



ORIGINAL ARTICLE | OPEN ACCESS

AI-Assisted Pedagogy and English Language Learning Gains among School Students: A Quasi-Experimental Study

Summiya Fasih* and Shahbaz Khan

School of Education, Beaconhouse National University, Lahore, Pakistan

*Corresponding author: summish11@gmail.com

Received: 09 September 2026 | Revised: 02 October 2026 | Accepted: 25 October 2026 | Published: 28 October 2026 |
Volume 3, Issue 2 (2026), Pages: 124-134 DOI: <https://doi.org/10.66558/jsshi.2026.2.8>

Keywords

Artificial intelligence
Generative AI
English language
learning
AI-assisted pedagogy
ChatGPT
Grok 4
Quasi-experimental
study

Abstract

Artificial intelligence (AI) is increasingly integrated into language education, yet empirical evidence on generative AI-assisted pedagogy remains limited in school settings. This study examined English language learning gains among school students receiving AI-assisted instruction using Grok 4 and ChatGPT compared with traditional instruction. A quantitative quasi-experimental pre-test–post-test design involved 73 students from Grades 5 and 6 in a Pakistani school. Two experimental groups received teacher-guided AI-assisted instruction using Grok 4 (Grade 5, $n = 18$) and ChatGPT (Grade 6, $n = 16$), while a combined control group ($n = 39$) received traditional instruction. The two-week intervention incorporated curriculum-aligned activities, interactive learning, feedback, and differentiated tasks. Learning gains were calculated as differences between post-test and baseline scores. One-way ANOVA revealed significant differences among groups, $F(2, 70) = 7.78, p = .001, \eta^2 = .182$. Tukey post-hoc comparisons showed significantly greater gains in both AI-assisted groups than in the control group, with no statistically significant difference between Grok 4 and ChatGPT. The findings suggest that teacher-guided AI-assisted pedagogy was associated with improved English achievement scores compared with traditional instruction. However, the short intervention, intact-group design, and assessment-related limitations warrant cautious interpretation.

Citation Fasih, S and Khan, S. (2026). AI-Assisted Pedagogy and English Language Learning Gains among School Students: A Quasi-Experimental Study. *Journal of Social Sciences, Humanities and Innovation*, 3 (2), 124-134 <https://doi.org/10.66558/jsshi.2026.2.8>

Introduction

English language proficiency is an important educational outcome for school students because English is widely used for academic communication, access to information, and participation in broader educational and professional contexts. However, developing English proficiency can be challenging when students have limited opportunities for meaningful language practice, individualized feedback,

and active classroom interaction. These challenges have increased interest in technology-supported approaches that can supplement conventional classroom instruction.

Artificial intelligence (AI) has emerged as one such approach. AI-supported educational technologies can provide learners with opportunities for practice, feedback, interaction, and adaptive learning. In English language education, research has identified potential applications of

License & Copyright

© The Author(s) 2026. Published by Scientific Scholarly Publishing.

This is an open-access article distributed under the terms of the **Creative Commons Attribution 4.0 International License (CC BY 4.0)**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are properly cited.

AI in areas including speaking, writing, reading, pedagogy, and learner self-regulation. At the same time, reported challenges include technical problems, limitations in AI-generated responses, concerns about standardization of language, and the need for appropriate teacher guidance.

The development of generative AI has expanded these possibilities. Large language model-based systems can respond to natural-language prompts, generate examples, provide explanations, support conversational practice, and offer feedback. In language education, ChatGPT in particular has been discussed as a tool for supporting teaching and learning, while researchers have also emphasized the need for teachers and learners to develop appropriate competencies for its ethical and effective use.

Empirical evidence generally indicates positive potential for AI-supported English language learning. A meta-analysis by Xu and Wang (2024), based on 40 empirical studies involving 3,290 participants and 55 effect sizes, found an overall positive effect of AI on English language learning achievement. A separate meta-analysis by Wu (2024) similarly examined AI-assisted second-language learning while considering contextual, instructional, and social-emotional moderators. These findings provide a broader empirical basis for investigating AI-assisted language instruction while also indicating that effects may depend on the conditions under which AI is implemented.

