ESTEEM Journal of English Study Programme
E-ISSN: 2622-2213
P-ISSN: 2622-9323

COMPARATIVE ANALYSIS OF LLM-BASED AI ASSISTANTS FOR EFL WRITING INSTRUCTION: CHATGPT, CLAUDE, GEMINI, AND DEEPSEEK

Aldha Williyan[1], Muhammad Aulia Taufiqi[2], Irma Ratna Ningsih[3], Muhammad Ilyas[4], Eka Pujiastuti[5]

[1]Universitsa Siliwangi, Tasikmalaya, Indonesia
[2]Institut Pesantren Babakan Cirebon, Cirebon, Indonesia
[3]Institut Seni Budaya Indonesia Bandung, Bandung, Indonesia
[4]STIT Al Hikmah Bumi Agung, Way Kanan, Lampung, Indonesia
[5]Politeknik Mitra Karya Mandiri, Ketanggungan Brebes, Indonesia

E-mail: alwilliyan@unsil.ac.id[1], auliataufiqi@ipeba.ac.id[2], irma.ratna@isbi.ac.id[3], ilyaspujakesuma@gmail.com[4], ekapujias123@gmail.com[5]

Accepted :
8 January 2026
Published :
8 January 2026
Corresponding Author:
Aldha Williyan
Email Corresponding :
alwilliyan@unsil.ac.id

ABSTRACT
This study investigates the comparative capabilities of four prominent LLM-based AI assistants (ChatGPT, Claude, Gemini, and DeepSeek) in supporting English as a Foreign Language (EFL) writing instruction, particularly in providing explanations of key linguistic concepts. Employing a three-phase research design, the study examines how these AI tools assist EFL teachers in explaining grammar, vocabulary usage, sentence structure, and cohesion/coherence. In Phase 1, standardized prompts were developed based on Khan's prompt engineering framework to ensure consistency across AI interactions. Phase 2 involved submitting these prompts to each AI system under controlled conditions using their free versions to reflect common educational constraints. In Phase 3, the AI-generated responses were analyzed thematically using Braun and Clarke's framework, focusing on clarity, conceptual depth, organization, and instructional value. The findings reveal distinct pedagogical profiles: ChatGPT offers structured and practice-oriented explanations, Claude emphasizes conceptual depth and academic reasoning, Gemini employs accessible and engaging explanatory strategies, and DeepSeek provides concise explanations suitable for quick reference. This study provides practical guidance for EFL teachers in selecting AI tools based on instructional context and learner needs.

Keywords: Artificial Intelligence, Educational Technology, English Language Teaching, Language Explanation, Writing Instruction.

1. INTRODUCTION

Large Language Model (LLM)-based AI assistants have rapidly emerged as influential tools in educational contexts, yet their pedagogical roles in English as a Foreign Language (EFL) writing instruction remain insufficiently examined from a comparative perspective. While existing research has documented the general potential of AI-driven systems in language education, much of the literature has focused on individual platforms or on outcome-oriented functions such as feedback provision and assessment, rather than on how these tools explain core linguistic concepts to support writing development (Lund et al., 2023). As a result, there is limited empirical guidance for EFL educators on how different LLM-based AI assistants function as explanation providers in writing instruction.

Against this backdrop, LLM-based AI assistants represent a transformative technological advancement in EFL education.

Vol 9, No 1 (2026): ESTEEM

Built on sophisticated deep learning architectures and trained on extensive text corpora, these systems have evolved to understand, generate, and explain human language with increasing precision (Bonner et al., 2023; Makridakis et al., 2023). Their ability to process natural language queries and deliver contextually relevant explanations has positioned them as promising pedagogical supports for EFL writing instruction, particularly in explaining grammatical concepts, vocabulary usage, sentence structure, and cohesion and coherence. The growing integration of these tools into EFL classrooms is driven by their capacity to provide on-demand, personalized explanations and immediate feedback, addressing challenges such as large class sizes and limited instructional time (Extance, 2023; Kasneci et al., 2023). For many EFL teachers, LLM-based AI assistants function as supplementary instructional resources that extend pedagogical reach rather than replace teacher expertise.

Among the expanding range of LLM-based AI assistants, four systems have become especially prominent in educational use: ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google), and DeepSeek (DeepSeek AI). Although all four demonstrate strong potential for supporting EFL writing instruction, they are built on different architectural designs, training strategies, and pedagogical orientations, which influence how they explain linguistic concepts (Liu et al., 2024). These differences are reflected in explanatory clarity, conceptual depth, instructional organization, and pedagogical alignment. Understanding such variation is essential for effective classroom integration, as the way an AI assistant explains grammar, vocabulary, sentence structure, or cohesion can significantly shape learners' comprehension and writing development (Evmenova et al., 2024; Kohnke et al., 2023).

Despite their increasing adoption, comprehensive comparative research examining how ChatGPT, Claude, Gemini, and DeepSeek function specifically as explanation providers in EFL writing instruction remains underexplored. Existing studies tend to emphasize isolated tools or limited comparison, leaving unanswered questions about how these systems differ in supporting key aspects of EFL writing pedagogy. Given the rapid diffusion of LLM-based AI tools in educational settings, this gap presents a critical need for systematic comparative analysis that can inform pedagogically grounded tool selection.

Given the lack of comparative research on how LLM-based AI assistants function as explanation providers in EFL writing instruction, this study addresses the following research questions:

1. In what ways do ChatGPT, Claude, Gemini, and DeepSeek differ when delivering explanations of grammatical concepts for EFL writing instruction?
2. What are the comparative capabilities of these four AI assistants in clarifying vocabulary usage principles for EFL writing teachers?
3. How do the four AI systems contrast in their approaches to explaining sentence structure for EFL writing pedagogy?
4. What distinguishes these four AI assistants from one another in their ability to elucidate cohesion and coherence principles for EFL writing instruction?

2. LITERATURE REVIEW

2.1 Prompt Engineering Essentials for LLM-Based AI Assistants

Khan (2024) presents a clear framework for constructing effective prompts when interacting with Large Language Model

Vol 9, No 1 (2026): ESTEEM

(LLM)-based AI tools (Figure 1). A well-structured prompt guides the AI to generate responses that are accurate, relevant, and tailored to user needs (Korzynski et al., 2023; Lin, 2023). Key elements include role assignment, which sets the persona or expertise the AI should adopt, such as a teacher, historian, or developer, so that the response aligns with the intended voice and perspective. Providing context or background information further helps the AI understand the situation or purpose, enabling it to generate more appropriate and specific responses. Equally important is the inclusion of a specific task, which clearly states what the AI is expected to do, such as summarizing a text, giving feedback, or translating a sentence. These three components, according to Khan (2024), lay the foundation for a more purposeful and targeted interaction.

