The Effect of AI-Enhanced Grammar Instruction on EFL University Students' Achievement, Attitudes, and Perceptions

P-ISSN: 1680-9300
E-ISSN: 2790-2129
Vol. (26), No. (2)
pp. 67-82

Ahmed A. Mohammed

College of Arts, Mosul University, Nineveh, Iraq.

Abstract:

This study investigates the effect of AI-enhanced grammar instruction on the achievement, attitudes, and perceptions of EFL students at university level. The participants of the study were 60 third-year EFL students (EG = 30 and CG = 30) enrolled in a grammar course at the University of Mosul. Both intact classes received the same instructional time, unit sequence, and teacher-led instruction. The experimental group was taught with simplified explanations, contextualized examples, and practice materials generated via Gemini AI tool and subsequently reviewed, edited, and aligned with the mandated textbook by the instructor, whereas the control group received textbook-based explanations and examples. Baseline equivalence was established for gender, age, a general grammar screening test, and pre-attitudes. Outcomes were measured using a unit-based grammar achievement test, a pre-post attitudes questionnaire, and perceptions questionnaire. The experimental group outperformed the control group on grammar achievement (M=34.26 vs 25.44; d=0.98; p<.001), reported more positive perceptions (M=88.35 vs 62.91; d=1.81; p<.001), and showed larger attitude gains (ΔM=27.59 vs 5.88; d=2.42; p<.001). The findings suggest that generative AI can enhance tertiary EFL grammar learning when used as an instructor-mediated materials-development tool within a teacher-led, curriculum-aligned instruction.

Keywords: Grammar, Computer-Assisted Language Learning, AI-Enhanced Materials, Perceptions, Attitudes.

1. Introduction

The English as a Foreign Language (EFL) program in higher education continue to teach students grammar through dedicated instructional methods. This phenomenon occurs in situations where students must demonstrate precise language use to succeed in academic writing and translation work and their examination results. Language teaching methods today still implement form-focused instruction because it helps students develop accurate language skills and attend to grammatical elements systematically (Ellis, 2006; Nassaji & Fotos, 2011). At the same time, research in second language acquisition (SLA) suggests that learners face mental challenges when they study grammar because they need to understand intricate rules and comprehensive metalinguistic terms and how language elements connect to meaning and language use (Larsen-Freeman, 2015). The cognitive challenges that students encounter in university settings can combine with their emotional obstacles which include anxiety and low confidence

Journal of Prospective Researches
Vol. (26), No. (2)
The paper was received on March 17, 2026; accepted on April 22, 2026; and published on April 29, 2026.
Corresponding author's e-mail: ahmad.mohammed@uomosul.edu.iq

10.61704/pr.616

and lack of motivation to diminish their ability to participate in university-based form-focused teaching (Dörnyei & Ryan, 2015; Horwitz, 2010).

The grammar instruction for tertiary EFL university programs depends on mandated textbooks which present organized content and structured topic sequences and assessment standards. The curriculum standardization through textbooks faces challenges because their explanations present a rule-heavy and terminology-driven style that creates dense content for advanced learners while offering insufficient scaffolding for beginners. Teachers need to use extra materials because they require more than what textbooks deliver through their standard content. These learning materials are regarded by the sociocultural framework as scaffolding which help students develop from rule knowledge to practical grammatical skills, as proposed by Lantolf and Thorne (2006). Thus, instructors are advised to spend additional time throughout the semester to create supplementary scaffolding resources for the courses.

Several scholars in the field have considered the Technology-enhanced Language Learning (TELL) to be a rich source of instructional support, including, but not limited to, grammar learning. In the same token, the studies on Computer-assisted Language Learning (CALL) evidently reveal that digital tools can help increase the exposure of students to language materials, structured practice opportunities, and additional learning time. As such these tools become part of the curriculum and teaching methods which can be employed by the teacher in the classroom (Chapelle, 2001). Researchers have repeatedly established that technology achieves its maximum effectiveness when it operates through educational teaching methods which teachers use to enhance their classroom instruction (Warschauer, 2011). The assessment-based grammar courses which use textbooks as their main resources need this particular issue because all students must demonstrate matching proficiency with the required standards through their assessments.

The current educational environment benefits from new generative artificial intelligence (AI) technologies which stem from developments in large language models. The system delivers three functions to users through its capacity to produce simplified explanations and alternative wording and context-specific examples on demand. The educational system needs teacher supervision because of the three main reasons which include system reliability, system accuracy and its intended educational setting. The teaching process uses AI as a supplementary tool because teachers control its application by selecting which content to use and verifying its accuracy while they associate it with required educational material and program objectives.

The learning outcomes of students depend on their responses to instructional approaches which determine their level of engagement and their ability to continue participating in the learning process. The study of technology acceptance demonstrates that people develop their educational technology attitudes based on their assessment of the technology's usefulness and its operational simplicity (Davis, 1989; Venkatesh et al., 2003). Students' motivation, confidence, and readiness to work on rule-based activities depend on their perceptions of grammar instruction which includes elements of clarity and practical value and supportive resources. For this reason, the evaluation of student achievement combined with their attitudes and perceptions about educational experiences provides complete assessment of the pedagogical effectiveness of AI-based grammar instruction in higher education.

