

A COMPARATIVE EVALUATION OF CHATGPT, GEMINI, AND PERPLEXITY FEEDBACK FOR B1-B2 EFL LEARNERS

Malika Abduvakilovna KASIMOVA

PhD in Pedagogical Sciences

State Security Service of the Republic of Uzbekistan
"Temurbeklar maktabi" Military-Academic Lyceum
Tashkent, Uzbekistan

Dilrabo Xamdamovna BABAKULOVA

Master's degree

Westminster International University in Tashkent
Tashkent, Uzbekistan

CHATGPT, GEMINI VA PERPLEXITY GRAMMATIK QAYTA ALOQASINING B1-B2 DARAJADAGI INGLIZ TILI O'RGANUVCHILARI UCHUN QIYOSIY BAHOLASH

Malika Abduvakilovna KASIMOVA

Pedagogika fanlari nomzodi (PhD)

O'zbekiston Respublikasi davlat xavfsizlik xizmati
"Temurbeklar maktabi" harbiy-akademik litseyi
Toshkent, O'zbekiston

Dilrabo Xamdamovna BABAKULOVA

Magistrant

Toshkent Xalqaro Vestminster Universiteti
Toshkent, O'zbekiston

СРАВНИТЕЛЬНАЯ ОЦЕНКА ГРАММАТИЧЕСКОГО ФИДБЭКА ЧАТГПТ, GEMINI И PERPLEXITY ДЛЯ ИЗУЧАЮЩИХ АНГЛИЙСКИЙ ЯЗЫК НА УРОВНЯХ B1-B2

Малика Абдувакилевна КАСИМОВА

Кандидат педагогических наук (PhD)

Военно-академический лицей «Темурбеклар мактаби» Службы государственной
безопасности Республики Узбекистан

Ташкент, Узбекистан malika.kasimova89@gmail.com,
<https://orcid.org/0009-0008-6645-903X>

Дилрабо Хамдамовна БАБАКУЛОВА

Магистрант

Международный Вестминстерский университет в Ташкенте
Ташкент, Узбекистан dbabakulova@wiut.uz, <https://orcid.org/0009-0009-2610-9198>

**For citation (iqtibos keltirish uchun, для
цитирования):**

Kasimova M.A., Babakulova D.X. A Comparative

Abstract. In recent years, Artificial Intelligence (AI) has become an increasingly influential component of educational practice, particularly in language learning. AI-powered tools are now capable

Evaluation of Chatgpt, Gemini, and Perplexity Feedback for B1-B2 EFL Learners.// O'zbekistonda xorijiy tillar. — 2025. — 11-jild, № 6. — B. 91-109.

<https://doi.org/10.36078/1767688041>

Received: September 30, 2025

Accepted: December 17, 2025

Published: December 20, 2025

Copyright © 2025 by author(s).

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



of offering immediate, personalized, and contextually sensitive support, helping learners develop writing accuracy, critical thinking, and revision strategies. In EFL contexts, AI enhances access to high-quality feedback and reduces teachers' workload by automating lower-level tasks, allowing instructors to focus on formative guidance. This study investigates the expanding role of AI in EFL writing instruction, with a special focus on grammar feedback generated by ChatGPT, Gemini, and Perplexity. The case study involved 16–17-year-old learners from a military-academic lyceum, representing B1–B2 proficiency levels. Using researcher-designed rubrics, AI-generated feedback was evaluated across three key dimensions: technical accuracy, pedagogical clarity, and linguistic appropriacy. The results show that ChatGPT provides human-like, detailed, and supportive explanations that foster understanding, while Gemini delivers concise and structured feedback that suits more autonomous learners. Perplexity, in contrast, offers factual but brief comments, emphasizing accuracy over elaboration. These findings suggest that while AI tools can enhance feedback speed and consistency, their pedagogical value depends heavily on teacher mediation and careful prompt design. The study concludes that AI should not replace the teacher's role but rather complement it acting as a co-pilot that supports learners through timely, personalized guidance. The research contributes to AI-informed pedagogy by linking feedback theory, rubric-based evaluation, and classroom practice.

Keywords: AI feedback; EFL writing; ChatGPT; Gemini; Perplexity; grammar correction; language learning.

Annotatsiya. So'nggi yillarda sun'iy intellekt texnologiyalari ta'lim jarayonining ajralmas qismiga aylana boshladi. Xususan, til o'qitish sohasida sun'iy intellekt vositalari o'quvchilarga tezkor, shaxsiylashtirilgan va kontekstga mos yordam ko'rsatish imkoniyatini yaratmoqda. Yozma nutqni o'rgatishda sun'iy intellektning qo'llanilishi nafaqat grammatik xatolarni tuzatish, balki o'quvchilarni mustaqil fikrashga, tahrirlash ko'nikmalarini rivojlantirishga va metakognitiv yondashuvga yo'naltirmoqda. Mazkur tadqiqot ingliz tili yozuv ko'nikmalarini o'qitishda sun'iy intellekt vositalarining ahamiyatini o'rganadi. Tadqiqotda ChatGPT, Gemini va Perplexity dasturlari orqali yaratilgan grammatik fikr-mulohazalar (feedback) tahlil qilindi. Tadqiqot harbiy-akademik litseyda tahsil olayotgan, ingliz tili darajasi B1–B2 bo'lgan 16–17 yoshli o'quvchilar ishtirokida o'tkazildi. Maxsus ishlab chiqilgan mezonlar asosida sun'iy intellekt tomonidan yaratilgan fikr-mulohazalar texnik anqlik, pedagogik ravshanlik va lingvistik moslik mezonlari bo'yicha baholandi. Natijalarga ko'ra, ChatGPT o'qituvchiga o'xshash tarzda batafsil va qo'llab-quvvatlovchi izohlar beradi, Gemini esa ixcham, aniq va mustaqil o'rganuvchilarga mos tahlil taqdim etadi. Perplexity asosan to'g'ri, ammo qisqa izohlar bilan cheklanadi. Tadqiqot shuni ko'rsatdiki, sun'iy intellekt vositalari tezkor va izchil fikr-mulohazalar berish imkonini yaratada, ularning haqiqiy ta'limiy qiymati o'qituvchi rahbarligi va puxta ishlab chiqilgan prompt dizayniga bog'liq. Sun'iy intellekt o'qituvchining o'rnini bosuvchi emas, balki hamkorlikdagi yordamchi sifatida qaralishi lozim.