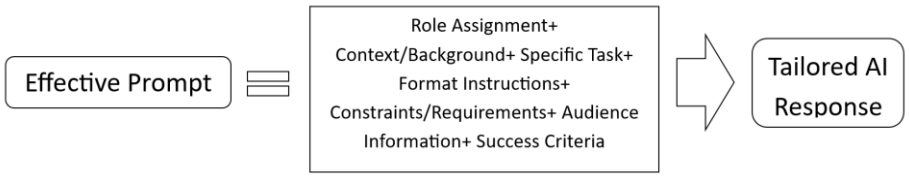

Figure 1. Visual representation of Khan's (2024) framework for effective prompt construction

To further refine AI-generated output, Khan (2024) emphasizes the importance of format instructions, which shape the tone, length, and structure of responses. Constraints or requirements, such as word limits or content boundaries, help maintain focus and relevance. Audience specification ensures that explanations are pitched at an appropriate level, while success criteria define what constitutes an effective response, such as clarity, completeness, or instructional usefulness. When integrated systematically, these prompt components enhance the consistency and pedagogical quality of AI-generated explanations. For this reason, Khan's framework is particularly suitable for studies that seek to examine explanatory functions of LLMs in educational contexts, including EFL writing instruction, where clarity, scaffolding, and audience alignment are essential.

2.2 Research novelty and contribution

This research presents a novel contribution to the growing body of literature on large language models (LLMs) in education by specifically examining their pedagogical utility as explanation providers in EFL writing instruction. While prior studies have investigated LLMs in areas such as automated essay assessment (Jiang et al., 2023; Yavuz et al., 2025), feedback generation ((Meyer et al., 2024), and educational content creation (Lang et al., 2024), the present study uniquely explores how multiple advanced LLMs (ChatGPT, Claude, Gemini, and DeepSeek) perform in delivering clear and pedagogically effective explanations of complex linguistic concepts central to EFL writing. By focusing on grammatical explanations, vocabulary usage, sentence structure, and cohesion/coherence principles, the study extends beyond output evaluation to examine the instructional depth and clarity of LLM-generated content, an area that remains underexplored in existing literature.

Previous studies have predominantly focused on single AI systems or on outcome-oriented functions such as feedback provision and assessment, offering limited insight into how different LLM-based tools explain core linguistic concepts to support EFL writing development. As a result, comparative evidence on the pedagogical quality of AI-

Vol 9, No 1 (2026): ESTEEM

generated explanations across multiple LLMs remains scarce.

Furthermore, this comparative analysis introduces a new dimension to evaluating AI-powered tools: their capacity to function as pedagogical assistants that support EFL educators in explicating intricate language features. Rather than emphasizing performance metrics alone, this research foregrounds the explanatory competence of LLMs, offering valuable insights into their potential for integration into language instruction.

Importantly, the use of Khan's (2024) prompt engineering framework directly informs the methodological design of this study by enabling the development of standardized, pedagogically grounded prompts across all AI systems. This methodological approach addresses a key limitation in prior comparative studies, where inconsistent prompting often constrains meaningful comparison of AI-generated explanations.

By systematically controlling prompt structure and focusing on explanatory clarity, depth, and instructional value, the present study responds directly to gaps in the existing literature and offers empirically grounded guidance for EFL educators, curriculum designers, and educational policymakers

METHODS

To systematically investigate how ChatGPT, Claude, Gemini, and DeepSeek support EFL teachers in delivering language explanations, this study follows a structured three-phase research flow. Each phase is designed to ensure methodological rigor while providing practical insights into AI-assisted language instruction.

3.1 Phase 1: Preliminary analysis and prompt design

Before evaluating the AI systems, a standardized set of prompts was created based on Khan's (2024) framework for language explanations, as shown in Table 1. These prompts ensure consistency across AI responses while addressing key aspects of writing instruction: grammar, vocabulary usage, sentence structure, and cohesion/coherence.

Table 1. Language explanation prompts per Khan's (2024) framework categories

Aspects of language explanations	Prompts
Grammar Explanation	**Role Assignment:** You are an experienced EFL writing instructor with expertise in teaching grammar to non-native English speakers. **Context/Background:** I am teaching intermediate-level EFL students who struggle with correctly using conditional sentences in their academic essays. Many make errors with conditional structures when discussing hypothetical situations. **Specific Task:** Explain the concept of conditional sentences (zero, first, second, and third conditionals) as they apply to academic writing. **Format Instructions:** Structure your explanation in clear sections with examples that demonstrate each type of conditional. Keep your explanation concise yet comprehensive. **Constraints/Requirements:** Use straightforward terminology appropriate

	for intermediate EFL learners. Include at least one example of each conditional type in an academic writing context. **Audience Information:** Your explanation will be presented to adult EFL learners at an intermediate level who need to improve their academic writing skills. **Success Criteria:** A successful explanation will help students distinguish between different conditional types, understand their structure and appropriate usage, and be able to apply them correctly in their own academic writing.
Vocabulary Usage	**Role Assignment:** You are an experienced EFL writing instructor with expertise in teaching academic vocabulary to non-native English speakers. **Context/Background:** I am teaching intermediate-level EFL students who struggle with using appropriate academic vocabulary in their essays. Many rely on basic, conversational vocabulary that lacks precision. **Specific Task:** Explain the concept of formal vs. informal vocabulary choices in academic writing and how to select appropriate academic vocabulary. **Format Instructions:** Structure your explanation in clear sections with examples that demonstrate vocabulary appropriateness. Keep your explanation concise yet comprehensive. **Constraints/Requirements:** Use straightforward terminology appropriate for intermediate EFL learners. Include at least five examples of informal words/phrases paired with their more appropriate academic alternatives. **Audience Information:** Your explanation will be presented to adult EFL learners at an intermediate level who need to improve their academic writing skills. **Success Criteria:** A successful explanation will help students understand the importance of vocabulary choice in academic contexts, recognize informal language in their writing, and know how to replace it with more appropriate academic alternatives.
Sentence Structure	**Role Assignment:** You are an experienced EFL writing instructor with expertise in teaching sentence construction to non-native English speakers. **Context/Background:** I am teaching intermediate-level EFL students who struggle with sentence variety in their academic essays. Many write primarily in simple sentences, making their writing repetitive and unsophisticated. **Specific Task:** Explain the concept of sentence variety and how to incorporate complex and compound sentences effectively in academic writing. **Format Instructions:** Structure your explanation in clear sections with

	examples that demonstrate different sentence structures. Keep your explanation concise yet comprehensive. **Constraints/Requirements:** Use straightforward terminology appropriate for intermediate EFL learners. Include examples of simple, compound, complex, and compound-complex sentences in an academic context. **Audience Information:** Your explanation will be presented to adult EFL learners at an intermediate level who need to improve their academic writing skills. **Success Criteria:** A successful explanation will help students understand different sentence types, recognize the impact of sentence variety on writing quality, and be able to incorporate more sophisticated sentence structures in their own academic writing.
Cohesion/Coherence Concepts	**Role Assignment:** You are an experienced EFL writing instructor with expertise in teaching paragraph and essay organization to non-native English speakers. **Context/Background:** I am teaching intermediate-level EFL students who struggle with creating cohesive and coherent paragraphs in their academic essays. Their writing often feels disconnected, with abrupt transitions between ideas. **Specific Task:** Explain the concepts of cohesion and coherence in academic writing, including how to use transitions and other linking devices effectively. **Format Instructions:** Structure your explanation in clear sections with examples that demonstrate cohesive writing techniques. Keep your explanation concise yet comprehensive. **Constraints/Requirements:** Use straightforward terminology appropriate for intermediate EFL learners. Include examples of cohesive devices (transitions, references, substitution) and how they contribute to overall coherence. **Audience Information:** Your explanation will be presented to adult EFL learners at an intermediate level who need to improve their academic writing skills. **Success Criteria:** A successful explanation will help students understand the difference between cohesion and coherence, recognize disconnected writing, and be able to implement cohesive devices to create smoother, more logical academic texts.