The current research examines how instructor-led, AI-enhanced grammar instruction affects EFL university students' grammar outcomes and their study attitudes and perceptions of learning in a textbook-based college environment. Two intact groups studied the same mandated textbook content under the same instructor and instructional time. The control group received textbook-based explanations and examples, whereas the experimental group received teacher-led instruction supported by AI-supported simplified materials derived from the same textbook and reviewed by the instructor for accuracy and alignment. Students took a unit-based achievement test which measured their grammar knowledge through course units whereas a general grammar screening test was used to measure baseline equivalency at the start of the semester. The same questionnaire was used to evaluate attitudes before and after instruction while perceptions were evaluated through a post-instruction assessment.

10.61704/pr.616

Specifically, the current research seeks answers to the following questions:

RQ1: Is there a difference between students receiving instructor-led AI-enhanced instruction and those receiving textbook-based instruction in grammar achievement?

RQ2: Is there a difference between the two groups in attitude development (post−pre) over the instructional period?

RQ3: Does the experimental group show a significant change in attitudes from pre- to post-instruction?

RQ4: Is there a difference between the two groups in post-instruction perceptions of instructional clarity, usefulness, and learning support?

These questions were tested using the following null hypotheses:

H01: There is no statistically significant difference between the experimental and control groups on the grammar achievement test.

H02: There is no statistically significant difference between the experimental and control groups in attitude development (gain scores: post−pre).

H03: There is no statistically significant difference between the experimental group's pre- and post-attitudes scores.

H04: There is no statistically significant difference between the experimental and control groups on the post-instruction perceptions measure.

2. Literature Review

2.1 Grammar in EFL Contexts

Many EFL programs especially those in higher education continue to teach grammar as a core subject because academic writing and translation and formal examinations require students to demonstrate their grammatical accuracy. The use of communicative approaches has affected teaching methods, yet teachers still dedicate time to teach explicit grammatical rules because educational institutions and assessment methods need students to demonstrate their rule knowledge and their ability to use language correctly (Ellis, 2006; Nassaji & Fotos, 2011). Universities maintain grammar courses as organized programs that follow specific syllabi, which present topics in a planned order and require students to master proper language usage.

The process of grammar acquisition in second language learning goes beyond the identification of grammatical rules. Researchers have established two types of knowledge: explicit knowledge which focuses on rule knowledge and metalinguistic skills while the second type enables speakers to use grammar correctly in actual speech (Ellis, 2006). University students learn grammar through classroom explanations and their writing exercises yet they face difficulties with grammar production because of their lack of confidence. The process of understanding grammatical structures requires mental resources. According to Larsen-Freeman (2015), if students encounter new words or complex grammatical rules, they must understand both the meaningful and the contextual elements of the language which create challenges for them.

One of the biggest challenges in EFL education is that teachers have to follow certain textbook requirements which govern what materials (topics) and exercises should be taught and used. In fact, textbooks support the consistency of curricular and provide the exams with a common reference. However, they may present dense and formal explanations that are not equally accessible to all learners. As a result, students struggle to connect rule statements with actual sentence construction when explanations contain excessive metalanguage and give them only basic support. Teachers in those environments deliver extra notes and different explanations together with additional examples to help students understand grammar better and use it correctly in their exercises and writing tasks. The learning process uses this support as instructional scaffolding which enables students to receive structured help that guides them from rule comprehension to accurate and self-assured application of the rules (Lantolf & Thorne, 2006).

The success of grammar learning depends on affective factors which influence students' progress. Students experience heightened anxiety distress because grammar instruction includes correction and assessment procedures which expose

their mistakes to public view (Horwitz, 2010). Students who see grammar as difficult will practice less while they may avoid risk-taking and stop learning activities (Dörnyei & Ryan, 2015). The classroom and self-study contexts of grammar practice will show different patterns of student attention and persistence and affective dimensions will impact how students learn grammar.

The majority of university grammar courses base their teaching on textbooks and their evaluation methods which creates a practical obstacle for teachers. The educators have to follow the necessary curriculum while they assist students who need better understanding through more examples and extra practice time. The process of creating this extra educational material requires a significant amount of time. It becomes especially challenging to maintain throughout an entire academic term because students have different skill levels and need various types of support. The teaching requirements have led educators to seek additional resources which help students understand material better through extra examples while using textbooks as their main teaching resource. The application of AI for generating simplified explanations and enriched examples functions as supplementary scaffolding which aids students in understanding material and practicing their skills without altering course content. The primary issue is whether AI-based assistance results in measurable advantages for grammar learning and students' assessment of teaching methods through their attitudes and perceptions.

2.2 AI and Large Language Models (LLMs) in Language Education

In the area of CALL, which assesses digital tools for their effectiveness in educational and classroom settings, discussions about AI in language education have started. The main claim of CALL is that technology improves language learning through two specific functions, which are providing increased access to language materials and practice sessions, and through its design alignment with educational outcomes and curriculum requirements.

Large language models (LLMs) have increased their technological support abilities. They can now create new content through their explanation and paraphrasing and example generation functions which operate without fixed content. The ability to generate content becomes essential for teaching grammar rules because students require both straightforward rule descriptions and various situational examples to understand how to use grammatical structures. CALL experts discuss new technologies as resources which provide learners with immediate language support and scaffolding assistance to help them understand and demonstrate their knowledge (Godwin-Jones, 2018).

CALL research has shown that technology achieves its best results when it becomes part of teaching methods instead of being used as an independent educational tool. The EFL teaching methods in higher education institutions that use textbooks and rely on assessment for learning outcomes face this particular challenge. The classroom model for most effective defense in this situation needs to use AI technology for creating extra materials. In the meantime, teachers should evaluate and modify these materials to match required educational standards and textbook content. AI technology enables teachers to create better educational frameworks and more detailed examples while they keep control over what students should learn and how educational content should proceed through the course/semester (Chapelle, 2003).