Kalit so'zlar: sun'iy intellekt; fikr-mulohazalar; ingliz tili yozuvi; ChatGPT; Gemini; Perplexity; grammatika; til o'rganish.

Аннотация. В последние годы искусственный интеллект (ИИ) активно внедряется в сферу образования, особенно в обучение

иностранным языкам. Современные ИИ-системы способны предоставлять мгновенную, персонализированную и контекстно-ориентированную помощь, что способствует развитию у учащихся навыков письма, корректировки текста и самостоятельного мышления. В преподавании английского языка как иностранного ИИ помогает снизить нагрузку на преподавателей, автоматизируя рутинные операции, и позволяет им уделять больше внимания формативной поддержке. Настоящее исследование посвящено изучению роли ИИ в обучении письму на английском языке как иностранному. Особое внимание уделяется грамматической обратной связи, создаваемой с помощью ChatGPT, Gemini и Perplexity. Кейстади проведено среди учащихся военно-академического лицея в возрасте 16–17 лет с уровнями владения английским языком B1–B2. Для оценки ИИ-обратной связи были разработаны специальные рубрики, включающие три критерия: техническая точность, педагогическая ясность и языковая уместность. Результаты показали, что ChatGPT предоставляет развернутые, «человеко-подобные» комментарии, Gemini — лаконичные и структурированные ответы, подходящие для самостоятельных учащихся, а Perplexity предлагает точные, но поверхностные замечания. Исследование подчеркивает, что эффективность ИИ-обратной связи во многом зависит от участия преподавателя и качественного проектирования запросов (prompt design). Делается вывод, что ИИ должен выступать не заменой, а партнером преподавателя — «вторым пилотом», который помогает учащимся получать своевременную и персонализированную поддержку. Работа способствует развитию педагогики, основанной на использовании ИИ в обучении иностранным языкам.

Ключевые слова: искусственный интеллект; письмо на иностранном языке; ChatGPT; Gemini; Perplexity; грамматическая коррекция; изучение языка.

Introduction

The integration of artificial intelligence in English as a Foreign Language instruction is rapidly accelerating across the globe, reflecting a shift from automation toward meaningful augmentation of learning processes [Amoush & Alhosban 2025]. As many countries actively experiment with AI-supported teaching, Uzbekistan is also entering a crucial stage where the purposeful and pedagogically sound use of AI can significantly enhance language learning outcomes. In our context, this shift is not only a technological trend but a genuine educational necessity, especially given increasing class sizes, exam-oriented teaching, and high expectations placed on foreign-language proficiency.

An essential factor shaping this need is the changing learner profile. The students in our academic lyceums represent Generation Z, a cohort with distinctive psychological and pedagogical characteristics. Born roughly between the late 1990s and the early 2010s, they are often described as “digital natives,” having grown up surrounded by smartphones, the internet, and constant connectivity [Beetham 2014]. Unlike previous generations, Gen-Z learners prefer visual, interactive, and fast-access information and expect immediate feedback and personalized

learning trajectories. This generation perceives technology as a natural and indispensable part of life, and they acquire most information online in rapid, multimodal formats [Prensky 2010].

Such characteristics transform their learning expectations: they value autonomy, personalization, interactivity, and opportunities to apply knowledge practically. D. Larsen-Freeman rightly emphasizes that “teachers working with modern learners must acknowledge their new ways of thinking and behaving; digital resources stimulate curiosity, analytical thinking, and discovery-based learning” [Larsen-Freeman 2003: 102]. However, these strengths also come with challenges: shorter attention spans, superficial information processing, and low motivation for traditional lecture-style instruction. These realities particularly affect foreign language teaching in academic lyceums, where learners are preparing for competitive university entrance exams and require intensive, high-quality feedback.

In this context, AI is not simply an innovative tool, it is a potential pedagogical bridge between traditional teaching demands and the learning habits of Z-generation students. AI can provide immediate feedback, personalized learning paths, interactive support, and alternative explanations that align with students’ expectations for speed and clarity. This makes AI-supported feedback especially relevant for B1–B2 learners, who stand at a critical stage between foundational accuracy (B1) and emerging academic fluency (B2). Teachers often struggle to deliver frequent, individualized feedback at these levels, yet students need such feedback the most to progress.

Another practical reason for exploring AI feedback lies in the specific environment of our study: a military-academic lyceum, where discipline, time constraints, and large academic demands limit opportunities for one-on-one teacher guidance. Many of the learners in this setting belong to the “Zoom generation,” shaped by post-pandemic online learning and accustomed to digital assistance and immediate explanations. As a result, their expectations regarding feedback speed and personalization are markedly different from earlier cohorts.

