



INTEGRATING ARTIFICIAL INTELEGENGE IN READING COMPREHENSION: CLASSROOM ACTION RESEARCH WITH EFL UNIVERSITY STUDENTS

Mahendra Puji Permana Aji¹, Jenny Ika Misela², Putri Novitasari³

¹Universitas Nusantara PGRI, Kediri, Indonesia

^{2,3}Universitas Wahidiyah, Kediri, Indonesia

Article info	ABSTRACT
Corresponding Author: Jenny Ika Misela jennyikamisela@gmail.com Universitas Wahidiyah	The use of generative artificial intelligence in teaching reading comprehension has expanded rapidly since late 2022. Despite growing interest among language educators, few studies have examined how teachers can systematically introduce AI tools into actual classroom reading instruction using iterative, reflective methods. This study aimed to answer two questions: first, whether structured AI integration across multiple instructional cycles improves EFL students' reading test scores, and second, how students' ways of interacting with AI change as they gain more experience. Forty second-year English majors at an Universitas Nusantara PGRI Kediri participated in three action research cycles over nine weeks. The researcher, who was also the classroom teacher, designed nine reading tasks where students used ChatGPT alongside academic passages. Data collection included pre-tests and post-tests measuring five reading sub-skills, screenshots of student-AI conversations (360 total), questionnaires administered after each cycle, semi-structured interviews with 12 students, and detailed classroom observation notes. Test scores rose from an average of 62.4% before the intervention to 79.25% at the end of Cycle 3. A repeated measures ANOVA confirmed that these improvements were statistically significant ($p < .001$) with a large effect size (partial $\eta^2 = 0.676$). The most notable gains appeared in two higher-order skills: synthesizing information across multiple text segments (+20.3 percentage points) and drawing inferences from implied content (+19.0 percentage points). An examination of the 360 ChatGPT interaction logs revealed a clear shift in how students used the tool. During Cycle 1, 62% of all prompts were either vocabulary definitions (34%) or translation requests (28%). By Cycle 3, these basic requests had dropped to 25% of total prompts, while more sophisticated prompts emerged: questioning an author's tone or bias (18%), asking for counterarguments (15%), and requesting feedback on the student's own understanding (12%). Interview data, analyzed using thematic analysis, produced four main themes. First, students consistently reported that ChatGPT helped them overcome comprehension barriers that would otherwise have stopped them from engaging with difficult academic texts. Second, participants described developing a more critical stance over time, learning to verify AI outputs and recognize potential inaccuracies. Third, students expressed ongoing tension between appreciating AI's helpfulness and worrying about becoming dependent on it. Fourth, those who worked in pairs during Cycles 2 and 3 reported that discussing AI responses with a partner reduced uncritical acceptance and deepened their understanding. When teachers integrate AI tools through sustained, reflection-driven classroom action research rather than one-time implementation, students can achieve measurable reading

	comprehension gains while simultaneously developing critical digital literacy. The findings suggest three practical principles for language educators: introduce basic prompting first, add collaborative structures to encourage peer discussion of AI outputs, and continuously adjust instruction based on observed student difficulties rather than following a fixed plan.
	Keywords: Artificial intelligence; reading comprehension; ChatGPT; classroom action
This article distributed under the terms of the Creative Commons Attribution-Share Alike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/)	

INTRODUCTION

Generative AI tools, particularly ChatGPT, have changed how university students approach academic texts since their public release in late 2022. What was once a niche area of educational technology has quickly become a daily reality in many language classrooms (Kasneci et al., 2023). For students learning English as a foreign language, reading comprehension remains one of the hardest skills to master (Grabe & Stoller, 2019). It is not just about knowing vocabulary or parsing sentence structures—learners must also make inferences, evaluate arguments, and synthesize information from multiple sources (Cromley & Azevedo, 2007). The problem is that many students struggle to coordinate all these cognitive tasks at once (Hyland, 2014). This is where AI tools enter the picture. They can function as translators, explainers of difficult concepts, and even strategic partners in the reading process (Joo & Belcher, 2025). However, the research base on how teachers should integrate these tools into regular classroom reading instruction is still quite thin (Zawacki-Richter et al., 2019; Holmes et al., 2019).

A close look at the published literature reveals two main gaps. First, most studies have concentrated on AI's role in supporting student writing, not reading (Law, 2024). When researchers have looked at reading, they have often let students use AI however they wanted, without much teacher guidance or structured classroom activity (Wang et al., 2024). That approach tells us something about what students do on their own, but it does not tell language teachers how to design lessons that make good use of AI (Wei, 2023). Second, while we know that collaborative learning—students working together, discussing texts, and co constructing meaning—can strengthen reading comprehension (Vygotsky, 1978), almost no studies have examined what happens when you combine collaborative reading tasks with AI scaffolding inside a single instructional design (Baadache & Latreche, 2025). These gaps matter because teachers need practical, evidence-based models, not just general suggestions about using technology (Mujiono & Weganofa, 2026).

This study tries to fill those gaps using a Classroom Action Research (CAR) design (Kemmis, McTaggart, & Nixon, 2014). Why CAR? Because integrating AI into reading instruction is not something you can do perfectly on the first try (Lee, 2025). It requires trying things out, seeing what works and what does not, and making adjustments based on real classroom evidence (Memarian & Doleck, 2023). A one off workshop or a single semester long experiment cannot capture that iterative process (Kemmis et al., 2014). CAR allows the teacher researcher to plan a lesson, teach it, observe what happens, reflect on the results, and then change the approach for the next cycle (Lee, 2025). Over three cycles, this builds a much

richer picture of how AI integration actually unfolds in a real classroom with real students (Kemmis et al., 2014).