Despite growing interest, research on LLMs in language education has not developed evenly in the field. Current educational research about LLMs in language education centers on three areas which include writing, translation, and feedback systems. However, grammar instruction remains underdeveloped as an instructional area. Most exploratory studies analyze how students interact with LLMs through autonomous use. Only a smaller number of studies examine teacher-led educational models which use AI as a built-in support system for teaching. The different methods of AI usage in learning environments create a fundamental teaching difference. In fact, autonomous access to AI technology lets students choose their own learning path while controlled access to AI technology requires students to follow specific rules and schedule their learning activities. The current research follows the second approach which positions LLMs as resources that provide both explanations and practical support instead of serving as complete teaching solutions.

2.3 Teacher-Mediated AI Integration

The discussion of generative AI in public settings concentrates on how students learn independently while research shows that educational software tools function better when teachers use them in classroom teaching. The effectiveness of technology increases when it creates learning outcomes through integrated tasks and teachers control its use instead of using it as an alternate teaching method (Chapelle, 2001; Garrett, 2009; Stockwell & Hubbard, 2013). Hence, teacher mediation becomes vital for EFL instruction in tertiary education because teachers need to maintain textbook requirements while they use additional materials to assist students.

Grammar instruction needs teacher mediation because students must develop precise language skills and understand how grammatical structures connect with their meanings and use consistent terminology throughout their language learning journey. Despite the fact that LLMs can generate useful explanations and examples, but still they can yield overgeneralized and/or inaccurate results that learners are not equipped to detect dependably. The teacher needs to choose proper outputs because their duties require them to check output correctness and keep texts at the appropriate learner level and maintain textbook content and course goals. The teacher functions as an epistemic gatekeeper for the classroom. This role supports Second Language Acquisition (SLA) theories which describe how teachers should help students learn language structures through classroom expertise and instructional methods (Long, 2015).

Students learn better when they use mediated AI systems since those systems operate as scaffolding for their educational development. Scaffolding should become responsive to student needs because students require different learning paths from their current knowledge state to their future proficient knowledge state according to sociocultural approaches (Lantolf & Thorne, 2006). AI tools help this process by providing simplified restatements together with context-specific additional examples and different ways to explain content. Teacher calibration determines how effectively scaffolding will function in educational settings. Teachers decide when to use simplification and which metalanguage terms should remain while they determine AI-supported content sequencing for both explanation and practice activities.

Teacher mediation impacts how students build trust and stay involved with learning activities. Students in technology-based learning environments depend on their teachers to confirm the authenticity of learning materials and to explain the correct usage methods of technological tools (Zheng et al., 2016). LLMs produce text that seems authoritative but sometimes contains unreliable information, which creates a major problem for users of the technology. Teachers who select and assess AI-created content help decrease doubts while establishing proper levels of trust to show students that AI functions as a secondary academic tool instead of an official research source (Godwin-Jones, 2023).

Finally, teacher mediation improves research design for classroom studies by preventing a typical confounding factor that affects LLM research. Students who use AI by themselves will produce different results because their research skills, self-control abilities and their verification methods differ from one another. The teacher-mediated model establishes uniform input standards while maintaining essential instructional material differences, which research needs to examine groups who share the same teacher and content delivery and contact time. The current study design tests three grammar topics with both groups receiving the same instruction but the experimental group using teacher-approved AI-supported content created from the same textbook material.

2.4 Perceptions and Attitudes as Constructs in Technology-Enhanced Language Learning

The research field of technology-enhanced learning investigates how learners develop their perceptions and attitudes as two separate yet interconnected constructs. The concept of perceptions describes how learners evaluate their experience with a learning tool through their judgment of its learning value and its ability to support their education. On the other hand, the concept of attitudes describes how learners evaluate things through their emotional reactions and their ability to show approval or preference, which then shapes their decision to use or recommend a specific method (Ajzen, 2005). This distinction is significant in classroom research. Teachers often

observe the perceptions of students through their direct experience with the teaching methods. In the same vein, they assess the attitudes of students through their authentic experiences and overall behavior patterns.

According to the Technology Acceptance Model (TAM), the cognitive beliefs of learners about technology can be developed through their assessments of the usefulness and the flexible use of the technology itself (Davis, 1989). That is to say, learners assess technology on the basis of two aspects: usefulness and usability. In CALL contexts, learners assess clarity, convenience, and learning support, specifically through these criteria, which directly impact their motivation and digital learning tool engagement level.

The study of grammar requires students to use their cognitive abilities because explicit grammar instruction depends on their understanding of grammatical rules. When explanations are dense or metalanguage-heavy, learners perceive instruction as difficult and less supportive, which results in lower engagement and persistence (Nassaji, 2017). Learners' responses depend on their emotional states. Students who face anxiety and lack confidence and motivation will participate less in language learning while avoiding form-focused tasks that assess their performance through visible errors (Dörnyei & Ushioda, 2011; Dörnyei & Ryan, 2015; Horwitz, 2010). Perceptions of clarity and usefulness work together with affective attitudes, which include enjoyment and confidence and anxiety, to create different learner experiences during grammar instruction.