This careful prompt design creates a controlled environment for obtaining comparable responses across all four AI systems, focusing on their explanatory capabilities in each language aspect relevant to EFL writing instruction.

3.2 Phase 2: Data collection

Vol 9, No 1 (2026): ESTEEM

The selection of ChatGPT, Claude, Gemini, and DeepSeek was guided by several interrelated considerations relevant to contemporary EFL instructional contexts. First, these AI assistants represent some of the most widely used and visible LLM-based tools currently accessible to educators, making them pedagogically relevant for comparative analysis. Second, each system reflects different design philosophies and developmental orientations, which are expected to influence how linguistic explanations are generated and structured. Third, all four tools provide free-access versions, a critical factor in many EFL settings where institutional or financial constraints limit access to premium technologies. Finally, the inclusion of these AI assistants allows the study to represent diverse LLM ecosystems, offering a balanced comparison across platforms with differing architectures, interaction styles, and instructional affordances.

The Data Collection Phase involves submitting the standardized prompts to four widely-used AI tools, ChatGPT, Claude, Gemini, and DeepSeek, under controlled conditions. To ensure consistency and eliminate external biases, all prompts are submitted simultaneously, using the same wording and within the same time frame. Moreover, only the free versions of these AI tools are used in this study. This decision reflects the practical realities of many educational settings, especially in under-resourced EFL contexts where premium subscriptions may be inaccessible (Pavlova et al., 2024; Tsai et al., 2023). Despite this constraint, the use of well-structured prompts and uniform conditions is expected to produce valid and reliable comparative data (Konet et al., 2024). Responses from each AI system are gathered for each category of language explanation, and all interactions, including potential limitations, inconsistencies, or failures, are meticulously documented to support the comparative analysis.

3.3 Phase 3: Comparative analysis

In the Comparative Analysis Phase, the responses generated by ChatGPT, Claude, Gemini, and DeepSeek AI are analyzed using thematic analysis adapted from Braun and Clarke's (2021) framework. The comparative analysis focused on clarity, conceptual depth, organizational structure, contextual appropriateness, and instructional value of the AI-generated explanations. The process begins with familiarization with the data, followed by generating initial codes related to clarity, accuracy, structure, contextual appropriateness, and instructional value. These codes are then organized into themes that highlight each AI's strengths and limitations. Themes are reviewed and refined to ensure coherence and distinctiveness, then defined and named to facilitate interpretation. The final analysis presents a comparative narrative that captures how each AI supports or falls short in addressing EFL learners' writing challenges. By employing Braun and Clarke's thematic analysis, this phase ensures a nuanced and replicable approach to understanding the role of free-access AI tools in supporting language explanations within EFL writing instruction.

3. RESULTS AND DISCUSSION

4.1 RQ1: Grammatical explanation differences in ChatGPT, Claude, Gemini, and DeepSeek for EFL writing

In examining how different LLM-based AI Assistants convey grammatical concepts, it is crucial to consider not only content accuracy but also the tone and style of delivery. These elements greatly impact learners' engagement, comprehension, and retention, particularly in language learning where tone affects accessibility. Each AI model reflects its design priorities and user experience goals (Centa Strahovnik, 2023; Munn & Henrickson, 2024). Thus,

Vol 9, No 1 (2026): ESTEEM

comparing their tone and style provides important insights into their pedagogical potential for supporting EFL learners' understanding of complex grammar points.

Table 2. Tone and Style in Grammar Instruction

Model	Tone	Style Characteristics
ChatGPT	Instructional, clear, and friendly	Step-by-step explanation with academic usage focus. Emphasizes usability with checkmarks and tips.
Claude	Academic and formal	Emphasizes conceptual depth and analytical thinking, offering meta-awareness (e.g., how conditionals support argumentation).
Gemini	Conversational and supportive	Directly addresses learners ("Alright everyone…"), balancing friendliness and clarity. Uses bulleted lists for accessibility.
DeepSeek	Concise and neutral	More succinct with less elaboration, suitable for quick reference rather than deep explanation.

As shown in Table 2, while each model has strengths, their tone and style reflect different pedagogical orientations (Adetayo et al., 2024; Alnasib & Alharbi, 2024). ChatGPT's user-centered, academic focus supports learners seeking clarity and practical guidance, while Claude's emphasis on analytical depth suits advanced users aiming to develop critical linguistic awareness. Gemini's casual tone lowers affective barriers, benefiting younger or less confident learners. DeepSeek's brevity aids quick reviews but may lack scaffolding for deeper understanding. These differences, as noted by Gao et al. (2024), highlight the importance of aligning AI tools with teaching goals and learner needs.

To further illustrate these differences, ChatGPT typically presented grammatical explanations in a step-by-step instructional format, often followed by clearly labeled examples and brief usage tips. For example, in explaining verb tense distinctions, ChatGPT explicitly stated the rule, demonstrated correct usage through multiple examples, and highlighted common learner errors. In contrast, Claude framed grammatical explanations within broader linguistic reasoning, emphasizing how grammatical choices function in argumentation and academic discourse rather than focusing solely on procedural rules.

Table 3. Depth of grammatical explanation

Model	Conceptual Depth	Nuance in Usage
ChatGPT	Moderate to high	Offers examples and small usage notes (e.g., deliberate mixing at advanced levels).
Claude	High	Expands into mixed concepts, counterfactual reasoning, and disciplinary applications.
Gemini	Moderate	Explains subtle differences using modals (e.g., might, could), includes interpretation of example implications.

Vol 9, No 1 (2026): ESTEEM

DeepSeek	Basic to moderate	Gives only basic descriptions; lacks deeper commentary or usage notes beyond structure and example.

The differing depth of explanation across models reflects distinct pedagogical priorities, impacting learners' mastery of complex grammar (Table 3). Claude offers rich conceptual elaboration, promoting language acquisition and metalinguistic awareness suited for advanced learners. ChatGPT and Gemini balance clarity and nuance, providing accessible insights ideal for intermediate learners. In contrast, DeepSeek's brief explanations, while helpful for quick reference, may limit deeper grammatical understanding. As Kayalı et al. (2023) and Lee et al. (2023) emphasize, selecting AI tools for grammar instruction requires careful attention to the depth of linguistic input to foster authentic, flexible language use.

Table 4. Structure and organization of grammar explanations

Model	Organization Style	Logical Flow
ChatGPT	Numbered breakdown with checklists, tips, and clear academic examples	Very structured and digestible
Claude	Structured by type, includes rationale and analytical insight	Hierarchical and scholarly
Gemini	Uses bullet points and embedded guidance per section	Student-friendly with integrated commentary
DeepSeek	Very direct and schematic; structure-focused	Efficient but lacks elaborative flow

The structural organization of each model, as shown in Table 4, reflects their communicative style and pedagogical assumptions about how learners process information (Holmes & Tuomi, 2022). ChatGPT uses a structured format with numbered steps and checklists, aiding learners who need clarity and explicit scaffolding. Claude's scholarly, hierarchical structure suits users seeking analytical depth and theoretical framing, ideal for advanced audiences. Gemini employs student-centered bullet points with contextual guidance, promoting intuitive understanding. Meanwhile, DeepSeek's schematic layout offers quick access but may hinder deeper learning by lacking transitional cues or reflective commentary. As Barrera Castro et al. (2024) and Walter (2024) note, aligning model delivery with learners' needs is key to instructional effectiveness.