Given these considerations, we selected ChatGPT, Gemini, and Perplexity for analysis because these three tools are the most commonly used platforms among our students and represent distinct AI features. ChatGPT is widely perceived as supportive and “human-like,” Gemini is integrated into the Google ecosystem and known for structured responses, and Perplexity is valued for its accuracy and up-to-date references. Their popularity among learners makes them important candidates for pedagogical evaluation within the Uzbek educational context.

While global research highlights the promise of AI-generated feedback, as shown in the studies conducted by A.H.Alsofyani and A.M.Barzani, J. Han. and M. Li, and J. Steiss et al., limited work has examined how such feedback aligns with the needs of Central Asian learners or how well it suits B1–B2 proficiency groups specifically. The study addresses the following research questions:

1. How do ChatGPT, Gemini, and Perplexity differ in the type and quality of feedback they provide to B1 and B2 EFL learners?

2. What pedagogical and linguistic features make AI feedback more valuable for learners?

3. How can teachers integrate AI feedback without losing their formative role?

By situating AI feedback within the broader frameworks of formative assessment and feedback literacy, this study aims to contribute both to AI-informed pedagogy and to the practical challenges of EFL teaching in Uzbekistan. Ultimately, we seek to illustrate how human-AI collaboration rather than AI alone can enhance feedback quality and better support learners' writing development.

Literature Review

The integration of artificial intelligence into EFL writing instruction has developed rapidly, evolving from simple automation tools such as grammar and spell checkers to adaptive, context-aware systems that support learner autonomy. Modern large language models (LLMs) such as ChatGPT, Gemini, Perplexity, and Copilot now function not merely as linguistic correctors but as intelligent learning companions capable of providing formative and personalized feedback. This shift has been examined in the works of Y. K. Dwivedi et al., O. Zawacki-Richter et al., as well as in the British Council's analytical report in 2024, all of whom highlight how LLMs help learners refine structure, coherence, and vocabulary. While these global developments are well documented, their relevance becomes even more urgent in the Uzbek context, where academic lyceum students must prepare for high-stakes assessments such as the CEFR-based national exam and the IELTS exam to enter higher education. These learners engage in extensive writing practice, yet the volume of their work often exceeds what teachers can reasonably provide feedback on. In this sense, AI tools present an opportunity not simply for innovation but for addressing a real pedagogical gap in Uzbekistan's resource-limited EFL classrooms.

A solid theoretical foundation in second-language feedback research provides a lens for interpreting the value of AI-generated responses. Effective feedback must be contextualized and audience-sensitive, aligning with the communicative purpose of the learner [Hyland 2004]. Feedback is a formative process centered on three core questions – “Where am I going?”, “How am I going?”, and “Where to next?” highlighting clarity, guidance, and learner motivation [Hattie & Timperley 2007]. Bitchener & Ferris differentiate between direct and indirect corrective feedback, showing how each affects learner uptake, while Carless & Boud extend this by arguing that students must develop *feedback literacy* - the ability to understand, evaluate, and act on feedback [Bitchener & Ferris 2012; Carless & Boud 2018].

In our teaching experience at a military-academic lyceum, these theoretical perspectives resonate strongly. Many B1-B2 learners struggle not with recognizing errors but with understanding *why* something is incorrect and *how* to improve it. They need feedback that is explicit, scaffolded, motivating, and developmentally appropriate - qualities that AI tools can either strengthen or fail to deliver depending on their design.

Empirical studies exploring AI-generated feedback show both promise and limitations. Positive findings reveal that AI systems deliver immediate, individualized responses that support iterative revision and autonomous learning. However, research conducted by A. Alnemrat et al. shows that such feedback often contains hallucinated corrections, overly general comments, and shallow explanations that do not sufficiently guide learners [Alnemrat et al. 2025]. In the Uzbek EFL context, these shortcomings are particularly problematic because many students rely heavily on external feedback to prepare for international exams such as IELTS or CEFR-based assessments. When AI-generated feedback is inaccurate or misleading, it can reinforce incorrect linguistic patterns or create a false sense of confidence. As a result, ensuring the quality and reliability of AI feedback becomes not only an academic concern but also a practical necessity for learners aiming to achieve competitive exam performance in a high-stakes environment.

Given these challenges, researchers have increasingly turned to rubric-based evaluation as a systematic way to judge the pedagogical usefulness of AI feedback. Rubrics allow educators to move beyond surface-level impressions and assess feedback according to clear, theory-driven criteria. Following this approach, studies draw on Hyland's principles of contextual appropriacy and tone, Hattie and Timperley's focus on clarity and actionability, Wu's emphasis on precision and scaffolding, and Liu's attention to metalinguistic explanation [Hyland 2004; Hattie & Timperley 2007; Wu et al. 2025; Liu 2024]. Together, these frameworks support three common analytical dimensions: technical accuracy, pedagogical clarity, and linguistic appropriacy.

1. Technical accuracy (identifying genuine errors and avoiding false positives),
2. Pedagogical clarity (providing explanations that promote transfer and self-regulation),
3. Linguistic appropriacy (ensuring tone, lexical choice, and terminology suit the learner's level).

The rubric developed for this study fills a gap in existing literature by adapting these theoretical principles to the realities of Uzbek EFL learners — particularly those in military-academic lyceums, where discipline, clarity, and efficiency are pedagogical priorities. Existing rubrics rarely address the local need for level-sensitive and culturally appropriate feedback analysis.