Law (2024), in a scoping literature review of 41 publications on generative artificial intelligence in language teaching and learning, identified potential benefits alongside important gaps in the emerging evidence base. The review highlighted the need for further empirical studies evaluating the effectiveness of generative AI tools, targeted interventions addressing specific language skills, and greater attention to the responsible integration of these technologies. These findings reinforce the importance of examining generative AI through classroom-based interventions rather than relying solely on its potential applications. In this context, the present study investigates learning gains following teacher-guided AI-assisted English instruction among Grade 5 and Grade 6 students in Pakistan.

Despite this growing evidence, important gaps remain. A recent systematic review of 144 empirical studies published during the first two years of generative AI research in language learning and teaching found that 86.7% of studies were conducted in higher education and that English as a foreign language context accounted for 86.1% of the literature. The review also identified limited research in K–

12 settings and relatively less attention to speaking, listening, and reading compared with writing. Similarly, Crompton et al. (2024), in a systematic review of 42 studies on AI and English language teaching and learning, identified a need for research across broader educational contexts and highlighted both affordances and challenges associated with AI use.

The present study addresses part of this gap by examining AI-assisted English pedagogy among school students in Pakistan. Rather than examining students' perceptions of AI or simply describing possible uses of generative AI, the study investigates measurable learning gains following a curriculum-aligned instructional intervention. It also compares two generative AI platforms, Grok 4 and ChatGPT, with traditional English instruction.

The study was guided by the following research question:

What is the effect of AI-assisted pedagogy on students' English language learning gains compared with traditional English instruction?

The study also examined whether learning gains differed between the Grok 4 and ChatGPT groups.

The study contributes empirical evidence from a school-level context that remains less represented in the rapidly expanding generative AI language-learning literature. It also provides evidence concerning the use of teacher-guided AI rather than positioning AI as a replacement for classroom instruction.

2. Literature Review

2.1 AI-Assisted Pedagogy in English Language Education

AI-assisted pedagogy refers in this study to the use of AI technologies as instructional supports within teacher-planned learning activities. The teacher remains responsible for lesson planning, curriculum alignment, classroom management, and evaluation, while AI is used to supplement explanation, practice, interaction, feedback, and differentiated learning opportunities.

The distinction between AI as pedagogical support and AI as an autonomous teacher is important. AI systems can generate responses rapidly and provide learners with additional opportunities for interaction, but their outputs may contain inaccuracies or may not always correspond to the instructional level or curricular requirements of learners. Crompton et al. (2024) identified both affordances and challenges in AI-supported English language teaching, including opportunities related to

speaking, writing, reading, pedagogy, and self-regulation, alongside technological and pedagogical limitations. Consequently, effective classroom integration requires the teacher to structure how AI is used. In the present study, AI was incorporated into teacher-prepared lesson plans rather than being treated as an independent instructional agent.

2.2 Generative AI and English Language Learning

Generative AI has introduced new forms of interaction into language learning. Unlike earlier computer-assisted language learning systems that often relied on predefined responses, generative AI can produce context-sensitive text and participate in extended conversational exchanges.

The growing empirical literature suggests that generative AI can support several dimensions of language learning. Li et al. (2025), in a systematic review of 144 empirical studies, identified six major research areas: perceptions and attitudes, psychological and cognitive aspects, teaching and learning strategies, language skills development, writing and feedback, and implementation and integration. However, the concentration of studies in higher education demonstrates that findings from university populations cannot automatically be assumed to apply to younger school students.

Similarly, the meta-analysis by Xu and Wang (2024) provides evidence of an overall positive relationship between AI use and English language learning achievement. Nevertheless, meta-analytic findings represent aggregated evidence across diverse technologies, populations, instructional designs, and durations. A classroom-based study involving younger learners therefore remains useful for examining how AI-assisted pedagogy functions in a specific school context.

2.3 ChatGPT in English Language Learning

ChatGPT has received substantial attention in language education because of its capacity to generate explanations, examples, questions, dialogue, and written language. Kohnke et al. (2023) describe ChatGPT as having potential applications in language teaching and learning while also emphasizing limitations and the competencies required for ethical and effective use.

For English language learners, conversational interaction can provide additional opportunities for language practice. AI-generated examples can also be adapted to students' learning activities, while feedback can be used as a starting point for revision and discussion. However, these

potential advantages depend on how the technology is embedded within pedagogy. The teacher's role remains important in checking responses, selecting appropriate activities, and ensuring that AI-supported tasks remain aligned with learning objectives.

