JoLIE 17:3/2024
doi : https://doi.org/10.29302/jolie.2024.17.3.8

TRANSFORMING ACADEMIC WRITING WITH AI: TOOLS FOR EFFECTIVE LEARNING

Teodora Popescu
1 Decembrie 1918 University of Alba Iulia, Romania

Abstract

Artificial Intelligence (AI), particularly through chatbots like ChatGPT, is increasingly integrated into education, especially writing instruction. While its potential to support writing development is acknowledged, its full pedagogical value remains underexplored. This paper examines how ChatGPT can be used in classrooms to simulate real-world writing tasks—such as argumentative and creative writing—and its role in enhancing proficiency across genres. It also addresses the ethical and instructional implications of integrating AI into the writing process.
This analysis examined revision data from eleven student texts to explore how various error types, word count changes, and revision behaviours relate to improvements in essay quality. The goal was to understand whether measurable features of revision correlate with outcomes such as improved essay scores and CAE grades. The paper also provides practical guidelines for teachers on incorporating ChatGPT into writing instruction.
Findings show that ChatGPT offers immediate, valuable feedback that supports revision and improves student writing. It assists with structural, grammatical, and stylistic issues, though its suggestions require critical evaluation to avoid over-reliance. While ChatGPT boosts engagement and fosters creativity, maintaining student independence remains essential.
In conclusion, when used thoughtfully, ChatGPT can be a powerful tool in writing education, offering interactive feedback and personalised support. However, educators must use it ethically and responsibly, ensuring it complements rather than replaces traditional instruction. Further research is needed to assess AI's long-term impact on writing and to develop best practices for its integration.

Keywords: Artificial Intelligence; ChatGPT; Writing education; Feedback; Pedagogical practices, EFL error correction.

Received: 20 June 2024 Revised: 25 September 2024 Accepted: 10 November 2024 Published: 15 December 2024

1 Introduction

The integration of Artificial Intelligence (AI) into education has introduced new opportunities for addressing long-standing challenges in both traditional and online learning environments. Among AI applications, chatbots have emerged as particularly promising tools, capable of enhancing student engagement, reducing

instructor workload, and supporting personalised learning. These benefits are especially relevant in large or remote classes, where students often experience delayed feedback and limited interaction, contributing to decreased motivation and academic disengagement. Simultaneously, educators are burdened with administrative tasks—such as grading and responding to repetitive queries—that detract from time spent on instruction and meaningful student interaction.

Conventional teaching methods frequently fall short in meeting the diverse needs of learners, particularly students from under-resourced backgrounds, non-native speakers, and individuals with disabilities. Moreover, the ability to monitor student performance in real time and identify learners at risk of falling behind remains a persistent concern, particularly in digital or blended learning contexts. Existing research underscores a growing crisis in student motivation in online education and increasing rates of burnout among educators tasked with maintaining quality instruction under significant pressure.

AI chatbots have been proposed as a scalable and adaptive solution to these problems. By offering immediate, interactive feedback, chatbots can act as virtual teaching assistants, supporting both students and educators (Gonda, &Chu 2019). For example, the *Chem Quest* chatbot (Jasin et al. 2023) used in an online chemistry course at the Singapore Institute of Technology improved student engagement and comprehension by providing targeted feedback and reinforcing key concepts. More broadly, chatbots are increasingly used in higher education to automate routine instructional tasks, support inquiry-based learning, and assist in teaching complex subjects such as programming, mathematics, and academic writing.

Writing, in particular, presents a compelling use case for AI integration. As a core competency across academic and professional contexts, writing enables individuals to communicate ideas clearly, think critically, and engage creatively with complex problems (Burkhardt, MacDonald, & Rathemarcher 2010). It underpins success in higher education and beyond, facilitating personal expression, civic participation, and professional advancement. Despite its centrality, the teaching of writing remains resource-intensive, requiring continuous feedback and individualised support—conditions that are difficult to sustain in large or underfunded classrooms.

AI-powered chatbots, such as ChatGPT, offer a novel avenue for supporting writing instruction. They can provide immediate, context-sensitive feedback on student work, guide revision, and simulate real-world writing scenarios. However, their integration into the classroom raises important pedagogical and ethical questions regarding their role in skill development, their impact on student autonomy, and the risk of over-reliance on automated systems. Understanding how such tools function in authentic educational settings, particularly in the development of advanced writing skills, is essential for informing responsible and effective use (Gill et al. 2024).