To make sense of what students need to do when they read, this study draws on the Direct and Inferential Mediation (DIME) model (Cromley & Azevedo, 2007). According to this framework, good comprehension depends on four interconnected abilities: knowing enough vocabulary, having relevant background knowledge, being able to draw inferences, and using effective reading strategies (Grabe & Stoller, 2019). None of these works in isolation. A student who knows every word in a passage might still fail to grasp the author's implied argument (Lea & Street, 2006). Another student with strong background knowledge might get stuck on unfamiliar vocabulary and give up (Hyland, 2014). The DIME model is useful here because it lets us think systematically about which parts of the reading process AI might support and which parts students need to do for themselves (Joo & Belcher, 2025).

Any discussion of AI in education must also address ethical concerns (Floridi & Cows, 2022). Researchers have identified several principles that matter in classroom settings: beneficence (doing good), non-maleficence (avoiding harm), justice (fair access), autonomy (student choice), and explicability (being transparent about how AI works) (Mujiono & Weganofa, 2026; Schiff, 2022). Students themselves have voiced worries about over reliance, privacy, data security, and the risk that using AI might weaken their own independent thinking (Yang et al., 2025; Chan, 2023). These are not abstract philosophical debates (Nguyen et al., 2024). In a classroom of 40 students, these concerns show up in real ways. Some students become anxious about cheating (Rahman & Watanobe, 2023). Others worry that their classmates with better internet access or more advanced digital skills will have an unfair advantage (Zhu et al., 2025). This study therefore built in several safeguards: explicit instruction on responsible AI use, equal access to technology, anonymized data collection, and a strong emphasis on verification skills (Chan, 2023).

What does existing research tell us about AI and reading? Several recent studies offer encouraging results. Joo and Belcher (2025) followed one international graduate student who used ChatGPT for academic reading. Over time, the student moved from simple translation requests toward more strategic uses like summarizing and checking inferences. Baadache and Latreche (2025) worked with 58 first year university students and found that combining AI tools with collaborative learning produced measurable gains in comprehension, vocabulary, and inference making. Another study involving 35 EFL students over eight weeks documented significant improvements in three specific critical reading skills: identifying main ideas, recognizing bias and tone, and generating counterarguments (Wang et al., 2024). Perhaps more interestingly, the same study found that students' patterns of AI interaction changed over time (Wei, 2023). In early weeks, they mostly asked for clarification. By the final weeks, they were asking more probing, evaluative questions (Joo & Belcher, 2025).

Despite these promising findings, serious challenges remain (Rahman & Watanobe, 2023). Some students become overly dependent on AI, accepting whatever it generates without question (Wang et al., 2024). Others struggle to verify whether the information ChatGPT provides is accurate (Wei, 2023). Researchers have also documented concerns about creativity loss, job displacement, algorithmic bias, and privacy risks (Yang et al., 2025; Zhu et al., 2025). These findings point to one clear conclusion: simply handing students access to an AI tool is not enough (Kasneci et al., 2023). Teachers must pair AI integration

with explicit instruction in critical AI literacy—teaching students how to evaluate AI outputs, spot potential biases, and maintain their own intellectual independence (Chan, 2023; Schiff, 2022).

This study addresses the gaps identified above through two research questions: 1) Over three cycles of classroom action research, how does structured AI integration affect the EFL students' of Universitas Nusantara PGRI Kediri scores on reading comprehension tests? 2) As instruction progresses across the three cycles, how do students' perceptions of AI change, and what patterns of interaction with ChatGPT emerge?

By answering these questions, the study aims to contribute both empirical evidence and practical pedagogical insights (Joo & Belcher, 2025). The iterative CAR design makes it possible to document not just whether AI integration works, but how it works—and how it can be improved through reflective adjustments (Kemmis et al., 2014; Lee, 2025). For language teachers who are wondering whether and how to bring AI into their reading classrooms, this study offers a detailed, cycle by cycle account of what happened with 40 real students over nine weeks of instruction (Baadache & Latreche, 2025; Wang et al., 2024).

METHOD

Generative AI tools, particularly ChatGPT, have changed how university students approach academic texts since their public release in late 2022. What was once a niche area of educational technology has quickly become a daily reality in many language classrooms (Kasneci et al., 2023). For students learning English as a foreign language, reading comprehension remains one of the hardest skills to master (Grabe & Stoller, 2019). It is not just about knowing vocabulary or parsing sentence structures—learners must also make inferences, evaluate arguments, and synthesize information from multiple sources (Cromley & Azevedo, 2007). The problem is that many students struggle to coordinate all these cognitive tasks at once (Hyland, 2014). This is where AI tools enter the picture. They can function as translators, explainers of difficult concepts, and even strategic partners in the reading process (Joo & Belcher, 2025). However, the research base on how teachers should integrate these tools into regular classroom reading instruction is still quite thin (Zawacki-Richter et al., 2019; Holmes et al., 2019).

A close look at the published literature reveals two main gaps. First, most studies have concentrated on AI's role in supporting student writing, not reading (Law, 2024). When researchers have looked at reading, they have often let students use AI however they wanted, without much teacher guidance or structured classroom activity (Wang et al., 2024). That approach tells us something about what students do on their own, but it does not tell language teachers how to design lessons that make good use of AI (Wei, 2023). Second, while we know that collaborative learning—students working together, discussing texts, and co constructing meaning—can strengthen reading comprehension (Vygotsky, 1978), almost no studies have examined what happens when you combine collaborative reading tasks with AI scaffolding inside a single instructional design (Baadache & Latreche, 2025). These gaps matter because teachers need practical, evidence-based models, not just general suggestions about using technology (Mujiono & Weganofa, 2026).

This study tries to fill those gaps using a Classroom Action Research (CAR) design (Kemmis, McTaggart, & Nixon, 2014). Why CAR? Because integrating AI into reading instruction is not something you can do perfectly on the first try (Lee, 2025). It requires trying

things out, seeing what works and what does not, and adjusting based on real classroom evidence (Memarian & Doleck, 2023). A one-off workshop or a single semester long experiment cannot capture that iterative process (Kemmis et al., 2014). CAR allows the teacher researcher to plan a lesson, teach it, observe what happens, reflect on the results, and then change the approach for the next cycle (Lee, 2025). Over three cycles, this builds a much richer picture of how AI integration unfolds in a real classroom with real students (Kemmis et al., 2014).