Empirical evidence further illustrates the potential of ChatGPT as a language-learning support tool. Mahapatra (2024) conducted a mixed-methods intervention study examining ChatGPT as a formative feedback tool for undergraduate English as a Second Language (ESL) students. The study reported statistically significant improvements in academic writing performance following the intervention and highlighted the potential value of ChatGPT-generated feedback in supporting students' writing development. Similarly, Song and Song (2023), in a mixed-methods study involving 50 English as a Foreign Language (EFL) students, found that learners receiving ChatGPT-assisted instruction demonstrated greater improvements in academic writing skills and writing motivation than students receiving traditional instruction. These studies provide empirical support for exploring ChatGPT as a supplementary resource for language practice and feedback. However, because both studies focused on academic writing among higher-education learners, their findings cannot be directly generalized to younger school students. The present study extends this line of inquiry by examining teacher-guided ChatGPT-assisted English instruction among Grade 6 students in a school setting in Pakistan.

The present study therefore did not conceptualize ChatGPT as an independent teacher. Instead, ChatGPT was incorporated into structured lessons designed and monitored by the teacher.

2.4 Grok and Emerging Generative AI Tools

Research on generative AI in language education has expanded rapidly, but literature has not developed equally across platforms. ChatGPT has received considerably more scholarly attention than newer or less extensively studied generative AI systems. Li et al. (2025) noted the need for greater diversity in research concerning specific AI tools and educational contexts.

Grok 4 represents an emerging generative AI platform whose pedagogical application in school-level English instruction has received substantially less empirical attention than ChatGPT. This creates a useful comparative dimension for the present study. Rather than assuming that one platform is inherently superior, the study examines

whether statistically distinguishable differences emerged between students who received Grok 4-assisted instruction and those who received ChatGPT-assisted instruction.

2.5 Constructivist and Social Constructivist Perspectives

The study was informed by constructivist and social constructivist perspectives. Constructivist learning theory views learners as active participants who construct knowledge through engagement with tasks and experiences rather than simply receiving information from an instructor (Piaget & Inhelder, 1972). Social constructivism further emphasizes the importance of interaction, dialogue, social participation, and guided support in learning (Vygotsky, 1978).

These perspectives provide a useful lens for AI-assisted pedagogy because generative AI can function as an additional source of interaction and feedback when incorporated appropriately into classroom activities. From a social constructivist perspective, learning does not result simply from access to a technological tool; it emerges through meaningful interaction and guided participation.

Sharpley (2023) similarly conceptualizes educational interaction with generative AI as a process of conversation and exploration rather than simply a sequence of prompts and answers. He emphasizes the possibility of social interaction involving learners, teachers, and AI systems while also highlighting the importance of ethical constraints and the continuing role of human educators.

In the present study, therefore, AI was positioned as a supplementary learning resource within teacher-guided activities. The theoretical framework supports the expectation that opportunities for interaction, scaffolding, feedback, and active participation may contribute to learning when AI is integrated into structured instruction.

2.6 Research Gap

The literature establishes a growing empirical basis for AI-assisted English language learning; however, several gaps remain relevant to the present investigation. First, research on generative AI in language education remains concentrated in higher education, limiting the extent to which existing findings can be generalized to younger school students. Li et al. (2025), in a systematic review of 144 empirical studies, found that 86.7% of the reviewed studies were conducted in higher education. Similarly, Wang et al. (2025), in a scoping review of 43 empirical studies, identified a strong concentration on higher-education contexts and writing-related applications. These

findings indicate the need for further classroom-based research involving school-level learners.

Second, although previous reviews report promising applications of generative AI in language learning, studies differ in their instructional approaches, technologies, participant groups, and outcome measures. Law (2024) highlighted the need for additional empirical evaluations of generative AI tools and targeted interventions addressing specific language-learning needs. Further context-specific studies are therefore required to examine learning outcomes under clearly described instructional conditions.

Third, comparatively limited attention has been given to generative AI platforms beyond ChatGPT. Wang et al. (2025) identified an overreliance on ChatGPT in the empirical literature and recommended investigating a wider range of AI models. In addition, direct comparisons between different generative AI platforms and traditional instruction remain important for understanding how AI-assisted pedagogy performs in particular educational settings.

