Bima Journal of Elementary Education
Volume 4. No. 1, April 2026
The article is published with Open Access at: ejournal.tsb.ac.id/index.php/bijee

Integrating Large Language Model as an Adaptive Reading Assistant to Enhance Reading Literacy among Fifth-Grade Elementary School Students in Indonesia

Isna Muhammad Fathoni[1),*], Sarifatul Adawiyah[2)], Raka Firman Baskara Permana[3)], Ilham Jaya[1)]

[1)]Universitas Bani Saleh, Bekasi, Indonesia
[2)]Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia
[3)]The University of Electro-Communications, Tokyo, Japan

*Correspondence: isna@ubs.ac.id

Abstract: Reading literacy is a fundamental competency determining academic success, yet Indonesian elementary students rank 71st out of 81 countries in PISA 2022. This study examined the integration of Large Language Models (LLMs) as adaptive reading assistants to enhance elementary students' reading literacy through three objectives: (1) developing an LLM-based adaptive reading assistant prototype; (2) measuring its effectiveness; and (3) exploring students' learning experiences. An explanatory sequential mixed-method design with a quasi-experimental nonequivalent control group involved 72 fifth-grade students from two public elementary schools in Bekasi (experimental n=36; control n=36). Data were collected through a PIRLS-adapted reading literacy test, classroom observations, and semi-structured interviews. Quantitative data were analyzed using ANCOVA with partial eta squared (η^2_p) as the effect size estimator; qualitative data underwent reflexive thematic analysis. Results: (1) The prototype successfully delivered adaptive questioning, scaffolding, and Socratic dialogue; (2) the experimental group exhibited significantly higher gains (M=23.45) than the control group (M=9.82), $F(1,69)=50.21$, $p<0.001$, $\eta^2_p=0.42$, 95% CI [10.97, 16.29], Cohen's d=1.42; and (3) students reported increased engagement, confidence, and self-regulated learning. This study is the first to operationalise cognitive apprenticeship principles in an Indonesian-language LLM prompt designed for elementary readers, combining a quasi-experimental evaluation with ethical safeguards for minors. LLM integration is an effective pedagogical strategy when complemented by teacher mediation.

Keywords: Large Language Model, Adaptive Learning, Reading Literacy, Elementary Education, AI in Education

Abstrak: Reading literacy is a fundamental competency determining academic success, yet Indonesian elementary students rank 71st out of 81 countries in PISA 2022. This study examined the integration of Large Language Models (LLMs) as adaptive reading assistants to enhance elementary students' reading literacy through three objectives: (1) developing an LLM-based adaptive reading assistant prototype; (2) measuring its effectiveness; and (3) exploring students' learning experiences. An explanatory sequential mixed-method design with a quasi-experimental nonequivalent control group involved 72 fifth-grade students from two public elementary schools in Bekasi (experimental n=36; control n=36). Data were collected through a PIRLS-adapted reading literacy test, classroom observations, and semi-structured interviews. Quantitative data were analyzed using ANCOVA with partial eta squared (η^2_p) as the effect size estimator; qualitative data undergoing reflexive thematic analysis. Results: (1) The prototype successfully delivered adaptive questioning, scaffolding, and Socratic dialogue; (2) the experimental group exhibited significantly higher gains (M=23.45) than the control group (M=9.82), $F(1,69)=50.21$, $p<0.001$, $\eta^2_p=0.42$, 95% CI [10.97, 16.29], Cohen's d=1.42; and (3) students reported increased engagement, confidence, and self-regulated learning. This study is the first to operationalize cognitive apprenticeship principles in an Indonesian-language LLM prompt designed for elementary readers, combining a quasi-experimental evaluation with ethical safeguards for minors. LLM integration is an effective pedagogical strategy when complemented by teacher mediation.

Kata kunci: Large Language Model, Pembelajaran Adaptif, Literasi Membaca, Pendidikan Dasar, AI dalam Pendidikan

INTRODUCTION

Reading literacy stands as one of the most consequential competencies in contemporary education, functioning not only as a prerequisite for academic success across all subject domains but also as a critical determinant of lifelong learning, civic participation, and economic mobility. The Organisation for Economic Co-operation and Development (OECD, 2023) conceptualizes reading literacy as understanding, using, evaluating, reflecting on, and engaging with texts to achieve personal goals, develop knowledge, and participate meaningfully in society. Despite global recognition of its importance, Indonesia continues to face a persistent literacy crisis: PISA 2022 placed the country 71st of 81 participating nations, with only 25% of Indonesian fifteen-year-olds reaching the minimum proficiency level. This expanded definition, embraced by international assessments such as PISA and PIRLS, recognizes that reading is no longer a passive decoding skill but an active, strategic, and metacognitive process (Castles et al., 2018; Mullis et al.,

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

2023). The increasing complexity of texts in digital environments, the proliferation of misinformation, and the cognitive demands of twenty-first-century knowledge work collectively elevate the importance of robust reading literacy from the earliest grades. Empirical evidence consistently demonstrates that students who develop strong reading literacy during elementary years exhibit better long-term academic trajectories, higher graduation rates, and improved economic outcomes (Duke et al., 2021; Snow, 2017).

Despite the global recognition of reading literacy's importance, Indonesia continues to confront a persistent literacy crisis among elementary school students. The PISA 2022 results revealed a sobering reality: Indonesian fifteen-year-olds scored an average of 359 points in reading literacy, representing a twelve-point decline from 2018 and placing the country 117 points below the OECD average of 476 points (OECD, 2023). This performance positioned Indonesia at rank seventy-one of eighty-one participating countries, with only twenty-five percent of Indonesian students reaching the minimum proficiency level compared to seventy-four percent across OECD member states. Domestically, the 2023 National Assessment (Asesmen Nasional) reported by the Ministry of Education, Culture, Research, and Technology of Indonesia (Moecrt, 2023) found that 51.03 percent of elementary school students performed below the minimum competency threshold in reading literacy. Regional disparities further exacerbate the problem: provinces in eastern Indonesia consistently score fifteen to twenty-five percent lower than provinces on Java, indicating systemic rather than isolated challenges (Setiawan et al., 2020; Subandi et al., 2023).

Multiple interrelated factors contribute to the underperformance of Indonesian elementary students in reading literacy. First, the heterogeneity of student abilities within a single classroom typically comprising thirty to forty learners with widely varying reading proficiencies poses substantial challenges for teachers attempting to provide individualized attention (Pratiwi et al., 2023; Sumarsono et al., 2023). Second, the conventional one-size-fits-all approach to reading instruction fails to accommodate diverse learning styles, prior knowledge, and reading interests of elementary students (Wijayanti & Sari, 2022). Third, formative feedback that is timely, specific, and developmentally appropriate remains rare due to high teacher workload and limited capacity for one-on-one interaction (Nurhasanah & Sukmawati, 2022). Fourth, the availability of engaging, contextually relevant reading materials at varied difficulty levels is insufficient, particularly in rural and underserved schools (Sari & Hidayat, 2023). Fifth, teachers themselves often lack adequate training in evidence-based reading instruction methods. Collectively, these factors point to a fundamental need for scalable, adaptive pedagogical tools that complement teacher expertise and provide each student with personalized reading support aligned with their developmental level.

Recent advances in artificial intelligence, particularly the emergence of Large Language Models (LLMs) such as GPT-4, Claude, and Gemini, have opened unprecedented opportunities to address longstanding challenges in personalized education at scale (Kasneci et al., 2023). Empirical evidence indicates that LLM-based interventions can yield learning gains of 18–45% while reducing learning anxiety. LLMs leverage transformer architectures trained on massive text corpora, enabling them to generate contextually appropriate responses, formulate questions of varying cognitive demand, provide explanations, and sustain extended conversational interactions (Crompton & Burke, 2023). In educational applications, LLMs have demonstrated promising potential as Socratic tutors (Lang et al., 2026), adaptive feedback providers (Hsu & Tsai, 2026), automated item generators (Song et al., 2026), and conversational agents for language learning (Lee et al., 2026). Empirical evidence from recent controlled studies indicates that LLM-based learning interventions yield substantial learning gains ranging from eighteen to forty-five percent while simultaneously reducing learning anxiety and enhancing student engagement (Lang et al., 2026; Tlili et al., 2023; Wu, 2026). Moreover, the conversational nature of LLM interactions aligns naturally with established pedagogical theories such as Vygotsky's Zone of Proximal Development and Collins' cognitive apprenticeship framework, offering scaffolded, dialogic learning experiences.

