Pendas : Jurnal Ilmiah Pendidikan Dasar,
ISSN Cetak : 2477-2143 ISSN Online : 2548-6950
Volume 11 Nomor 03, September 2026

THE EFFECTS OF AI-ASSISTED READING, GRAPHIC ORGANIZERS, AND CONVENTIONAL QUESTION-AND-ANSWER INSTRUCTION ON GRADE 5 STUDENTS' READING COMPREHENSION

Tiffany Eunike Patricia[1],Pratama Irwin Talenta[2]
[1,2] Universitas Nasional Karangturi
[1]euniketiffany45@gmail.com, [2]pratama.irwin@unkartur.ac.id

ABSTRACT

This preliminary classroom study examines the effects of three reading-instruction conditions - AI-assisted reading, a graphic-organizer strategy, and conventional question-and-answer instruction - on Grade 5 students' reading comprehension. The study was planned as a quantitative comparative investigation involving three intact classes. The supplied dataset listed 67 student records and contained 65 numeric baseline scores. However, post-test scores were available only for Class 5B and Class 5C, while the Class 5A post-test column was absent. Baseline scores were statistically comparable across the three classes, $F(2, 62) = 0.006$, $p = .994$. Among complete cases, Class 5B improved from $M = 73.44$ to $M = 88.13$, $t(15) = 2.59$, $p = .021$, with a medium within-group effect ($d_z = 0.65$). Class 5C improved from $M = 77.22$ to $M = 83.89$, but the change was not statistically significant, $t(17) = 0.92$, $p = .372$. The available post-test means for Class 5B and Class 5C did not differ significantly, Welch's $t = 0.71$, $p = .481$. These findings suggest that structured instructional support may facilitate reading performance, but they do not yet allow a definitive conclusion about the AI-assisted condition. Completion and verification of the Class 5A post-test data are required before the original three-group research question can be fully tested.

Keywords: *AI-assisted reading; reading comprehension; graphic organizer; elementary EFL; classroom research*

A. Introduction

English is a key international language that plays a major role in education and communication around the world. Students need to learn English because many academic materials such as books, journals, and digital resources are written in English. As Harmer (2007) points out, learning English helps students improve their communication abilities and access a broader range of knowledge. Hence, developing strong English skills is vital for students' academic progress. Among the four main language skills, reading is considered very important because it allows students to gain information and understand different topics.

English reading is an important part of school learning because students increasingly encounter information through textbooks, digital resources, and other academic materials written in English. Reading comprehension is more than recognizing individual words. It requires readers to identify main

ideas, connect details, interpret vocabulary in context, draw inferences, and build an integrated understanding of a text (Grabe & Stoller, 2011; Snow, 2002). For elementary learners of English as a foreign language, these processes can be difficult because limited vocabulary and developing reading strategies increase the cognitive demands of a task (Woolley, 2011).

Classroom instruction therefore needs to provide sufficient support while encouraging students to process a text actively. Brown (2001) argues that reading instruction should develop purposeful skills such as identifying key information, predicting meaning, locating details, and interpreting unfamiliar words. From a sociocultural perspective, this support can be understood as scaffolding: temporary assistance that enables learners to complete tasks that would otherwise be difficult and that is gradually reduced as competence develops (Vygotsky, 1978).

Conventional question-and-answer activities remain widely used because they are easy to organize and allow teachers to check whether students understand a passage. Nevertheless, when the activity is limited to reading a text and answering questions, students may participate passively or rely on surface-level recall. Graphic organizers offer a more visible form of scaffolding by helping learners arrange main ideas, supporting details, sequences, causes, and relationships. This type of external representation can guide students to notice text structure and summarize information more systematically.

Some students have difficulty with texts due to limited vocabulary and a lack of good reading strategies. In addition, traditional reading activities often make students less interested because they only read the text and answer questions without any type of interactive support. Because of this, students may become passive and show little interest in reading activities. Woolley (2011) suggests that students need suitable reading strategies to help them understand texts more effectively. Digital technology has expanded the range of reading supports available in the classroom. Technology-enhanced language learning can provide quick access to explanations, multimedia resources, and interactive practice (Chapelle, 2003). Artificial intelligence extends these possibilities by generating explanations, offering immediate feedback, simplifying difficult language, supporting vocabulary learning, and responding to individual questions. Luckin et al. (2016) describe educational AI as potentially useful for personalizing learning support, although the instructional value of any tool depends on how it is integrated into teaching.

Recent studies have reported promising outcomes for AI-supported reading. Chea and Xiao (2024) discuss both the potential benefits and risks of AI tools for academic English reading, while Wang et al. (2024) found that AI-powered tools can support vocabulary acquisition in EFL instruction. Interactive chatbot activities have also been associated with reading interest and engagement (Liu et al., 2024), and AI tools may enhance enjoyment and motivation

when learners use them within purposeful tasks (Yuan & Liu, 2024). These findings suggest that AI can function as an additional scaffold rather than a replacement for teacher guidance.