To make sense of what students need to do when they read, this study draws on the Direct and Inferential Mediation (DIME) model (Cromley & Azevedo, 2007). According to this framework, good comprehension depends on four interconnected abilities: knowing enough vocabulary, having relevant background knowledge, being able to draw inferences, and using effective reading strategies (Grabe & Stoller, 2019). None of these works in isolation. A student who knows every word in a passage might still fail to grasp the author's implied argument (Lea & Street, 2006). Another student with strong background knowledge might get stuck on unfamiliar vocabulary and give up (Hyland, 2014). The DIME model is useful here because it lets us think systematically about which parts of the reading process AI might support and which parts students need to do for themselves (Joo & Belcher, 2025).

Any discussion of AI in education must also address ethical concerns (Floridi & Cowls, 2022). Researchers have identified several principles that matter in classroom settings: beneficence (doing good), non-maleficence (avoiding harm), justice (fair access), autonomy (student choice), and explicability (being transparent about how AI works) (Mujiono & Weganofa, 2026; Schiff, 2022). Students themselves have voiced worries about over reliance, privacy, data security, and the risk that using AI might weaken their own independent thinking (Yang et al., 2025; Chan, 2023). These are not abstract philosophical debates (Nguyen et al., 2024). In a classroom of 40 students, these concerns show up in real ways. Some students become anxious about cheating (Rahman & Watanobe, 2023). Others worry that their classmates with better internet access or more advanced digital skills will have an unfair advantage (Zhu et al., 2025). This study therefore built in several safeguards: explicit instruction on responsible AI use, equal access to technology, anonymized data collection, and a strong emphasis on verification skills (Chan, 2023).

What does existing research tell us about AI and reading? Several recent studies offer encouraging results. Joo and Belcher (2025) followed one international graduate student who used ChatGPT for academic reading. Over time, the student moved from simple translation requests toward more strategic uses like summarizing and checking inferences. Baadache and Latreche (2025) worked with 58 first year university students and found that combining AI tools with collaborative learning produced measurable gains in comprehension, vocabulary, and inference making. Another study involving 35 EFL students over eight weeks documented significant improvements in three specific critical reading skills: identifying main ideas, recognizing bias and tone, and generating counterarguments (Wang et al., 2024). Perhaps more interestingly, the same study found that students' patterns of AI interaction changed over time (Wei, 2023). In early weeks, they mostly asked for clarification. By the final weeks, they were asking more probing, evaluative questions (Joo & Belcher, 2025).

Despite these promising findings, serious challenges remain (Rahman & Watanobe, 2023). Some students become overly dependent on AI, accepting whatever it generates

without question (Wang et al., 2024). Others struggle to verify whether the information ChatGPT provides is accurate (Wei, 2023). Researchers have also documented concerns about creativity loss, job displacement, algorithmic bias, and privacy risks (Yang et al., 2025; Zhu et al., 2025). These findings point to one clear conclusion: simply handing students access to an AI tool is not enough (Kasneci et al., 2023). Teachers must pair AI integration with explicit instruction in critical AI literacy—teaching students how to evaluate AI outputs, spot potential biases, and maintain their own intellectual independence (Chan, 2023; Schiff, 2022).

RESULT AND DISCUSSION

Effects of AI Integration on Reading Comprehension Performance

The first research question asked how the integration of AI tools into reading comprehension instruction affected students' reading comprehension performance across three cycles of classroom action research. Quantitative data from pre-test and post-test assessments across the three cycles are presented in Table 1.

Table 1. Descriptive Statistics and Repeated Measures ANOVA for Reading Comprehension Scores Across Cycles

Assessment	N	Mean (%)	Sd	Gain from Previous
Pre-test (baseline)	40	62.40	7.82	-
Cycle 1 post-test	40	68.95	7.43	+6.55
Cycle 2 post-test	40	74.80	6.91	+5.85
Cycle 3 post-test	40	79.25	6.48	+4.45

Note: Total possible score = 100%

Looking at the numbers in Table 1, a clear upward trend appears across the three cycles. Students started with an average score of 62.4% before any AI-integrated instruction. After the first three-week cycle, that average climbed to 68.95%, a gain of just over six and a half percentage points. By the end of Cycle 2, the mean score had reached 74.8%, adding nearly another six points. The final post-test after Cycle 3 showed an average of 79.25%, bringing the total improvement from beginning to end to 16.85 percentage points (Wang et al., 2024).

To determine whether these improvements were statistically meaningful, one-way repeated measures ANOVA were conducted. The analysis revealed a statistically significant effect of time (cycle) on reading comprehension scores, Wilks' Lambda = 0.324, $F(3, 37) = 25.73$, $p < .001$, partial $\eta^2 = 0.676$ (Cromley & Azevedo, 2007). This large effect size indicates that the AI-integrated intervention accounted for approximately 68% of the variance in the score improvements students demonstrated. In practical terms, this means the changes observed were not random fluctuations but were systematically related to the instructional approach (Grabe & Stoller, 2019). Post-hoc pairwise comparisons using Bonferroni correction were then performed to examine differences between each successive assessment point. The mean difference between pre-test and Cycle 1 post-test was 6.55 percentage points ($p < .001$, 95% CI [3.89, 9.21]). This statistically significant gain suggests that even the initial, relatively basic integration of AI tools in Cycle 1 produced measurable benefits for students' reading comprehension (Baadache & Latreche, 2025).

The improvement from Cycle 1 to Cycle 2 was also statistically significant, with a mean difference of 5.85 percentage points ($p < .001$, 95% CI [3.42, 8.28]). This finding is

worth noting because the instructional approach in Cycle 2 differed from Cycle 1 in two important ways: students received explicit instruction on advanced prompting strategies, and they worked in pairs rather than individually (Vygotsky, 1978). The continued significant gain suggests that these modifications supported further learning (Wei, 2023).