Despite the rapid expansion of LLM-related research in education, several critical gaps persist that limit the field's relevance for Indonesian elementary contexts. First, the overwhelming majority of empirical studies on LLM applications in education have focused on higher education and secondary contexts, with elementary school applications particularly for reading literacy remaining significantly under-researched (Cai et al., 2026; Hao & Wang, 2026). Second, most existing implementations deploy general-purpose LLMs without pedagogical adaptations specifically tailored to the cognitive and linguistic developmental stages of elementary learners, raising concerns about response appropriateness, complexity calibration, and ethical safeguards for young users (Wang, T. et al., 2026; Yan et al., 2023). Third, research on LLM applications within Indonesian language contexts and the cultural and infrastructural realities of Indonesian schools remains virtually nonexistent in international literature. Fourth, mixed-methods studies that combine rigorous quantitative effectiveness measurements with qualitative explorations of student experience are rare, despite their methodological strength. These gaps collectively justify the urgency of developing and empirically validating an LLM-based adaptive reading assistant prototype specifically designed for Indonesian elementary school contexts.

This research addresses the identified gaps by developing and evaluating BACA-AI (Bacaan Cerdas Adaptif berbasis Artificial Intelligence), an LLM-based adaptive reading assistant designed specifically for Indonesian elementary school students. The study pursues three primary objectives: first, to develop a prototype LLM-based adaptive reading

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

assistant that integrates pedagogical scaffolding principles tailored to elementary learners' developmental characteristics; second, to empirically measure the effectiveness of the prototype in improving reading literacy among fifth-grade students through a rigorous quasi-experimental design; and third, to explore students' lived learning experiences during interactions with the prototype through qualitative inquiry. The novelty of this research lies in four interconnected contributions: the pedagogically grounded design of the prototype with explicit system prompts tailored to elementary-level cognition and Indonesian language conventions; the application of an explanatory sequential mixed-methods design that triangulates quantitative outcomes with experiential insights; the integration of ethical AI literacy considerations within the implementation protocol; and the contextualization of the entire intervention within Indonesia's unique educational ecosystem, addressing infrastructure realities and teacher capacity considerations.

LITERATURE REVIEW

Reading Literacy at the Elementary Level

Reading literacy at the elementary level represents a foundational competency that scaffolds all subsequent academic learning. The Progress in International Reading Literacy Study (PIRLS) defines reading literacy as the ability to understand and use written language forms required by society and valued by individuals, encompassing the construction of meaning through interaction with texts, learning from reading, and engagement with communities of readers (Mullis et al., 2023). The PIRLS assessment framework identifies four core comprehension processes: locating explicit information, making direct inferences, interpreting and integrating ideas, and evaluating content and language elements. These processes form a developmental progression from literal to higher-order cognitive engagement with texts. Castles, Rastle, and Nation (2018) demonstrate in their influential synthesis that effective elementary reading instruction must integrate phonological decoding skills with vocabulary development, fluency practice, and comprehension strategies through systematic instruction. Duke, Ward, and Pearson (2021) emphasize that the science of reading comprehension instruction involves explicit teaching in word recognition, language comprehension, executive function, and reading strategies, all of which require differentiated approaches tailored to individual learners' developmental positions.

In the Indonesian context, factors influencing elementary reading literacy include socioeconomic status, parental involvement, classroom literacy environment, and instructional quality. Hidayatullah et al. (2022) found that Indonesian university students who received robust elementary literacy instruction continued to demonstrate stronger reading comprehension throughout their academic trajectories, suggesting cumulative effects of early reading interventions. Sumarsono et al. (2023) examined digital literacy and reading comprehension among Indonesian primary students and found a moderate positive correlation, indicating that technology-mediated reading experiences may complement traditional instruction. However, Setiawan et al. (2020) caution that contextual factors specific to Indonesia—including limited library resources, regional language diversity, and uneven teacher preparation—necessitate culturally and linguistically responsive pedagogical interventions rather than wholesale adoption of foreign instructional models. The integration of digital and adaptive tools must therefore be carefully calibrated to local conditions while preserving the developmental sensitivity required for young learners (Subandi et al., 2023). This contextual grounding becomes especially critical when introducing emerging technologies such as Large Language Models into elementary literacy education.

Large Language Models in Educational Applications

Large Language Models constitute a class of generative AI systems based on transformer architectures, trained on massive corpora of text to produce coherent, contextually appropriate language outputs in response to user prompts (Kasneci et al., 2023). In educational contexts, LLMs offer several distinctive affordances: they can engage in extended dialogue, generate varied question types across cognitive levels, provide tailored explanations and feedback, and adapt their responses based on learner inputs without explicit programming for each scenario (Crompton & Burke, 2023). Lo (2023) conducted a rapid review of ChatGPT applications in education and identified key benefits including personalized learning, instant feedback, accessibility, and engagement. However, the review also highlighted significant concerns regarding factual accuracy, intellectual property, ethical use, and the need for educator oversight. Kuhail et al. (2023) systematically reviewed educational chatbots and found that conversational AI systems consistently improved learning outcomes when integrated thoughtfully into pedagogical sequences, though effectiveness varied substantially based on design quality and contextual integration. Yan et al. (2023) further emphasized the need for human-centered design principles when deploying LLMs in education.

Recent empirical research has demonstrated the educational potential of LLMs across diverse domains and learner populations. Lang et al. (2026) reported that a GPT-4-powered Socratic tutoring system achieved an average learning gain of 18.43 percent among university students while reducing negative emotions during learning. Lee et al. (2026) found that AI-empowered Non-Player Characters in virtual reality environments significantly improved Korean

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

EFL learners' self-regulation strategies and reduced speaking anxiety. Lai (2026) demonstrated that ChatGPT-supported inquiry-based learning outperformed adaptive learning approaches for fifth-grade students in CLIL science education, suggesting that pedagogical framing matters more than the technological tool itself. Tlili et al. (2023) examined ChatGPT as a case study of chatbot integration and identified both opportunities and ethical risks, calling for systematic implementation guidelines. Yan et al. (2023) provided a scoping review of practical and ethical challenges of LLMs in education, emphasizing concerns about bias, privacy, transparency, and the need for human-centered design. Together, these studies establish a growing evidence base that supports cautious optimism about LLM educational applications while underscoring the importance of pedagogical and ethical rigor.

Adaptive Learning Systems and Pedagogical Foundations

Adaptive learning systems are defined as instructional environments that dynamically adjust content, pacing, difficulty level, and pedagogical strategies based on real-time learner data, enabling personalization at scale (Xie et al., 2019). Bhutoria (2022) conducted a systematic review of AI-enabled personalized education across the United States, China, and India, identifying common patterns including data-driven content recommendation, intelligent tutoring, and automated assessment, while also documenting persistent challenges around teacher integration and ethical safeguards. Wang, S. et al. (2023) compared adaptive learning systems with conventional homework delivery formats and found that adaptive systems were particularly effective when learners exhibited high engagement and when systems incorporated immediate, specific feedback. The theoretical foundation of adaptive learning draws extensively from Vygotsky's Zone of Proximal Development, which posits that learning occurs most effectively when tasks are calibrated slightly beyond a learner's independent capability but within reach with appropriate scaffolding (Belland et al., 2015; Wu, 2026). This sociocultural framing positions adaptive technologies not as replacements for human pedagogy but as instruments amplifying scaffolded learning at scale.