The present study addresses these gaps by examining English language learning gains among Grade 5 and Grade 6 students in a Pakistani school who received teacher-guided Grok 4-assisted instruction, ChatGPT-assisted instruction, or traditional English instruction. By comparing learning gains across these three groups, the study contributes school-level evidence from a context that remains less represented in the generative AI language-learning literature. However, because the two AI platforms were used with different grade levels, the comparison between platforms is interpreted cautiously rather than as a controlled test of platform superiority.

3. Theoretical Framework

The study integrates Constructivist Learning Theory and Social Constructivist Theory.

Constructivism emphasizes active learning and the learner's role in constructing knowledge through engagement with experiences and tasks (Piaget & Inhelder, 1972). Within AI-assisted English instruction, this perspective is reflected in activities that require students to interact with examples, practice language, respond to questions, revise their work, and apply newly learned concepts.

Social constructivism places greater emphasis on learning through interaction and guided participation. Vygotsky (1978) proposed that learning is shaped by social interaction and that appropriate support can help learners

perform tasks that they could not initially complete independently. Concepts such as the zone of proximal development and scaffolding are particularly relevant to instructional activities involving teacher guidance and feedback.

The present study applies these perspectives by conceptualizing AI as a supplementary interactive resource rather than an autonomous instructor. AI-generated examples, explanations, questions, and feedback can provide additional learning opportunities, while teachers organize, monitor, and contextualize those opportunities. This approach is also consistent with Sharples's (2023) view of generative AI as potentially supporting social processes of conversation and exploration in education. The theoretical framework therefore proposes that AI-assisted pedagogy may support English language development through several interconnected instructional processes:

AI-assisted pedagogy → interaction and active participation → feedback and scaffolding → differentiated practice → English language learning gains

The framework does not assume that AI automatically produces learning. Rather, it proposes that learning benefits depend on the pedagogical use of technology.

4. Methodology

4.1 Research Design

A quantitative quasi-experimental pre-test–post-test control-group design was used. Random assignments of individual students was not feasible because the study was conducted within existing school sections. Instead, intact classroom groups were used.

Three groups were included in the final analysis: a Grade 5 Grok 4 group, a Grade 6 ChatGPT group, and a combined control group consisting of Grade 5 and Grade 6 students who received traditional English instruction.

The primary outcome was the learning gain score, calculated by subtracting the baseline score from the post-test score:

Gain score = Post-test score – Baseline score

This approach allowed the analysis to focus on change in English achievement rather than relying solely on post-test performance.

4.2 Participants and Sampling

The study included 73 students from Grades 5 and 6 in a Pakistani school.

Because intact sections were used, convenience sampling was employed. The school administration assigned available Grade 5 and Grade 6 sections that met the study requirements. Students were therefore not individually randomized to treatment conditions

(Table 1).

Table 1: Distribution of Participants by Grade Level and Instructional Group

Group	Grade	Instructional condition	n	Percentage
Grok 4	5	AI-assisted pedagogy	18	24.7%
ChatGPT	6	AI-assisted pedagogy	16	21.9%
Control	5 and 6	Traditional instruction	39	53.4%
Total			73	100%

The unequal group sizes reflect the composition of the available intact sections.

4.3 Baseline and Post-Test Measures

The baseline measure was derived from students' most recent English final-term examination scores, while a researcher-developed post-test was administered following the instructional intervention. Both assessments were scored out of 20 marks and were aligned with the English curriculum relevant to the respective grade levels. Separate grade-appropriate post-test versions were developed for Grade 5 and Grade 6 students. Although the assessments were designed to reflect relevant curricular learning objectives, they were not identical instruments, and equivalence in assessment difficulty was not independently established. Consequently, differences between baseline and post-test scores were interpreted as observed achievement-score changes rather than definitive measures of language proficiency development.

The assessment covered five domains:

- 1. reading comprehension;
- 2. writing;
- 3. grammatical competence;
- 4. vocabulary; and
- 5. speaking.

Writing and speaking components were evaluated using established scoring criteria and rubrics. The assessment materials and instructional plans were reviewed by the research supervisor, an English subject teacher, and the school coordinator to establish content relevance and age-

level appropriateness.

The study used the overall achievement score for statistical analysis. Although the assessment incorporated multiple English language domains, the present analysis does **not** claim separate statistically significant effects for each individual skill.