The evidence is not yet conclusive, particularly for elementary -school contexts. Results can be affected by students' digital literacy, the design of prompts, teacher mediation, novelty effects, and the quality of automated feedback. In addition, many studies focus on students' perceptions rather than direct classroom comparisons of learning outcomes. Research that compares AI-assisted reading with other structured teaching approaches is therefore needed to clarify whether the technology provides benefits beyond well-designed non-digital scaffolding.

The original proposal for the present study compares three intact Grade 5 classes: Class 5A receiving AI-assisted reading, Class 5B using a graphic organizer, and Class 5C receiving conventional question-and-answer instruction. All classes were intended to read the same material and complete a common 20-item reading comprehension test. The test covered main ideas, specific information, vocabulary meaning, inference, and moral value.

The present manuscript reports a preliminary analysis of the supplied score file. It addresses three questions: (1) Were the classes comparable on the available baseline scores? (2) Did scores change from baseline to post-test in the classes with complete paired data? and (3) Did the available post-test scores differ between Class 5B and Class 5C? Because the dataset does not yet contain Class 5A post-test scores, the central comparison involving AI-assisted reading remains provisional.

B. Research Method

The study used a quantitative, quasi-experimental classroom design with three intact classes and no random assignment. The proposal described the study as a post-test-only comparison. The supplied spreadsheet, however, contained a baseline or initial score for each class and post-test data for only two classes. The statistical analysis was therefore adapted to the observed data structure and is reported as a preliminary analysis rather than as a complete test of the planned three-group design.

Participants and Instructional Conditions

The spreadsheet listed 67 student records: 23 in Class 5A, 21 in Class 5B, and 23 in Class 5C. Numeric baseline scores were available for 65 students. The proposal assigned the instructional conditions shown in Table 1. Student names were present in the source file but were not reproduced in the analysis or manuscript.

Table 1. Planned instructional conditions

Class	Planned instructional condition	Available score structure
5A	AI-assisted reading	Baseline: n = 22; Post-

		test: n = 0
5B	Graphic organizer	Baseline: n = 21; Post-test: n = 16
5C	Conventional question-and-answer	Baseline: n = 22; Post-test: n = 19

Note. *Counts refer to numeric scores in the supplied spreadsheet. Blank cells and entries marked X were treated as missing. A numeric score of 0 was retained as an observed score.*

Instrument

Reading comprehension was assessed using a 20-item multiple-choice test. According to the proposal, the items measured identification of the main idea, retrieval of specific information, vocabulary meaning, inference, and moral value. Raw totals were converted to a 0-100 scale. The proposal states that an English teacher would review the items for content validity. Item-level responses were not included in the spreadsheet; consequently, internal-consistency reliability could not be estimated from the available data.

Procedure

Each class was intended to receive the same reading text while experiencing a different instructional strategy. Class 5A used AI-assisted reading support, Class 5B used a graphic organizer, and Class 5C used conventional question-and-answer activities. After instruction, all students were expected to complete the same post-test. The available files did not specify the number of instructional meetings, treatment duration, AI platform, prompt sequence, or exact graphic-organizer format, so these details should be added before journal submission.

Data Preparation and Analysis

Data were read directly from the three worksheet tabs. Numeric values in the NILAI or PRE-TEST column were harmonized as baseline scores. Numeric values in the POST-TEST column were treated as post-test scores. Blank cells and nonnumeric X entries were coded as missing; no missing score was replaced or imputed. Descriptive statistics included the sample size, mean, standard deviation, minimum, and maximum. Baseline equivalence was examined using a one-way analysis of variance (ANOVA), with Levene's test used to inspect homogeneity of variance. Paired-samples t-tests were conducted for Classes 5B and 5C using students with both baseline and post-test scores. A Welch independent-samples t-test compared the available post-test scores of Classes 5B and 5C. The significance level was set at .05. Cohen's d_z was reported for paired comparisons, and Hedges' g was reported for the between-group comparison.

Ethical Considerations

The proposal indicates that the activities were part of regular classroom practice, that participation would not disadvantage students academically, and that the data would be used only for research. The present manuscript reports only aggregated scores and does not

disclose student identities. Formal documentation of school permission and parental or guardian consent should be described according to the requirements of the intended journal and institution.