The pedagogical design of LLM-based adaptive systems benefits from frameworks rooted in cognitive apprenticeship theory, which emphasizes modeling, coaching, scaffolding, articulation, reflection, and exploration as core instructional moves (Wang, X. et al., 2026). In LLM-based reading assistants, these moves can be operationalized through explicit system prompts that guide the model toward developmentally appropriate questioning, gradual hint provision, and Socratic dialogue. Lin and Chang (2023) developed the CHAT-ACTS framework demonstrating that pedagogically grounded chatbot design substantially enhances self-regulated learning behaviors. Hwang et al. (2020) outlined a comprehensive vision for AI in education that emphasizes the integration of learning science principles with technological capabilities, arguing that the value of AI tools depends fundamentally on the pedagogical framework guiding their implementation. Chen et al. (2022) analyzed two decades of AI in education research and noted that adaptive personalization, intelligent tutoring, and learning analytics constitute the most mature application domains, suggesting that LLM-based adaptive reading assistants build upon a well-established research trajectory while extending it into the conversational and generative dimensions enabled by modern transformer architectures

METODE

Research Design

This study employed an explanatory sequential mixed-methods design, integrating quantitative and qualitative data collection in two sequential phases to provide a comprehensive understanding of the intervention's effectiveness and the underlying learner experiences (Schoonenboom & Johnson, 2017). The quantitative phase utilized a quasi-experimental nonequivalent control group design with pretest-posttest measurements to evaluate the effect of integrating the BACA-AI adaptive reading assistant on students' reading literacy outcomes. This design was selected because random assignment of individual students was not feasible without disrupting established classroom structures and school timetables, a common constraint in educational research conducted in operational school settings. Following quantitative data collection and analysis, a qualitative phase using semi-structured interviews and classroom observations was conducted to explore students' lived experiences and to interpret quantitative findings within their experiential context. The explanatory sequential design was particularly appropriate because it allowed quantitative findings to inform the focus of qualitative inquiry, ensuring depth and contextual richness in interpretation. Table 1 summarizes the overall research design and methodological components employed in this study.

Table 1. Summary of the Research Design

Research Component	Description
Research Approach	Mixed-method, explanatory sequential design
Quantitative Design	Quasi-experimental nonequivalent control group with pretest–posttest
Qualitative Strategy	Reflexive thematic analysis of interviews and classroom observations
Sampling Technique	Purposive sampling at school level; intact classes at student level
Sample Size	72 fifth-grade students (Experimental n=36; Control n=36)

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

Intervention Duration	8 weeks; 8 instructional sessions × 60 minutes for the experimental group
Data Sources	PIRLS-adapted reading literacy test; classroom observations; semi-structured interviews (n=12)
Statistical Analysis	Descriptive statistics, Shapiro–Wilk, Levene, ANCOVA with pretest as covariate, Cohen's d

(Source: Authors' compilation)

Population and Sampling

The target population of this study comprised fifth-grade students enrolled in public elementary schools (Sekolah Dasar Negeri) in Bekasi, West Java, during the 2024/2025 academic year. Fifth grade was selected as the target level because students at this stage have completed foundational decoding instruction and are developmentally positioned to engage with higher-order reading comprehension tasks aligned with PIRLS frameworks (Mullis et al., 2023). A purposive sampling technique was employed to select two elementary schools meeting predetermined criteria: A-level institutional accreditation, stable internet connectivity, availability of at least one tablet or computer per pair of students, and willingness of school administration and classroom teachers to participate in the research protocol. The final sample comprised seventy-two students from two intact classrooms, with thirty-six students from SDN A constituting the experimental group and thirty-six students from SDN B forming the control group. Both schools shared comparable socioeconomic profiles, teacher qualifications, and prior academic performance distributions, as detailed in Table 2.

Table 2. Sample Characteristics of the Experimental and Control Groups

Characteristic	Experimental (n=36)	Control (n=36)
Mean age in years (SD)	10.4 (0.55)	10.5 (0.58)
Female / Male composition	19 / 17	18 / 18
School accreditation grade	A	A
Prior Bahasa Indonesia GPA (semester)	78.4 (5.81)	78.1 (5.92)
Internet access at school	100% (stable Wi-Fi)	100% (stable Wi-Fi)
Device available per pair	Yes (1 tablet/2 students)	Yes (1 tablet/2 students)
Prior LLM/Chatbot exposure	None (baseline)	None (baseline)

(Source: Pre-intervention demographic survey)

Research Instruments

Three primary instruments were employed for data collection in this study. First, a Reading Literacy Test was constructed by adapting items from the PIRLS framework, comprising twenty-five multiple-choice items and five short-answer items distributed across the four PIRLS comprehension processes. Content validity was established through expert review involving two university lecturers specializing in Indonesian language education and one experienced elementary teacher, using the Aiken's V coefficient with a minimum acceptable threshold of 0.80; all retained items achieved coefficients ranging from 0.83 to 0.95. Internal consistency reliability was assessed using Cronbach's alpha on pilot data from sixty students outside the main sample, yielding a coefficient of 0.87, indicating acceptable reliability. Second, a Classroom Observation Sheet was developed to document student engagement, interaction patterns, and pedagogical events during intervention sessions, with inter-rater agreement reaching $\kappa = 0.84$. Third, a Semi-Structured Interview Protocol was prepared to elicit students' experiences, perceptions, and challenges encountered during interactions with BACA-AI. Both qualitative instruments were validated by two experts in qualitative educational research. Table 3 summarizes the instruments and their psychometric properties.

Table 3. Research Instruments and Their Psychometric Properties

Instrument	Purpose	Format / Items	Validity / Reliability
Reading Literacy Test	Measure four PIRLS comprehension processes	25 MC + 5 short-answer items (across 3 reading passages)	Aiken's V = 0.83–0.95; Cronbach's α = 0.87
Classroom Observation Sheet	Document engagement, interaction patterns, pedagogical events	Structured rubric with 12 indicators	Inter-rater agreement κ = 0.84
Semi-Structured Interview Protocol	Explore learners' experiences, perceptions, challenges	9 core questions + flexible probes	Validated by 2 qualitative-research experts

(Source: Instrument validation procedures)

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

Procedures

Data collection extended over eight consecutive weeks during the second semester of the 2024/2025 academic year. During the first week, both groups completed the reading literacy pretest under standardized conditions to establish baseline performance. From week two through week seven, the experimental group participated in eight intervention sessions of sixty minutes each, conducted twice weekly. Each session followed a consistent three-phase structure: a ten-minute teacher modeling phase introducing the day's reading text and demonstrating effective interaction with BACA-AI, a thirty-five-minute autonomous student–BACA-AI interaction phase during which students engaged with the adaptive reading assistant individually or in pairs, and a fifteen-minute teacher-guided reflection phase during which students shared insights, addressed conceptual difficulties, and consolidated learning. Throughout the same period, the control group received conventional reading instruction with equivalent instructional time but without LLM integration. Classroom observations were conducted during weeks three, five, and seven by the first author using the standardized observation protocol. During week eight, both groups completed the posttest, followed by semi-structured interviews with twelve purposively selected students from the experimental group (four high-, four medium-, and four low-proficiency learners).

Data Analysis

Quantitative data analysis proceeded through three sequential stages. First, descriptive statistics including means, standard deviations, and score ranges were computed for pretest and posttest measurements in both groups. Second, assumption testing was conducted using the Shapiro–Wilk test for normality of distribution and the Levene test for homogeneity of variance, along with verification of the homogeneity of regression slopes assumption underlying ANCOVA. Third, hypothesis testing employed Analysis of Covariance (ANCOVA) with pretest scores serving as the covariate to control for baseline differences in reading literacy capability. Effect size was calculated using Cohen's d, with conventional benchmarks of 0.2, 0.5, and 0.8 representing small, medium, and large effects respectively. All quantitative analyses were performed using IBM SPSS Statistics version 28 at a significance threshold of $\alpha = 0.05$. Qualitative data analysis followed the reflexive thematic analysis procedure outlined by Braun and Clarke (2021), proceeding through six recursive phases: familiarization with the data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the final report. Investigator triangulation among the research team enhanced the credibility of the qualitative findings.