The improvement from Cycle 2 to Cycle 3 remained statistically significant as well, with a mean difference of 4.45 percentage points ($p = .002$, 95% CI [1.56, 7.34]). Although the magnitude of gain was slightly smaller than in previous cycles, the continued upward trajectory indicates that students were still developing their skills by the end of the nine-week intervention (Wang et al., 2024). One possible explanation for the slightly reduced gain is that students were approaching a performance ceiling, or that the reading tasks in Cycle 3 were deliberately more challenging, requiring synthesis and critical evaluation rather than basic comprehension (Cromley & Azevedo, 2007).

Performance on Reading Sub-skills

Table 2 breaks down student performance by the five reading sub-skills measured in the study. This disaggregated view reveals which aspects of reading comprehension benefited most from AI-integrated instruction.

Table 2. Performance on Reading Sub-skills Across Cycles

Sub-Skill	Pre-test (%)	Cycle 1 (%)	Cycle 2 (%)	Cycle 3 (%)	Total Gain
Identifying main idea	70.5 (8.2)	75.3 (7.6)	79.8 (7.1)	83.5 (6.5)	+13.0
Vocabulary in context	64.2(9.1)	70.8 (8.4)	76.6 (7.8)	80.8 (7.2)	+16.6
Making inferences	58.5 (10.2)	65.5 (9.5)	72.3 (8.9)	77.5 (8.3)	+19.0
Evaluating Arguments	60.8 (9.8)	66.2 (9.1)	72.0 (8.5)	76.8 (8.0)	+16.0
Synthesizing Information	58.2 (10.5)	65.8 (9.8)	73.5 (9.2)	78.5 (8.5)	+20.3

Note: Values presented as Mean (Standard Deviation).

The data in Table 2 tells an interesting story. Two sub-skills stand out as showing the largest gains: synthesizing information across multiple text segments gained 20.3 percentage points, and making inferences from implied content gained 19.0 percentage points. Both are higher-order cognitive skills that typically pose significant challenges for EFL learners (Grabe & Stoller, 2019; Hyland, 2014).

Why might AI integration have been particularly effective for these two sub-skills? One plausible explanation comes from the nature of how students used ChatGPT. As documented in the AI interaction logs, students in later cycles increasingly asked the AI to compare information across different parts of a text or to identify implicit claims that were not directly stated (Joo & Belcher, 2025). In other words, students were using AI as a "thinking partner" that helped them practice precisely the cognitive moves required for synthesis and inference (Wei, 2023).

The smallest gain among the five sub-skills was in identifying main ideas, which increased by 13.0 percentage points. However, it is worth noting that students started with a much higher baseline score on this sub-skill (70.5%) compared to inference (58.5%) or synthesis (58.2%). This higher starting point may reflect previous instructional emphasis on main idea identification in earlier courses, as well as the possibility of a ceiling effect where

students had less room to improve (Cromley & Azevedo, 2007). Despite the smaller gain, the increase from 70.5% to 83.5% still represents meaningful progress.

Vocabulary in context (+16.6 percentage points) and evaluating arguments (+16.0 percentage points) showed moderate but substantial gains. The vocabulary improvement is interesting because students actually reduced their direct vocabulary definition requests to ChatGPT over time (from 34% of prompts in Cycle 1 to 15% in Cycle 3, as shown later in Table 4). This suggests that vocabulary knowledge improved not because students asked for more definitions, but because repeated exposure to academic texts—supported by strategic AI use—helped them acquire vocabulary more naturally (Grabe & Stoller, 2019).

Discussion

Discussion of research question 1 findings

The significant improvements documented across all three cycles support the theoretical position that AI can function as a strategic partner in reading development rather than merely a shortcut or crutch (Joo & Belcher, 2025). Within the DIME framework proposed by Cromley and Azevedo (2007), ChatGPT appeared to support multiple components of reading comprehension simultaneously. It helped with vocabulary knowledge through definitions and contextual explanations. It supported background knowledge by elaborating on unfamiliar concepts when students asked for clarification. Perhaps most importantly, it scaffolded inference-making and strategic reading behaviors through guided questioning and modeling of comprehension monitoring (Wei, 2023).

These findings align with previous research by Baadache and Latreche (2025), who found that combining AI tools with collaborative learning produced substantial gains in comprehension and inference-making among first-year university students. The results are also consistent with Wang and colleagues' (2024) study of ChatGPT-based reading instruction, which documented significant improvements across multiple reading sub-skills over an eight-week period.

However, the continued improvement across all three cycles in the present study suggests something that single-cycle or pre-test/post-test designs cannot capture: effective AI integration requires sustained, reflective implementation rather than one-time exposure (Lee, 2025; Kemmis et al., 2014). The iterative CAR design enabled the researcher to progressively refine instructional scaffolding based on observed student needs. For example, the introduction of collaborative learning in Cycle 2 directly addressed the tendency toward passive AI use documented during Cycle 1 observations (Vygotsky, 1978). Similarly, the emphasis on critical evaluation and synthesis in Cycle 3 built on students' developing strategic competence, which would not have been possible if the same instructional approach had been used throughout (Memarian & Doleck, 2023).

Another noteworthy finding is the pattern in which sub-skills improved most. The largest gains occurred in synthesis and inference—precisely the skills that are most difficult to teach through traditional, teacher-centered methods (Grabe & Stoller, 2019). In a conventional reading classroom, a teacher might model inference-making or provide guided practice, but each student receives limited individual feedback. With AI integration, however, each student could practice inference generation repeatedly, receive immediate responses, and refine their questions based on AI output (Joo & Belcher, 2025). This individualized, low-stakes practice environment may explain why these higher-order skills showed such substantial improvement (Wei, 2023).