LLMs create additional assessment criteria for students because they must assess model clarity and practical value together with their ability to generate trust and verified information. Early studies of CALL highlight the fact that students prefer AI-generated explanations because they provide immediate access but students still need to verify output accuracy and reliability (Godwin-Jones, 2023). In classroom contexts, teachers serve as assessment tools for students who need to determine which information sources have credibility and which digital tools they should use properly. Teacher mediation establishes two types of influence on students because it determines their assessment of material clarity and usefulness, and as a result, shaping their opinion about AI technology usage in education (Zheng et al., 2016).

For these reasons, the current study operationalizes the two concepts according to their exact conceptual definitions and their specific measurement methods. Perceptions represent students' cognitive assessments of their particular instructional experience which they developed after receiving instruction. Attitudes represent students' assessment of grammar learning and AI in language learning which researchers measured through identical assessment tools before and after their instructional period. This separation enables the research to investigate how AI-enhanced teaching affects student performance and their assessment of the learning experience while researchers measure student attitude changes throughout the teaching time.

2.5 Previous Studies on Generative AI in EFL Learning

Empirical research on generative AI in EFL context has expanded rapidly, but most studies have examined learner use of tools such as ChatGPT for general language support, writing, or speaking, rather than grammar instruction as a primary instructional target. Across contexts, students recurrently describe generative AI as a supportive learning resource and report positive perceptions. They think that it offers instant responses, simplified explanations, and flexible practice prospects. For example, studies of student perceptions toward ChatGPT in English learning contexts generally indicate high perceived usefulness, while also raising concerns about reliability and overdependence (Phosa, 2024).

Researchers have conducted multiple studies to understand how people perceive and feel about their use of ChatGPT for particular skill development. For example, Saudi students perceived ChatGPT positively as a speaking support tool according to Alsalem's (2024) study which showed that students perceived benefits from the tool while they actually needed to use it. Similarly, Esteban et al. (2025) found that higher education students held positive views about ChatGPT for English learning yet students evaluated the tool based on their doubts regarding its output accuracy and learning benefits from AI. Together, the research findings of these two studies demonstrate that students use AI tools based on their perception of worth and their degree of trust in those tools and

their educational suitability.

Grammar-focused studies exist, but they remain comparatively limited. Küçük studied AI-supported instruction together with ChatGPT instruction in Iraq and found that students considered AI technology as useful for learning but they also identified research limits and educational dangers that needed special teaching methods to handle (Küçük, 2023, 2024). The research demonstrates how regional interest in AI technology for grammar learning has increased but it also shows the literature's main issue which states that AI technology needs proper teaching methods to prevent unplanned use in educational settings that match the curriculum.

Another related aspect concerns AI use as a grammar support or correction tool. Schmidt-Fajlik (2023) studied ChatGPT as a grammar checking tool and showed its language support advantages and its danger of users believing AI results without doubt. The findings establish grammar instruction requirements because they identify epistemic trust as the primary challenge which students need to learn grammar accurately. However, they lack the skills to judge which incorrect explanations and deceptive corrections are false.

Research on Gemini is even more limited than research on ChatGPT. Students who used the Gemini app showed positive attitudes toward its accessibility features and its ability to give them personalized feedback according to the findings from Nguyen's (2025) mixed-methods research study of MA student experiences with the app. This study includes more than grammar instruction yet its results show how students who respect AI tools will only use them when they think the tools match academic standards and their course requirements.

Taken together, existing studies indicate that generative AI functions as a beneficial learning tool according to users, yet trust issues, accuracy problems, and curriculum compatibility issues continue to exist as major obstacles. Along these lines, this claim stems from the studies which track students' AI usage outside the supervised classroom environments where teachers regulate the reasonable access to AI resources. This point is substantial since teacher-mediated integration changes the ecology of learning. In other words, it can decrease the accuracy problems and create homogenous educational content that achieves the course requirements and, at the same time, maintains proper AI usage, in alignment with educational standards and institutional evaluation methods.

2.6 Research Gap and Justification

Research on generative AI in EFL has grown quickly but there is still a lack of proof about grammar teaching methods that universities use in classrooms which depend on textbooks and educational assessments. Existing research mainly studies how students use AI technology which creates a research gap because there are still few studies that examine teacher-led AI models which use AI to create explanations and examples that teachers check for accuracy before using them in their planned lessons. That is why there is a growing need for designs that connect achievement outcomes with learners' attitudes and perceptions, since evaluations of clarity, usefulness, and trust can influence engagement and sustained learning.

Against this background, the present study provides evidence from classroom research by testing a teacher-led, curriculum-aligned model of Gemini-supported materials enhancement and examining its relationship with grammar achievement as well as students' attitudes and perceptions.

3. Methodology

3.1 Research Design

iThis study adopts a quasi-experimental, non-equivalent groups design to examine the effect of AI-enhanced grammar instruction on EFL university students' grammar achievement, attitudes, and perceptions. Two intact classes enrolled in the same grammar course were assigned to two instructional groups: a control group (CG) using textbook-based materials and an experimental group (EG) using AI-enhanced materials curated by the instructor. Because intact classes were used, random assignment was not feasible; therefore, the study is treated as quasi-experimental (Shadish, Cook, & Campbell, 2002). Grammar achievement was assessed using a test administered to both groups at the end of the semester to compare outcomes between instructional groups. To reduce selection concerns associated with intact groups, a short general grammar screening test not tied to the course units was administered at the beginning of the semester to examine

baseline comparability. Students' attitudes were measured using the same attitude questionnaire administered before and after instruction, while perceptions were measured after conducting the experiment.