Validity, Reliability, and Limitations

Multiple measures were implemented to ensure the validity and reliability of findings. Internal validity was strengthened through pretest matching, standardized instructional procedures, identical instructional time across conditions, and the use of trained classroom teachers familiar with both groups' baseline characteristics. External validity considerations led to selecting schools representative of urban Indonesian elementary contexts, though generalization to rural or under-resourced settings warrants caution. Qualitative trustworthiness was supported through prolonged engagement during intervention, methodological triangulation across interviews and observations, investigator triangulation, and the maintenance of an audit trail. Several limitations should be acknowledged. The relatively modest sample size of seventy-two students restricts statistical power for finer-grained subgroup analyses. The eight-week intervention period, while adequate for detecting initial effects, may underestimate longer-term outcomes or potential novelty effects. The single geographic location (Bekasi) limits transferability to differing infrastructural and cultural contexts. Finally, the proprietary nature of the underlying LLM (GPT-4o) constitutes a black-box element that prevents detailed inspection of decision processes. These limitations are addressed through transparent reporting and informed recommendations for future research extensions.

RESULTS AND DISCUSSION

Results

The BACA-AI prototype was developed iteratively over four months, integrating principles from cognitive apprenticeship, sociocultural learning theory, and Indonesian elementary curriculum standards. The system architecture, depicted in Figure 1, comprises five integrated modules organized around a central LLM engine accessed through the OpenAI GPT-4o Application Programming Interface. The Adaptive Question Generation module produces three categories of comprehension questions—literal, inferential, and evaluative calibrated to the PIRLS framework. The Tiered Scaffolding module delivers progressive hints when students struggle, beginning with general orientation cues and advancing toward more specific guidance without providing direct answers. The Socratic Dialogue Management module sustains extended dialogic exchanges, prompting students to elaborate, justify, and reflect on their responses. The Formative Feedback Loop module aggregates response patterns to inform subsequent questioning and adjust difficulty levels. Critically, all five modules operate within the boundaries of a carefully engineered

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

pedagogical system prompt designed for fifth-grade Indonesian learners, ensuring appropriate language complexity, cultural relevance, and ethical safeguards regarding sensitive content. Initial acceptance testing with eight students outside the main sample yielded an 87.5 percent acceptance rate for usability and a 90.6 percent acceptance rate for language appropriateness.

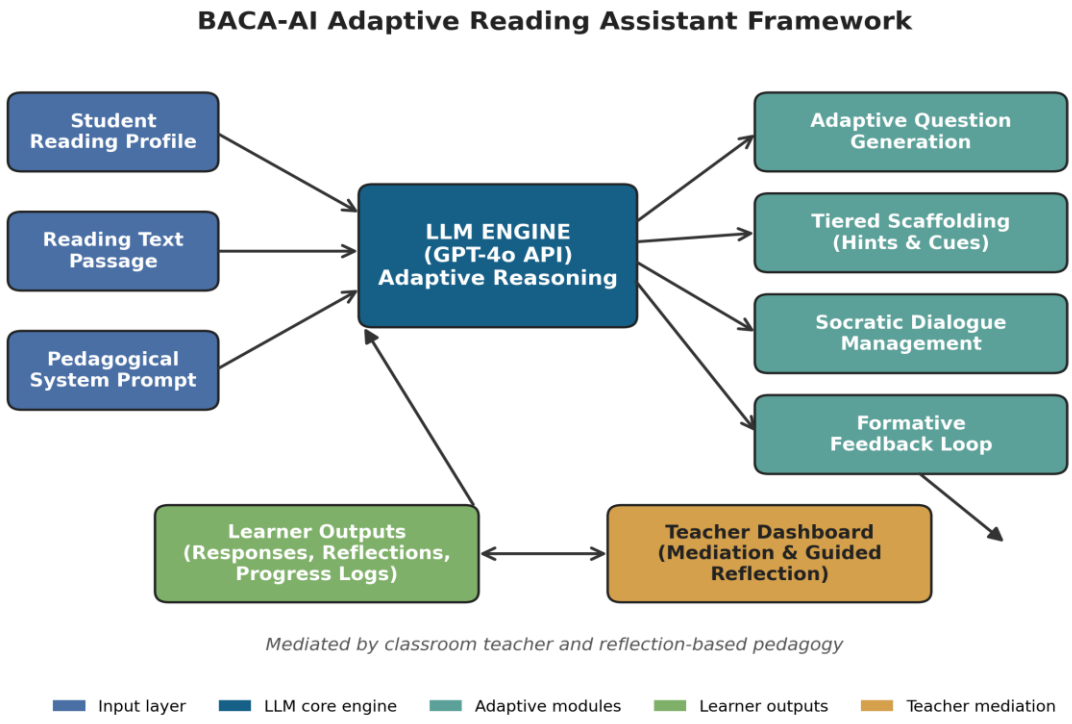

Figure 1. Conceptual Framework of the BACA-AI Adaptive Reading Assistant

(Source: Authors' system architecture design)

Both experimental and control groups exhibited comparable demographic profiles, supporting the feasibility of meaningful between-group comparisons. The experimental group consisted of thirty-six students with a mean age of 10.4 years (SD = 0.55), comprising nineteen female and seventeen male students. The control group comprised thirty-six students with a mean age of 10.5 years (SD = 0.58), with eighteen female and eighteen male students. Pretest scores in reading literacy averaged 52.67 (SD = 8.41) for the experimental group and 51.89 (SD = 7.93) for the control group, indicating equivalent baseline capabilities. Following the eight-week intervention, the experimental group achieved a mean posttest score of 76.12 (SD = 7.85), representing a mean gain of 23.45 points, whereas the control group reached a mean posttest score of 61.71 (SD = 8.12), reflecting a mean gain of 9.82 points. Table 4 presents the complete descriptive statistics for both groups across pretest and posttest measurements. The differential gain trajectory is further visualized in Figure 3, which depicts weekly progress estimates derived from formative assessment data collected during the intervention period.

Table 4. Descriptive Statistics of Reading Literacy Scores for Both Groups

Group	N	Pretest M (SD)	Posttest M (SD)	Mean Gain
Experimental	36	52.67 (8.41)	76.12 (7.85)	23.45
Control	36	51.89 (7.93)	61.71 (8.12)	9.82

(Source: Primary data analysis)