Despite the growing number of international studies, several gaps remain. Few empirical investigations have directly compared the quality of grammar feedback produced by major AI tools such as ChatGPT, Gemini, and Perplexity. Even fewer examine how feedback varies according to learner proficiency, particularly across B1 and B2 levels. Additionally, research from non-Western, underrepresented contexts is largely absent, meaning that the experiences of Central Asian learners — especially those navigating high-stakes entry requirements are missing from global discussions [Amoush & Alhosban 2025; Alpar 2025; Zeevy-Solovey 2024].

This study aims to address these gaps by analyzing feedback through a

rubric grounded in established theory but applied to a uniquely demanding educational environment in Uzbekistan. It provides insight into how AI tools perform not in idealized conditions but in real classrooms where learners depend on accurate and meaningful feedback to progress.

Methodology

This study adopted a qualitative descriptive research design to examine how three AI tools - ChatGPT, Google Gemini, and Perplexity differ in the type, tone, and pedagogical usefulness of the grammar feedback they provide to EFL learners. Rather than relying on numerical or statistical measures, the research focused on interpretive and comparative textual analysis, with the aim of understanding not only *what* kinds of feedback these tools generate but also *how* that feedback might support or hinder language learning in authentic classroom settings.

The study was conducted at a military-academic lyceum in Uzbekistan, an institution that integrates intensive English instruction within both academic and military-preparatory contexts. In this environment students must handle extensive writing practice while also preparing for competitive entrance exams. For many of these learners, teacher feedback is essential but limited by time constraints. This contextual challenge motivated the choice to explore AI-supported feedback in this specific setting.

Participants were 32 male students aged 16–17, preparing for admission to higher military institutions. They were divided equally into two proficiency groups: sixteen B1 learners working on grammatical accuracy and sentence construction, and sixteen B2 learners developing cohesion, fluency, and academic writing tone. All students wrote an argumentative essay on the topic *“Should students use smartphones in the classroom?”* under classroom conditions without access to digital tools. This topic was intentionally chosen because it is relatable for teenagers and reliably elicits common grammar issues typical of B1–B2 levels.

To provide a clearer understanding of the AI systems evaluated in this study, it is important to outline the distinctive purposes, learning affordances, and limitations of each tool. Although ChatGPT, Gemini, and Perplexity are all large language models, they differ significantly in how they generate information, structure responses, and support learning tasks. These differences help explain the variation observed in their grammar feedback and highlight why comparing them is pedagogically relevant.

ChatGPT is designed primarily as a generative conversational model that creates text, lesson ideas, prompts, and explanations. It is widely used in educational settings because of its ability to simulate teacher-like interactions, generate dialogues, role-plays, and produce coherent drafts. Its learning focus is aligned especially with language arts, making it effective for English, creative writing, literature-based tasks, and feedback on learner essays.

A core strength of ChatGPT is its capacity to engage in extended conversations and personalize its tone, which often makes feedback feel encouraging and human-like. However, the model may produce hallucinations or inaccurate information, especially when prompts are

vague or when it attempts to justify corrections. This limitation makes teacher moderation essential to prevent learners from internalizing incorrect structures.

Gemini offers a multimodal ecosystem that integrates Google Search, Docs, YouTube, and visual data, making it well-suited for tasks that combine text with images, data, or cross-subject research. Educationally, Gemini is strong in supporting multimedia learning, generating structured explanations, and providing precise grammar corrections.

Its learning focus extends beyond language instruction to broader academic contexts, particularly project-based learning where students analyze visuals, charts, or extended textual or visual information. Gemini's responses tend to be well-organized, analytical, and concise, which benefits upper-intermediate learners who prefer clear and direct guidance. However, it can be overly dependent on structured templates, and its tone may sound somewhat formal or academic, making it less emotionally engaging for teenage learners. The tool's functionality is also narrower outside the Google ecosystem, limiting adaptability for institutions that do not rely on Google services.

Perplexity combines real-time search capabilities with conversational generation, making it distinctive from ChatGPT and Gemini. It provides verified, up-to-date information, pulling references from credible sources — a feature particularly useful for factual subjects such as history, geography, and science. For writing tasks, Perplexity's feedback is typically concise, factual, and efficient, reflecting the model's emphasis on accuracy over elaboration.

This tool is strongest when learners need quick corrections or reliable factual updates. However, it offers little personalization and its feedback often resembles editorial notes rather than pedagogical explanations. This makes it less effective for lower proficiency learners who rely on scaffolding. Additionally, its reliance on online content raises concerns regarding copyright, summarization accuracy, and inconsistent depth depending on source availability.

Each student's essay was then processed through the three AI tools - ChatGPT, Gemini, and Perplexity — using the same TRAPF-based prompt (Task, Role, Audience, Purpose, Format). This procedure generated 96 feedback outputs (32 essays × 3 tools). All feedback samples were anonymized to maintain confidentiality and reduce researcher bias.

A deliberate contrast was made between a minimal prompt ("Check my grammar and give me feedback") and the structured TRAPF prompt. This choice was informed both by classroom observations where students often rely on short, vague prompts that result in superficial feedback and by prior research confirming that the specificity of a prompt strongly shapes the usefulness of AI-generated feedback, particularly in formative contexts. For this reason, the TRAPF framework was adopted as a standardized and pedagogically aligned prompting method, reflecting the kind of structured guidance typically provided by an experienced teacher. The TRAPF elements included:

- T — Task: identify key grammar errors, explain them briefly, and suggest one model sentence.
- R — Role: act as an experienced ESL teacher and academic writing coach.
- A — Audience: write feedback suitable for a B1 or B2 teenage learner.
- P — Purpose: help the learner revise more accurately and effectively.
- F — Format: three clear bullet points + one model sentence + one actionable next step.