Nevertheless, it would be premature to conclude that AI alone caused these improvements. Three factors likely interacted to produce the observed gains. First, the AI tool itself provided scaffolding that reduced cognitive load, allowing students to focus on comprehension strategies rather than getting stuck on individual words or sentences (Wang et al., 2024). Second, the explicit instruction students received on prompting strategies, particularly in Cycles 2 and 3—taught them how to use the tool more effectively (Joo & Belcher, 2025). Third, the collaborative learning structures introduced in Cycle 2 created opportunities for peer discussion and negotiation of meaning, which are known to support deeper comprehension (Vygotsky, 1978). The CAR design, with its progressive refinements, allowed all three factors to work together synergistically.

In summary, the quantitative findings for Research Question 1 demonstrate that structured, sustained AI integration across three CAR cycles produced statistically significant and practically meaningful improvements in EFL students' reading comprehension. The largest effects appeared in higher-order sub-skills—synthesis and inference—that are typically most challenging for language learners to develop. These improvements were not a one-time jump but accumulated progressively across cycles, suggesting that iterative refinement of AI-integrated instruction is essential for maximizing learning outcomes.

Student Perceptions and Interaction Patterns with AI (research question 2)

The second research question explored two interrelated aspects of the AI integration process: first, how students perceived the use of AI tools for reading comprehension across the three cycles, and second, what patterns of interaction with ChatGPT emerged as instruction progressed.

Questionnaire Findings on Student Perceptions

Table 3 presents students' perceptions of AI integration measured through the 5-point Likert-scale questionnaire administered at the end of each cycle. The data show a clear trend of increasingly positive perceptions across the three cycles, alongside declining concerns about over-reliance.

Table 3. Student Perceptions of AI Integration Across Cycles (Mean Scores, 1-5)

Statement	Cycle 1	Cycle 2	Cycle 3
AI tools helped me understand difficult texts	38.5 (0.73)	4.13 (0.69)	4.35 (0.62)
I felt confident using AI for reading tasks	3.42 (0.88)	3.98 (0.77)	4.28 (0.68)
Using AI improved my reading skills	3.68 (0.79)	4.05 (0.72)	4.33 (0.66)
I was concerned about over-relying on AI	4.08 (0.76)	3.85 (0.83)	3.45 (0.91)
I verified AI-generated information	3.33(0.89)	3.85 (0.80)	4.20 (0.72)
AI made reading more motivating	3.80 (0.82)	4.10 (0.74)	4.25 (0.67)
I prefer reading with AI support	3.55 (0.93)	3.98 (0.85)	4.15 (0.78)

Note: Standard deviations in parentheses.

Looking at the numbers in Table 3, several patterns stand out. Students' agreement that AI tools helped them understand difficult texts rose steadily from 3.85 in Cycle 1 to 4.35 in Cycle 3. Confidence in using AI for reading tasks showed even larger growth, jumping from 3.43 to 4.28 over the same period. The item measuring perceived skill improvement followed a similar upward trajectory, moving from 3.68 to 4.33 (Wang et al., 2024).

Perhaps the most interesting pattern involves the relationship between two items: concern about over-reliance and verification behavior. At the beginning of the study (Cycle 1), students expressed relatively high concern about becoming too dependent on AI ($M = 4.08$). At the same time, their reported verification of AI-generated information was modest ($M = 3.33$). By Cycle 3, a notable shift occurred. Concern about over-reliance had dropped to 3.45, while verification behavior had climbed to 4.20 (Joo & Belcher, 2025).

This inverse relationship suggests something important about how students' relationship with AI matured over time. Rather than simply becoming less worried and less careful, students appeared to develop a more nuanced stance. They became more confident in their ability to use AI strategically while simultaneously becoming more disciplined about checking AI outputs (Wei, 2023). One student captured this shift in an open-ended response: "At first I just copied what ChatGPT said. Now I ask for ideas, but I check and think myself. I know the AI can be wrong sometimes" (Baadache & Latreche, 2025).

Friedman's ANOVA confirmed that the changes in perception scores across cycles were statistically significant for all questionnaire items ($p < .01$). This indicates that the three-cycle intervention produced measurable and meaningful shifts in how students viewed and used AI for academic reading (Wang et al., 2024).

Patterns of AI Interaction: Analysis of ChatGPT Logs

To understand how students' actual behavior with AI changed—not just their self-reported perceptions, the researcher analyzed 360 screenshots of ChatGPT interactions (9 tasks $\times$ 40 students). Table 4 presents the frequency distribution of prompt types across the three cycles.

Table 4. Frequency of Prompt Types Across Cycles (Percentage of Total Prompts)

Prompt Type	Cycle 1	Cycle 2	Cycle 3
Vocabulary definition	34%	22%	15%
Translation request	28%	18%	10 %
Main idea verification	18%	20%	18%
Clarification seeking	12%	15%	12%
Tone/bias questioning	4%	12%	18%
Counterargument generation	2%	8%	15%
Meta-cognitive reflection	2%	5%	12%

The data in Table 4 reveal a dramatic transformation in how students engaged with ChatGPT over the nine-week intervention.

Cycle 1: Basic Comprehension Support.

During the first cycle, student prompts were dominated by two types of requests. Vocabulary definitions made up 34% of all prompts, and translation requests added another 28%. Together, these two categories accounted for nearly two-thirds of all student-AI interactions (62%). Typical prompts from this stage were simple and direct, such as: "What does 'ubiquitous' mean?" or "Translate this paragraph to Indonesian" (Joo & Belcher, 2025).

Only a small minority of less than 10%—demonstrated of critical engagement, such as questioning an author's bias or generating alternative viewpoints (Wei, 2023).

This pattern suggests that in the absence of explicit instruction, students defaulted to using AI as a basic comprehension aid or what might be called a "digital dictionary" (Wang et al., 2024). While this use did support basic understanding—as evidenced by the significant gains from pre-test to Cycle 1 post-test shown earlier in Table 1—it did not push students toward deeper, more analytical engagement with texts.

Cycle 2: Emergence of Critical Engagement.