This article investigates the application of ChatGPT in academic writing instruction, focusing on its effectiveness in providing feedback on argumentative writing tasks. Drawing on a qualitative analysis of student interactions and revisions,

the study explores the chatbot's pedagogical potential and limitations, with the aim of offering practical insights for educators and contributing to broader discussions on AI in education.

2. Literature Review

Artificial Intelligence (AI) refers to the simulation of human cognitive functions—such as learning, reasoning, and problem-solving—by computer systems (Bellman 1978). As AI technologies advance, they increasingly influence teaching and learning methodologies, with notable applications in writing instruction. AI-driven systems leverage machine learning to provide adaptive learning environments, allowing for personalised feedback and support for learning and professional development (De Laat et al. 2020).

In education, AI is increasingly integrated into instructional design, offering new approaches to teaching and learning across various disciplines (Chodorow, et al. 2010), including language learning [Al-Obaydi, Pickhart, & Klimova 2023]). Among its most noted applications is the capacity to personalise instruction, as AI systems can adapt content and pacing to individual student needs. Intelligent tutoring systems extend this functionality beyond the classroom, providing learners with on-demand feedback and clarification (Zawacki-Richter et al., 2019). AI is also used to streamline assessment by automating grading processes and generating real-time feedback, potentially increasing efficiency for both students and educators (Luckin, et al. 2016). Additionally, AI tools contribute to administrative management, helping reduce the workload associated with routine tasks and allowing educators to allocate more time to instructional responsibilities (Delen 2010).

Despite these advantages, the implementation of AI in education is not without challenges. Limitations include concerns about data privacy, the risk of over-reliance on automated systems, and questions regarding the pedagogical depth of AI-generated feedback. Moreover, AI tools are often developed within specific linguistic, cultural, or technological contexts, which can limit their accessibility or relevance in diverse educational settings. As such, while AI holds considerable promise, its integration into educational practice must be approached critically and supported by empirical research that evaluates both its benefits and its constraints.

ChatGPT, in particular, is a versatile tool that helps students accelerate their learning (Boskabai, Nikfar, Ugwuoke &Ali 2024, Strzelecki et al. 2024, Habeb Al-Obaydi et al. 2023) by offering a range of services, such as writing assistance, language translation, conversation practice, tutoring, research help, and more. It can support students in expanding on notes, explaining complex concepts, and improving skills like writing and language learning. For writing, students can use ChatGPT to draft, revise, and receive feedback on essays and assignments, assisting with grammar, vocabulary, and organisation (Skrabut 2023; Ferlazzo 2023). ChatGPT can also help enhance students' creative writing. For instance, a student can submit a story or essay and get instant, detailed feedback on its structure, grammar, content,

and other key elements. By leveraging its language knowledge, ChatGPT can pinpoint strengths and weaknesses in the student's work and provide suggestions for improving writing techniques. Additionally, the student can submit specific sections of their writing to receive feedback on how to refine their style, rhythm, and fluency (Ramos 2023).

While AI tools such as ChatGPT offer multiple advantages, they are not without their drawbacks. Hicks, Humphries, and Slater (2024) critique large language models like ChatGPT by arguing that the inaccuracies often attributed to these systems—such as "AI hallucinations"—are better understood as "bullshit" in the philosophical sense defined by Frankfurt (2005). They suggest that these models are designed to generate text that appears truth-apt without any genuine concern for truth. This aligns with Frankfurt's concept of "bullshit", in which a speaker or writer is less concerned with the factual accuracy of their statements and more focused on producing something that seems true, regardless of its veracity. The authors argue that this understanding offers a more precise framework for discussing and predicting the behaviour of AI systems in generating misleading or false information.

Historically, the landscape of academic writing instruction has remained relatively consistent, with formal writing courses designed to improve the writing skills of an increasingly diverse undergraduate population. Over the past few decades, teaching writing has evolved into a distinct profession, separate from other academic disciplines (Gottschalk, & Hjortshoj 2004). These dedicated courses and pedagogical methods have played a pivotal role in shaping the foundational skills needed for effective academic writing. However, as technologies continue to evolve, the teaching and practice of academic writing are being reshaped by the introduction of artificial intelligence (AI) in education.