Table 5. Academic writing application

Model	Relevance to EAP (English for Academic Purposes)
ChatGPT	Strong. Links to academic reasoning and writing conventions (e.g., causes/effects).
Claude	Very strong. Emphasizes argumentation, research analysis, and disciplinary language use.
Gemini	Strong. Makes usage practical and considers writing improvement strategies (e.g., clarity, subtlety).

Vol 9, No 1 (2026): ESTEEM

DeepSeek	Moderate. Provides academic examples but lacks expansion on academic writing strategies.

When applying grammatical instruction in English for Academic Purposes (EAP), the alignment between explanation and academic discourse is crucial, as shown by Table 5. Claude and ChatGPT go beyond correctness by linking grammar to argumentation, cause-effect reasoning, and discipline-specific language, supporting persuasive and critical writing. Gemini, though less analytical, connects grammar to clarity and precision, key traits in academic writing. DeepSeek, however, provides correct examples but lacks deeper integration with academic discourse. As Malik et al. (2023) note, the more a model connects grammar to rhetorical and communicative goals, the more effective it is for EAP instruction.

Table 6. Learner Engagement and Accessibility

Model	Pedagogical Engagement	Suitability for Learners
ChatGPT	Interactive tone, invites practice, includes visuals (✓) and tips	Ideal for intermediate EFL learners
Claude	Focuses on depth over engagement, more reflective than interactive	Best for advanced EFL or academic learners
Gemini	Friendly and accessible, balances clarity with example-based learning	Suitable for upper beginner to intermediate learners
DeepSeek	Minimal interactivity, assumes quick assimilation	Better for self-directed or quick-reference learners

As shown in Table 6, the four AI models differ in their engagement and suitability for various EFL proficiency levels. ChatGPT, with its interactive tone and practical tips, is ideal for intermediate learners. Claude, offering depth and reflection, is better suited for advanced learners. Gemini strikes a balance of clarity and friendliness, making it accessible for upper beginner to intermediate learners. DeepSeek, with its straightforward, minimal-interactive approach, caters to self-directed learners. These differences align with studies highlighting the importance of adapting tools to learners' needs and proficiency levels (Huang et al., 2025; Xu & Li, 2024).

From a classroom perspective, these findings suggest that ChatGPT may be particularly suitable for structured grammar instruction where learners require explicit guidance and practice-oriented explanations. Claude may be more appropriate for advanced EFL or academic writing contexts that emphasize grammatical reasoning and critical language awareness. Gemini appears well suited for learners who benefit from an accessible and supportive tone, while DeepSeek may function best as a supplementary or quick-reference tool rather than a primary source of grammatical instruction.

4.2 RQ2: Comparative capabilities of AI tools in vocabulary clarification for EFL writing

While all tools demonstrate capacity to distinguish formal from informal language, they exhibit notable differences in their explanatory frameworks and theoretical

Vol 9, No 1 (2026): ESTEEM

grounding. This analysis examines how ChatGPT, Claude, Gemini, and DeepSeek approach vocabulary clarification through four distinct dimensions, beginning with the conceptual depth each assistant provides when explaining vocabulary usage principles.

Table 7. Vocabulary explanation depth

Assistant	Strengths	Weaknesses
ChatGPT	Explains the concept clearly with a concise definition and rationale. The distinction between formal and informal vocabulary is simple and accessible.	Slightly surface-level; doesn't deeply explore the grammatical features or register awareness beyond vocabulary substitution.
Claude	Most in-depth explanation of formal vs. informal features, including phrasal verbs, contractions, connotations, and disciplinary context. Offers nuanced advice on vocabulary moderation and tone.	Might be slightly advanced for lower-level EFL teachers without scaffolding.
Gemini	Offers a friendly metaphor (pajamas at a party) to explain tone/register. Lists linguistic features clearly and connects word choice to tone, precision, and professionalism.	Slightly conversational and lacks technical vocabulary that advanced teachers might expect (e.g., lexical density, nominalization).
DeepSeek	Concise and accurate breakdown of vocabulary types. Shows a straightforward substitution method and practice exercise.	Lacks elaboration on why and how these substitutions affect academic tone or structure. Too brief for in-depth teacher training.

The varying depths of explanation across AI tools, as shown in Table 7, highlight a tension between accessibility and linguistic sophistication. Claude's detailed coverage of register features provides strong theoretical grounding for advanced instructors, while Gemini's metaphorical style favors immediate comprehension. Optimal AI selection thus depends on instructor expertise and pedagogical goals (Walter, 2024; Kim et al., 2022), advanced users may prefer Claude's nuanced approach, whereas those needing practical, classroom-ready explanations might favor ChatGPT or Gemini. DeepSeek's clear but minimalist style may require supplementation for effective teacher development.

Illustrative responses further demonstrate these contrasts. Gemini frequently explained vocabulary usage through relatable metaphors and context-rich examples, such as associating formal language with professional settings and informal language with casual situations. ChatGPT combined concise definitions with practical usage notes and classroom-oriented examples, while DeepSeek provided brief substitution-based explanations with minimal elaboration. Claude, by comparison, offered extended discussions of register, connotation, and disciplinary appropriateness, reflecting a more theory-driven approach to vocabulary explanation.

Table 8. Practical Support (Examples + Exercises) in vocabulary explanation

Assistant	Strengths	Weaknesses

Vol 9, No 1 (2026): ESTEEM

ChatGPT	Offers a side-by-side table and an optional sentence rewriting task. Great for EFL classroom modeling.	Limited variety in examples; doesn't cover many academic genres or disciplines.
Claude	Rich example set (5 formal/informal sentence transformations), and encourages contextual analysis of language use.	No hands-on activity or prompts for learners to practice directly.
Gemini	Broad list of substitutions and tips for checking tone/context. Encourages using a thesaurus/dictionary.	No structured activity or interactive task.
DeepSeek	Includes two rewrite practice sentences with modeled answers. Useful for quick drills.	Minimal variety; lacks scaffolding or explanation beyond the rewrites.

The use of practical examples and exercises shows differing priorities among AI assistants in balancing explanation and application, as presented in Table 8. ChatGPT and DeepSeek emphasize direct practice through rewriting tasks, supporting practice-oriented teaching, while Claude focuses on extensive examples without structured activities, requiring teachers to create their own. This highlights a key gap in AI tools: the inconsistent integration of explanation and scaffolded practice. As noted by Ruiz-Rojas et al. (2023) and Williyan et al. (2024), teachers may need to supplement AI content or combine outputs from multiple platforms for balanced instruction.