Structured TRAPF prompt used in the study: *Write constructive, targeted, and encouraging feedback on a B1 teenage learner's writing to effectively improve their grammar and structure. Review the following learner writing sample for key errors in grammar, (e.g. verb tense, syntax, and clarity, article usage). Demonstrate student's mistakes in the essay by using bold letters. Act as a highly experienced, supportive ESL teacher mentor who specializes in providing actionable feedback for B1 teenage learners. Provide a clear, active task (a "next step") for the learner to perform with their writing.*

The structured prompt consistently generated clearer, more precise, and level-appropriate feedback, while the minimal prompt often produced brief and generic comments. This finding confirms that effective prompting is essential for AI feedback to be pedagogically meaningful.

The evaluation of AI feedback was guided by a researcher-designed rubric synthesizing principle from established feedback theories: contextual appropriacy, formative clarity, scaffolding and precision, and metalinguistic support. The rubric assessed three dimensions:

- Technical accuracy: Did the AI identify real errors and avoid false positives?
- Pedagogical clarity: Were explanations clear, scaffolded, and supportive?
- Linguistic appropriacy: Was the feedback natural, level-appropriate, and easy to understand?

Data were analyzed through thematic analysis involving iterative readings of each feedback output. Recurring patterns in error detection, explanation quality, tone, and engagement were identified and compared across tools and proficiency levels. Clear differences emerged: ChatGPT tended to provide detailed, teacher-like scaffolding; Gemini offered accurate, structured, but more formal comments; Perplexity gave fact-based but minimal feedback. As educators, we found these distinctions particularly meaningful because they illustrate how each tool responds differently to the real pedagogical needs of B1 versus B2 learners.

Findings

The analysis of AI-generated grammar feedback across ChatGPT, Gemini and Perplexity revealed notable variation in depth, tone, and pedagogical focus. While all tools successfully identified grammatical errors and provided corrective suggestions, the style, clarity, and adaptability of their feedback for learners differed substantially. ChatGPT

produced detailed and interactive teacher-like responses; Gemini emphasized clarity and structure; and Perplexity provided concise, fact-oriented feedback with minimal elaboration. These patterns suggest that each tool reflects a distinct “feedback persona” aligned with different pedagogical needs and learner profiles.

The evaluation rubric was designed to examine the technical, pedagogical, and linguistic dimensions of feedback quality.

Table 1
Comparative Analysis of AI Feedback

Criterion	Focus of Evaluation	Examples of Indicators
Technical Accuracy	Precision of error detection; avoidance of false positives	Correct grammar identification, consistency, reliable corrections
Pedagogical Clarity	Explicitness, scaffolding, and constructive tone	Clear explanations, learner-oriented advice, step-by-step guidance
Linguistic Appropriacy	Suitability of feedback tone and phrasing for learner level	Formality, naturalness, and register adaptation (B1 vs B2)

ChatGPT produced the most elaborated and human-like feedback, mirroring a supportive ESL teacher. It typically provided metalinguistic explanations (e.g., “After modal verbs like *may*, use the base verb form”) and model rewrites. This depth fostered learner understanding, especially for B1-level students.

Table 2
Strengths and limitations of ChatGPT Grammar feedback

Criterion	Key Strengths	Limitations
Accuracy & Reliability	<ul style="list-style-type: none"> • Detects nearly all grammatical errors with high precision. • Maintains the original meaning and tone of the sentence. • Uses precise metalinguistic terminology when explaining errors. 	<ul style="list-style-type: none"> • Some explanations are brief or insufficiently developed. • Certain corrections lack elaboration (e.g., “<i>a great deal of benefits</i>” → “<i>many benefits</i>” without further justification).
Pedagogical Value & Utility	<ul style="list-style-type: none"> • Provides clear, level-appropriate explanations suitable for B1–B2 learners. 	<ul style="list-style-type: none"> • Occasionally overgeneralizes grammar rules. • Examples include:

Clarity, Tone & Format	<ul style="list-style-type: none"> • Focuses on recurring, high-impact grammar issues. • Encourages reflection and rewriting rather than simple correction. • Supports transfer of grammar rules to future tasks. • Maintains a supportive, motivating, and teacher-like tone that is especially helpful for adolescent learners. • Produces visually organized, easy-to-follow feedback. • Helps students improve coherence and overall flow in revised texts. 	<ul style="list-style-type: none"> advising learners to avoid starting sentences with “<i>Because</i>”; recommending “<i>present simple after if</i>” without conditional context; or suggesting “<i>-ly for all adverbs</i>” all of which oversimplify real usage. • May include overly positive or informal comments (e.g., “<i>Excellent work! Keep it up!</i>”) that do not always match academic writing conventions.
-----------------------------------	---	---

However, ChatGPT occasionally overgeneralized its grammar explanations or repeated similar comments (“Use articles correctly”) without contextual differentiation. The tone was highly motivational, sometimes approaching informality (“Excellent work! Keep it up!”), which could be engaging for teenagers but less appropriate for formal academic contexts. Overall, ChatGPT’s strength lies in comprehensive and empathetic feedback, though it may require teacher moderation to balance tone and depth.

Our B1 learners reacted especially positively to ChatGPT’s supportive tone. Several students described the feedback as “motivating” and said it made revision less stressful. In the military-academic lyceum context, where students often expect strict and authoritative feedback, ChatGPT’s human-like warmth provided a rare “safe space” to make mistakes without judgment. However, teachers must moderate its overly supportive tone to maintain academic rigor.