A notable shift occurred during the second cycle. Vocabulary and translation requests decreased substantially, dropping to 22% and 18% respectively. At the same time, new types of prompts began appearing with meaningful frequency. Tone and bias questioning rose to 12% of prompts, and counterargument generation reached 8% (Baadache & Latreche, 2025).

Students started formulating more sophisticated inquiries. Examples from the interaction logs included: "What is the author's attitude toward renewable energy? Is the article biased?" and "Can you provide an opposing viewpoint to this argument?" (Joo & Belcher, 2025). These were not questions students had asked during the first three weeks of the study. Several factors likely contributed to this shift: explicit instruction on advanced prompting strategies delivered at the beginning of Cycle 2, the introduction of collaborative pair work that encouraged discussion of AI outputs, and reading tasks specifically designed to require critical evaluation (Vygotsky, 1978; Wei, 2023).

Cycle 3: Strategic and Meta-Cognitive Use.

By the final cycle, the transformation in student-AI interaction patterns had fully consolidated. Vocabulary requests fell further to 15%, and translation prompts dropped to only 10%—a dramatic reduction from the 62% combined total seen in Cycle 1. Meanwhile, tone and bias questioning climbed to 18%, and counterargument generation reached 15% (Wang et al., 2024).

Perhaps the most interesting development in Cycle 3 was the rise of meta-cognitive reflection prompts. These were almost non-existent in Cycle 1 (2%) and remained modest in Cycle 2 (5%), but by Cycle 3 they made up 12% of all prompts. Students wrote things like: "Here is my summary. What did I miss?" and "What reading strategies should I use for this type of text?" (Joo & Belcher, 2025). These prompts are qualitatively different from earlier ones because they ask the AI not for information about the text but for feedback on the student's own understanding and learning process (Wei, 2023).

This progression from surface-level clarification toward meta-cognitive reflection is significant for several reasons. First, it demonstrates that students can develop from passive recipients of AI-generated information to active, strategic users who deploy AI for specific cognitive purposes (Baadache & Latreche, 2025). Second, the emergence of meta-cognitive prompting suggests that AI can support not only text comprehension but also learners' awareness and regulation of their own reading processes—a key component of autonomous learning (Grabe & Stoller, 2019). Third, the increasing proportion of critical questioning prompts indicates that students were developing what might be called "critical AI literacy": the ability to interrogate AI outputs, recognize potential biases, and maintain an evaluative stance toward algorithmically generated content (Chan, 2023; Schiff, 2022).

Thematic Analysis of Interview Data

Thematic analysis of semi-structured interviews (n = 36, representing 12 students interviewed after each cycle) yielded four major themes characterizing students' experiences with AI-integrated reading instruction. Each theme is presented below with representative quotes and connections to existing literature.

Theme 1: AI as Scaffold for Comprehension.

Across all three cycles, students consistently reported that ChatGPT helped them overcome comprehension barriers that would otherwise have prevented meaningful engagement with academic texts. This finding aligns with the concept of "easification" in AI-mediated learning, where technology reduces the cognitive load of reading, enabling learners to access texts above their independent proficiency level (Joo & Belcher, 2025).

One student (S24) explained: "When I don't know many words or the sentences are very long, I ask ChatGPT to explain simply. Then I can understand and continue reading. Without AI, I might just stop." Another student (S08) added: "It's helpful for difficult parts, but I try not to ask for everything. Otherwise, I don't learn." This distinction between appropriate and inappropriate scaffolding is important. Students who used AI selectively—for specific difficulties rather than comprehensive support—expressed greater confidence in their own developing abilities (Wei, 2023).

The scaffolding function of AI aligns with Vygotsky's (1978) concept of the Zone of Proximal Development (ZPD). When students encountered texts slightly above their independent reading level, ChatGPT provided the "expert assistance" needed to bridge the gap between what they could do alone and what they could achieve with support. Over time, as students internalized the strategies modeled by AI, their independent reading capacity expanded (Grabe & Stoller, 2019).

Theme 2: Developing Critical Stance Toward AI.

A prominent theme across interviews was students' growing awareness that AI outputs require verification and critical evaluation. This critical stance was particularly evident among students interviewed after Cycles 2 and 3, suggesting that explicit instruction on AI limitations and ethical use contributed to more sophisticated engagement (Chan, 2023).

One student (S17) reflected: "In the beginning I thought ChatGPT knows everything. Now I know it can be wrong or biased. I ask, 'Are you sure?' or check with my friend." Another student (S31) shared a specific example: "I asked about the causes of pollution and ChatGPT gave only industrial causes. I remembered from class that household waste is also important. So, I told ChatGPT and it agreed and added more information." (Baadache & Latreche, 2025).

This developing critical stance is encouraging because it suggests that students were not simply becoming more accepting of AI outputs over time. Instead, they were learning to treat AI-generated information as provisional and subject to verification—a key component of critical digital literacy (Schiff, 2022; Yang et al., 2025). The finding also aligns with research by Rahman and Watanobe (2023), who documented that students who received explicit instruction on AI limitations demonstrated greater skepticism and verification behavior compared to those who did not.

Theme 3: Autonomy and Dependence Tensions.

Students expressed ambivalence about the impact of AI on their independent reading abilities. While most appreciated the support AI provided, many voiced legitimate concerns about becoming dependent on the technology (Wang et al., 2024).

One student (S05) captured this tension well: "I'm worried that if I always use AI, my English will not improve by itself. But if I don't use it, I struggle more." Another student (S22) described her strategic approach to managing this tension: "I try to read first by myself. Then I use AI only for parts I really cannot understand. This way I still practice my own reading." (Wei, 2023).

This tension between autonomy and algorithmic support reflects broader debates in AI education literature regarding the conditions under which AI scaffolds versus replaces human cognition (Holmes et al., 2019; Kasneci et al., 2023). Students who reported more strategic AI use—deploying the tool for specific difficulties rather than comprehensive support—expressed less concern about dependency. This suggests that explicit instruction in strategic AI use may help mitigate the risks of over-reliance (Zhu et al., 2025).