The advent of AI tools has introduced both opportunities and challenges in the realm of academic writing. Technologies such as AI-powered grammar checkers, including Grammarly, and writing assistance tools like ChatGPT (Barrot 2023; Ramos 2023), have dramatically altered the way students and educators approach writing. These tools can provide immediate feedback, correct grammatical errors, and suggest improvements to enhance the clarity and coherence of academic texts. However, their increasing prevalence raises concerns regarding issues like plagiarism and the authenticity of student work. Writing produced or heavily aided by AI might be perceived as a form of "automated writing," which poses challenges for traditional notions of authorship and originality.

While AI tools can certainly enhance the writing process by offering guidance and improving writing quality, it is crucial to strike a balance. Educators must consider how to effectively integrate AI assistance without undermining the development of critical thinking and independent writing skills. Striking this balance ensures that AI serves as a tool to enhance, rather than replace, the academic writing process. As such, the integration of AI into writing instruction necessitates a re-evaluation of both teaching methodologies and assessment practices.

In this evolving context, it is worth drawing attention to prior research on student errors and language competence, such as Popescu's (2013) corpus-based

study on translation errors among Romanian EFL learners. Popescu identifies a range of linguistic and comprehension errors—particularly in morphology, collocation, and lexical choice—many of which overlap with challenges observed in student writing more broadly. While her study focused specifically on translation, the insights regarding learner difficulties and self-correction strategies are highly relevant to academic writing instruction. Her findings also underscore the importance of guiding students in developing collocational competence and morphological awareness—areas where AI tools like ChatGPT often provide surface-level corrections without addressing deeper patterns of error or cultural-linguistic nuance. Thus, extending Popescu's work to the domain of writing, particularly in light of AI assistance, could yield important pedagogical insights into how such tools interact with learner interlanguage and affect the development of language proficiency.

As AI tools, particularly writing assistants like ChatGPT, become increasingly embedded in academic writing practices, new concerns emerge around the integrity and originality of student work. With the rise of such tools, issues related to plagiarism have taken on a new dimension. Traditionally, plagiarism detection tools such as Turnitin, SafeAssign, and PlagScan have played a central role in identifying instances of academic dishonesty. These tools are particularly valued for their ability to detect electronic sources, allowing instructors to efficiently pinpoint copied content (Alexander, Savvidou, &Alexander 2023).

However, the widespread use of AI in writing raises new challenges in the fight against plagiarism. While these tools remain crucial in detecting traditional forms of plagiarism, such as copy-pasting from online sources, they are less effective at identifying AI-generated content. As Pecorari (2008) points out, plagiarism detection software has limitations, particularly in detecting plagiarism from print sources or password-protected databases, which students may commonly use. Additionally, the digital fingerprinting techniques used by these tools may overlook some forms of plagiarised text, potentially allowing AI-assisted content to go undetected.

Moreover, as plagiarism detection tools become more accessible to students, they can pre-emptively check their work for potential plagiarism, adjusting it to avoid detection. This can lead to manipulative behaviours aimed at circumventing the software, further complicating the issue. Some educators also risk over-relying on these tools, using them to assess not only the originality of content but also aspects like grammar and formatting. This over-reliance detracts from the critical judgment required in evaluating student work and risks missing subtle nuances that plagiarism detection software might overlook.

In this evolving landscape, educators must recognise the limitations of plagiarism detection tools, particularly in the age of AI, and strike a balance between utilising technology and applying critical, human judgment to assess academic work effectively.

3 Research Methodology

This study employs a mixed-methods approach, combining quantitative experimental design and qualitative analysis to assess the effectiveness of ChatGPT in enhancing the quality of argumentative essays. The research includes a pre-test (initial essay assessment) and a post-test (re-evaluated essay after receiving ChatGPT's feedback) to observe how the AI tool influences the writing quality. The independent variable is the feedback provided by ChatGPT, while the dependent variable is the quality of the students' argumentative essays. The combination of pre- and post-test results offers both numerical evidence of improvement and qualitative insights into the nature of changes in student writing.