3.2 Context and Participants

The current study is conducted within an EFL grammar course delivered over a 15-week semester at the University of Mosul, College of Arts. Participants are undergraduate EFL students enrolled in two intact classes of the same course (60 in total). Both classes follow the same mandated grammar textbook, cover the same units in the same sequence, and receive the same number of instructional hours. Participation in the research components is voluntary and does not affect course grades.

3.3 Instructional Conditions

The research study tests two different teaching methods which provide equal learning content and identical educational pace and classroom time and teacher instructional time. The only planned difference between the two approaches involves the specific materials used to teach grammar rules and provide examples.

- **Control Group: Textbook-Based Materials**

Students learn grammar through the textbook which serves as their main resource for grammar explanations and examples. The instructor explains grammar points following the textbook presentation and provides practice aligned with the textbook activities and course objectives.

- **Experimental Group: AI-Enhanced Materials Curated by the Instructor**

Students receive instruction on the same grammar topics and complete comparable practice. The instructor presents content through AI-enhanced materials which he specially created for his students. The instructor uses Gemini to create basic textbook explanations and additional material which shows practical examples of the same textbook content. All AI-generated outputs undergo a process which includes review, correction, editing and alignment to textbook terms and course goals before their classroom use. Students do not interact directly with Gemini. The teachers maintain control over instruction while using AI for materials development to support their established curriculum.

- **Time-on-task control**

To avoid confusing the results with unequal exposure, both groups receive comparable instructional time, practice quantity, and homework load. While the EG receives additional AI-generated examples, the CG receives an equivalent number of instructor-prepared examples and practice items.

3.4 Instruments

3.4.1 General Grammar Screening Test (Baseline)

The baseline grammar ability assessment in Week 1 used a brief screening test which measured general grammatical knowledge beyond the specific unit content taught during the academic semester. The screening test functioned solely to examine the initial comparability between the two intact classes and the test results served no purpose as outcome measurements. The test used an objective scoring method which assigned 0-1 points to each test item while total scores served as the basis for descriptive statistics and between-group analysis. The baseline screening test included 20 multiple-choice questions which measured grammar knowledge that was not specifically taught in the course content.

3.4.2 Grammar Achievement Test

Grammar achievement was measured using a researcher-developed test administered to both groups in Week 15. The test content was based on the grammar points taught during the semester and was distributed across topics using a table of specifications to ensure balanced coverage and content validity. The test included three item formats (total 40 items): 24 multiple-choice items, 10 fill-in-the-blank items, and 6 sentence-level tasks (error correction and or sentence transformation). The test consisted of items which received scores through a system that assigned 1 point for correct answers and 0 points for incorrect answers with the highest

possible score being 40 points. The assessment of content validity was conducted through syllabus alignment and expert assessment by five EFL specialists who evaluated item relevance, clarity, and level. The test reliability assessment used KR-20 for the complete test while item analysis after administration provided additional support for the results.

3.4.3 Attitudes Questionnaire (Pre/Post)

Students' attitudes are measured using a 20-item Likert-scale questionnaire administered twice: Week 1 (pre) and Week 15 (post). The same items and wording are used for both groups at both time points. Items are written in general terms to ensure they are meaningful before and after instruction, focusing on:

1. Attitudes toward learning grammar؛
2. Attitudes toward using AI and technology to support learning, including willingness to use AI and trust when AI output is checked by the instructor.

A 5-point scale ranging from 1 (Strongly disagree) to 5 (Strongly agree) was employed to record the responses. Cronbach alpha was used to assess the internal consistency. The subscale scores were computed and supported by reliability evidence, and the intended construct structure.

3.4.4 Perceptions Questionnaire

The perceptions of students were measured at the end of the semester only (i.e., Week 15). This is because perceptions of any instructional approach necessitate direct experience of the course. The aim of the perceptions instrument was to evaluate the cognitive judgments of the instruction which the students have experienced. This included clarity of explanations, examples' practicality, evident learning engagement, support, and effectiveness.

Two versions of the perceptions questionnaire were used. They included identical structure and near-identical wording in which the CG version refers to "textbook-based explanations and examples, while the EG version refers to "AI-enhanced explanations and examples reorganized by the instructor." This wording emphasizes the fact that the intervention is teacher-led in both groups. It keeps the instructional contrast focused on the materials taught.

Students' responses were recorded on the same 5-point Likert scale. Internal consistency was assessed by means of Cronbach's alpha. Subscale scoring was used if justified by the construct structure.

3.5 Procedure

In Week 1 of the study, students enrolled in both classes received the instructions of the study's purpose and provided informed consent. After that, both groups completed a short general grammar screening test to examine their level and then completed the (pre) attitudes questionnaire. The two groups were taught by the same instructor, followed the same mandated textbook, covered the same chapters in the same sequence, and received the same number of instructional hours and comparable practice and homework. The CG used textbook-based materials, whereas the EG used AI-enhanced materials curated by the instructor, derived from the same textbook content, and checked and adapted by the instructor before classroom use. In Week 15, both groups completed the grammar achievement test, completed the (post) attitudes questionnaire, and then completed the perceptions questionnaire, with parallel versions administered to match each group's instructional materials.

3.6 Data Analysis

Quantitative data were analysed using both descriptive and inferential statistics. The researcher used general grammar screening test scores to assess baseline equivalence between groups which they tested through independent samples t test and Mann–Whitney U test depending on assumption violations. Grammar achievement was assessed through test scores while using independent samples t test to compare between different student groups. For attitudes, internal consistency was assessed using Cronbach's alpha, and pre to post change was examined within each group using paired samples t tests. The researcher compared post-instruction attitudes between different groups through independent samples t tests. For perceptions, post-instruction perception

scores were compared between the two groups using an independent samples t test. Effect sizes were reported (Cohen's d for parametric comparisons and r for non-parametric tests), and statistical significance was set at $p < .05$.

