International Journal of Multidisciplinary Research and Development
www.allsubjectjournal.com
E-ISSN: 2349-4182
Received: 22-09-2026, Accepted: 22-10-2026, Published: 06-11-2026
Volume 13, Issue 4, 2026, Page No. 110-116

AI-scaffolded FIVES instruction and literacy transformation in Indonesian Elementary Schools: What integrated data reveal about learning mechanisms

Sri Dewi Nirmala
Master of Elementary Education, Graduate School, Universitas Terbuka, Indonesia
DOI: https://doi.org/10.66856/ijmrd.2026.13.4.13415

Abstract
Many reading literacy programs in primary schools raise student scores, yet few explain why those gains occur. This study examines how the FIVES model (Fact, Inference, Vocabulary, Experiences, Summary) supported by artificial intelligence (AI) works for Grade 5 students in twelve primary schools across five Indonesian provinces. A convergent mixed methods design was used. Quantitative data came from a reading literacy pretest and posttest (N=172), a self-regulated learning questionnaire (N=277), and a reflective learning instrument (N=208). Qualitative data came from classroom observations of seven teachers, interviews with nine teachers, and interviews with 27 students. Both strands were analysed separately and then brought together in a joint display. Mean literacy scores rose from 54.11 to 81.31 (N-Gain 0.584, Cohen's d 2.46), and the spread of scores between students narrowed. The integration produced five findings. AI helped students see the reasoning process at the Inference stage. The FIVES structure offered different entry points for students of different abilities. The way teachers guided AI use mattered more than the availability of AI itself. Reflective skills proved learnable for all students. AI also offered a safe space for students who had been reluctant to ask questions. These findings suggest that strengthening teachers' pedagogical capacity to guide AI use should be a priority in technology-supported literacy programs in primary schools.

Keywords: FIVES model, reading literacy, artificial intelligence, self-regulated learning, reflective learning, mixed methods, primary school

Introduction

Reading remains a serious challenge for Indonesian students. In PISA 2022, Indonesia scored 359 in reading, far below the OECD average of 476 (OECD, 2023) [20]. The greatest difficulties appear in tasks that ask students to draw conclusions, interpret, and evaluate information, rather than simply locate what is written. At the primary level, Abidin *et al*. (2025) [1] found that reading motivation and interest have a clear effect on students' literacy, while support from teachers and schools has done little to raise that motivation. In other words, the way reading is taught in classrooms still needs attention. The problem also reaches beyond language lessons. Prabowo *et al*. (2023) [20] showed that the reading literacy of Grade 5 students in Indonesia and Malaysia shapes their comprehension in mathematics, so weak reading skills carry over into other subjects.

Many programs have tried to address this, and a good number report encouraging score gains. A meta-analysis by Silverman *et al*. (2025) [24] on technology interventions for literacy in primary schools found positive effects, but the size of those effects depended heavily on program features and on how the programs were carried out. This is where the problem lies. Score gains tell us that something has changed, but they do not tell us what happened in the classroom, who benefited most, or why results in one school differed from those in another.

The FIVES model developed by Shea and Roberts (2016) [23] offers a staged framework for reading comprehension. At the Fact stage, students locate information stated explicitly in the text. At the Inference stage, they draw conclusions from implied information. The Vocabulary stage helps students understand word meanings in the context of a sentence. The Experiences stage invites students to connect the text with their own lives, and the Summary stage trains them to restate the main ideas in their own words. The sequence moves from the easiest task to the most cognitively demanding one. The Inference stage deserves particular attention. A meta-analysis of 56 studies by Rice and Wijekumar (2024) [21] showed that explicit inference instruction has a moderate effect on both inference skills and general reading comprehension (g = 0.495).