Subjects

The sample consists of eleven second-year students enrolled in a philology program at *1 Decembrie 1918* University of Alba Iulia. These students, who are preparing to become future English teachers, were purposefully selected for their relevance to the research focus on academic writing and language education. Each participant was asked to write a 350-word argumentative essay on the topic "Crime Does Not Pay", which simulates a Cambridge Advanced English (CAE) writing task. This group was selected to ensure the relevance of their academic standing and their familiarity with writing instruction.

Instruments

1. CAE Assessment grid: the primary instrument for evaluating the quality of the students' essays is the CAE assessment grid (Cambridge University Press & Assessment, 2024). The grid evaluates essays across four criteria: Content, Communicative Achievement, Organisation, and Language. This standardised tool provides a reliable framework for comparing the quality of the essays before and after feedback from ChatGPT.
2. ChatGPT AI tool: The experimental intervention in this study was ChatGPT, which provided feedback on the students' essays. The feedback focused on multiple aspects, including grammar, sentence structure, clarity, consistency, vocabulary usage, and overall argument development. The AI-generated revisions aimed to help students improve their argumentative writing by enhancing the logical flow and stylistic quality of their essays.

Methods

1. Error categorisation: A qualitative analysis of the essays was conducted to categorise errors into six types: spelling, lexical, morphological, syntax, semantic, and textual errors. This categorisation enabled the identification of common writing errors, as well as the specific areas where ChatGPT's feedback made the most significant impact.
2. Word count analysis: Both the pre- and post-revision word counts of each essay were recorded to measure the extent of text expansion or modification resulting from ChatGPT's revisions. The Percentage of Words Added was calculated to assess the magnitude of changes in each essay following the AI's intervention.

3. Word Error Rate (WER): The Word Error Rate (WER) was computed to evaluate the quality and accuracy of ChatGPT's revisions. WER measures the number of changes made by the AI relative to the number of words, providing a quantitative indication of the AI's impact on the quality of writing.

Procedure

Each student initially wrote an argumentative essay on the topic "Crime Does Not Pay", which was assessed using the CAE grid. Following this initial assessment, each essay was submitted to ChatGPT for feedback. The AI tool focused on improving the following aspects of the essays:

- Clarity and grammar: Correcting grammatical mistakes and enhancing clarity.
- Sentence structure: Improving the fluency and coherence of sentences.
- Consistency and phrasing: Ensuring consistent tone and appropriate phrasing throughout the essay.
- Refining ideas: Suggesting improvements in the articulation and development of ideas.
- Rephrasing for flow: Offering recommendations for smoother transitions between sentences and paragraphs.
- Improving tone and meaning: Ensuring that the argumentative tone was appropriate for academic writing and conveying the message effectively.

After revisions, the essays were re-evaluated using the CAE assessment grid. The number and types of errors were compared to determine how much the feedback from ChatGPT improved the writing. The word count and percentage of words added were also analysed to gauge the level of expansion or modification in the revised essays.

Each essay was scored using the CAE scale both before and after ChatGPT's intervention. These scores were then compared to identify changes in the quality of the essays. The quantitative data on error types, word count, and Word Error Rate (WER) were used to measure improvements in writing accuracy and quality. In addition, a qualitative analysis of the revisions provided insights into the specific areas where ChatGPT had the most significant impact on student writing.

4 Results and Interpretation

The dataset comprises 11 student essays revised with ChatGPT assistance, evaluated using error categories aligned with the Cambridge English: Advanced (CAE) writing assessment grid. The data reveal several patterns indicating the impact of AI-assisted revision on linguistic accuracy, text expansion, and grading outcomes.

Essays were analysed across six error categories: spelling, lexical choice, morphology, syntax, semantics, and textual organisation. Textual errors were consistently the most frequent, reflecting persistent challenges with coherence and cohesion. This supports the conclusion that while ChatGPT effectively addresses surface-level language issues (e.g., grammar, vocabulary), structural improvements

still require explicit instruction and possibly additional prompting. Global results are presented in Table 3 below.