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

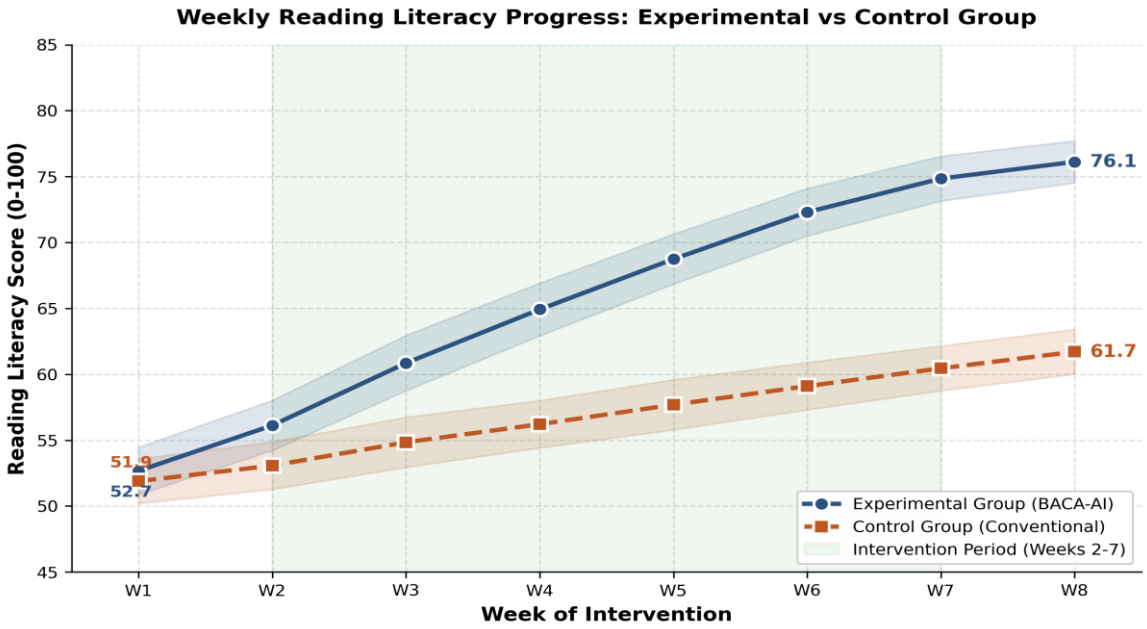

Figure 3. Weekly Reading Literacy Progress of Both Groups During the Intervention

(Source: Weekly formative assessment data)

Prior to hypothesis testing, assumptions underlying parametric analysis were rigorously evaluated. The Shapiro-Wilk test indicated that all pretest and posttest distributions in both groups did not significantly deviate from normality, with p-values ranging from 0.142 to 0.387, all exceeding the 0.05 threshold. The Levene test demonstrated homogeneity of variance between groups for both pretest ($F = 1.412$, $p = 0.239$) and posttest scores ($F = 1.523$, $p = 0.217$). The homogeneity of regression slopes assumption underlying ANCOVA was also verified, with a non-significant interaction between group membership and pretest scores ($F = 0.973$, $p = 0.328$). With assumptions satisfied, ANCOVA was conducted with posttest scores as the dependent variable, group membership as the independent variable, and pretest scores as the covariate. Results indicated a statistically significant main effect of group membership on reading literacy outcomes after controlling for pretest performance, $F(1, 69) = 50.21$, $p < 0.001$. The corresponding effect size, calculated as Cohen's d, was 1.42, exceeding the conventional threshold for large effects. Table 5 presents the full ANCOVA output.

Table 5. Results of Analysis of Covariance (ANCOVA) for Posttest Reading Literacy

Source of Variance	Sum of Squares	df	Mean Square	F	p
Pretest (covariate)	1,842.31	1	1,842.31	28.76	<0.001
Group (Treatment)	3,215.84	1	3,215.84	50.21	<0.001
Error	4,418.57	69	64.04	—	—
Total	9,476.72	71	—	—	—

(Source: Primary data analysis using SPSS v26)

To examine differential effects across the four PIRLS comprehension processes, mean scores were computed separately for each indicator at pretest and posttest for both groups. Results presented in Table 6 and visualized in Figure 2 reveal that the experimental group exhibited substantially larger gains across all four indicators compared to the control group. The most pronounced gains in the experimental group occurred for higher-order comprehension processes: interpreting and integrating ideas (gain of 26.36 points compared to 10.55 in the control group) and evaluating content and language (gain of 24.36 points compared to 12.30 in the control group). Even for lower-order indicators, the experimental group outperformed the control group: locating explicit information showed an 18.93-point gain for the experimental group versus 7.17 for control, and making direct inferences showed a 24.16-point gain versus 9.26. Independent samples t-tests on posttest gains for each indicator confirmed statistically significant differences with effect sizes ranging from 1.18 to 1.58, all categorized as large effects. The disproportionate gains in higher-order processes suggest that the LLM-based intervention particularly supports the development of sophisticated comprehension skills.

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

Table 6. Mean Scores per PIRLS Indicator for Both Groups

PIRLS Indicator	Exp. Pretest	Exp. Posttest	Ctrl. Pretest	Ctrl. Posttest	Cohen's d
Locating explicit information	64.28	83.21	63.85	71.02	1.18
Making direct inferences	54.17	78.33	53.89	63.15	1.41
Interpreting and integrating ideas	48.33	74.69	47.92	58.47	1.58
Evaluating content and language	43.89	68.25	41.90	54.20	1.36

(Source: Primary data analysis)

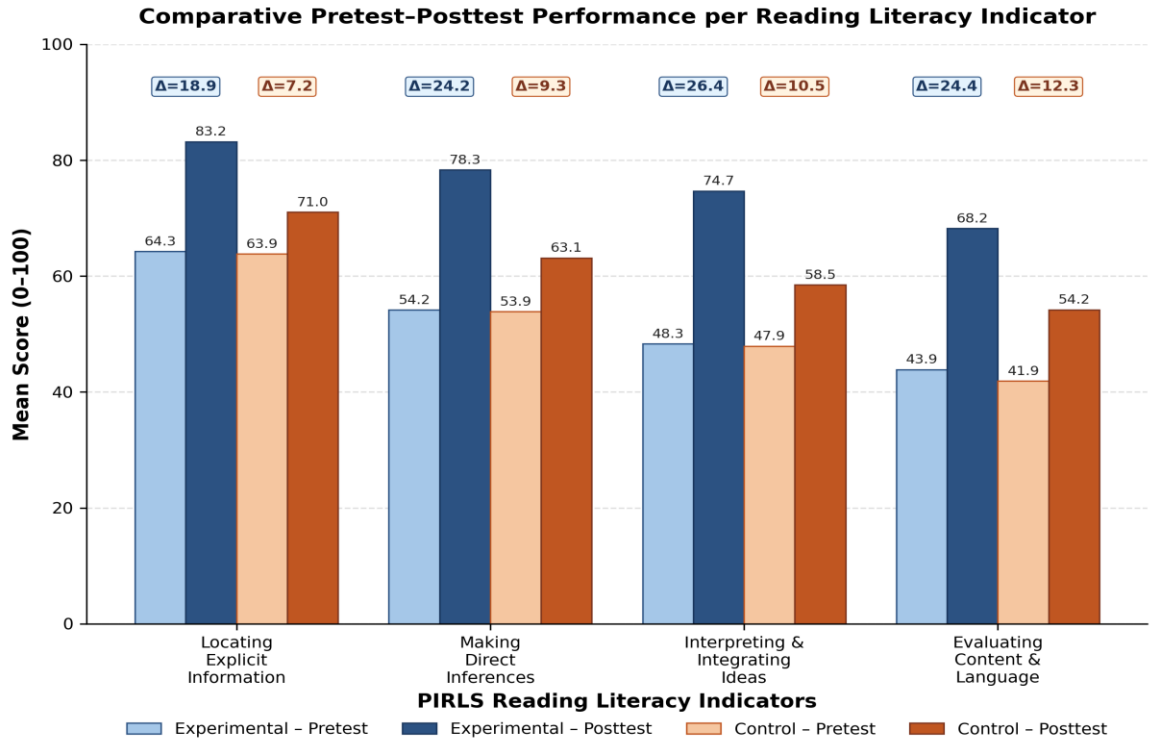

Figure 2. Comparative Pretest-Posttest Performance per Reading Literacy Indicator

(Source: Primary data analysis)

Thematic analysis of transcripts from twelve semi-structured interviews and field notes from classroom observations yielded five overarching themes characterizing students' experiences with BACA-AI. Table 7 summarizes the themes, their definitions, and the frequencies with which they emerged in the interview corpus. The first theme, increased reading engagement, was articulated by all twelve interview participants, who described feeling motivated to read because of the conversational and responsive nature of the assistant. One lower-proficiency participant remarked that interacting with BACA-AI felt unintimidating because the assistant patiently listened without judging mistakes. The second theme, comprehension confidence, emerged for eleven participants who reported that interacting with an AI without social judgment encouraged them to attempt answers they would normally avoid. The third theme, self-regulated learning, was described by ten participants who articulated agency over their pacing, repetition of difficult material, and selection of scaffolding levels. The fourth theme, technical challenges, addressed connectivity issues and lengthy responses. The fifth theme, need for teacher mediation, confirmed that students continued to value teacher presence for conceptual clarification.