Generative AI opens new possibilities for reading instruction, but it also carries risks. Kasneci *et al*. (2023) and Lo (2023) [14] describe the potential of AI to offer more personalised support, while warning about inaccurate answers and dependence. The evidence of benefit is strong. A meta-analysis of 69 experimental studies by Deng *et al*. (2025) [10] found that ChatGPT improved academic performance, motivation, and higher-order thinking tendencies, but it also reduced students' mental effort. Wang and Fan (2025) [29] reported a large effect of ChatGPT on learning performance (g = 0.867) and a moderate effect on higher-order thinking (g = 0.457), noting that the size of the effect depended on the role ChatGPT was given in the lesson. Experimental work has also confirmed the risks. Bastani *et al*. (2025) [4] found that students with unrestricted access to GPT-4 did better during practice but worse once that access was removed, while an AI tutor with pedagogical safeguards largely prevented this harm. Darvishi *et al*. (2024) [9] observed that students tended to rely on AI help rather than learn from it, and a systematic review by Zhai *et al*. (2024) [31] showed that over-reliance on AI dialogue systems can weaken students' thinking skills. A review of 118 studies by Yan *et al*. (2024) [30] likewise noted that many uses of large language models in education are not yet mature in practical or ethical terms. Taken together, these studies suggest that what matters is not whether AI is present, but the instructional structure that

frames its use. The five-stage FIVES framework has the potential to provide that structure.

Teachers play a key role here. The review by Celik *et al*. (2022) [6] showed that the benefits of AI in the classroom depend largely on teachers' pedagogical decisions. Kim (2024) [15] mapped several forms of teacher and AI collaboration and found that teachers continue to steer the direction of learning. A qualitative study by Tripathi *et al*. (2025) [27] showed that teachers have begun to distinguish between cheating with AI and thinking with AI, and they train students to edit AI output critically. To do this well, teachers need digital competencies that include an understanding of AI, which Ng *et al*. (2023) notes are still largely missing from existing teacher competency frameworks.

Beyond comprehension, good instruction also builds self-regulated learning, that is, students' ability to plan, monitor, and evaluate their own learning (Schunk & Greene, 2018) [22]. A meta-analysis by Achuthan (2025) [2] found that AI-based interventions had a large effect on overall self-regulated learning and a moderate effect on cognitive and metacognitive regulation. The ability to reflect on one's learning is the most visible form of this metacognitive awareness. Neshaei *et al*. (2025) [17] found that large language models can help learners write deeper reflections, and the review by Tsakeni *et al*. (2025) [24] concluded that AI is most useful when it is designed to support metacognition rather than replace it. In the Indonesian primary school context, Arianto and Hanif (2024) [3] showed that metacognitive strategies and self-regulated learning both predict students' self-efficacy and problem-solving skills.

Although evidence on AI in learning keeps growing, most studies have been carried out in secondary and higher education, and most rely on a single type of data. Studies in Indonesian primary schools that combine test results, questionnaires, observations, and interviews in a single analysis are still rare. Yet explaining the mechanism behind an outcome requires bringing quantitative and qualitative data together on purpose, not merely presenting them side by side (Creswell & Plano Clark, 2018) [8]. This study addresses that gap through three research questions.

1. What can be learned about how the AI-supported FIVES model works when quantitative and qualitative data are analysed together?
2. Where do the two types of data confirm each other, complement each other, or reveal new understanding?
3. What mechanisms explain the success of AI-supported FIVES and the differences in results between schools?

Research Method

a. Research Design

This study used a convergent mixed methods design (Creswell & Plano Clark, 2018; Creswell & Creswell, 2023) [7, 8]. Quantitative and qualitative data were collected during the same period, analysed separately, and then brought together to see whether they confirmed each other, complemented each other, or diverged. Quantitative data were collected through a one-group pretest and posttest design for reading literacy and a descriptive survey for self-regulated learning and reflection. Qualitative data were collected through classroom observations, teacher interviews, and student interviews.

b. Research Sites and Participants

The study took place in twelve primary schools in five provinces: West Java (eight schools) and Banten, South Kalimantan, Bali, and Bangka Belitung (one school each). This spread of locations made it possible to check whether findings held across different cultural, linguistic, and social settings. Participants were Grade 5 students and their class teachers. The teachers came from partner schools and were also enrolled as students at Universitas Terbuka while teaching at those schools. Their participation was requested through a formal letter. The number of participants varies across data types because data completeness differed from school to school. A summary is given in Table 1.