Theme 4: Collaborative AI Use Enhances Learning.

Students who participated in paired tasks during Cycle 2 and Cycle 3 reported that working with a peer while using AI improved their learning outcomes compared to working alone (Vygotsky, 1978). One student (S14) explained: "When I work alone with ChatGPT, I just accept answers. But with my partner, we discuss what ChatGPT says. Sometimes we disagree and have to check again." Another student (S36) added: "My friend didn't understand what ChatGPT meant, so I had to explain in my own words. That helped me understand better too." (Baadache & Latreche, 2025).

This finding suggests that collaborative scaffolding may mitigate the risks of uncritical AI acceptance. When students work with partners, peer dialogue creates natural opportunities for negotiation, questioning, and co-construction of meaning (Vygotsky, 1978). The partner serves as a "second pair of eyes" that can catch errors or biases the individual might miss, and the requirement to explain AI outputs to a peer deepens the explainer's own understanding (Grabe & Stoller, 2019).

This theme aligns with the work of Baadache and Latreche (2025), who found that combining AI tools with collaborative learning produced larger effect sizes than AI use alone. The present study's findings extend this by showing that collaborative AI use also changed the quality of student-AI interaction, shifting students from passive acceptance toward active negotiation and critical dialogue.

Discussion of Research Question 2 Findings

The qualitative findings for Research Question 2 paint a picture of progressive transformation in how students perceived, used, and thought about AI for reading comprehension. Three key insights emerge from the analysis. First, the data show that students' relationship with AI is not static but evolves substantially with experience and instruction (Joo & Belcher, 2025). The shift from vocabulary and translation requests in Cycle 1 to tone/bias questioning and meta-cognitive reflection in Cycle 3 demonstrates that students can develop sophisticated, strategic AI literacy when given appropriate scaffolding. This finding challenges the concern that AI tools inevitably lead to passive, uncritical use (Rahman & Watanobe, 2023). Under the right instructional conditions, the opposite can occur.

Second, the inverse relationship between declining over-reliance concerns and increasing verification behaviour suggests that critical AI literacy and confidence in AI use are not opposing forces. Students did not become less concerned about dependence by becoming less careful. Rather, they became more confident because they developed strategies for verifying and critically evaluating AI outputs (Chan, 2023; Schiff, 2022). This finding has practical implications for instruction: teachers should emphasise verification skills and critical questioning alongside tool training, rather than treating ethics as a separate topic.

Third, the emergence of meta-cognitive prompts in Cycle 3 represents a qualitatively different level of engagement. When students ask ChatGPT "What did I miss in my summary?" or "What strategies should I use?" they are no longer using AI as a source of information about the text. They are using AI as a reflective partner in their own learning process (Wei, 2023). This kind of meta-cognitive engagement is associated with deeper learning and greater long-term retention in reading development research (Grabe & Stoller, 2019). The fact that students developed this capacity over nine weeks suggests that AI tools, when integrated thoughtfully, can support not just short-term comprehension but also the development of autonomous, self-regulated reading strategies.

However, it is important to acknowledge that not all students progressed at the same rate. Analysis of individual interaction logs revealed variability in how quickly students adopted more sophisticated prompting strategies. Students with higher initial reading proficiency (based on pre-test scores) generally showed faster progression toward critical and meta-cognitive prompts. Lower-proficiency students tended to take longer to reduce their reliance on vocabulary and translation requests (Wang et al., 2024). This finding suggests that differentiated instruction may be necessary in AI-integrated classrooms, with lower-proficiency students potentially needing more explicit scaffolding in basic prompting strategies before moving to critical engagement.

Integration of Research Question 1 and Research Question 2 Findings

Taken together, the quantitative findings from Research Question 1 and the qualitative findings from Research Question 2 paint a coherent picture. The significant improvements in reading comprehension scores—particularly in higher-order sub-skills like synthesis and inference—correspond with the documented shift in student-AI interaction patterns from basic clarification toward critical questioning and meta-cognitive reflection (Joo & Belcher, 2025; Wei, 2023).

Students did not simply use AI more often as the cycles progressed. They used it differently, and those differences in use were associated with larger learning gains. This finding supports the theoretical position that the quality of AI integration matters more than the quantity of AI use (Wang et al., 2024). A student who asks ChatGPT for 20 vocabulary definitions may show less improvement than a student who asks five well-crafted questions about author bias, argument strength, and missing perspectives.

The collaborative learning structures introduced in Cycle 2 appear to have played a role in accelerating this shift toward critical engagement. Interview data indicated that students who worked in pairs were more likely to question AI outputs and to articulate their reasoning to partners. This aligns with Vygotsky's (1978) sociocultural theory, which posits that higher cognitive functions develop first in social interaction before being internalized by the individual. In the context of AI integration, peer collaboration may provide the social

scaffold that helps students move from passive AI acceptance toward active, critical AI engagement (Baadache & Latreche, 2025).

CONCLUSION

This classroom action research has demonstrated that when AI tools are integrated through sustained, reflective, pedagogically grounded instruction, they can significantly enhance EFL students' reading comprehension while simultaneously fostering the critical digital literacy skills necessary for responsible AI use. The three-cycle CAR design allowed for progressive refinement of instruction based on classroom evidence, resulting in meaningful improvements in both quantitative outcomes (reading comprehension scores) and qualitative outcomes (student-AI interaction patterns and perceptions).

The findings support a vision of AI not as a replacement for teacher instruction or learner effort, but as a strategic partner that, under thoughtful pedagogical direction, expands the possibilities for meaningful engagement with academic texts. Students in this study did not become passive consumers of AI-generated content. On the contrary, over nine weeks they developed into more active, critical, and reflective readers who used AI strategically to support specific cognitive purposes while maintaining their own intellectual autonomy.