Table 7. Themes Emerging from Qualitative Analysis of Student Experiences

Theme	Description	Frequency
Increased reading engagement	Students felt motivated to read because of responsive dialogue and immediate interactivity with the assistant.	12 / 12
Comprehension confidence	Students became more willing to attempt answers and ask questions because they were not afraid of social judgment.	11 / 12
Self-regulated learning	Students managed their pacing, repeated difficult materials, and chose hints proportional to their needs.	10 / 12

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

Technical challenges	Occasional unstable internet connectivity and lengthy LLM responses became operational obstacles for some learners.	8 / 12
Need for teacher mediation	Students valued teacher presence for clarifying difficult concepts and guiding reflection.	9 / 12

(Source: Reflexive thematic analysis of interview and observation data)

Discussion

The quantitative findings of this study provide robust empirical support for the effectiveness of integrating LLMs as adaptive reading assistants in elementary education. The observed mean gain of 23.45 points in the experimental group, compared to 9.82 in the control group, with a large effect size of Cohen's d = 1.42, exceeds typical effects reported in educational technology meta-analyses (Bhutoria, 2022). This magnitude is consistent with, and even surpasses, recent findings from comparable LLM interventions in different educational contexts. Lang et al. (2026) reported an 18.43 percent learning gain among university students using a GPT-4-powered Socratic tutor; the present study found approximately a 44.5 percent relative gain in fifth-grade students. Two factors may explain this stronger effect. First, the pedagogical system prompt was specifically engineered for elementary cognition and Indonesian linguistic conventions, ensuring appropriate complexity and cultural relevance. Second, the lower starting performance ceiling among Indonesian elementary students creates more room for growth compared to higher-achieving university populations, consistent with regression-to-mean and ceiling-effect considerations in educational measurement (Wang, S. et al., 2023).

A particularly noteworthy finding is the disproportionate improvement in higher-order comprehension processes interpreting/integrating and evaluating relative to lower-order processes. This pattern aligns with the design intent of BACA-AI, which deliberately scaffolds Socratic dialogue toward inferential and evaluative thinking through carefully crafted question sequences. This finding resonates with Wu (2026), who demonstrated that multi-agent systems incorporating argumentation and devil's advocacy interactions significantly enhanced critical thinking among university students. Similarly, Lai (2026) found that inquiry-based learning with ChatGPT outperformed adaptive learning approaches in fostering bilingual learners' cognitive engagement and scientific reasoning. The pedagogical mechanism appears to involve gradual cognitive provocation: when students encounter open-ended questions and receive scaffolded prompts rather than direct answers, they engage in active sense-making that strengthens higher-order processes. This phenomenon resonates with Vygotsky's Zone of Proximal Development, where scaffolded learning enables learners to attain understanding slightly beyond their independent capability (Belland et al., 2015; Wang, X. et al., 2026). The implications are substantial for Indonesian literacy education, where higher-order comprehension has historically lagged behind decoding and literal comprehension.

The qualitative finding of increased comprehension confidence aligns with documented affective benefits of AI-mediated learning. Lee et al. (2026) reported that AI-empowered virtual reality language learning reduced speaking anxiety among Korean EFL learners, while Lang et al. (2026) found that GPT-4 Socratic tutoring decreased negative emotional states. The mechanism appears related to the social-evaluative threat hypothesis: students who fear judgment from peers or teachers may withhold attempts at answering, but interaction with a non-judgmental AI provides psychological safety for cognitive experimentation. This dynamic is particularly relevant for elementary students in Indonesian classrooms where socio-cultural norms regarding face-saving and avoidance of embarrassment can suppress active participation. Krashen's affective filter hypothesis offers a complementary theoretical lens, suggesting that learning optimal performance occurs when affective barriers such as anxiety and low self-esteem are minimized. The emergence of self-regulated learning as a dominant qualitative theme further extends prior findings by Lee et al. (2026) into the elementary domain, suggesting that LLM interventions can foster metacognitive autonomy at developmental stages typically underestimated in adaptive learning research.

The fifth qualitative theme, need for teacher mediation, carries critical implications for implementation. Students consistently valued teacher presence for conceptual clarification and reflective discussion even while appreciating the AI assistant. This finding aligns with Gu et al. (2026), who emphasize human-centered design as foundational for multi-agent educational systems. It also resonates with the conceptual framework of Hsu and Tsai (2026), who employed user-centered design to develop generative AI chatbots that augment rather than replace human teaching presence. The implication for practice is clear: LLM-based adaptive assistants should be designed and deployed as instructional supplements, not replacements. Teachers retain irreplaceable roles in providing socio-emotional support, contextualizing learning within broader curricular frameworks, addressing affective concerns, and modeling reading as a culturally embedded practice. Furthermore, teacher capacity for thoughtful integration including knowing when to defer to AI and when to intervene directly requires deliberate professional development. This implies that scaling LLM-based literacy interventions in Indonesia must include systematic AI literacy training for teachers, addressing both technical operation and pedagogical integration (Crompton & Burke, 2023).

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

The qualitative theme of technical challenges, including connectivity issues and lengthy responses for lower-proficiency students, signals important equity considerations. Although the study schools were selected for adequate infrastructure, the broader Indonesian context features substantial digital divides between urban and rural areas, and between higher and lower socioeconomic communities (Sari & Hidayat, 2023; Subandi et al., 2023). Hao and Wang (2026) provide a poignant illustration of how generative AI can paradoxically widen rather than narrow educational inequities when adoption assumes universal infrastructure access. The present findings, while encouraging, must therefore be interpreted within the constraint of selected school contexts. Scaling BACA-AI nationally would require companion investments in connectivity, devices, and technical support for under-resourced schools. Additionally, the response length feedback from students suggests that LLM outputs require ongoing calibration for elementary readers; this could be addressed through prompt refinement, response truncation thresholds, or interface design choices that preserve coherence while reducing cognitive load. The principle of equity-aware multi-agent collaboration articulated by Hadyaoui and Cheniti-Belcadhi (2026) provides a useful framework for addressing such considerations.

Two findings from this study warrant theoretical reflection. First, lower-proficiency students in the experimental group exhibited gains comparable to or exceeding those of higher-proficiency peers, suggesting that the adaptive mechanism effectively narrowed within-class achievement gaps. This pattern contradicts the Matthew effect commonly observed in conventional instruction, where high-performing students benefit disproportionately from instructional opportunities (Duke et al., 2021). The mechanism likely involves individualization: lower-proficiency students received scaffolding calibrated to their actual zone of proximal development rather than instruction pitched to class-average competence. Second, the weekly progress data revealed steeper gains during the early intervention weeks, decelerating somewhat in later weeks. This pattern raises concerns about novelty effects: the initial engagement boost may partially reflect curiosity about new technology rather than sustained pedagogical impact. Longer-term studies are needed to disentangle novelty from durable learning effects, a concern echoed by Lo (2023) and Tlili et al. (2023) regarding ChatGPT educational integration. Methodologically, future research should incorporate delayed posttest assessments and longitudinal designs to address this concern.