3.7 Ethical Considerations

Students are informed that responses are anonymous, and participation or non-participation does not affect grades. Data are stored securely and used only for research purposes.

4. Results

4.1 Group Equivalence (Baseline Comparability)

The EG and CG were examined for equivalence on main background variables (e.g., gender, age), a general grammar screening test, and the pre-attitudes measure before testing the research hypotheses. The results indicated that students in the two groups were statistically equivalent. This suggested that any following differences are more likely due to the instructional settings rather than previous group differences.

Table 1. Group Equivalence by Gender (Chi-Square Test).

Group	Male	Female	χ^2 (calculated)	χ^2 (tabulated)	df	Sig. (α = .05)
Experimental	17	13	0.26	3.84	1	n.s.
Control	15	15				

Note. n.s. = not significant

Table 2. Group Equivalence in Age (Independent Samples T-Test)

Group	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Experimental	30	21.32	12.08	0.203	2.002	58	n.s.
Control	30	21.93	11.12				

Table 3. Group Equivalence in General Grammar Screening (Independent Samples T-Test)

Group	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Experimental	30	13.62	10.46	0.449	2.002	58	n.s.
Control	30	12.18	13.98				

10.61704/pr.616

Table 4. Group Equivalence in Pre-Attitudes (Independent Samples T-Test)

Group	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Experimental	30	58.67	10.07	0.656	2.002	58	n.s.
Control	30	57.03	9.32				

4.2 Testing the Research Hypotheses

4.2.1 Grammar Achievement Test (Experimental vs Control)

4.2.1.1 Null Hypothesis 1

H01: There is no statistically significant difference between the mean scores of the EG and the CG on the grammar achievement test.

To examine whether AI-enhanced grammar instruction produced higher achievement than textbook-based instruction, an independent samples t-test was conducted on the grammar achievement test scores.

Table 5. Grammar Achievement Test: Experimental vs Control (Independent Samples T-Test)

Group	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Experimental	30	34.26	9.2	3.8	2.002	58	Significant
Control	30	25.43	8.74				

As shown in Table 5, the experimental group achieved higher post-test scores (M = 34.26, SD = 9.20) than the control group (M = 25.43, SD = 8.74). This difference was statistically significant, $t(58) = 3.81$, $p < .001$, $d = 0.98$; thus, H01 was rejected.

4.2.2 Development in Attitudes (Gain Scores: Experimental vs Control)

To examine whether attitudes developed more strongly in the EG, the difference (post-pre) was computed for each group and compared using an independent samples t-test.

4.2.2.1 Null Hypothesis 2

H02: There is no statistically significant difference between the EG and the CG in attitude development (gain scores: post−pre).

10.61704/pr.616

Table 6. Attitudes Development (Gain Scores): Experimental vs Control

Group	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Experimental	30	27.58	9.2	9.36	2.002	58	Significant
Control	30	5.87	8.74				

As clearly shown in Table 6, attitude gain scores were larger for the EG (M = 27.58, SD = 9.20) than for the CG (M = 5.87, SD = 8.74). This difference was statistically significant, $t(58) = 9.36$, $p < .001$, $d = 2.42$, indicating greater attitude improvement in the experimental condition; therefore, H02 was rejected.

4.2.3 Pre–Post Attitudes within the Experimental Group

A paired-samples t-test was conducted to examine whether the EG's attitudes improved significantly from pre to post.

4.2.3.1 Null Hypothesis 3

H03: There is no statistically significant difference between the EG's mean scores on the pre-attitudes and post-attitudes measures.

Table 7. Experimental Group Attitudes: Pre vs Post (Paired Samples T-Test)

Test	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Pre	30	58.67	10.07	16.4	2.04	29	Significant
Post	30	86.26	9.2				

As indicated in the table above, in the EG, attitudes increased significantly from pretest (M = 58.67, SD = 10.08) to posttest (M = 86.26, SD = 9.20), $t(29) = 16.40$, $p < .001$, $dz \approx 3.00$, indicating a strong positive shift over the intervention period. Therefore, H03 was rejected.

4.2.4 Perceptions Questionnaire (Experimental vs Control)

4.2.4.1 Null Hypothesis 4

H04: There is no statistically significant difference between the mean scores of the EG and the CG on the perceptions scale.

To test whether students taught with AI-enhanced materials reported more positive perceptions than those taught with textbook-based instruction, an independent samples t-test was conducted on the perceptions scores.

10.61704/pr.616

Table 8. Perceptions: Experimental Vs Control (Independent Samples T-Test)

Group	N	Mean	SD	t (calculated)	t (tabulated)	df	Sig. (α = .05)
Experimental	30	88.35	14.4	7.005	2.002	58	Significant
Control	30	62.9	13.71				

As indicated in Table 8 above, perceptions were significantly more positive in the EG (M = 88.35, SD = 14.40) than in the CG (M = 62.90, SD = 13.71), t(58) = 7.01, p < .001. Therefore, H04 was rejected.

5. Discussion

This study investigated whether AI-enhanced grammar materials curated by the instructor produce differences in achievement, attitudes and perceptions towards learning compared to traditional textbook-based instruction in a university EFL grammar course. Overall, the findings indicate consistent advantages for the AI-enhanced condition across the study outcomes. The results are interpreted below in relation to the study hypotheses and research questions on generative AI in EFL.