4.4 Intervention

The intervention was implemented over a two-week instructional period. The experimental groups received the same school curriculum and were taught through teacher-guided lesson plans that incorporated AI-supported activities. Grade 5 students used Grok 4, while Grade 6 students used ChatGPT. The control students received traditional English instruction.

The AI-assisted lesson plans incorporated student-centered approaches including guided practice, think-pair-share, cooperative learning, peer learning, differentiated tasks, and opportunities for interaction and feedback. AI was therefore used as a supplementary resource within teacher-directed classroom instruction.

During a grammar lesson on the present indefinite tense, students engaged with AI-generated sentence examples and guided practice questions related to the target grammatical structure. Students then practised forming affirmative sentences, discussed their responses with peers, and applied the grammatical structure in their own sentences. The teacher monitored the activity, clarified misunderstandings, and checked the accuracy and appropriateness of AI-generated material. This example illustrates how AI was positioned as a supplementary resource within teacher-guided instruction rather than as a replacement for the teacher.

The lesson plans were reviewed by the English subject teacher and school coordinator for appropriateness to the students' level and curriculum. The researcher and supervisor also reviewed the plans before implementation.

The instructional duration was constrained by the school calendar. During the intervention period, the school schedule was affected by changes in the number of instructional days, online instruction on Friday, and an extended Eid break. These conditions limited the duration of exposure to the intervention and are considered when interpreting the findings.

4.5 Data Analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS).

Descriptive statistics were calculated for gain scores, including sample size, mean, standard deviation, minimum, and maximum values.

Before conducting the one-way analysis of variance (ANOVA), the homogeneity of variance assumption was examined using Levene's test. A one-way ANOVA was then conducted to determine whether mean gain scores differed among the Grok 4, ChatGPT, and control groups. Because the ANOVA tests whether at least one group differs from another, Tukey's honestly significant difference (HSD) procedure was used for post-hoc pairwise comparisons.

Statistical significance was evaluated at $\alpha = .05$. Eta squared (η^2) was calculated to estimate the proportion of variance in gain scores associated with instructional group membership. Given the quasi-experimental design, the use of intact classroom groups, and differences in grade-level composition, the statistical findings were interpreted as evidence of differences in observed achievement-score changes rather than definitive causal effects of the instructional interventions.

4.6 Ethical Considerations

The study was conducted in accordance with the ethical considerations applicable to school-based educational research involving minors. The research activities were undertaken with permission from the participating school administration and under the supervision of the research team. AI-assisted instructional activities were incorporated into teacher-guided classroom lessons, with the teacher responsible for monitoring the appropriateness of learning materials and student interactions. Student assessment data were handled confidentially, and findings were reported in aggregated form without identifying individual participants.

5. Results

5.1 Descriptive Statistics

Table 2 presents the descriptive statistics for the gain scores.

Table 2: Descriptive Statistics of English Language Learning Gain Scores

Group	n	Mean	SD	Minimum	Maximum
Grok 4	18	1.83	1.15	−1	4
ChatGPT	16	2.13	2.19	−2	7

Group	n	Mean	SD	Minimum	Maximum
Control	39	0.13	2.21	-3	6
Total	73	0.99	2.18	-3	7

The ChatGPT group recorded the highest mean gain score ($M = 2.13$, $SD = 2.19$), followed by the Grok 4 group ($M = 1.83$, $SD = 1.15$). The control group showed a substantially smaller mean gain ($M = 0.13$, $SD = 2.21$).

The descriptive results therefore indicate greater average improvement in both AI-assisted groups than in the traditional instruction group.

5.2 Homogeneity of Variance

Levene's test was used to examine whether the variances of the three groups were sufficiently homogeneous for one-way ANOVA.

The result was not statistically significant, $F(2, 70) = 2.19$, $p = .120$. Therefore, the assumption of homogeneity of variance was considered acceptable for the ANOVA.

5.3 One-Way Analysis of Variance

A one-way ANOVA was conducted on the gain scores to determine whether the three instructional groups differed significantly (Table 3).

Table 3: One-Way ANOVA of English Language Learning Gain Scores

Source	SS	df	MS	F	p
Between groups	62.377	2	31.189	7.780	.001
Within groups	280.60				
Total	342.98				

The ANOVA revealed a statistically significant difference in gain scores among the three groups, $F(2, 70) = 7.78$, $p = .001$.