Total error counts per text ranged from 23 to 40. Despite variability in raw errors, the percentage of mistakes relative to word count remained low, averaging approximately 0.16 across texts. Texts with negative growth in word count (e.g., Texts 3, 7, 10, and 11) still saw notable score improvements, suggesting that conciseness and clarity, not just expansion, played a role in higher evaluations.

Most essays grew in length post-revision, with word count increases ranging from 7.6% to 57.7%. This expansion often resulted in more fully developed arguments and clearer articulation of ideas. Notably, Texts 2, 5, and 8 more than doubled their original lengths, correlating with substantial reductions in lexical and semantic errors, as well as improved overall structure. Below is an example of a revised sequence from Text 2. Improvements were mainly made in terms of verb tense, clarity, lexical enhancement, register and formality, coherence and improved flow.

Original text (Text 2):

> *I understood phrase "crime doesn't pay" like what you did one day will be with you the whole life.*

ChatGPT-improved text:

> *I understand the phrase "Crime doesn't pay" to mean that the consequences of your actions, especially wrongful ones, will stay with you for the rest of your life.* (OpenAI, 2024)

Table 1. Comparison of original and revised sentences: linguistic improvements in clarity, grammar, and formality

Area of improvement	Original phrase/sentence	Revised phrase/sentence	Explanation of revision
Verb tense	*I understood* (past tense)	*I understand* (present tense)	The revised sentence correctly uses the present tense ("I understand") as this is a general statement or explanation.
Clarity of meaning	*like what you did one day will be with you the whole life*	*that the consequences of your actions, especially wrongful ones, will stay with you for the rest of your life*	The original sentence is vague and somewhat colloquial. The revision specifies the meaning more clearly and formally.
Use of "like"	*like*	*to mean*	"Like" is informal and imprecise in this context. The revision uses "to mean", which is more appropriate for academic writing and idiomatic explanations.

Area of improvement	Original phrase/sentence	Revised phrase/sentence	Explanation of revision
Lexical choice	*the whole life*	*for the rest of your life*	"The whole life" is less precise and somewhat informal. The revised phrase "for the rest of your life" is more precise and formal.
Structure and flow	*what you did one day will be with you the whole life*	*the consequences of your actions, especially wrongful ones, will stay with you for the rest of your life*	The revised sentence breaks down the original vague structure into a clearer and more formal explanation with appropriate details ("consequences of your actions" and "especially wrongful ones").
Idiomatic precision	*I understood phrase "crime doesn't pay" like...*	*I understand the phrase "Crime doesn't pay" to mean that...*	The revision corrects the incorrect structure "I understood phrase" by adding "the" and adjusting the sentence to be grammatically correct and clearer in meaning.
Formal tone	Casual, somewhat informal phrasing	More formal, academic tone	The revision raises the level of formality by using terms like "consequences", "actions", and "especially wrongful ones", making it more appropriate for academic or professional contexts.

Conversely, Texts 3, 7, 10, and 11 were shortened after revision, with word count reductions up to **10.2%**. Nevertheless, these texts also showed improved scores, indicating that ChatGPT not only adds content but can effectively streamline writing when necessary. This suggests a dual functionality: expansion where development is needed and condensation where redundancy or off-topic content exists. Below is an example of a revised sequence from Text 11. Improvements were made in terms of lexical precision, grammatical accuracy, idiomatic expressions, cohesion and style and conceptual clarity.

Original text (Text 11):

> *Crime itself is the prime version of anarchism. It existed from the dawn of the times, the times when not a single law was in charge, when people used to think they're just as animals, protecting their families, as wolves, protecting their pride.*

ChatGPT-improved text:

> *Crime itself is the prime example of anarchism. It has existed since the dawn of time, during an era when no laws governed society—when people behaved like animals, protecting their families like wolves defending their pack.*
>
> (OpenAI, 2024)

Table 2. Comparative analysis of original and revised sentence: improvements in grammar, lexical choice, and metaphoric accuracy