Table 1: Participants by Data Type

Data Type	Participants	Schools	Instrument
Reading literacy pretest and posttest	172 students	6	FIVES-based literacy test, 15 items
Self-regulated learning	277 students	10	Four-dimension questionnaire, 1-4 scale
Reflective learning	208 students	9	Reflective journal with a four-level rubric
Classroom observation	7 teachers	7	Seven-aspect observation sheet
Teacher interviews	9 teachers	9	Semi-structured interviews
Student interviews	27 students	9	34 sessions; high, medium, and low ability

c. Instructional Procedure

Teachers taught reading through the five FIVES stages. The AI tool used was ChatGPT. At each stage, AI served as a support tool, for example to show how a conclusion can be drawn from two pieces of information in a text, to explain the meaning of a word in its sentence context, or to respond to a summary written by a student. At the end of each lesson, students wrote a reflective journal guided by prompts that became more demanding from week to week.

d. Instruments

The reading literacy test consisted of 15 items built on the FIVES framework: two Fact items, three Inference items, two Vocabulary items, three open-ended Experiences items, three Summary items, and two critical thinking items combining Inference and Summary. The maximum raw score of 39 was converted to a 100-point scale. The self-regulated learning questionnaire measured four dimensions: setting goals and planning, carrying out and monitoring learning, seeking help when facing difficulty, and evaluating one's own learning outcomes (Schunk & Greene, 2018) [22]. Each statement was answered on a four-point scale from Never to Always.

The quality of student reflection was rated as a four-level rubric adapted from Hatton and Smith (1995) [11]. At Level 1, students only describe what they did. At Level 2, they begin to name what felt difficult. At Level 3, they link difficulties to learning strategies. At Level 4, they analyse

the causes of their difficulties and plan how to improve. The observation sheet assessed seven aspects: the sequence of FIVES stages, the quality of teacher questions, the way AI was used, students' cognitive engagement, adjustment of support, assessment during learning, and classroom atmosphere. The teacher and student interview guides were organised in three time frames: the situation before the intervention, experiences during it, and changes noticed afterwards.

e. Data Analysis

Quantitative data were analysed using descriptive statistics and paired-samples t-tests. Gains were measured with the normalised gain (N-Gain), categorised as low (< 0.3), medium (0.3 to 0.7), and high (> 0.7), and effect size was calculated with Cohen's d. Interview data and observation notes were analysed thematically following the six phases described by Braun and Clarke (2022) [5], from familiarisation and coding to developing, reviewing, and writing up themes. Trustworthiness was supported through source triangulation, member checking, and peer debriefing. Integration took place after both analyses were complete. Quantitative and qualitative findings were placed side by side in a joint display, and each pair was judged as confirming, complementing, or diverging (Creswell & Plano Clark, 2018) [8]. This process produced five integration points, which are discussed in the results.

f. Research Ethics

The study received ethical approval from LPPM Universitas Terbuka (No. B/685/UN31.LPPM/PT.01.04/2026) and a research permit from the Graduate School of Universitas Terbuka (No. 647/UN31.SPS/PT.01.04/2026). Teachers gave written consent. Parents were informed through the class WhatsApp groups managed by the teachers and gave permission, and students agreed before being interviewed. School names were replaced with regional codes, teachers were coded GT-01 to GT-09, and no student identities are reported. The data are stored by the lead researcher and can be accessed only by members of the research team.

Results and Discussions

a. Reading Literacy, Self-Regulated Learning, and Reflection

Among the 172 students with complete test data, the mean score rose from 54.11 on the pretest to 81.31 on the posttest, a gain of 27.20 points. The mean N-Gain of 0.584 falls in the medium category, but the effect size was very large (d = 2.46). The paired t-test showed a significant difference, $t(171) = 32.260$, $p < .001$. The narrowing of scores is also worth noting. The standard deviation fell from 12.61 to 8.34, which means the distance between weaker and stronger students shrank. In total, 93.6% of students reached a medium or high N-Gain. Results by location are shown in Table 2.