The calculated eta squared was $\eta^2 = .182$, indicating that approximately 18.2% of the variance in gain scores was associated with differences among the instructional groups

in this sample.

The significant omnibus result indicates that at least one group differed from another. It does not, by itself, establish that every pair of groups differed. Therefore, Tukey HSD post-hoc comparisons were examined.

5.4 Post-Hoc Comparisons

Tukey HSD comparisons showed that the Grok 4 group differed significantly from the control group, with the Grok 4 group demonstrating greater learning gains. The ChatGPT group also differed significantly from the control group, with the ChatGPT group demonstrating greater learning gains.

The comparison between the Grok 4 and ChatGPT groups was not statistically significant. Thus, the pattern of results can be summarized as follows:

- Grok 4-assisted instruction produced significantly greater gain scores than traditional instruction.
- ChatGPT-assisted instruction produced significantly greater gain scores than traditional instruction.
- The difference between Grok 4 and ChatGPT was not statistically significant.

These findings indicate that the statistically significant overall ANOVA was primarily attributable to the differences between each AI-assisted group and the control group rather than to a statistically significant difference between the two AI platforms.

The Tukey HSD analysis identified two homogeneous subsets based on mean gain scores. The traditional instruction group ($M = 0.13$, $n = 39$) formed the first subset, while the Grok 4-assisted group ($M = 1.83$, $n = 18$) and ChatGPT-assisted group ($M = 2.13$, $n = 16$) formed the second subset. Both AI-assisted groups demonstrated statistically significantly higher mean gain scores than the traditional instruction group. However, the difference between the Grok 4 and ChatGPT groups was not statistically significant ($p > .05$). These results indicate that the overall difference among instructional groups was primarily associated with the higher observed gains in the AI-assisted groups relative to the control group.

6. Discussion

6.1 AI-Assisted Pedagogy and English Language Learning Gains

The principal finding of this study was that students receiving AI-assisted English instruction demonstrated greater learning gains than students receiving traditional instruction under the conditions of the study. Both AI

groups had higher mean gain scores than the control group, and the post-hoc analysis confirmed statistically significant differences between each AI-assisted group and the control group.

This finding is consistent with the broader empirical literature suggesting that AI-supported approaches can contribute positively to English language learning. Xu and Wang's (2024) meta-analysis of 40 empirical studies found an overall positive effect of AI on English language learning achievement. The present study adds to this evidence by examining younger school students in a classroom-based Pakistani context.

The findings should nevertheless be interpreted in relation to the instructional conditions. The intervention did not involve students simply using an AI application independently. Instead, AI was embedded within teacher-prepared lessons involving guided practice, interaction, differentiated tasks, peer learning, and feedback. Consequently, the findings support the use of AI-assisted pedagogy under the conditions studied rather than establishing that access to AI alone causes improved English achievement.

This distinction is important because AI technologies can provide useful instructional affordances while also presenting limitations. Crompton et al. (2024) identified potential benefits of AI in speaking, writing, reading, pedagogy, and self-regulation, but also reported challenges including technical problems and limitations in AI capabilities. Therefore, pedagogical design and teacher oversight remain important components of classroom implementation.

The findings can also be considered in light of more recent reviews of generative AI in language education. Law (2024) highlighted the potential applications of generative AI in language teaching and learning while emphasizing the need for further empirical investigation. Similarly, Lee et al. (2026) reviewed research on generative AI in language classrooms, providing a broader context for examining how these tools are incorporated into language instruction. The present study contributes to this developing evidence base by examining learning gains following teacher-guided AI-assisted English instruction among Grade 5 and Grade 6 students in Pakistan. The higher mean gain scores in both AI-assisted groups suggest that integrating generative AI into structured classroom activities may support English learning under the conditions examined. However, given the short intervention period, the use of intact classroom groups, and the study's single-school context, the findings should be

interpreted as preliminary classroom-based evidence rather than as proof that generative AI is consistently more effective than traditional instruction across settings.

6.2 Grok 4 and ChatGPT

The ChatGPT group had the highest mean gain score ($M = 2.13$), while the Grok 4 group had a mean gain of $M = 1.83$. However, the difference between these two groups was not statistically significant.