Table 9. Pedagogical tone and vocabulary explanation suitability

Assistant	Strengths	Weaknesses
ChatGPT	Very approachable and student-friendly. Encourages teachers to use tools like the AWL and learner dictionaries.	Might be slightly too simplified for teachers needing deeper theoretical grounding.
Claude	Offers high-level meta-awareness for teachers (e.g., connotation, moderation, audience consideration).	Might be too formal or abstract for some novice EFL teachers.
Gemini	Most engaging tone for less-experienced teachers. Uses relatable analogies and empathetic phrasing.	Could be too informal for professional teacher development contexts.
DeepSeek	Straightforward and minimalistic. Helpful for quick reference or summary slides.	Less supportive in explaining underlying pedagogical principles.

As presented in Table 9, each AI assistant's pedagogical tone reflects assumptions about teacher expertise and development needs. Gemini's accessible, metaphor-rich style aids immediate comprehension but may limit professional vocabulary growth, while Claude's academic tone supports linguistic development yet may challenge novice instructors. This tension between accessibility and rigor complicates AI tool use in teacher education, where both practical application and theoretical growth are crucial. Consistent with Triberti et al. (2024) and Watanabe (2024), institutions should strategically select tools based on pedagogical goals and teacher experience.

Table 10. Vocabulary organization and clarity

Vol 9, No 1 (2026): ESTEEM

Assistant	Strengths	Weaknesses
ChatGPT	Clearly structured into 5 sections with headings and bullet points. Visual table aids understanding.	Slightly linear; could include more cross-references (e.g., AWL list usage tied to examples).
Claude	Systematically lays out comparison, reasoning, and examples. Best logical flow.	Dense at times, especially for beginners.
Gemini	Clear use of lists, conversational structure, and examples.	Organization less formal, more like lecture notes than a training guide.
DeepSeek	Concise with numbered sections and clear contrasts.	Sparse headings and transitions; may feel abrupt or rushed.

The organizational approaches of these AI assistants highlight different priorities in information structure and cognitive support for teachers (Table 10). ChatGPT offers accessible entry points but may not sufficiently challenge experienced teachers, while Claude fosters deeper pedagogical reasoning but risks overwhelming beginners. This suggests that optimal AI use in teacher education could involve sequential exposure, starting with accessible tools like ChatGPT or Gemini, then advancing to Claude's more complex framing (Cordero et al., 2024). DeepSeek's minimalistic style, though limited for comprehensive training, remains valuable for immediate, actionable support.

Pedagogically, these findings indicate that Gemini may be especially effective for introducing vocabulary concepts and supporting engagement through contextualized explanation. ChatGPT may better support systematic vocabulary instruction that integrates definition, usage, and classroom application. Claude appears most suitable for teacher development or advanced instruction focusing on register awareness and academic vocabulary use, while DeepSeek may serve as a concise reference tool when time or instructional scope is limited.

4.3 RQ3: AI systems' contrasts in sentence structure explanation for EFL writing

AI writing assistants differ notably in how they teach sentence structure to EFL learners, reflecting distinct educational philosophies (Table 11). These differences go beyond style, revealing varied emphases on formal structure, relational understanding, or practical engagement, each influencing learner comprehension and writing development.

Table 11. Pedagogical approach to sentence structure

Theme	ChatGPT	Claude	Gemini	DeepSeek
Approach	Structured and formal, with clear headings and logical progression.	Organized with clear headings and academic focus on sentence relationships.	Informal and engaging with analogies to make sentence variety relatable.	Simple, straightforward approach, focusing on core concepts.
Organization	Organized into sections: sentence types,	Similar to ChatGPT but with a focus on	Emphasizes flow and clarity through practical examples	Brief sections with basic definitions and examples.

	tips for improvement, practice.	academic writing relationships.	and strategies.	
Focus	Academic, emphasizing clarity and logical flow of ideas.	Academic sophistication, showing relationships between ideas.	Clarity, engagement, and smooth flow, with an emphasis on readability.	Clear, concise, and practical for beginners.

Analysis of these AI systems reveals a spectrum of instructional philosophies. ChatGPT and Claude emphasize academic, structure-focused instruction, reflecting a cognitive academic language proficiency (CALP) framework (Tariq, 2025). In contrast, Gemini and DeepSeek adopt more accessible, simplified approaches, Gemini aligns with conceptual metaphor theory (Mizumoto et al., 2024), while DeepSeek's minimalist style suggests a focus on basic interpersonal communicative skills. These foundational differences shape how sentence structure is explained and which elements are prioritized. The varying depth and complexity in their explanations reflect differing assumptions about learner needs and syntactic development (Krajka & Olszak, 2024), highlighting the ongoing tension between detailed instruction and accessible simplification in EFL writing pedagogy.

For example, when explaining complex sentence structures, ChatGPT typically decomposed sentences into constituent clauses and explained their functions sequentially. Claude emphasized syntactic relationships and rhetorical purpose, often explaining why certain structures enhance meaning or argumentation. Gemini relied on simplified examples and analogies to illustrate sentence variety and flow, whereas DeepSeek provided brief definitions and examples with limited elaboration.

Table 12. Explanation of Sentence Types

Theme	ChatGPT	Claude	Gemini	DeepSeek
Detail in Explanation	Detailed explanation with examples and specific sentence structures.	Provides clear examples, but with a deeper focus on the relationships between clauses.	Brief yet clear explanation, with an emphasis on flow and clarity.	Basic definitions and examples without delving deeply into sentence relationships.
Types Covered	Simple, compound, complex, compound-complex with clear structure breakdown.	Simple, compound, complex, compound-complex with emphasis on relationships between ideas.	Simple, compound, complex, compound-complex; focuses more on flow and variety.	Simple, compound, complex, compound-complex with minimal elaboration.
Complexity	Provides nuanced insight	Focus on why certain sentence	Less focus on nuances, but	Straightforward, aimed at beginner

Vol 9, No 1 (2026): ESTEEM

into the use of each sentence type.	types express different relationships.	encourages variety in structure.	level learners with minimal complexity.

The variations in explanatory depth reflect different perspectives on effective sentence structure instruction (Table 12). Claude's focus on relationships between clauses suggests a cohesion-centered approach, while ChatGPT's structural breakdown aligns with traditional grammar-translation methods. Gemini emphasizes flow and variety, reflecting a communicative orientation that prioritizes reader experience over strict grammar. DeepSeek's minimalist style follows a basic English for Specific Purposes framework, offering functional guidance without theoretical complexity. These contrasts raise important questions about whether EFL writing instruction should prioritize structural correctness, rhetorical effectiveness, or communicative clarity, and whether these priorities can be balanced. The practical methods used by these AI systems highlight divergent views on bridging the gap between theoretical understanding and applied writing skills (Combs et al., 2023), revealing different assumptions about skill acquisition and practice.

Table 13. Applied Use & Engagement in Sentence Structure

Theme	ChatGPT	Claude	Gemini	DeepSeek
Practice Encouragement	Provides sentence transformation exercises for hands-on learning.	Offers practical exercises and suggests varied ways to combine sentences.	Encourages observation of academic texts and offers practical examples.	Includes simple exercises to combine sentences into varied structures.
Engagement Methods	Uses clear, structured tips and examples to guide learning.	Suggests combining sentences with varied conjunctions and beginnings.	Uses engaging analogies, like "spices in cooking," to make learning enjoyable.	Focuses on basic exercises and applications with clear, actionable steps.
Use of Examples	Provides several detailed examples to demonstrate sentence transformation.	Focuses on showing relationships between clauses with varied examples.	Provides engaging examples with varied sentence structures for clarity.	Provides clear examples, but less variety in terms of sentence structures.