Gemini

Gemini offered concise, systematically structured feedback that was accurate but less personalized. It tended to highlight errors efficiently with short corrective examples rather than extended explanations. For example, it corrected “*in classroom*” → “*in the classroom*” without elaborating on the grammatical rule. This minimalism favored B2 learners, who possess higher metalinguistic awareness, but left B1 students with limited scaffolding for understanding why corrections were necessary.

Table 3

Strengths and Limitations of Gemini Grammar Feedback

Criterion	Key Strengths	Limitations
Accuracy & Reliability	<ul style="list-style-type: none"> Provides comprehensive error detection supported by detailed examples and structured tables. Consistently preserves the learner's original meaning and tone. Uses highly precise metalinguistic explanations that enhance clarity. 	<ul style="list-style-type: none"> May rely heavily on structured, template-like responses, reducing flexibility and personalization.
Pedagogical Value & Utility	<ul style="list-style-type: none"> Follows a clear “rule → correction → why” explanation sequence. Prioritizes recurring, high-impact grammar areas such as tense, pronouns, and articles. Offers strong scaffolding through task-based engagement. Promotes generalization and pattern awareness, helping learners notice repeated structures. 	<ul style="list-style-type: none"> Tone can be somewhat formulaic and overly academic, making it less conversational and potentially less accessible for younger or lower-level learners.
Clarity, Tone & Format	<ul style="list-style-type: none"> Maintains a polite, professional, and coherent tone throughout the feedback. Presents information in an organized, visually structured format that is easy to follow. Highlights stylistic clarity and contributes to improved coherence. 	<ul style="list-style-type: none"> May feel less emotionally engaging for teenage learners; dense text or structured formats may overwhelm some students.

Gemini's professional tone was consistent and clear, though somewhat template-based and impersonal. The tool proved strongest in accuracy and coherence but weaker in formative, dialogic support.

During the study, B2 students expressed that Gemini's structured, academic tone matched their expectations for exam preparation (IELTS, CEFR B2). However, B1 learners found the feedback "dry" or "too short," indicating that affective resonance is an important component of effective feedback — something rarely discussed in AI literature. For exam-oriented Uzbek learners, Gemini aligns well with structured revision cycles but requires teacher-guided explanations for lower- proficiency students.

Perplexity

Perplexity generated brief, factual feedback emphasizing correctness and vocabulary precision. Its responses often resembled editorial notes (e.g., "Change *lost* to *lose* after *may*"), providing accurate corrections without explicit grammatical reasoning.

Table 4
Strengths and Limitations of Perplexity AI Grammar Feedback

Criterion	Key Strengths	Limitations
Accuracy & Reliability	<ul style="list-style-type: none"> Provides reliable factual accuracy in grammar correction. Detects common surface-level grammatical issues effectively. Produces relatively few false corrections compared with other tools. 	<ul style="list-style-type: none"> Misses complex or context-dependent errors (e.g., no correction for "when ever" → "whenever"). Offers limited metalinguistic explanation (e.g., "Use the correct articles," "Watch verb forms").
Pedagogical Value & Utility	<ul style="list-style-type: none"> Provides concise, fact-oriented feedback suitable for quick revisions. Identifies multiple grammar issues within a single text. 	<ul style="list-style-type: none"> Lacks clear prioritization or scaffolding—treats different grammar issues (articles, connectors, syntax, phrasing) with equal weight. Encourages minimal reflection or engagement; offers correction without teaching underlying rules. Corrects the text efficiently but does not facilitate long-term learning transfer.
Clarity, Tone & Format	<ul style="list-style-type: none"> Maintains a friendly, clear, and neutral tone. Produces functional, linear feedback often accompanied by verified references. 	<ul style="list-style-type: none"> Feedback lacks personalization, emotional engagement, and motivation for younger learners.

The feedback was efficient but sometimes over-summarized and lacked pedagogical guidance. For instance, while it detected errors reliably, it did not explain *why* they occurred or *how* to avoid them in future writing tasks. Perplexity's style was therefore informative but limited in transferability — students could fix errors but might struggle to internalize grammatical concepts. Its academic tone, however, aligned well with B2-level learners requiring quick, reference-style feedback.

Perplexity was well-received by several B2 learners who preferred “quick correction without extended explanations.” However, for B1 learners this brevity created confusion. In our context, many Uzbek students rely heavily on teacher approval; therefore, Perplexity’s impersonal tone sometimes reduced confidence rather than increasing autonomy.

Table 5
Differences between B1 and B2 learners

Dimension	B1 Learners (Intermediate)	B2 Learners (Upper-Intermediate)
Focus of Feedback	Grammar accuracy, article use, verb forms, and basic syntax	Cohesion, style, tone, and lexical precision
Feedback Type	Directive (“Use <i>the</i> before specific nouns”)	Reflective (“This phrase could be more formal”)
Tone	Encouraging and scaffolded	Professional and concise
AI Tool Fit	ChatGPT most supportive; Gemini less accessible; Perplexity too brief	Gemini most efficient; ChatGPT slightly redundant; Perplexity accurate but minimal
Pedagogical Implication	Needs step-by-step explanations and examples	Prefers autonomy-oriented, concise feedback

A clear divergence emerged in how AI tools addressed learners at different proficiency levels. These contrasts mirror our classroom observations. B1 learners at the military-academic lyceum often require emotional support and step-by-step explanations due to high exam pressure and fear of making mistakes. B2 learners, conversely, value time-saving, exam-like conciseness.

These findings demonstrate that AI feedback effectiveness is level-dependent. B1 learners benefit from detailed, teacher-like responses that explain rules and model usage, whereas B2 learners prefer focused, succinct guidance that respects their linguistic independence.