This result should not be interpreted as evidence that the two platforms are identical in all respects. Rather, within this sample, intervention period, assessment structure, and instructional design, there was insufficient statistical evidence to conclude that learning gains differed between the two AI groups.

The finding is relevant because much of the language-learning literature has focused on ChatGPT and comparatively less attention has been given to other generative AI platforms. The recent systematic review by Li et al. (2025) highlights the need for more diverse research concerning generative AI tools and educational settings.

The present findings therefore provide preliminary comparative evidence involving Grok 4 and ChatGPT at the school level. However, because the two platforms were implemented with different grade levels, the comparison cannot be interpreted as a controlled head-to-head platform experiment. Grade level is potentially confounded with platform assignment. Consequently, future studies should compare the platforms within the same grade, curriculum, duration, and assessment conditions and, where feasible, use random assignment or matched groups.

6.3 Interpretation Through Constructivist and Social Constructivist Perspectives

The findings can be interpreted through the study's constructivist and social constructivist framework.

From a constructivist perspective, learning occurs through active engagement with tasks and experiences. The AI-assisted lessons required students to participate in activities rather than merely receive information. Students interacted with examples, practiced language, responded to questions, received feedback, and applied concepts in guided activities.

From a social constructivist perspective, the role of interaction and scaffolding is particularly relevant. Vygotsky (1978) emphasized the importance of social interaction and guided support in learning. In the present

study, AI created an additional opportunity for interaction, while the teacher remained responsible for organizing and monitoring the learning process.

Sharpley (2023) similarly proposes that generative AI can be understood as part of a social process of conversation and exploration in educational settings. The present findings are compatible with this interpretation because the intervention incorporated AI within classroom pedagogy rather than separating AI use from teacher-mediated instruction.

The findings should not, however, be treated as direct empirical confirmation of the theoretical mechanisms. The study measured English learning gains rather than independently measuring scaffolding, interaction, engagement, or knowledge construction. Therefore, the theoretical framework provides an interpretive lens for understanding the intervention rather than a direct test of each theoretical construct.

6.4 Contribution to Existing Knowledge

The study makes three principal contributions.

First, it provides empirical evidence from a school-level context. Recent reviews show that generative AI language research remains concentrated in higher education, making school-based evidence particularly relevant.

Second, the study examines AI-assisted pedagogy through a classroom-based intervention rather than relying exclusively on learner perceptions. The use of gain scores and an experimental-control comparison provides quantitative evidence concerning changes in achievement.

Third, the study includes two generative AI platforms. While the design does not permit a definitive platform superiority claim, it provides preliminary evidence that both Grok 4 and

ChatGPT-assisted instruction were associated with greater gains than the traditional instruction group, while their gains did not differ significantly from one another.

6.5 Practical Implications

The findings have implications for English language teachers and schools considering the integration of generative AI.

First, AI can be incorporated as a supplementary instructional resource rather than replacing teachers. Teachers can use AI to generate examples, create practice activities, support differentiated tasks, facilitate language interaction, and provide additional

feedback opportunities.

Second, teacher oversight remains essential. Generative AI can produce inaccurate, inappropriate, or poorly calibrated responses. Teachers should therefore review AI-generated material and ensure that activities correspond to curriculum objectives and students' developmental levels. Kohnke et al. (2023) similarly emphasizes the competencies required for effective and ethical use of ChatGPT in language education.

Third, schools should consider teacher professional development when introducing AI-supported pedagogy. Effective implementation requires teachers to understand not only how to operate AI tools but also how to integrate them into sound pedagogical activities.

Finally, AI should be integrated around clear learning objectives rather than introduced simply because technology is available. The present study's intervention combined AI with established instructional approaches such as guided practice, cooperative learning, peer interaction, and differentiated instruction.

6.6 Limitations

Several limitations should be considered.

First, the study used a convenience sample and intact classroom sections rather than random assignment. Consequently, pre-existing differences between groups cannot be completely excluded.

Second, the intervention lasted only two weeks. The results therefore represent short-term learning gains and cannot establish whether the observed differences would persist over a longer period.

Third, the study involved students from a single school. The findings may therefore not generalize to schools with different student populations, resources, curricula, or technological conditions.

Fourth, the two AI platforms were associated with different grade levels: Grok 4 with Grade 5 and ChatGPT with Grade 6. This means that the comparison between platforms is potentially confounded by grade-level differences. The non-significant difference between the two AI groups should therefore be interpreted cautiously.

