) Implementation of this module has proven effective in developing students' critical thinking skills, with the achievement of grades *N-Gain* average of 0.66 (medium-high category), especially in the evaluation and text analysis aspects. Synergy between teacher collaboration (*Lesson Study*) and the cognitive challenges of critiquing technology (GenAI) create a learning ecosystem that is relevant to the demands of the 21st century, even in regional contexts such as Bungo Regency.

For education practitioners, it is recommended to adopt this module while still contextualizing it to local issues in their respective school environments. Teachers should be provided with brief training on the ethics and limitations of using GenAI to ensure optimal classroom support. For future researchers, it is recommended to develop similar research with a broader scope (large-scale field trials with a control group) or explore the integration of GenAI into other language skills, such as listening or speaking, using different development models.

REFERENCES

Anggrayni, M., Friska, S. Y., & Retnawati, E. (2023). Pengembangan modul ajar mata pelajaran IPAS dalam Kurikulum Merdeka pada kelas IV sekolah dasar. *Innovative: Journal of Social Science Research*, *3*(2), 14504–14516.

Badan Standar, Kurikulum, dan Asesmen Pendidikan (BSKAP). (2022). *Capaian pembelajaran mata pelajaran Bahasa Indonesia Fase A-F*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment.

Indhra, F. M., Fitrah, Y. F. Y., Nazurty, N., & Kusmana, A. (2025). Educational values in Bungo community Palabe and its relevance as Indonesian language teaching materials in preschools. *Jurnal Konseling dan Pendidikan*, *13*(1), 475–490.

Kasneci, E., Seßler, K., Küchemann, S., Tanieva, M., Ilter, D., & Webersinke, F. (2023). ChatGPT: Opportunities, challenges, and limitations in education. *Computers and Education: Artificial Intelligence*, *5*, Article 100164.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). *ChatGPT for language teachers: Educate them and integrate it*. TESOL Press.

Lewis, C., & Hurd, J. (2011). *Lesson study step by step: How teacher learning communities improve instruction*. Heinemann.

Mollick, E. R., & Mollick, L. (2023). *Using AI to implement effective teaching strategies in classrooms: Agent-based simulations and AI as a cognitive tool* (Wharton School Research Paper). The Wharton School.

Pradana, A. R., & Setyawan, F. (2024). Integrasi kecerdasan buatan dalam pembelajaran literasi di sekolah dasar: Peluang dan tantangan pedagogis. *Jurnal Pendidikan Dasar Indonesia*, *9*(1), 45–58.

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Journal of Educational Innovation*, *1*(3), 1–15.

Susanti, D., & Wibowo, A. (2025). Analisis kemampuan berpikir kritis siswa SD dalam pembelajaran berbasis proyek di daerah berkembang. *Cakrawala Pendidikan*, *44*(2), 312–325.

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Center for Innovation in Teaching the Handicapped, Indiana University.

Widodo, A., & Riandi. (2022). Model pembelajaran lesson study untuk meningkatkan profesionalisme guru di era digital. *Jurnal Ilmiah Sekolah Dasar*, *6*(3), 401–410.

Williamson, B. (2023). The AI classroom: Critical perspectives on generative models in education. *Learning, Media and Technology*, *48*(1), 1–15.

Yusuf, M., & Fitriani, A. (2026). Evaluasi kritis terhadap output generative AI oleh siswa sekolah dasar: Sebuah studi eksperimental di Jambi. *Jurnal Teknologi Pendidikan dan Pembelajaran*, *14*(1), 88–102.