C. Research Results

Data Availability and Baseline Performance

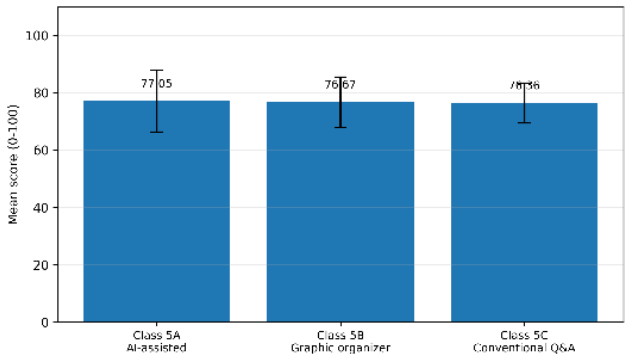
The three classes had nearly identical baseline means. Class 5A had a mean of 77.05, Class 5B had a mean of 76.67, and Class 5C had a mean of 76.36. The baseline distributions were relatively wide, particularly in Class 5A, where scores ranged from 25 to 100. Table 2 summarizes the descriptive statistics.

Table 2. Descriptive statistics for available baseline scores

Class	n	Mean	SD	Min	Max
5A - AI-assisted	22	77.05	25.62	25	100
5B - Graphic organizer	21	76.67	20.51	35	95
5C - Conventional Q&A	22	76.36	16.77	45	100

Levene's test was not significant, W = 0.70, p = .501, indicating no evidence of unequal baseline variances. A one-way ANOVA likewise showed no statistically significant difference among the baseline means, F(2, 62) = 0.006, p = .994. Thus, the available baseline data do not indicate an initial achievement advantage for any class.

Figure 1. Mean baseline reading scores across the three classes (error bars show 95% confidence intervals)



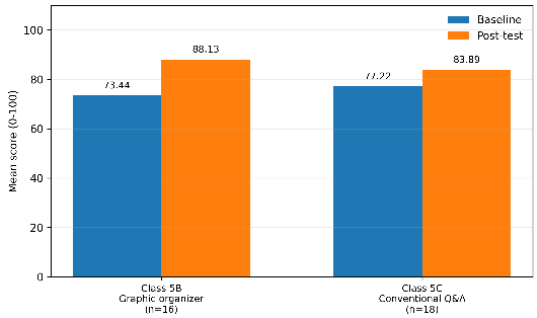
Within-Class Changes for Complete Cases

Paired baseline and post-test scores were available for 16 students in Class 5B and 18 students in Class 5C. As shown in Table 3, the mean gain in Class 5B was 14.69 points. This improvement was statistically significant, t(15) = 2.59, p = .021, and the within-group effect was medium, d_z = 0.65. Class 5C improved by an average of 6.67 points, but the change was not statistically significant, t(17) = 0.92, p = .372, d_z = 0.22.

Table 3. Complete-case baseline-to-post-test changes

Class	Paired n	Baseline M (SD)	Post-test M (SD)	Mean gain	t (df)	p	d_z
5B - Graphic organizer	16	73.44 (22.04)	88.13 (29.49)	14.69	2.59 (15)	.021	0.65
5C - Conventional Q&A	18	77.22 (18.17)	83.89 (26.82)	6.67	0.92 (17)	.372	0.22

Figure 2. Complete-case baseline and post-test means for Classes 5B and 5C



Available Post-Test Comparison

Using all numeric post-test scores, Class 5B had a mean of 88.13 (SD = 29.49, n = 16), while Class 5C had a mean of 81.05 (SD = 28.85, n = 19). Welch's independent-samples t-test showed that the 7.07-point mean difference was not statistically significant, t = 0.71, p = .481. The estimated between-group effect was small, Hedges' g = 0.24. Therefore, the available data do not establish that the graphic-organizer condition produced higher post-test performance than conventional question-and-answer instruction.

The post-test distributions also showed substantial clustering at the maximum score: 13 of 16 numeric Class 5B post-test scores and 11 of 19 numeric Class 5C post-test scores were 100. This ceiling concentration reduces the test's ability to distinguish among higher-performing students and contributes to wide standard deviations because a smaller number of students received low scores.

Discussion

Interpretation of Baseline Equivalence

The near-identical baseline means are important because they suggest that the classes entered the study with comparable average reading performance. This finding supports the intended comparative design and reduces concern that one class had a substantial initial advantage. Nevertheless, the classes were intact rather than randomly assigned, so equivalence on a single total score does not rule out differences in motivation, attendance, digital familiarity, or prior reading experience.

Structured Scaffolding and Reading Performance

The available complete-case data show a statistically significant improvement in Class 5B, the class assigned to the graphic-organizer condition. Graphic organizers may support comprehension by making relationships among ideas visible and by directing learners to identify text structure, main ideas, and supporting details. This interpretation is consistent with scaffolding theory: an external structure can reduce unnecessary processing demands and help learners allocate attention to meaning construction (Vygotsky, 1978; Woolley, 2011).

Class 5C also improved, although the mean gain was smaller and statistically nonsignificant. Conventional question-and-answer instruction can still promote learning when teachers provide explanation and feedback. The result therefore should not be interpreted as evidence that conventional instruction is ineffective. Rather, the available data suggest that an additional organizing scaffold may have supported more consistent improvement in Class 5B.