As generative AI tools become increasingly accessible and capable, the question facing language educators is not whether to use AI, but how to use it well. This study offers one empirically grounded answer: through iterative classroom action research that treats AI integration as a process of continuous improvement, informed by student evidence and guided by explicit instruction in critical AI literacy. For language teachers willing to engage in this reflective, evidence-based approach, AI tools offer not a shortcut around the hard work of learning to read, but a scaffold that can help more students reach higher levels of comprehension than they might achieve alone.

BIBLIOGRAPHY

- Baadache, H., & Latreche, K. (2025). *AI and collaborative learning in EFL reading instruction: A classroom-based study with first-year English students at Batna 2 University*. *Revue des Sciences Humaines*, *18*(2), 45-68.
- Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology*. *Qualitative Research in Psychology*, *3*(2), 77-101.
- Chan, C. K. Y. (2023). *A comprehensive AI policy education framework for university teaching and learning*. *International Journal of Educational Technology in Higher Education*, *20*, Article 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Cromley, J. G., & Azevedo, R. (2007). *Testing and refining the direct and inferential mediation model of reading comprehension*. *Journal of Educational Psychology*, *99*(2), 311-325. <https://doi.org/10.1037/0022-0663.99.2.311>
- Floridi, L., & Cowls, J. (2022). *A unified framework of five principles for AI in society*. *Harvard Data Science Review*, *1*(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading (3rd ed.)*. Routledge.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Hyland, K. (2014). *English for academic purposes*. Routledge.

- Joo, D., & Belcher, D. (2025). *Human-AI collaborative reading in academic contexts: An exploratory case study*. Journal of English for Academic Purposes, *73*, 102-118. <https://doi.org/10.1016/j.jeap.2024.102118>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... 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>
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer.
- Law, L. (2024). *Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping review*. Computers and Education: Artificial Intelligence, *6*, 100212. <https://doi.org/10.1016/j.caeai.2024.100212>
- Lea, M. R., & Street, B. V. (2006). *The "academic literacies" model: Theory and applications*. Theory into Practice, *45*(4), 368-377. https://doi.org/10.1207/s15430421tip4504_11
- Lee, M. (2025). *Developing and evaluating an interactive reading tool with teachers in the loop: Action research approach*. CEUR Workshop Proceedings, *3927*, 1-15.
- Memarian, B., & Doleck, T. (2023). *Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) in education: A systematic review*. Education and Information Technologies, *28*, 14265-14300. <https://doi.org/10.1007/s10639-023-11734-7>
- Mujiono, & Weganofa, R. (2026). *Ethical and legal challenges in implementing artificial intelligence (AI) and metaverse in education*. In S. K. Behera, A. K. Das, S. Mohapatra, & S. K. Mishra (Eds.), Artificial intelligence and the metaverse in education (pp. 45-68). Wiley.
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2024). *Ethical principles for artificial intelligence in education*. Education and Information Technologies, *29*, 9353-9382. <https://doi.org/10.1007/s10639-023-12200-0>
- Rahman, M. M., & Watanobe, Y. (2023). *ChatGPT for education and research: A review of benefits and risks*. Computers and Education: Artificial Intelligence, *5*, 100161. <https://doi.org/10.1016/j.caeai.2023.100161>
- Schiff, D. (2022). *Education for AI, not AI for education: The role of education and ethics in national AI policy*. AI and Society, *37*, 1107-1123. <https://doi.org/10.1007/s00146-021-01179-1>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, Y., Liu, C., & Tu, Y. F. (2024). *Effects of ChatGPT-based reading instruction on EFL students' reading comprehension and motivation*. Interactive Learning Environments, 1-18. Advance online publication. <https://doi.org/10.1080/10494820.2024.2308094>
- Wei, L. (2023). *Artificial intelligence in language instruction: Impact on English learning achievement, self-regulation, and anxiety*. Computer Assisted Language Learning, 1-25. Advance online publication. <https://doi.org/10.1080/09588221.2023.2284123>

- Yang, H., Gao, C., & Shen, H. (2025). *Student perspectives on AI ethics in higher education: A mixed-methods study*. Education and Information Technologies, *30*, 125-148. <https://doi.org/10.1007/s10639-024-13054-w>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). *Systematic review of research on artificial intelligence applications in higher education*. International Journal of Educational Technology in Higher Education, *16*, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhu, C., Sun, M., Luo, J., & Li, Y. (2025). *How to use generative AI ethically in higher education: A systematic literature review*. International Journal of Educational Technology in Higher Education, *22*, Article 15. <https://doi.org/10.1186/s41239-025-00465-w>
- Gray, D. E. (2022). *Doing research in the real world* (5th ed.). Sage Publications, Ltd.
- Rose, H., McKinley, J., & Galloway, N. (2020). Global Englishes and language teaching: A review of pedagogical research. *Language Teaching*, 54(2), 157–189. <https://doi.org/10.1017/S0261444820000518>
- Subekti, A. S. (2022). Teaching Indonesian L2 learners academic writing: A proposed instructional model. *SAGA: Journal of English Language Teaching and Applied Linguistics*, 3(2), 91–102. <https://doi.org/10.21460/saga.2022.32.134>
- Subekti, A. S. (2025). Gender differences in academic motivation among rural Indonesian junior high school students. *Revista Brasileira de Educação Do Campo (Brazilian Journal of Rural Education)*, 10(e19634), 1–19. <https://doi.org/10.70860/ufnt.rbec.e19634>
- Subekti, A. S., & Kurniawati, L. A. (2022). Indonesian high school English teachers' attitudes towards the use of instructional technology during the Covid-19 pandemic. *Jurnal IQRA': Kajian Ilmu Pendidikan*, 7(2), 76–88. <https://doi.org/https://doi.org/10.25217/ji.v7i2.2394>
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese, and Iranian learners of English: A comparative study. In Z. Dornyei & E. Ushioda (Eds.), *Motivation, language identity, and the L2 Self* (pp. 66–97). Multilingual Matters.