Area of improvement	Original phrase/sentence	Revised phrase/sentence	Explanation of revision
Collocation	*prime version of anarchism*	*prime example of anarchism*	"Example" is the correct collocation in this context; "version" is awkward and semantically imprecise.
Verb tense	*It existed*	*It has existed*	Present perfect tense ("has existed") is more appropriate when referring to something continuing from the past to the present.
Time adverbial	*from the dawn of the times*	*since the dawn of time*	The revised version uses the correct idiomatic expression. The original is non-standard and awkward.
Repetition	*the times, the times when...*	*during an era when...*	Redundant repetition is replaced with a more elegant, formal alternative.
Grammar/Syntax	*not a single law was in charge*	*no laws governed society*	"No laws governed society" is more academic and conceptually precise than "not a single law was in charge", which sounds informal and imprecise.
Awkward expression	*people used to think they're just as animals*	*people behaved like animals*	Simplifies and corrects the expression. The original mixes tenses ("used to think" with "they're") and is conceptually confusing.
Animal metaphor	*as wolves, protecting their pride*	*like wolves defending their pack*	Wolves live in packs, not prides (lions do). The revision corrects the metaphor and improves fluency.
Punctuation & flow	Overuse of commas leads to choppiness	Dash adds fluidity and rhythm	A dash helps organise the sentence and guide the reader through the contrast and metaphor more smoothly.

Area of improvement	Original phrase/sentence	Revised phrase/sentence	Explanation of revision
Register/Formality	Generally informal with grammatical inconsistencies	More formal, academically appropriate	The revised version aligns with the conventions of formal argumentative writing, particularly in exams like CAE.

A notable outcome of the revision process was the variation in essay length, with some texts increasing in length and others becoming shorter. These shifts reflect the different strategies employed by ChatGPT in attempting to improve clarity, coherence, and formal appropriateness.

In several cases, the essays became longer due to elaboration and the addition of contextual information. Where the original texts contained vague or compressed ideas, the model tended to expand on these, breaking down complex or ambiguous statements and rephrasing them using more explicit or formal language. This sometimes led to clearer thematic development and a more structured argumentative flow. However, these additions were not always necessary; in some instances, the expansions risked introducing redundancy or diluting the original conciseness of the student's message.

In contrast, other texts were shortened as ChatGPT removed repetitive phrasing or imprecise wording. These edits aimed to streamline the expression and improve the overall focus of the writing. While this often resulted in clearer and more concise prose, the risk in such reductions is the potential loss of nuance or tone, particularly when stylistic or rhetorical elements are interpreted as superfluous.

These differing outcomes illustrate the model's dual tendency to either expand or condense based on perceived linguistic efficiency, rather than a deep understanding of communicative intent. Although many of the changes contributed to improved readability and grammatical correctness, they were not always aligned with higher-order concerns such as argument structure, rhetorical impact, or authorial voice.

Overall, the variation in essay length reflects both the strengths and constraints of AI-assisted revision. While the model can aid in surface-level improvements and support more readable and formally appropriate texts, it operates with limited sensitivity to discourse-level decisions and the subtleties of individual expression. These limitations underscore the importance of human mediation in interpreting, accepting, or rejecting AI-generated suggestions in the writing process.

An initial descriptive analysis revealed that textual errors were the most frequent across the corpus, followed by morphological and lexical errors. Spelling and syntactic errors were comparatively less prevalent, suggesting that while learners generally demonstrated competence at the morpho-phonological level, challenges remained in managing cohesion, coherence, and lexical precision. This pattern aligns with prior findings in second language writing research, where discourse-level and lexical choices often represent more advanced stages of acquisition.

Table 3. Mean frequency of error types

Error Type	Mean Errors (across 11 texts)
Spelling	3.64
Lexical	4.45
Morphological	5.73
Syntax	3.64
Semantic	3.27
Textual	10.00

Total error counts ranged from 23 to 40 per text. Notably, texts with the highest frequency of errors did not always correspond to the lowest holistic scores, highlighting the complexity of writing assessment and the possible mitigating role of content, structure, or rhetorical effectiveness. This observation motivated a closer examination of correlations between quantifiable revision features and assessment outcomes.

Table 4. Descriptive statistics for key writing and assessment metrics

Metric	Mean	Min	Max
Total errors	30.09	23	40
Final word count	251.73	97	381
Percentage of words added (%)	22.04	-10.19	57.69
Percentage of mistakes	0.17	0.09	0.27
Initial essay grade	2.18	2	3
Improved essay grade	4.18	4	5

Taken together, the results summarised in Table 5 below demonstrate that ChatGPT can assist in improving both surface-level accuracy and overall clarity in student writing. However, the variation in error profiles and the persistence of textual coherence issues suggest that AI intervention is most effective when paired with explicit pedagogical guidance. Teachers should help students critically assess AI-generated changes and continue to focus on discourse-level skills that remain challenging for current language models.