Fifth, the baseline and post-test scores were obtained from different assessment instruments. The baseline measure was based on the most recent English final-term

examination, whereas the post-test was researcher-developed. Although both assessments were scored out of 20 marks and aligned with relevant curricular content, their equivalence in difficulty and measurement characteristics was not independently established. Therefore, the calculated gain scores may partly reflect differences in assessment construction rather than changes in English language proficiency alone. This limitation should be considered when interpreting the magnitude of the observed learning gains.

Finally, although the assessment included reading comprehension, writing, grammatical competence, vocabulary, and speaking, the statistical analysis examined the overall gain score. The study therefore does not establish separate statistically significant effects for each individual language skill.

7. Conclusion and Future Research

7.1 Conclusion

This study examined the effect of AI-assisted pedagogy on English language learning gains among Grade 5 and Grade 6 school students. Using a quasi-experimental design with 73 students, the study compared teacher-guided Grok 4-assisted instruction, ChatGPT-assisted instruction, and traditional English instruction.

The results showed a statistically significant overall difference in gain scores among the three groups. Both AI-assisted groups achieved significantly greater gains than the traditional instruction group, while the difference between the Grok 4 and ChatGPT groups was not statistically significant.

The findings indicate that, within the context of this study, teacher-guided AI-assisted English instruction was associated with significantly greater observed achievement-score gains than traditional instruction. Both the Grok 4 and ChatGPT groups demonstrated higher mean gains than the control group, while no statistically significant difference was detected between the two AI-assisted groups. However, because the platforms were implemented with different grade levels, the findings do not establish their relative effectiveness or equivalence.

The results highlight the potential value of integrating generative AI as a supplementary resource within structured, teacher-guided English instruction. The observed outcomes suggest that AI-supported activities involving interaction, practice, feedback, and differentiated learning may offer useful pedagogical opportunities. Nevertheless, the findings should be

interpreted in light of the short intervention period, intact-group design, and assessment-related limitations.

Given the quasi-experimental design, convenience sampling, short intervention period, and school-specific context, the findings should be interpreted cautiously. They provide evidence concerning the study conditions rather than a universal claim about the effectiveness of AI in all English language classrooms.

7.2 Future Research

Future research should extend the intervention over longer periods to determine whether learning gains are sustained. Studies involving multiple schools and larger samples would improve the generalizability of findings.

Future investigations should also compare generative AI platforms within the same grade level and under equivalent instructional conditions. Randomized or matched-group designs would provide stronger evidence concerning differences between platforms. Further research could examine individual language skills separately, particularly reading, writing, speaking, grammar, and vocabulary. Such analyses would help determine whether AI-assisted pedagogy benefits particular dimensions of English language development differently. Longitudinal research could investigate whether students retain improvements after AI support is reduced or removed. This is important because short-term performance during technology-supported instruction does not necessarily establish independent long-term language development.

Finally, future research should examine the mechanisms through which AI-assisted pedagogy may influence learning, including interaction, feedback, scaffolding, engagement, learner autonomy, and teacher mediation. Such studies would provide a stronger empirical test of the constructivist and social constructivist processes proposed in the present framework.

References

- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503–2529. <https://doi.org/10.1111/bjet.13460>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550.

<https://doi.org/10.1177/00336882231162868>

intelligence on English language learning achievement. *System*, 125, 103428.
<https://doi.org/10.1016/j.system.2024.103428>

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, Article 100174.
<https://doi.org/10.1016/j.caeo.2024.100174>

Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 34(1), 335–359.
<https://doi.org/10.1080/10494820.2025.2498537>

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445.
<https://doi.org/10.1016/j.caeai.2025.100445>

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9.
<https://doi.org/10.1186/s40561-024-00295-9>

Piaget, J., & Inhelder, B. (1972). *The psychology of the child*. Basic Books.

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice*, 9(2), 159–167.

<https://doi.org/10.1080/23735082.2023.2261131>

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, Article 1260843.
<https://doi.org/10.3389/fpsyg.2023.1260843>

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

Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*.
<https://doi.org/10.1080/17501229.2025.2509759>

Wu, X.-Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System*, 126, 103498.

<https://doi.org/10.1016/j.system.2024.103498>

Xu, T., & Wang, H. (2024). The effectiveness of artificial