These differences in application reflect contrasting theories of language skill development (Table 13). ChatGPT's transformation exercises align with cognitive skill acquisition theory, emphasizing deliberate practice (Cherukunnath & Singh, 2022). Claude's sentence combination approach reflects systemic functional linguistics, focusing on cohesion through conjunctions (Herman et al., 2023). Gemini's use of analogy and authentic texts draws from genre-based pedagogy and the comprehensible input hypothesis (Hamman-Ortiz et al., 2023). DeepSeek's pattern-based

Vol 9, No 1 (2026): ESTEEM

drills suggest a behaviorist model emphasizing repetition (Almohawes, 2024). These varied strategies raise important questions about balancing explicit instruction, guided practice, and authentic use in EFL writing. Each system's communicative style reveals assumptions about learners and contexts, shaping comprehension, engagement, and instructional impact.

These patterns suggest that ChatGPT may be most effective for explicit instruction in sentence construction and form-focused writing pedagogy. Claude may be better suited for higher-level writing instruction where sentence structure is closely linked to rhetorical meaning and academic style. Gemini may support learners who benefit from accessible explanations and illustrative examples, while DeepSeek may be appropriate for introductory instruction or rapid skill reinforcement.

Table 14. Tone and Learner Fit in sentence explanation

Theme	ChatGPT	Claude	Gemini	DeepSeek
Tone	Formal, instructional, and professional.	Formal, academic, with emphasis on academic writing sophistication.	Informal, conversational, and engaging with analogies.	Simple, direct, and practical.
Target Learner	EFL learners who need clear, structured, and professional academic writing guidance.	EFL learners looking to understand complex relationships in academic writing.	EFL learners who benefit from an engaging, easy-to-understand approach with an emphasis on clarity.	EFL beginners or learners who need straightforward, no-frills guidance.
Fit for Learners	Best for learners aiming for academic writing proficiency in structured settings.	Ideal for intermediate to advanced learners focused on academic sophistication.	Suitable for learners who enjoy a more relaxed and engaging learning environment.	Best for beginners or those who need quick, actionable insights.

These tonal and targeting differences underscore the importance of aligning AI tools with learner context in EFL writing instruction (Table 14). ChatGPT and Claude suit academic settings with formal writing demands, while Gemini's interactive style may support self-directed or communicative learning. DeepSeek's simplicity can aid lower-proficiency learners or quick remediation. As Pack and Maloney (2024) and Praphan and Praphan (2023) note, no single AI tool fits all contexts, effectiveness depends on how well the tool matches learner needs, instructional goals, and educational settings, highlighting the need for thoughtful integration over wholesale adoption.

4.4 R4: Distinctive AI approaches to cohesion and coherence in EFL writing

The comparative analysis of ChatGPT, Claude, Gemini, and DeepSeek shows that while all four AI tools emphasize the importance of cohesion and coherence in EFL writing, they differ in definitions, explanations, and pedagogical focus. These differences offer EFL teachers valuable

Vol 9, No 1 (2026): ESTEEM

options for selecting tools that best match their instructional goals and student needs. As AI becomes more integrated into language learning, understanding these distinctions is essential to ensure effective support for students' writing development.

Table 15. Definitions of Cohesion and Coherence

AI Assistant	Difference in Definition of Cohesion	Difference in Definition of Coherence
ChatGPT	Focuses on cohesion as the connection between sentences and ideas using cohesive devices.	Emphasizes coherence as the logical organization of ideas and clarity in the writing process.
Claude	Defines cohesion as the "glue" that links ideas within sentences, focusing on sentence-level devices.	Describes coherence as clarity in thought and the overall organization of ideas within the text.
Gemini	Describes cohesion as linking sentences using cohesive devices like transitions and reference words.	Explains coherence as the overall flow of ideas that are logically connected, ensuring the reader can easily follow the argument.
DeepSeek	Emphasizes cohesion as related to grammar and vocabulary tools that connect ideas at the sentence level.	Defines coherence as the logical flow of ideas where the overall argument is clear and easy to follow.

The analysis shows that although AI tools define cohesion and coherence differently, they consistently highlight their importance in writing quality (Table 15). This suggests that teachers must understand how each tool delivers feedback, as it affects instructional alignment. Supporting prior research, the findings highlight the need for teachers' critical engagement with technology (Pleasants et al., 2024). Pedagogically, teachers should make informed choices when integrating AI tools to ensure alignment with their instructional goals and feedback practices.

In practice, Claude tended to approach cohesion and coherence from a discourse-level perspective, emphasizing logical progression and rhetorical organization across paragraphs. ChatGPT operationalized these concepts by explicitly listing cohesive devices and demonstrating their use through before-and-after examples. Gemini often embedded cohesion and coherence within simplified sample texts, while DeepSeek focused on concise contrasts between weak and revised writing.

Table 16. Cohesion devices and explanation

AI Assistant	Cohesion Devices	Explanation of Cohesion
ChatGPT	Lists devices like transitions, references, substitution, ellipsis; provides clear examples.	Focuses on smooth sentence connections with simple explanations and before–after writing comparisons.
Claude	Includes transitions, references, repetition, substitution with detailed	Explains how transitions and references improve flow; uses clear examples to show

Vol 9, No 1 (2026): ESTEEM

	examples.	impact.
Gemini	Covers transitions, references, substitution, lexical cohesion with engaging examples.	Highlights sentence connection and flow, with examples for clarity.
DeepSeek	Lists transitions, references, repetition with brief examples.	Stresses how devices link ideas; contrasts weak and strong writing to show effect.

Table 16 shows that all AI assistants recognize key cohesion devices, such as transitions, reference words, and repetition, but differ slightly in emphasis and explanation style. ChatGPT and Gemini offer more accessible, example-driven descriptions, while Claude provides detailed categorization, and DeepSeek emphasizes contrast between weak and cohesive writing. These variations, aligned with Huang et al. (2025), suggest that teachers can choose specific AI tools depending on students' needs, whether for clarity, structure, or detailed understanding of cohesive writing.

Table 17. Coherence Tips and Examples

AI Assistant	Coherence Tips	Example Type
ChatGPT	Focus on one idea per paragraph, use logical order (e.g., time, cause-effect).	Weak vs. improved paragraph showing structure.
Claude	Arrange ideas logically, stay on topic, use clear topic sentences.	Before–after examples emphasizing organization.
Gemini	Organize logically, maintain focus, use transitions.	Disjointed vs. coherent writing comparison.
DeepSeek	Ensure relevance to main idea, use logical sequence (e.g., general → specific).	Non-coherent vs. revised coherent version.

The analysis highlights that AI writing assistants offer varied yet consistent strategies for promoting coherence, such as logical sequencing and clear topic development (Table 17). For teachers, this suggests the importance of understanding how each tool structures and presents coherence-related feedback. Rather than relying solely on AI-generated suggestions, teachers should play an active role in interpreting and contextualizing this feedback to align with their instructional approach (Evmenova et al., 2024). This enables more targeted writing instruction and supports students in developing a clearer understanding of coherence beyond automated responses.