Table 2: Reading Literacy Results Before and After Instruction

Location	N	Pretest Mean	Posttest Mean	Gain	N-Gain	Category	t	p
Bandung Regency, West Java	25	60.41	84.31	23.90	0.592	Moderate	10.155	<.001
Garut, West Java	27	51.00	75.97	24.97	0.508	Moderate	13.724	<.001
Cimahi, West Java	25	47.08	82.64	35.56	0.678	Moderate	17.427	<.001
Bandung Regency, West Java	42	49.19	81.53	32.34	0.630	Moderate	27.987	<.001
Lebak, Banten	25	63.72	82.67	18.94	0.538	Moderate	10.668	<.001
Bandung City, West Java	28	56.59	81.04	24.44	0.536	Moderate	9.850	<.001
Total / mean	172	54.11	81.31	27.20	0.584	Moderate	32.260	<.001

An analysis by FIVES stage was carried out in one school in Garut, which had the most complete item-level records. Inference showed the largest gain (+38.3 points) even though it had the lowest starting score (40.7). The other stages also improved steadily: Fact (+24.7), Vocabulary (+26.3), Experiences (+21.9), and Summary (+23.8). Because these figures come from one school, they are treated as early signals and were examined further through qualitative data.

Self-regulated learning data from 277 students showed a mean of 74.0%, with 60.0% of students in the high to very high category. Help-seeking when facing difficulty had the highest mean (76.2%), while evaluating one's own learning outcomes was the lowest (71.1%). Reflection data from 208 students showed a mean attainment of 71.5%. Most students (59.1%) were at Level 3, and 12.5% had reached Level 4. The weekly journal records showed a steady upward trend. This profile is summarised in Table 3.

Table 3: Profile of Self-Regulated Learning and Reflection Levels

Dimension	Attainment	Category	Notes
Setting goals and planning	74.8%	High	Developed steadily during instruction
Carrying out and monitoring learning	73.9%	High	Students actively checked their own understanding
Seeking help when facing difficulty	76.2%	High	Highest dimension; linked to AI as a learning partner
Evaluating own learning outcomes	71.1%	Medium-high	Lowest dimension; most demanding of reflection
Reflection Level 3	59.1% of students	Developing	Largest group; rose from week to week
Reflection Level 4	12.5% of students	Advanced	Notable for a first experience of reflective writing

b. Patterns from Observations and Interviews

Interviews with nine teachers revealed four patterns that appeared across schools and provinces. First, all teachers noticed a change in their role, from mostly explaining to mostly guiding how students think. This change felt awkward at first but was later seen as positive. Second, seven of the nine teachers named Inference as the hardest stage to facilitate, because this was where students most often got stuck. Even so, when students got past it, teachers described it as the most memorable moment of learning. Third, teachers in the best-performing schools routinely asked students to answer on their own before opening the AI, then compared and discussed the differences. Fourth, several teachers observed that as students' vocabulary

improved, their anxiety about reading also seemed to ease, including at other stages.

The 27 student interviews yielded five patterns that were also consistent across schools and ability levels. Students described AI as a discussion partner that can be wrong and needs checking, not as a place to copy answers. They also reported the same order of difficulty: Fact was easiest because the answers are in the text, while Inference and Summary were hardest. Most students said they now try to answer on their own before asking, something they rarely did before. On the reflective journals, students from different schools told a similar story, moving from not knowing what to write to being able to name which parts they understood and which they did not. Finally, each ability group had its own favourite stage. Lower-ability students felt confident at Fact and Vocabulary, middle-ability students were most engaged at Vocabulary, and higher-ability students enjoyed the challenge of Inference most.

c. Data Integration: Five Findings on How AI-Supported FIVES Works

When the quantitative and qualitative data were brought together, five integration points emerged. Table 4 summarises what each type of data showed and the understanding that came from combining them.