Comparing all three tools holistically revealed complementary strengths:

- ChatGPT — excels in *depth, personalization, and pedagogical warmth* but tends to overelaborate and occasionally mix formal/informal tones.
- Gemini — excels in *clarity, consistency, and academic tone* but lacks scaffolding and creative adaptation.
- Perplexity — excels in *accuracy and factual reliability* but offers limited elaboration and learner engagement.

This analysis shows that none of the AI tools can give optimal feedback on their own, but each of them offers something useful. ChatGPT gives motivating and detailed comments, Gemini focuses on accuracy and structure, and Perplexity provides quick, factual suggestions. When teachers combine these strengths and guide students in using the feedback, it becomes much more effective. In this way, the teacher's role is not replaced by AI but expanded — enabling students connect different types of feedback and learn from them more deeply.

The findings highlight that AI feedback works best when it supports teachers rather than replaces them. Teachers play a key role in giving meaning to AI suggestions with adding context, checking accuracy, and keeping the feedback personal and human. Using rubrics based on established feedback theories [Hyland, 2004; Hattie & Timperley, 2007] helped reveal how AI can contribute to real learning, not just correction. When thoughtfully guided by teachers, these tools can ease workload, give students more varied feedback, and help them become more independent learners.

In Uzbekistan's examination-driven educational system particularly in military lyceums teachers face heavy workloads and large volumes of writing assignments. AI feedback offered students a rare opportunity to revise independently before receiving teacher input, reducing repetitive teacher correction and improving learners' revision habits. Importantly, students used AI feedback to *experiment* with grammar without fear of immediate teacher judgment, increasing willingness to revise which is a psychologically meaningful outcome.

Discussions

The findings of this study deepen our understanding of how AI-generated feedback can support foreign language writing instruction, especially in settings where teacher workload and large class sizes limit opportunities for individualized attention. In the Uzbek military-academic lyceum context, where academic rigor and discipline are strongly emphasized, students tend to rely heavily on teacher approval and may hesitate to take risks in writing. Notably, AI feedback provided a "safe space" for experimentation: learners could revise without fear of immediate judgment. This psychological comfort appeared especially important for B1 learners, many of whom expressed that ChatGPT's encouraging tone made them "less afraid of making mistakes." Such emotional responses, rarely discussed in AI feedback research, suggest that the affective dimension of feedback is inseparable from its pedagogical impact.

The comparative analysis of ChatGPT, Gemini, and Perplexity demonstrates that these tools operate less as evaluators and more as *pedagogical partners*, each offering distinctive strengths. ChatGPT's elaborated, conversational style closely resembled teacher scaffolding and was particularly effective for lower-proficiency learners who require step-by-step explanation. Gemini provided concise, well-structured corrections appropriate for more independent B2 learners, although several students commented that its brevity sometimes made them uncertain about how to revise effectively. Perplexity offered fact-based and technically accurate comments but lacked interpretive depth, limiting its value for learners who depend on explanation for improvement.

These differences reinforce earlier insights from researchers: effective feedback must be contextual, actionable, and responsive to learner needs. In this study, no single AI tool demonstrated the full range of these qualities, but each contributed a meaningful element. ChatGPT provided emotional support and clear explanations; Gemini contributed precision; Perplexity offered efficiency. We found that the most pedagogically effective approach is not to select one "best" tool but to strategically combine them depending on learning objectives for example, using ChatGPT for formative scaffolding, Gemini for targeted correction, and Perplexity for factual clarity.

One of the most important findings relates to differences in how B1 and B2 learners interacted with and benefited from AI feedback. B1 students needed explicit corrections, simplified explanations, and motivational tone, elements provided most consistently by ChatGPT. In contrast, B2 learners were ready for more autonomous revision and responded well to Gemini's concise, matter-of-fact style. This suggests that effective AI integration must account for developmental trajectories in feedback needs. However, none of the tools automatically adapted their feedback strategies based on proficiency level; this pedagogical sensitivity still requires human mediation. AI can imitate aspects of teacher feedback, but it cannot yet fully replicate the nuanced, learner-specific decisions teachers make.

Across the tools, the study also identified a recurring tension between technical accuracy and pedagogical clarity. Gemini and Perplexity often produced correct but minimalistic comments, technically sound yet pedagogically thin. ChatGPT, in contrast, offered richer explanations but occasionally overgeneralized or included unnecessary elaboration. These patterns highlight that AI feedback is not a self-contained instructional solution; it becomes most effective when embedded within teacher-led revision cycles. During micro-observations, many students sought teacher confirmation even after receiving AI feedback, asking questions such as "Is this correction really correct?" This demonstrates that AI feedback can initiate revision, but teacher guidance remains essential for verification and deeper learning.

From a professional development perspective, the findings underscore the urgent need to strengthen AI literacy among EFL teachers in Uzbekistan. Teachers must learn not only how to use AI tools but how to evaluate their pedagogical value critically. The rubric developed in this

study, which is grounded in clarity, accuracy, and appropriacy, provides a practical model that can help teachers guide learners in using AI responsibly. If applied more broadly in teacher training programs, such rubrics may also enhance teachers' confidence in integrating AI into formative assessment practices.

Practically, this study suggests several useful applications for AI-supported feedback. AI can serve as a first layer of response to student writing, reducing teacher workload and providing students with immediate insights before class discussion. AI-generated drafts can also be used for peer-review sessions, encouraging learners to discuss and critique feedback together. However, the findings make clear that uncritical or unsupervised use of AI risks misunderstandings or superficial correction. Therefore, AI feedback should be integrated into a structured cycle involving drafting, AI-supported revision, peer discussion, and teacher validation.