Table 18. Interaction between Cohesion and Coherence

AI Assistant	Interaction between Cohesion and Coherence
ChatGPT	Cohesive devices help create coherence by logically linking ideas and maintaining flow.

Vol 9, No 1 (2026): ESTEEM

Claude	Cohesion supports coherence by organizing ideas clearly and logically.
Gemini	Cohesive devices ensure logical flow and contribute to overall coherence.
DeepSeek	Cohesion enhances coherence by providing smooth transitions and clarity in organization.

The findings in Table 18 show that all AI tools recognize the relationship between cohesion and coherence, offering guidance on how cohesive devices enhance the flow of ideas. These systems stress the importance of clear organization and logical transitions. For teachers, this underscores the need to select AI tools that provide context-sensitive feedback, helping students improve their writing skills. By understanding how each AI system links cohesion and coherence, teachers can adjust their strategies to align with the AI's strengths, ensuring their methods complement the feedback provided (Seo et al., 2024). This approach supports prior research, such as Walter (2024), emphasizing the need for teachers to critically assess AI tools in the classroom.

Table 19. Practice Exercises

AI Assistant	Practice Exercises
ChatGPT	Offers a worksheet or practice paragraph for cohesion and coherence.
Claude	Provides an activity to identify cohesive devices in a paragraph.
Gemini	Offers a task to rewrite a disjointed paragraph using cohesive devices.
DeepSeek	Provides a task to rewrite a disjointed paragraph, focusing on logical flow and coherence.

The practice exercises in Table 19 demonstrate how each AI tool supports teachers by providing different types of activities to reinforce cohesion and coherence in students' writing. This variety of practice exercises offers flexibility for teachers in how they integrate AI into their classroom activities. By understanding the specific types of practice exercises each AI system provides, teachers can choose AI tools that align with their teaching objectives, ensuring students engage in appropriate practice to enhance their writing skills (Ng et al., 2023). This approach empowers educators to use AI more effectively in supporting students' writing development.

From an instructional standpoint, these findings suggest that Claude may be particularly effective for teaching advanced writing skills related to discourse organization and coherence development. ChatGPT may be more suitable for explicit instruction on cohesive devices and paragraph structure, while Gemini can support introductory discussions of text flow through accessible exemplification. DeepSeek may be best used as a supplementary tool for quick revision or reinforcement rather than comprehensive instruction.

5. CONCLUSION

This comparative study of ChatGPT, Claude, Gemini, and DeepSeek reveal distinct pedagogical profiles that align with different EFL instructional needs and contexts. Each AI assistant demonstrates unique strengths in explaining writing concepts: ChatGPT offers structured, accessible explanations with practical exercises that balance theory and application; Claude provides deeper conceptual analysis

and sophisticated linguistic frameworks ideal for advanced academic contexts; Gemini employs engaging metaphors and conversational approaches that lower affective barriers for less confident learners; and DeepSeek delivers concise, straightforward explanations suited for quick reference and self-directed learning. Rather than identifying a single optimal system, this research underscores the importance of strategic selection based on specific pedagogical objectives, learner profiles, and instructional contexts. As these technologies continue to evolve, EFL educators would benefit from a differentiated approach, leveraging each system's comparative advantages to address diverse aspects of writing instruction, from grammatical concepts and vocabulary usage to sentence structure and cohesion principles. Future research might explore how these complementary capabilities could be systematically integrated into comprehensive EFL writing curricula that maximize technological affordances while addressing the nuanced needs of language learners.

From a practical standpoint, the findings suggest that the pedagogical value of LLM-based AI assistants in EFL writing instruction depends strongly on learner level, class size, and instructional purpose. ChatGPT appears particularly effective for intermediate learners and larger classes that benefit from structured explanations and practice-oriented support. Claude is best suited for advanced learners or smaller, academically focused classes where deeper linguistic reasoning and discourse-level understanding are emphasized. Gemini offers advantages for lower to intermediate learners by reducing affective barriers through accessible and engaging explanations, while DeepSeek may serve as a supplementary tool for quick review or self-directed learning. These distinctions underscore the importance of strategic, context-sensitive AI integration rather than uniform adoption across instructional settings.

6. DECLARATION OF AI ASSISTANCE

This research article was developed solely by the authors, including the selection of data collection and analysis techniques, the process of gathering and analyzing data, interpretation of findings, and overall research design. The conceptualization, critical analysis, and presentation of results are entirely the authors' original work. ChatGPT was utilized solely to enhance the grammar, clarity, and readability of the manuscript. The use of AI assistance did not influence the intellectual content, arguments, or originality of the study.

7. REFERENCES

Adetayo, A. J., Aborisade, M. O., & Sanni, B. A. (2024). Microsoft Copilot and Anthropic Claude AI in education and library service. *Library Hi Tech News*. https://doi.org/10.1108/LHTN-01-2024-0002

Almohawes, M. (2024). Second language acquisition theories and how they contribute to language learning. *World Journal of English Language*, *14*(3), 181. https://doi.org/10.5430/wjel.v14n3p181

Alnasib, B. N. M., & Alharbi, N. S. (2024). Challenges and motivation: Assessing Gemini's impact on undergraduate EFL students in classroom settings. *World Journal of English Language*, *14*(5), 501. https://doi.org/10.5430/wjel.v14n5p501

Barrera Castro, G. P., Chiappe, A., Becerra Rodriguez, D. F., & Sepulveda, F. G. (2024). Harnessing AI for education 4.0: Drivers of personalized learning. *Electronic Journal of E-Learning*, *22*(5), 01–14. https://doi.org/10.34190/ejel.22.5.3467

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English With Technology*, *2023*(1). https://doi.org/10.56297/BKAM1691/WIEO1749

Braun, V., & Clarke, V. (2021). Thematic

analysis: A practical guide. In *Sage Publications* (p. 320). https://books.google.se/books?id=Hr11DwAAQBAJ&hl=sv

Centa Strahovnik, M. (2023). Identiteta in pogovorni sistemi umetne inteligence. *Bogoslovni Vestnik*, *83*(4). https://doi.org/10.34291/BV2023/04/Centa

Cherukunnath, D., & Singh, A. P. (2022). Exploring cognitive processes of knowledge acquisition to upgrade academic practices. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.682628

Combs, K., Lu, H., & Bihl, T. J. (2023). Transfer learning and analogical inference: A critical comparison of algorithms, methods, and applications. *Algorithms*, *16*(3), 146. https://doi.org/10.3390/a16030146

Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2024). Integration of generative artificial intelligence in higher education: Best practices. *Education Sciences*, *15*(1), 32. https://doi.org/10.3390/educsci15010032

Evmenova, A. S., Regan, K., Mergen, R., & Hrisseh, R. (2024). Improving writing feedback for struggling writers: Generative AI to the rescue? *TechTrends*, *68*(4), 790–802. https://doi.org/10.1007/s11528-024-00965-y

Extance, A. (2023). ChatGPT has entered the classroom: How LLMs could transform education. *Nature*, *623*(7987), 474–477. https://doi.org/10.1038/d41586-023-03507-3