Table 4: Joint Display of Quantitative and Qualitative Findings

Integration Point	Quantitative Data	Qualitative Data	Emerging Understanding
1. Inference as the turning point	Largest gain (+38.3) from the lowest starting score (40.7)	7 of 9 teachers: AI showed how to reason; students called Inference the hardest and most satisfying stage	AI makes reasoning visible so students can imitate and learn it
2. The gap narrowed	SD fell from 12.61 to 8.34; 93.6% of students at medium to high N-Gain	Each ability group found the FIVES stage that suited it best	The FIVES structure offers different entry points without differentiated materials
3. How teachers used AI shaped results	Two schools with lower N-Gain	Both schools had lower observation scores on AI use; confirmed by teacher interviews	How teachers guide AI use matters more than access to AI
4. Reflection as a skill	71.6% of students at Levels 3-4; rising each week	Students across schools described the same path, from confused to capable	Reflection can be learned by all students given structure and time
5. Daring to ask	Help-seeking was the highest dimension (76.2%)	Lower-ability students felt more comfortable asking AI because classmates could not see	AI becomes a safe space to ask without fear of being judged

1. Inference: the hardest stage with the largest gain

The test data showed that Inference had the lowest starting score and the largest gain. This is striking, but it does not explain how the hardest stage grew the most. The explanation came from the teachers. Seven of the nine described the same thing. Students who could already find facts easily often stalled when asked to draw a conclusion, not because they had not read the text, but because they did not yet know how to reason from the information available. AI helped at this point by showing the reasoning steps, for example how two pieces of information in a text can be connected to reach a conclusion. Students watched this example, tried to imitate it, questioned it, and gradually began to use it on their own. Lower-ability students even felt that a stage which once seemed impossible was now within reach.

This finding is in line with the meta-analysis by Rice and Wijekumar (2024) [21], which showed that inference skills can be taught explicitly and that doing so improves general reading comprehension. The difference here is that the modelling of reasoning came from the teacher and also from an AI that students could consult as often as they needed. This benefit appeared only because AI was placed inside a clear structure. Bastani *et al*. (2025) [4] have shown that AI without pedagogical safeguards can harm learning. The Inference stage in FIVES provides such a safeguard, because students are asked to reason first rather than simply receive an answer. This also agrees with the review by Tsakeni *et al*. (2025) [24], which found AI to be most useful when it supports students' thinking. Gains at the Vocabulary stage probably also supported progress in Inference. The meta-analysis by Sterpin *et al*. (2021) [25] found that vocabulary knowledge has a moderate effect on children's ability to make inferences. Students who understand words in context have more material to reason with.

2. The gap narrowed because each student found a foothold

The drop in standard deviation from 12.61 to 8.34 is easy to overlook, yet it matters. Besides raising the class average, the intervention brought the students who had been furthest behind closer to those who were furthest ahead. The interview data explain how this happened. Lower-ability students built confidence at the Fact stage, where answers can be found directly in the text, and then felt ready to try the next stage. Middle-ability students were most engaged at Vocabulary, where AI helped them grasp word meanings quickly without taking away the chance to think. Higher-ability students were challenged most at Inference and Summary.

This narrowing did not come from making all students the same. It came from the five FIVES stages offering different entry points for different abilities. Teachers did not need to prepare separate materials for each group. The pattern matches the finding by Thomas *et al*. (2024) [26] that lower-achieving students tend to benefit more from tutoring that combines teachers and AI. Jauhiainen and Garagorry Guerra (2024) [13] showed that AI can adapt materials to the knowledge level of primary school pupils. The present study adds that this kind of adaptation can also arise from the structure of instruction itself, without depending on AI to personalise content.

3. How teachers guided AI shaped the results

All schools showed significant gains, but two schools had lower N-Gains than the rest. The observation data showed that these two schools also scored lower on the way AI was used. The difference was not whether AI was used, since every school used it, but how teachers placed AI within learning activities. In the best-performing schools, students first wrote their own answers, then compared them with the AI response and discussed the differences in class. In the

schools with lower results, students more often read the AI response straight away and copied it. Teacher interviews confirmed this pattern. Teachers in the best-performing schools deliberately encouraged students to treat AI as a discussion partner that can make mistakes, while teachers in other schools admitted how hard it was to keep students thinking for themselves when AI answers were so easy to get.