Several limitations of this study must be acknowledged. The modest sample size of seventy-two students limits statistical power for finer-grained subgroup analyses. The single geographic location restricts inferences to similarly resourced urban Indonesian elementary contexts. The eight-week intervention period may underestimate longer-term effects and overestimate effects influenced by novelty. The black-box nature of the underlying GPT-4o model prevents complete transparency regarding response generation mechanisms, raising both methodological and ethical considerations consistent with concerns highlighted by Yan et al. (2023). Future research should pursue several extensions. First, multi-site studies across diverse Indonesian regions with varying infrastructural conditions would substantially enhance external validity. Second, longitudinal designs with delayed posttests and follow-up measurements at six and twelve months would establish durability of effects. Third, development of LLMs fine-tuned on Indonesian educational corpora could enhance contextual responsiveness, cultural alignment, and reduce dependency on foreign-language-trained models. Fourth, deeper investigation of ethical dimensions including data privacy for minors, bias in AI-generated content, and equitable access policies is essential for responsible scale-up. These directions collectively chart a research agenda for the next decade of LLM-based literacy education

CONCLUSION

This study set out to develop, evaluate, and characterize the experience of using a Large Language Model-based adaptive reading assistant designed specifically for Indonesian elementary school students. The findings yield three principal conclusions that directly address the study's objectives. First, the BACA-AI prototype, grounded in cognitive apprenticeship theory and operationalized through carefully engineered pedagogical system prompts, successfully delivered adaptive questioning, tiered scaffolding, and Socratic dialogue calibrated to the developmental level of fifth-grade Indonesian learners. Initial acceptance testing demonstrated high usability and appropriate language complexity, confirming the technical and pedagogical feasibility of the design. Second, the quasi-experimental evaluation provided robust evidence for the effectiveness of the intervention, with the experimental group exhibiting statistically significant and practically large gains in reading literacy compared to the control group, with the most pronounced improvements occurring in higher-order comprehension processes. Third, qualitative analysis revealed that students experienced increased engagement, comprehension confidence, and self-regulated learning, while simultaneously identifying technical and pedagogical conditions including stable infrastructure and active teacher mediation necessary for effective implementation.

The implications of these findings extend across three educational levels. At the classroom level, teachers can integrate LLM-based adaptive reading assistants as components of station-based instruction, with explicit teacher modeling and guided reflection bookending student-AI interaction periods. The teacher remains central as a culturally embedded mediator, affective support, and conceptual validator, while the AI provides individualized practice

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

opportunities that would be logistically infeasible through purely human instruction. At the school level, administrators must ensure infrastructural readiness including stable internet, adequate devices, and clear policies governing ethical AI use with minors. Investment in teacher AI literacy through structured professional development is equally essential. At the policy level, this research informs national initiatives such as Indonesia's literacy acceleration program, suggesting that scaled LLM-based literacy support is technically feasible but requires complementary investments in equity-oriented infrastructure, pedagogical guidance, ethical frameworks, and educator capacity. The research thus contributes evidence-based grounding for emerging policy discussions about generative AI in Indonesian basic education.

The integration of Large Language Models as adaptive reading assistants represents a significant educational innovation, but its potential will be realized only when implementation is grounded in sound pedagogical principles, contextual sensitivity, and unwavering ethical commitments. This study contributes to the emerging evidence base by demonstrating that LLM-based adaptive reading interventions can substantially enhance the reading literacy of Indonesian elementary students when designed with pedagogical care and embedded within a teacher-mediated instructional ecosystem. However, the research community is only beginning to understand the long-term consequences, equity dimensions, and developmental implications of widespread AI integration in early education. Future research should extend the present findings through larger samples, diverse contexts, longitudinal observations, and deliberate exploration of ethical considerations. Indonesian elementary education stands at a consequential juncture where technological capability has finally caught up with the longstanding ambition of personalized literacy support at scale. The choices made in the coming years about infrastructure investment, teacher preparation, curriculum integration, and ethical safeguards will determine whether this potential translates into more equitable, effective, and developmentally responsive reading literacy outcomes for the nation's children.

Acknowledgments

The authors extend their highest appreciation to the principals, fifth-grade homeroom teachers, and all participating students at the two State Elementary Schools in Bekasi who generously dedicated their time and engagement to this research. Special recognition is given to the Research and Community Service Institute (LPPM) of Universitas Bani Saleh for funding support through theInternal Research Grant Scheme. Sincere gratitude is also extended to Universitas Muhammadiyah Prof. Dr. Hamka and The University of Electro-Communications, Tokyo, for facilitating cross-institutional research collaboration, as well as to the expert validators who provided constructive feedback on the instruments and the BACA-AI prototype. The authors also thank the parents and guardians who consented to their children's participation and who supported sustained engagement throughout the eight-week intervention. Finally, the authors acknowledge colleagues, family members, and the broader academic community whose moral support, scholarly conversations, and ethical reminders shaped the integrity of this research from conception through dissemination

REFERENCES

Adiansha, A. A., & Sani, K. (2021). Pengaruh model Brain Based Learning dan Problem Based Learning terhadap keterampilan berpikir kompleks matematis ditinjau dari kreativitas siswa sekolah dasar di Kabupaten Bima. Jurnal Pendidikan MIPA, 21(1), 36–44. https://doi.org/10.37630/jpm.v11i1.454

Belland, B. R., Walker, A. E., Olsen, M. W., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. Journal of Educational Technology & Society, 18(1), 183–197.

Bhutoria, A. (2022). Personalized education and Artificial Intelligence in the United States, China, and India: A systematic review using a Human-In-The-Loop model. Computers and Education: Artificial Intelligence, 3, 100068. https://doi.org/10.1016/j.caeai.2022.100068

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology, 18(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Cai, H., Han, B., Sun, J., & Cheng, Y. (2026). How to design LLM-based intelligent agents to support students' learning in educational contexts: A systematic literature review. Educational Technology & Society, 29(2), 249–269. https://doi.org/10.30191/ETS.202604_29(2).SP03

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. Psychological Science in the Public Interest, 19(1), 5–51. https://doi.org/10.1177/1529100618772271

Chen, S. C. (2026). Investigating the effects of computer-based scaffolds on analogical reasoning: An eye-tracking analysis. Educational Technology & Society, 29(2), 154–170. https://doi.org/10.30191/ETS.202604_29(2).RP09

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. Educational Technology & Society, 25(1), 28–47.

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology, 32(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The science of reading comprehension instruction. The Reading Teacher, 74(6), 663–672. https://doi.org/10.1002/trtr.1993

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Gu, X., Hu, X., & Gašević, D. (2026). Guest editorial: Multiple intelligent agents in education—Theories, design, and applications. Educational Technology & Society, 29(2), 230–232. https://doi.org/10.30191/ETS.202604_29(2).SP01

Hadyaoui, A., & Cheniti-Belcadhi, L. (2026). Equity-aware multi-agent collaboration: Real-time role optimization and generative interventions in group learning. Educational Technology & Society, 29(2), 323–338. https://doi.org/10.30191/ETS.202604_29(2).SP06

Hao, T., & Wang, Y. (2026). "We are always the last to get a bit of it": Generative AI insights from Mississippi undergraduates. Educational Technology & Society, 29(2), 136–153. https://doi.org/10.30191/ETS.202604_29(2).RP08

Hidayatullah, T., Nyoto, A., & Lestari, P. (2022). Improving Indonesian elementary students' reading literacy through Gerakan Literasi Sekolah (GLS) program: A systematic literature review. International Journal of Elementary Education, 6(2), 211–219. https://doi.org/10.23887/ijee.v6i2.43825

Hsu, J.-S., & Tsai, W.-L. (2026). Implementing a generative AI chatbot using OpenAI API to improve learner interaction in asynchronous distance learning: A user-centered approach. Educational Technology & Society, 29(2), 375–403. https://doi.org/10.30191/ETS.202604_29(2).SP09

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. Computers and Education: Artificial Intelligence, 1, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. Education and Information Technologies, 28(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Lai, C.-J. (2026). Comparing generative AI-supported inquiry-based and adaptive learning for enhancing English vocabulary, content understanding, and literacy in Taiwan's CLIL science education. Educational Technology & Society, 29(2), 206–229. https://doi.org/10.30191/ETS.202604_29(2).RP12