Ultimately, this study reframes AI not as an automated corrector but as a co-pilot - a collaborative assistant that extends the teacher's reach while respecting the teacher's role as the primary decision-maker. When human insight meets machine precision, feedback becomes faster, more accessible, and more personalized. Yet meaningful language development still depends on the teacher's empathy, contextual judgment, and pedagogical intention. This partnership approach aligns with global trends in AI-augmented education, emphasizing technology as an amplifier of teaching rather than a replacement.

Limitations and future directions

While this case study offers valuable insights, several limitations remain. The sample size was small and context-specific, limited to a military-academic lyceum in Uzbekistan. The analysis was qualitative, focusing on textual patterns rather than statistical correlations. Future research could expand the dataset, include diverse proficiency levels, and apply quantitative or mixed-method approaches to measure learning gains over time. Further studies might also explore prompt design, comparing how structured versus open-ended prompts affect AI feedback quality - a variable this study identified as crucial but only partially addressed.

Conclusion

This study demonstrates that AI-generated grammar feedback can meaningfully enhance EFL writing instruction when used with pedagogical intention and human mediation. While ChatGPT, Gemini, and Perplexity each offer valuable yet distinct strengths, their effectiveness becomes most visible when aligned with specific learner profiles and real classroom needs. In the Uzbek context - particularly within exam-oriented institutions such as military-academic lyceums - students often rely heavily on teacher approval and fear making mistakes. AI feedback, as observed in this study, provided a psychologically safe space for learners to experiment, revise, and take risks with their writing. Several B1 students shared that ChatGPT's warm, supportive tone made them feel "encouraged to write more," while others noted that Gemini's structured clarity helped

them “see errors more clearly,” even if the tone felt more distant. These reactions point to an important, often overlooked dimension: feedback is not only cognitive — it is emotional.

The comparative findings show that AI tools do not replace the teacher’s role but reshape it. AI becomes a first layer of individualized feedback, offering students immediate guidance before teachers step in to contextualize, refine, and validate the suggestions. This dynamic proved particularly meaningful in an environment where teachers face heavy workload and large amounts of writing practice due to CEFR/IELTS preparation demands. With AI support, students engaged in more revision cycles, showed greater willingness to self-correct, and developed early elements of feedback literacy — skills essential for lifelong learning.

From a broader pedagogical perspective, the study illustrates that AI is most powerful when understood not as an evaluator but as a co-pilot. ChatGPT contributed motivational depth, Gemini brought precision and structure, and Perplexity offered factual, efficient correction. Together, these tools enrich the feedback landscape by addressing different learner needs: B1 students benefited from scaffolded explanations, while B2 learners preferred succinct, autonomy-supportive guidance. When these insights are combined with teacher expertise, AI-assisted feedback creates a more responsive, human-centered learning environment — one where students feel supported, challenged, and empowered to improve.

Ultimately, this research underscores a central message: AI can accelerate correction, but only teachers can transform feedback into learning. When thoughtfully integrated, AI expands what teachers can accomplish, enabling them to focus on higher-order instruction while still ensuring that learners receive timely, meaningful, and personalized feedback. This partnership between human judgment and machine efficiency holds significant promise for shaping the future of EFL writing instruction in Uzbekistan and beyond.

References

Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlShareefen, R. (2025). AI vs. teacher feedback on EFL argumentative writing. *Frontiers in Education*, 10, Article 1614673.

<https://doi.org/10.3389/feduc.2025.1614673>

Alpar, Ö. (2025). Evaluating generative AI tools for improving English writing skills: A preliminary comparison of ChatGPT-4, Google Gemini, and Microsoft Copilot. *European Journal of Educational Research*, 12(3), 1775–1787.

Alsofyani, A. H., & Barzanji, A. M. (2024). The effects of ChatGPT-generated feedback on Saudi EFL learners’ writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*. Advance online publication.

Amoush, K., & Alhosban, A. A. (2025). Mastering EFL writing with ChatGPT: A systematic review of benefits, challenges, and best practices. *Theory and Practice in Language Studies*, 15(8), 3843–3852.

Beetham, H., & Sharpe, R. (Eds.). (2013). *Rethinking pedagogy for a digital age: Designing for 21st century learning*. Routledge.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

British Council. (2024). *AI and language learning: The British Council's vision*.
<https://www.orsys.fr/orsys-lemag/en/ia-language-learning-vision-british-council/>

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
<https://doi.org/10.1080/02602938.2018.1463354>

Dwivedi, Y. K., Hughes, D. L., Ismagilova, E., Aarts, G., Coombs, C., & Sharma, S. K. (2023). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Journal of Business Research*, 154, Article 113353.
<https://doi.org/10.1016/j.jbusres.2023.113353>

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, Article 103401.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
<https://doi.org/10.3102/003465430298487>

Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.

Larsen-Freeman, D. (2003). *Techniques and principles in language teaching* (2nd ed.). Oxford University Press.

Liu, S. (2024). Effects of metalinguistic corrective feedback on novice EFL writers. *British Journal of Educational Technology*. Advance online publication.

Prensky, M. R. (2010). *Teaching digital natives: Partnering for real learning*. Corwin Press.

Steiss, J., et al. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894.

Wu, J., Li, J., Ge, Z., Xu, M., Lin, L., & Zhang, R. (2025). Effectiveness of generative AI in automated written corrective feedback with prompting. *Journal of Educational Computing Research*. Advance online publication.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), Article 39.
<https://doi.org/10.1186/s41239-019-0171-0>

Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, 6(2), Article 1482.