This conclusion could not have been drawn from any single type of data. Test scores only recorded differences between schools, observations only gave implementation ratings, and interviews only offered accounts. Only when all three were combined did it become clear that the way teachers guide AI use is the most decisive factor. This finding agrees with Kim (2024) [15], who stressed that teachers should keep their pedagogical role when working with AI, and with Tripathi *et al*. (2025) [27], who found that teachers have begun teaching students to edit AI answers critically. The tendency of students to rely on AI described by Darvishi *et al*. (2024) [9] also explains what happened in the schools with lower results: without the teacher's direction, AI easily becomes a shortcut. Wang and Fan (2025) [29] likewise found that the effect of ChatGPT on higher-order thinking varies with the role it is given. Celik *et al*. (2022) [6] noted early on that the benefits of AI in the classroom depend on teachers' pedagogical decisions, and the data in this study show what that looks like in practice.

4. Reflection is a skill that can be learned

By the end of the intervention, 71.6% of students had reached Levels 3 and 4, and attainment kept rising from week to week. This is a meaningful result, given that most students had never written a structured reflective journal before. The interviews revealed the process behind the numbers. At first, the guiding prompts felt abstract and hard to answer. Over time, because the FIVES stages gave them a clear and memorable learning experience to look back on, students began to say precisely which parts they understood, and which still confused them. Some even started writing their reflections before being asked.

Bringing these two sources together shows that reflective ability is not a talent that only some students have. It can grow in every student when they are given structure and time. This agrees with Neshaei *et al*. (2025) [17], who showed that AI support can deepen reflective writing when AI models how to reflect rather than simply correcting. The fact that evaluating one's own learning outcomes remained the lowest dimension in Table 3 suggests that this ability needs more time to mature. That is to be expected, since self-evaluation is the part of self-regulated learning that demands the most metacognitive awareness (Arianto & Hanif, 2024; Schunk & Greene, 2018). Achuthan (2025) [2, 3, 22] also found that the effect of AI on metacognitive regulation is smaller than its effect on self-regulated learning as a whole.

5. AI provided a safe space to ask questions

Help-seeking had the highest mean of all dimensions (76.2%). The simplest explanation is that AI is always available, while a teacher may not have time to answer every question. The qualitative data point to a deeper reason. Lower-ability students from West Java, Banten, and Bangka Belitung described a similar experience. Before the intervention, they often held back from asking the teacher or their classmates because they were afraid of looking incapable. Asking AI felt safer because their classmates could not see, so they were willing to ask again and again until they truly understood. Their questions also changed, moving from simply asking for answers to testing their understanding and comparing different views.

This pattern is consistent with the finding by Hou *et al*. (2024) [12] that students feel more comfortable asking AI than asking in front of others, because they do not worry about being a burden or being judged. The present study shows that primary school students, especially those with lower ability, have a similar experience. This sense of safety is linked to reading motivation, which Abidin *et al*. (2025) [1] identified as a strong influence on the literacy of Indonesian students. Even so, this safe space needs to be managed so that it does not turn into dependence (Zhai *et al*., 2024) [31]. For that reason, the teacher's role discussed in the third finding remains the necessary counterweight.

d. Implications for Practice

The findings carry three practical implications. First, teacher training should focus on how to guide students in using AI, rather than mainly on how to operate the application (Ng *et al*., 2023) [20]. The same technology produces different results in different teachers' hands. A simple habit, such as asking students to answer on their own before opening the AI, clearly separated stronger from weaker results. Second, the FIVES framework can be used in mixed-ability classes without grouping students separately or preparing different materials, because its five stages already provide points of engagement for students at different levels. Third, reflection should be a regular part of every lesson, on a par with the five FIVES stages, because this skill develops gradually and its benefits go beyond test scores. For schools and local education authorities, this suggests that funding for teachers' pedagogical capacity is as important as, and perhaps more important than, funding for devices and applications.

e. Limitations

This study has several limitations. The one-group pretest and posttest design without a comparison group limits causal claims, since part of the gain may reflect students' natural development or the enthusiasm that comes with a new program. Deng *et al*. (2025) [10] also caution that novelty effects often make the impact of AI look larger in short-term studies. The stage-by-stage FIVES analysis was carried out in only one school, the one with the most complete data. The integration points were identified after each strand had been analysed, rather than planned from the start. The number of schools outside West Java was small, and the student interview sample did not fully reflect diversity in gender, regional language, and family economic background. Future studies should use a comparison group, plan integration points from the outset, and include more schools outside Java.