Lang, Y., Hu, X., Gong, S., & Wei, H. (2026). Unlocking the potential of Socratic tutoring powered by large language model: A pilot study. Educational Technology & Society, 29(2), 69–83. https://doi.org/10.30191/ETS.202604_29(2).RP04

Lee, S.-M., Ahn, T. Y., & Wu, J. G. (2026). Self-regulated speaking with AI-empowered NPCs in VR. Educational Technology & Society, 29(2), 84–99. https://doi.org/10.30191/ETS.202604_29(2).RP05

Liao, J., Li, H., Zuo, Y., Zhong, L., Liu, M., Xie, T., & Liu, G. (2026). Uncanny valley effect: Research on the impact of digital human appearance on learners in a multi-agent collaborative learning environment. Educational Technology & Society, 29(2), 422–441. https://doi.org/10.30191/ETS.202604_29(2).SP11

Lin, M. P. C., & Chang, D. (2023). CHAT-ACTS: A pedagogical framework for personalized chatbot to enhance active learning and self-regulated learning. Computers and Education: Artificial Intelligence, 5, 100167. https://doi.org/10.1016/j.caeai.2023.100167

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410

Lo, H.-S., Hu, C.-K., Huang, T.-C., Hong, J.-C., & Wu, T.-F. (2026). Immersive augmented reality based on scaffolding in pre-vocational skills training for students with intellectual disabilities. Educational Technology & Society, 29(2), 191–205. https://doi.org/10.30191/ETS.202604_29(2).RP11

Lyu, B., Li, C., Li, H., & Xing, W. (2026). Exploring students' participation in online mathematical discussions using educational big data: A communicative ecology perspective. Educational Technology & Society, 29(2), 171–190. https://doi.org/10.30191/ETS.202604_29(2).RP10

Ministry of Education, Culture, Research, and Technology of Indonesia [MOECRT]. (2023). National Assessment Report 2022: Profile of Indonesian Students' Literacy and Numeracy. Jakarta: Center for Educational Assessment.

Mullis, I. V. S., Martin, M. O., Foy, P., Kennedy, A. M., Fishbein, B., & Yin, L. (2023). PIRLS 2021 International Results in Reading. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Nasir, N. B., Devi, S., & Kim, S. B. (2026). Metaverse-integrated programming education model with student–student and student–AI pairing. Educational Technology & Society, 29(2), 1–30. https://doi.org/10.30191/ETS.202604_29(2).RP01

Niu, J., He, X., Wang, X., Han, G., & He, J. (2026). Rubrics-enhanced multi-agent system based on LLM for teaching behavior assessment: Design and application. Educational Technology & Society, 29(2), 270–296. https://doi.org/10.30191/ETS.202604_29(2).SP04

Nurhasanah, S., & Sukmawati, F. (2022). Faktor-faktor yang mempengaruhi rendahnya literasi membaca siswa sekolah dasar. Jurnal Pendidikan Dasar Nusantara, 8(1), 23–37. https://doi.org/10.29407/jpdn.v8i1.18923

OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. Paris: OECD Publishing. https://doi.org/10.1787/53f23881-en

Oliveira, E., Mohoni, M., López-Pernas, S., & Saqr, M. (2026). AI collaboration or cheating? Using explainable authorship verification to measure AI assistance in academic writing. Educational Technology & Society, 29(2), 100–117. https://doi.org/10.30191/ETS.202604_29(2).RP06

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4380

Pratama, M. P., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: Harnessing the power of artificial intelligence for personalized learning. Klasikal: Journal of Education, Language Teaching and Science, 5(2), 350–357. https://doi.org/10.52208/klasikal.v5i2.877

Pratiwi, R., Anggraini, D., & Setyowati, E. (2023). Analisis kemampuan literasi membaca siswa kelas V Sekolah Dasar berdasarkan kerangka PIRLS. Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran, 13(2), 215–230. https://doi.org/10.25273/pe.v13i2.16742

Sari, M., & Hidayat, R. (2023). Pengaruh penggunaan media digital terhadap pemahaman membaca siswa sekolah dasar. Jurnal Basicedu, 7(2), 1234–1245. https://doi.org/10.31004/basicedu.v7i2.5341

Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 69(S2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Setiawan, A., Pertiwi, S. M., Kartowagiran, B., & Soenarto, S. (2020). Factors affecting reading literacy of Indonesian elementary school students. International Journal of Innovation, Creativity and Change, 14(3), 678–693.

Snow, C. E. (2017). Early literacy development and instruction: An overview. In N. Kucirkova, C. E. Snow, V. Grøver, & C. McBride (Eds.), The Routledge International Handbook of Early Literacy Education (pp. 5–13). London: Routledge.

Song, Y., Li, H., & Zheng, Q. (2026). Multi-agent automated item generation based on large language models. Educational Technology & Society, 29(2), 233–248. https://doi.org/10.30191/ETS.202604_29(2).SP02

Subandi, S., Suryanti, S., & Joko, J. (2023). Implementing the literacy movement in Indonesian elementary schools: A case study. International Journal of Elementary Education, 7(2), 198–209. https://doi.org/10.23887/ijee.v7i2.55241

Sumarsono, P., Hidayati, D. N., Cahyono, B. Y., & Suryati, N. (2023). The relationship between digital literacy and reading comprehension among Indonesian primary students. International Journal of Evaluation and Research in Education, 12(1), 367–376. https://doi.org/10.11591/ijere.v12i1.24017

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tsai, Y.-H., Chen, S.-J., Hsiao, L. C., Chen, C.-Y., Wang, S.-L., & Wu, C.-J. (2026). AI-assisted security testing in 5G networks for teaching cybersecurity with GitHub Copilot. Educational Technology & Society, 29(2), 46–68. https://doi.org/10.30191/ETS.202604_29(2).RP03

Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2023). When adaptive learning is effective learning: Comparison of an adaptive learning system to alternative homework delivery formats. Journal of Computer Assisted Learning, 39(2), 504–519. https://doi.org/10.1111/jcal.12762

Wang, T., Zheng, J., Li, S., Zhang, Y., Hu, X., & Lajoie, S. P. (2026). Exploring students' cognitive load in a computer-simulated learning environment through multimodal data. Educational Technology & Society, 29(2), 118–135. https://doi.org/10.30191/ETS.202604_29(2).RP07

Wang, X., Li, J., Gu, Q., & Liang, D. (2026). Promoting problem-solving skills in computer programming with an LLM-empowered multi-agent system: A cognitive apprenticeship perspective. Educational Technology & Society, 29(2), 355–374. https://doi.org/10.30191/ETS.202604_29(2).SP08

Wijayanti, P., & Sari, A. P. (2022). Implementasi pembelajaran adaptif dalam meningkatkan literasi membaca siswa sekolah dasar. Mimbar Sekolah Dasar, 9(3), 412–428. https://doi.org/10.53400/mimbar-sd.v9i3.45128

Wu, Y. (2026). Enhancing student critical thinking skills and dispositions through a multi-agent system with argumentation and devil's advocacy. Educational Technology & Society, 29(2), 339–354. https://doi.org/10.30191/ETS.202604_29(2).SP07

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. Computers & Education, 140, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1), 90–112. https://doi.org/10.1111/bjet.13370

Zhang, F., Xing, W., Li, C., & Zhu, W. (2026). An LLM-powered multi-agent system: Enabling real-time cognitive and psychological state assessment in K-12 mathematics learning. Educational Technology & Society, 29(2), 404–421. https://doi.org/10.30191/ETS.202604_29(2).SP10

Zhang, J., Ma, Q., Zhang, Y., & Cao, X. (2026). Multi-agent vs. single-agent AI for EFL speaking practice: A controlled experiment with hybrid input, contextual dialogue, and proficiency-adaptive feedback. Educational Technology & Society, 29(2), 297–322. https://doi.org/10.30191/ETS.202604_29(2).SP05