Implications for AI-Assisted Reading

The original research purpose was to evaluate AI-assisted reading in relation to non-AI approaches. The theoretical rationale remains strong. AI tools can provide immediate vocabulary explanations, simplified paraphrases, questions, summaries, and feedback that respond to learners' needs. Research in EFL contexts has linked AI-supported reading with vocabulary development, reading engagement, and motivation (Chea & Xiao, 2024; Liu et al., 2024; Wang et al., 2024; Yuan & Liu, 2024). These functions can operate as contingent scaffolds, particularly when a teacher controls the learning goals and verifies the accuracy of AI-generated content.

However, the current score file does not include a Class 5A post-test measure. Consequently, the findings cannot show whether AI-assisted reading improved Class 5A performance or whether it was more effective than the graphic-organizer and conventional conditions. Any claim that AI caused superior comprehension would exceed the available evidence. The Class 5A post -test scores should be entered and checked before the planned one-way post-test ANOVA or a stronger pre-test-adjusted analysis is conducted.

Methodological Implications

Several data-management issues should be addressed in the final study. First, all classes should use the same column labels and scoring rules. Second, attendance and missing-test reasons should be documented separately from test scores. Third, the study should retain item-level responses so that reliability and item difficulty can be evaluated. Fourth, the large number of perfect post-test scores suggests that some items may have been too easy or that the test did not provide enough difficult items to discriminate among high-performing students. A revised test should include a broader range of item difficulty while continuing to measure main ideas, details, vocabulary, inference, and moral value.

Once complete data are available, analysis of covariance (ANCOVA)

could compare post-test scores while controlling for baseline performance. Alternatively, gain scores could be compared across all three groups if their assumptions are satisfied. Reporting confidence intervals and effect sizes alongside p-values would make the practical meaning of the differences clearer. A follow-up qualitative component, such as student interviews or observation notes, would also help explain how learners used the AI tool or graphic organizer during reading.

Limitations

This preliminary analysis has several limitations. The Class 5A post-test data were unavailable; post-test participation was incomplete in Classes 5B and 5C; treatment duration and implementation fidelity were not documented in the supplied files; the sample consisted of intact classes from one setting; and only total test scores were available. The ceiling concentration and the presence of missing values further limit statistical precision. These limitations mean that the results should be treated as an evidence-based working draft rather than as a final causal evaluation of the three instructional approaches.

D. Consulcions

The available baseline scores show that the three Grade 5 classes began with almost identical average reading comprehension. Among the two classes with paired post-test data, the graphic-organizer class demonstrated a statistically significant and moderately sized improvement, while the conventional question-and-answer class showed a smaller, nonsignificant gain. The available post-test difference between those two classes was not statistically significant.

The dataset does not yet support a conclusion about AI-assisted reading because the Class 5A post-test scores are absent. The most important next step is therefore to complete and verify the Class 5A post-test data, document the instructional procedure in detail, and rerun the planned three-group analysis. With complete data and consistent administration, the study can make a useful contribution to understanding whether AI-based scaffolding provides advantages beyond graphic organizers and conventional classroom questioning in elementary EFL reading.

REFERENCES

Jurnal Ilmiah:

Brown, H. D. (2001). Teaching by principles: An interactive approach to language pedagogy (2nd ed.). Longman.

Chapelle, C. A. (2003). English language learning and technology: Lectures on applied linguistics in the age of information and communication technology. John Benjamins.

Chea, P., & Xiao, Y. (2024). Artificial intelligence in higher education: The power and damage of AI-assisted tools on academic English reading skills. Journal of General Education and Humanities, 3(3), 287-306. https://doi.org/10.58421/gehu.v3i3.242

Grabe, W., & Stoller, F. L. (2011). Teaching and researching reading (2nd ed.). Routledge.

Harmer, J. (2007). The practice of English language teaching (4th ed.). Pearson Longman.

Liu, C.-C., Chen, W.-J., Lo, F., Chang, C. -H., & Lin, H.-M. (2024). Teachable Q&A agent: The effect of chatbot training by students on reading interest and engagement. Journal of Educational Computing Research. https://doi.org/10.1177/0735633 1241236467

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson.

Snow, C. E. (2002). Reading for understanding: Toward an R&D program in reading comprehension. RAND Corporation.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J., & Wang, L. (2024). Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction. IEEE Access. https://doi.org/10.1109/ACCESS. 2024.3446657

Woolley, G. (2011). Reading comprehension: Assisting children with learning difficulties. Springer.

Yuan, L., & Liu, X. (2024). The effect of artificial intelligence tools on EFL learners' engagement, enjoyment, and motivation. Computers in Human Behavior, 162, 108474. https://doi.org/10.1016/j.chb.202 4.108474