Conclusion

This study shows that the question of why a teaching approach works can only be answered well when quantitative and qualitative data are brought together. The AI-supported FIVES model raised the reading literacy of Grade 5 students substantially, from a mean of 54.11 to 81.31, and at the same time narrowed the gap between weaker and stronger students. Students' self-regulated

learning was in the high category, and most students were able to write reflections that linked their learning difficulties to the strategies they used.

In answer to the first research question, the integrated analysis showed that AI-supported FIVES works through five mechanisms: AI makes reasoning at the Inference stage visible, the FIVES structure offers different entry points for different abilities, teachers guide how AI is used, reflection is a skill all students can learn, and AI provides a safe space for students who are reluctant to ask. In answer to the second question, the two types of data confirmed each other on Inference and reflection, complemented each other on the narrowing gap, and revealed new understanding on the sense of safety in asking questions, which a questionnaire alone would not have shown. In answer to the third question, differences in results between schools were best explained by how teachers placed AI in learning activities, not by whether AI was available.

Theoretically, this study broadens our understanding of the role of AI in reading instruction. AI should not be seen only as a source of answers or a personalisation tool. Within a clear structure and under the teacher's guidance, AI can serve as a model of reasoning that students can imitate and as a learning partner that eases the social pressure students feel when they do not yet understand. The findings also confirm that the structure of instruction, here the five stages of FIVES, acts as a counterweight that keeps AI from becoming a shortcut.

Practically, several recommendations follow. Teachers should get students used to answering on their own before opening the AI, then comparing and discussing the differences. Principals and supervisors should include the way AI use is guided in their classroom supervision, rather than only checking whether AI is being used. Local education authorities and teacher education institutions, including graduate programs for in-service teachers, should provide training that focuses on the pedagogical use of AI. Reflection should also become a regular part of reading lessons, because its benefits grow over time and go beyond higher test scores.

Future research should use designs with a comparison group, measure effects at each FIVES stage in more schools, and follow students over a longer period to make sure that the gains are not simply a novelty effect. Studies should also include more schools outside Java and look more closely at how the sense of safety in asking AI can be kept from turning into dependence. For education researchers in Indonesia, this study offers an example of how test, questionnaire, observation, and interview data can be analysed together to explain learning mechanisms as well as to report outcomes.

References

1. Abidin Y, Rakhmayanti F, Undayasari D. Factors affecting literacy ability of elementary school students in Indonesia. Mimbar Sekolah Dasar, 2025, 12(2). https://doi.org/10.53400/mimbar-sd.v12i2.86486
2. Achuthan K. Artificial intelligence and learner autonomy: A meta-analysis of self-regulated and self-directed learning. Frontiers in Education,2025:10:1738751. https://doi.org/10.3389/feduc.2025.1738751
3. Arianto F, Hanif M. Evaluating metacognitive strategies and self-regulated learning to predict primary school students' self-efficacy and problem-solving skills in science learning. Journal of Pedagogical Research,2024:8(3):301-319. https://doi.org/10.33902/JPR.202428575
4. Bastani H, Bastani O, Sungu A, Ge H, Kabakcı Ö, Mariman R. Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences,2025:122(26):e2422633122. https://doi.org/10.1073/pnas.2422633122
5. Braun V, Clarke V. Thematic analysis: A practical guide. SAGE Publications, 2022.
6. Celik I, Dindar M, Muukkonen H, Järvelä S. The promises and challenges of artificial intelligence for teachers: A systematic review of research. TechTrends,2022:66(4):616-630. https://doi.org/10.1007/s11528-022-00715-y
7. Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). SAGE Publications, 2023.
8. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research (3rd ed.). SAGE Publications, 2018.
9. Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impact of AI assistance on student agency. Computers & Education,2024:210:104967. https://doi.org/10.1016/j.compedu.2023.104967
10. Deng R, Jiang M, Yu X, Lu Y, Liu S. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education,2025:227:105224. https://doi.org/10.1016/j.compedu.2024.105224
11. Hatton N, Smith D. Reflection in teacher education: Towards definition and implementation. Teaching and Teacher Education,1995:11(1):33-49. https://doi.org/10.1016/0742-051X(94)00012-U
12. Hou I, Mettille S, Li Z, Man O, Zastudil C, MacNeil S. The effects of generative AI on computing students' help-seeking preferences. In Proceedings of the 26th Australasian Computing Education Conference (ACE '24). ACM, 2024. https://doi.org/10.1145/3636243.3636248
13. Jauhiainen JS, Garagorry Guerra A. Generative AI and education: Dynamic personalization of pupils' school learning material with ChatGPT. Frontiers in Education,2024:9:1288723. https://doi.org/10.3389/feduc.2024.1288723
14. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, *et al*. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences,2023:103:102274. https://doi.org/10.1016/j.lindif.2023.102274
15. Kim J. Types of teacher-AI collaboration in K-12 classroom instruction: Chinese teachers' perspective. Education and Information Technologies,2024:29(13):17433-17465. https://doi.org/10.1007/s10639-024-12523-3
16. Lo CK. What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences,2023:13(4):410. https://doi.org/10.3390/educsci13040410
17. Neshaei SP, Mejia-Domenzain P, Davis RL, Käser T. Metacognition meets AI: Empowering reflective