5.1 Achievement (RQ1 / H01)

In relation to the first hypothesis, the findings showed that students taught with AI-enhanced materials outperformed their peers who studied using traditional textbook methods. A likely explanation is that the AI-enhanced materials served as cognitive scaffolding that helped students process textbook metalanguage more easily while establishing clearer connections between form and meaning and usage in their learning. Generally speaking, grammar is explained through a set of rules and terminology in EFL contexts. One of the negative outcomes of this approach is that it can overload students' working memory and hinder uptake, especially for learners at mixed proficiency levels. The adoption of AI-generated materials to help present a simplified content improved understanding and facilitated students practice more effectively.

This interpretation is in alignment with the widespread evidence which emphasize that learners perceive generative AI as helpful. It provides accessible explanations and flexible examples, despite the fact that some previous studies have focused on general language support rather than grammar as the ultimate target (Phosa, 2024; Esteban et al., 2025). Viewed from another perspective, the present study differs from self-access models. The students did not interact with AI independently. Alternatively, the instructor organized and aligned the AI output to course objectives. It is believed that this may have helped explain the reliable achievement advantage opposed to contexts in which accuracy and alignment remain undefined.

5.2 Attitude Development Between Groups (RQ2 / H02)

The AI-enhanced group demonstrated better attitude development than the control group according to the results of the second hypothesis of the study. The intervention affected both performance and perceived clarity according to the study results and it also changed how students evaluated their learning of grammar and their use of AI in teacher-led courses. The results imply that better understanding together with organized assistance eliminated the emotional challenges which students face when learning grammar especially the anxiety and frustration that result from learning abstract rules and receiving evaluation based on errors. When learners experience progress and understand the content more easily, their attitudes toward learning material will become more positive

because they find the learning experience to be easier and more rewarding.

This interpretation is compatible with evidence that learners' attitudes toward generative AI when they believe the tool helps them learn and builds their confidence, which proves to be useful but requires responsible use in suitable educational contexts (Alsalem, 2024; Esteban et al., 2025). The present study demonstrates that attitude improvements occur most effectively when instructors use AI to create teaching materials instead of giving students direct access to AI tools, which helps educators manage their concerns about student dependence on technology and its accuracy.

5.3 Pre–Post Attitude Change in Experimental Group (RQ3 / H03)

After implementation, this hypothesis was validated because the EG showed significant improvement in their attitudes from pretest to posttest. This within-group alteration indicates that constant exposure to the AI-enhanced materials across the interventional period produced an obvious change in learners' attitude toward the learning experience. This result can be interpreted as an "experience-driven" attitude change. In other words, attitudes become more positive when learners recurrently encounter an approach that is clearer, more supportive, and more associated with their learning needs.

5.4 Perceptions (RQ4 / H04)

Based on this hypothesis, students in the EG who were taught by means of the AI-enhanced materials revealed positive perceptions of the instruction. This outcome can be interpreted through the so-called "instructional fit". In other words, learners are more likely to perceive an approach positively if it is clear, relevant to assessment prospects, and consistent with course routines. For instance, in grammar courses which are taught based on mandated textbooks, students frequently evaluate additional support grounded in whether it helps them meet course loads and examination requirements or not. Thus, it is seen that AI-enhanced explanations which prove to be aligned with the textbook may be perceived as both legitimate and useful. This result is consistent with previous research results in which students rate generative AI tools as useful for explanation and support, but, at the same time, express concerns about reliability if such tools are used autonomously (Esteban et al., 2025; Phosa, 2024). It is also in line with previous studies which highlight the fact that if learners use AI output unverifiably, misconceptions may arise (Schmidt-Fajlik, 2023). Finally, it can be argued that the present study's design reduced this risk because the AI output, used in the intervention, was filtered and presented through instructor-led teaching to support students perceive higher trustworthiness and clarity.

6. Conclusions

The current study investigated the effect of instructor-curated, AI-enhanced grammar topics on EFL university students' achievement, attitudes, and perceptions within a mandated textbook university grammar course. Across outcomes, the findings of the study showed a statistically significant difference in favour of the AI-enhanced intervention. Students who learned through AI achieved significantly higher scores in the grammar test than those who followed the textbook-only instruction. Furthermore, students' attitudes improved in two ways: (a) the EG who studied the AI-enhanced materials showed stronger attitude development than the CG, and (b) within the EG, attitudes also improved significantly from pretest to posttest. These results suggest that long-term exposure to the AI-enhanced materials contributed to a more positive attitude toward learning grammar.

Complementary to achievement and attitudes, learners in the EG also reported more positive perceptions of the instruction, mainly in relation to clarity and learning support. Taken together, these results suggest that generative AI may add pedagogical value in EFL grammar instruction. This is the case especially when it is used as a teacher-led, curriculum-aligned materials-development tool rather than as an unlimited self-access model. The study's interventional design highlights the important role of instructor as a mediator. Namely, AI outputs should be reviewed and edited regularly to ensure accuracy, terminological consistency, and alignment with the mandated textbook and assessment expectations. In doing so, this could reduce any risks associated with unverified AI-generated

explanations.