Gao, L. (Xuehui), López-Pérez, M. E., Melero-Polo, I., & Trifu, A. (2024). Ask ChatGPT first! Transforming learning experiences in the age of artificial intelligence. *Studies in Higher Education*, *49*(12), 2772–2796. https://doi.org/10.1080/03075079.2024.2323571

Hamman Ortiz, L., Santiago Schwarz, V., Hamm- Rodríguez, M., & Gort, M. (2023). Engaging teachers in genre-based pedagogy for writing arguments: A case study of shifts in practice and understanding. *TESOL Quarterly*, *57*(2), 402–432. https://doi.org/10.1002/tesq.3156

Herman, H., Rafiek, M., Agustina, T., Saddhono, K., Malabar, S., Saputra, N., & Purba, R. (2023). Exploring the metafunctions to improve EFL learners' writing ability in the perspective of systemic functional linguistics. *Research Journal in Advanced Humanities*, *4*(2). https://doi.org/10.58256/rjah.v4i2.1195

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Huang, F., Peng, D., & Teo, T. (2025). AI affordances and EFL learners' speaking engagement: The moderating roles of gender and learner type. *European Journal of Education*, *60*(1). https://doi.org/10.1111/ejed.70041

Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages*, *8*(4), 247. https://doi.org/10.3390/languages8040247

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kayalı, B., Yavuz, M., Balat, Ş., & Çalışan, M. (2023). Investigation of student experiences with ChatGPT-Supported online learning applications in higher education. *Australasian Journal of Educational Technology*, *39*(5), 20–39. https://doi.org/10.14742/ajet.8915

Khan, I. (2024). *The quick guide to prompt engineering: Generative AI tips and tricks for ChatGPT, Bard, Dall-E, and Midjourney*. Wiley.

Kim, J., Lee, H., & Cho, Y. H. (2022).

Vol 9, No 1 (2026): ESTEEM

Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

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

Konet, A., Thomas, I., Gartlehner, G., Kahwati, L., Hilscher, R., Kugley, S., Crotty, K., Viswanathan, M., & Chew, R. (2024). Performance of two large language models for data extraction in evidence synthesis. *Research Synthesis Methods*, *15*(5), 818–824. https://doi.org/10.1002/jrsm.1732

Korzynski, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, *11*(3), 25–37. https://doi.org/10.15678/EBER.2023.110302

Krajka, J., & Olszak, I. (2024). "AI, will you help?" How learners use Artificial Intelligence when writing. *XLinguae*, *17*(1), 34–48. https://doi.org/10.18355/XL.2024.17.01.03

Lang, G., Triantoro, T., & Sharp, Ja. (2024). Large language models as ai-powered educational assistants: Comparing GPT-4 and Gemini for writing teaching cases. *Journal of Information Systems Education*, *35*(3), 390–407. https://doi.org/10.62273/YCIJ6454

Lee, D., Kim, H., & Sung, S.-H. (2023). Development research on an AI English learning support system to facilitate learner-generated-context-based learning. *Educational Technology Research and Development*, *71*(2), 629–666. https://doi.org/10.1007/s11423-022-10172-2

Lin, Z. (2023). Why and how to embrace AI such as ChatGPT in your academic life. *Royal Society Open Science*, *10*(8). https://doi.org/10.1098/rsos.230658

Liu, Z.-M., Hwang, G.-J., Chen, C.-Q., Chen, X.-D., & Ye, X.-D. (2024). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1–25. https://doi.org/10.1080/09588221.2024.2389923

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence- written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570–581. https://doi.org/10.1002/asi.24750

Makridakis, S., Petropoulos, F., & Kang, Y. (2023). Large language models: Their success and impact. *Forecasting*, *5*(3), 536–549. https://doi.org/10.3390/forecast5030030

Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective. *International Journal of Educational Research Open*, *5*(October), 100296. https://doi.org/10.1016/j.ijedro.2023.100296

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, *3*(2), 100116. https://doi.org/10.1016/j.rmal.2024.100116

Munn, L., & Henrickson, L. (2024). Tell me a story: A framework for critically

investigating AI language models. *Learning, Media and Technology*, 1–17. https://doi.org/10.1080/17439884.2024.2327024

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, *71*(1), 137–161. https://doi.org/10.1007/s11423-023-10203-6

Pack, A., & Maloney, J. (2024). Using artificial intelligence in TESOL: Some ethical and pedagogical considerations. *TESOL Quarterly*, *58*(2), 1007–1018. https://doi.org/10.1002/tesq.3320

Pavlova, A., Gerazov, B., & Barreiro, A. (2024). Large language models and OpenLogos: An educational case scenario. *Open Research Europe*, *4*, 110. https://doi.org/10.12688/openreseurope.17605.1

Pleasants, J., Gui, X., Krutka, D. G., Logan, C., & Heath, M. K. (2024). Coming to critical technology consciousness: A phenomenological study of educators. *Learning, Media and Technology*, 1–14. https://doi.org/10.1080/17439884.2024.2438925

Praphan, P. W., & Praphan, K. (2023). AI technologies in the ESL/EFL writing classroom: The villain or the champion? *Journal of Second Language Writing*, *62*, 101072. https://doi.org/10.1016/j.jslw.2023.101072

Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*, *15*(15), 11524. https://doi.org/10.3390/su151511524

Seo, K., Yoo, M., Dodson, S., & Jin, S. H. (2024). Augmented teachers: K–12 teachers' needs for artificial intelligence's complementary role in personalized learning. *Journal of Research on Technology in Education*, *0*(0), 1–18. https://doi.org/10.1080/15391523.2024.2330525

Tariq, U. (2025). Nexus of essay writing and computer-assisted language learning (CALL) in English language classroom. *Interactive Technology and Smart Education*, *22*(1), 103–133. https://doi.org/10.1108/ITSE-12-2023-0246

Triberti, S., Di Fuccio, R., Scuotto, C., Marsico, E., & Limone, P. (2024). "Better than my professor?" How to develop artificial intelligence tools for higher education. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1329605

Tsai, M. L., Ong, C. W., & Chen, C. L. (2023). Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT. *Education for Chemical Engineers*, *44*(April), 71–95. https://doi.org/10.1016/j.ece.2023.05.001

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

Watanabe, A. (2024). Have courage to use your own mind, with or without AI: The relevance of Kant's enlightenment to higher education in the age of artificial intelligence. *Electronic Journal of E-Learning*, 00–00. https://doi.org/10.34190/ejel.21.5.3229

Williyan, A., Fitriati, S. W., Pratama, H., & Sakhiyya, Z. (2024). AI as co-creator: Exploring Indonesian EFL teachers' collaboration with AI in content development. *Teaching English With Technology*, *24*(2), 5–21. https://doi.org/10.56297/vaca6841/LRDX3699/RZOH5366

Xu, J., & Li, J. (2024). Effects of AI affordances on student engagement in EFL classrooms: A structural equation modelling and latent profile analysis. *European Journal of Education*, *59*(4). https://doi.org/10.1111/ejed.12808

Yavuz, F., Çelik, Ö., & Yavaş Çelik, G. (2025). Utilizing large language models

Vol 9, No 1 (2026): ESTEEM

for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *British Journal of Educational Technology*, *56*(1), 150–166. https://doi.org/10.1111/bjet.13494