writing with large language models. British Journal of Educational Technology,2025:56(5):1864-1896. https://doi.org/10.1111/bjet.13601
18. Ng DTK, Leung JKL, Su J, Ng RCW, Chu SKW. Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. Educational Technology Research and Development,2023:71(1):137-161. https://doi.org/10.1007/s11423-023-10203-6
19. OECD. PISA 2022 results (Volume I): The state of learning and equity in education. OECD Publishing, 2023. https://doi.org/10.1787/53f23881-en
20. Prabowo A, Suparman S, Li CS, Janan D, Damayanti TD. The effect of reading literacy to mathematics comprehension of elementary school students in Indonesia and Malaysia. International Journal of Evaluation and Research in Education,2023:12(1):546-554. https://doi.org/10.11591/ijere.v12i1.25714
21. Rice M, Wijekumar K. Inference skills for reading: A meta-analysis of instructional practices. Journal of Educational Psychology,2024:116(4):569-589. https://doi.org/10.1037/edu0000855
22. Schunk DH, Greene JA. Handbook of self-regulation of learning and performance (2nd ed.). Routledge, 2018. https://doi.org/10.4324/9781315697048
23. Shea M, Roberts N. The FIVES strategy for reading comprehension. Learning Sciences International, 2016.
24. Silverman RD, Keane K, Darling-Hammond E, Khanna S. The effects of educational technology interventions on literacy in elementary school: A meta-analysis. Review of Educational Research,2025:95(5):972-1012. https://doi.org/10.3102/00346543241261073
25. Sterpin LF, Ortiz SS, Formoso J, Barreyro JP. The role of vocabulary knowledge on inference generation: A meta-analysis. Psychology of Language and Communication,2021:25(1):168-193. https://doi.org/10.2478/plc-2021-0008
26. Thomas DR, Lin J, Gatz E, Gurung A, Gupta S, Norberg K, *et al*. Improving student learning with hybrid human-AI tutoring: A three-study quasi-experimental investigation. In Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24). ACM, 2024. https://doi.org/10.1145/3636555.3636896
27. Tripathi T, Sharma SR, Singh V, Bhargava P, Raj C. Teaching and learning with AI: A qualitative study on K-12 teachers' use and engagement with artificial intelligence. Frontiers in Education,2025:10:1651217. https://doi.org/10.3389/feduc.2025.1651217
28. Tsakeni M, Nwafor SC, Mosia M, Egara FO. Mapping the scaffolding of metacognition and learning by AI tools in STEM classrooms: A bibliometric-systematic review approach (2005-2025). Journal of Intelligence,2025:13(11):148. https://doi.org/10.3390/jintelligence13110148
29. Wang J, Fan W. The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. Humanities and Social Sciences Communications,2025:12:621. https://doi.org/10.1057/s41599-025-04787-y
30. Yan L, Sha L, Zhao L, Li Y, Martinez-Maldonado R, Chen G, *et al*. Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology,2024:55(1):90-112. https://doi.org/10.1111/bjet.13370
31. Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments,2024:11:28. https://doi.org/10.1186/s40561-024-00316-7