6.1 Implications

Based on the findings of the current study, it can be argued that selecting AI-based supplementary resources for teaching grammar will help improve explanation clarity through better contextual example development without changing the existing curriculum. The implementation process requires educators to treat AI-generated content as draft materials. These materials need their assessment to check for correct grammar and proper use of metalanguage and whether or not they match the course learning outcomes and evaluation criteria. Equally important, educators can increase students' participation by presenting AI-enhanced materials as teacher-created resources. To this end, they demonstrate their authority, foster trust and promote responsible classroom behavior during assessment periods.

6.2 Limitations

When interpreting the findings, several limitations should be taken into consideration. Firstly, the current study used a quasi-experimental design with intact classes and no random assignment, which limits generalizable inferences. Secondly, the grammar achievement was measured by means of an end-of-course test without a course-aligned pretest. Baseline comparability was established using a general screening test. Thus, pre-existing differences in course-specific knowledge cannot be completely excluded. Thirdly, the study was conducted in a single institution with one instructor, which may yield generalizability bounds and introduce the possibility of teacher and/or expectancy effects. Finally, self-report questionnaires were used to measure attitudes and perceptions, which may be influenced by response tendencies and the instructional approach framing.

6.3 Future Research

The current study opens a number of directions for future research. For example, this intervention can be replicated across different universities or colleges, levels, and course types to test the effects of generalizability and context. Research designs could focus on the inclusion of course pretests, delayed post-tests to examine retention and additional variables. Further research is also needed for attitudes and perceptions in generative-AI context for the purpose of instruments' validation and refinement. Finally, comparative studies could test different integration models and investigate which features of AI generated materials, such as the simplification level, example types, and practice density, are strongly associated with grammar learning outcomes.

References

Ajzen, I. (2005). Attitudes, personality and behavior (2nd ed.). Open University Press.

Alsalem, L. A. (2024). Investigating EFL students' perceptions and attitudes towards ChatGPT in promoting speaking skill. International Journal of English Language Education, 12(1), 2–24. https://doi.org/10.5296/ijele.v12i1.22358

Chapelle, C. A. (2001). Computer Applications in Second Language Acquisition. Cambridge University Press. https://doi.org/10.1017/cbo9781139524681

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Dornyei, Z., & Ryan, S. (2015). The psychology of the language learner revisited. Routledge. https://doi.org/10.4324/9781315779553

Dörnyei, Z., & Ushioda, E. (2011). Teaching and researching motivation (2nd ed.). Routledge. https://doi.org/10.4324/9781315833750

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. TESOL quarterly, 40(1), 83-107. https://doi.org/10.2307/40264512

Esteban, A. J., Park, I., Nga, N. T., Perunovic, S., Park, S. E., & Yi, J. I. (2025). Undergraduate Students' Perceptions on the Use of ChatGPT for English Learning at a Korean University. rEFLections, 32(2), 994-1016. https://so05.tci-thaijo.org/index.php/reflections/article/view/282889

Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. The modern language journal, 93, 719-740. https://doi.org/10.1111/j.1540-4781.2009.00969.x

Godwin-Jones, R. (2018). Using mobile technology to develop language skills and cultural understanding. Language Learning & Technology, 22(3), 3–17.

Horwitz, E. K. (2010). Foreign and second language anxiety. Language Teaching, 43(2), 154–167. https://doi.org/10.1017/S026144480999036X

Küçük, T. (2023). AI-integrated grammar teaching in language preparatory school. International Journal of Social Sciences & Educational Studies, 11(1), 1–17. https://doi.org/10.23918/ijsses.v11i1p1

Küçük, T. (2024). ChatGPT-integrated grammar teaching: EFL learners' perspectives at a private university in Erbil. In Arab World English Journal (AWEJ) Special Issue on ChatGPT (pp. 100–111). https://doi.org/10.24093/awej/chatgpt.8

Lantolf, J. P. (2006). Sociocultural theory and L2: State of the art. Studies in second language acquisition, 28(1), 67-109. https://doi.org/10.1017/S0272263106060037

Larsen-Freeman, D. (2015). Research into practice: Grammar learning and teaching. Language teaching, 48(2), 263-280. https://doi.org/10.1017/S0261444814000408

Long, M. H. (2015). Second language acquisition and task-based language teaching. Wiley Blackwell.

Nassaji, H., & Fotos, S. (2011). Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context. Routledge. https://doi.org/10.4324/9780203850961

Nguyen, H. A. (2025). MA Students' Perceptions and Experiences with the Gemini App in English Language Learning: A Mixed-Methods Study at Van Lang University, Vietnam. International Journal of AI in Language Education, 2(2), 1-19. https://doi.org/10.54855/ijaile.25221

Phosa, S. (2024). The study of EFL students' attitudes and behaviors toward using ChatGPT in English learning. Journal of English Language and Linguistics, 5(3), 332–344. https://doi.org/10.62819/jel.2024.656

Schmidt-Fajlik, E. (2023). ChatGPT as a grammar checker: The efficacy of ChatGPT in correcting English writing. AsiaCALL Online Journal, 14(1), 127–153. https://doi.org/10.54855/acoj.231417

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Stockwell, G., & Hubbard, P. (2013). Some emerging principles for mobile-assisted language learning. The international research foundation for english language education, 2013, 1-15.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Warschauer, M. (2011). Learning in the cloud: How (and why) to transform schools with digital media. Teachers College Press.

Zheng, B., Warschauer, M., Lin, C. H., & Chang, C. (2016). Learning in One-to-One Laptop Environments: A Meta-Analysis and Research Synthesis. Review of Educational Research, 86, 1052-1084. https://doi.org/10.3102/0034654316628645