While ChatGPT clearly demonstrates potential in supporting student writing—especially by offering immediate feedback, improving fluency, and correcting grammatical errors—it is not without significant limitations. These shortcomings must be acknowledged to avoid overestimating the tool's pedagogical value and to ensure its responsible integration into instructional practice.

One of the most consistent limitations observed across revised texts is ChatGPT's tendency to focus primarily on surface-level linguistic corrections (e.g., grammar, syntax, vocabulary), often at the expense of deeper critical engagement and argumentative development. The model can rephrase ideas with elegance and fluency, but it rarely enhances—or even interrogates—the underlying logic, depth,

Table 5. Linguistic error analysis, Word count variation, and Score improvement in student essays revised with ChatGPT assistance

	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Text 9	Text 10	Text 11
Spelling	4	0	2	5	2	3	0	2	7	8	0
Lexical	4	2	8	4	5	5	4	5	4	6	7
Morphological	5	7	3	9	10	4	5	5	5	3	4
Syntax	3	3	2	7	3	4	3	3	5	2	5
Semantic	2	6	3	3	2	2	2	6	3	2	3
Textual	6	5	11	12	10	12	14	10	11	12	9
TOTAL no of errors	24	23	29	40	32	30	28	31	35	33	28
Initial number of words	132	182	323	263	260	193	272	177	131	198	108
Final number of words	160	287	320	329	381	263	263	268	141	190	97
Percentage of words added	21,21%	57,69%	-0,93%	25,10%	46,54%	36,27%	-3,31%	51,41%	7,63%	-4,04%	-10,19%
Percentage of errors	0,18%	0,13%	0,09%	0,15%	0,12%	0,16%	0,10%	0,18%	0,27%	0,17%	0,26%
Initial essay grade	3	3	2	2	2	2	2	2	2	2	2
Improved essay grade	5	5	4	4	4	4	4	4	4	4	4

or originality of a student's argument. As a result, it remains an unreliable substitute for human feedback when it comes to evaluating reasoning, strengthening evidence, or guiding rhetorical sophistication. Consider the following example:

Original text (Text 1):

> *"Crime doesn't pay", I'm agree with that, because a crime means death, and subject "death" scared me. […]*
> *Like a girl it's too hard, because it's very dangerous to walk at night, because exists very strange people in this world.*

ChatGPT-improved text:

> *"Crime doesn't pay," because crime often leads to death, and the concept of death scares me.*
> *As a woman, it is especially difficult, as it feels dangerous to walk alone at night because of the unpredictable and sometimes threatening people in the world.* (OpenAI, 2024)

The revised version is undeniably more polished: grammatical errors are corrected, vocabulary is refined, sentence structure is improved, and the transitions between ideas are smoother. For example, "I'm agree with that" is corrected to "I agree with the saying," and "because exists very strange people in this world" becomes "because of the unpredictable and sometimes threatening people in the world." These revisions undeniably enhance fluency and coherence at the sentence and paragraph level.

However, what remains unchanged is the superficiality of the argument. The core ideas are neither interrogated nor expanded. For instance:

- The essay does not define or critically examine what the expression "Crime doesn't pay" truly entails (e.g., consequences for perpetrators, societal cost, justice systems).
- It relies heavily on emotive generalisations (e.g., "crime means death"), which are retained without prompting further reflection or factual support.
- The rhetorical strategy is predominantly personal and affective, but ChatGPT does not guide the writer toward incorporating more analytical or evidence-based reasoning that is typically expected in argumentative writing.

In this sense, ChatGPT improves the form while leaving the content underdeveloped. It corrects how things are said but does not challenge what is being said—or how well it is being argued. This example reinforces the need for teacher-led guidance to help students move beyond expression toward critical thinking, conceptual clarity, and argumentation skills. Without such oversight, AI-assisted revisions risk giving a false sense of academic adequacy while leaving key cognitive and rhetorical gaps unaddressed.

A second limitation would be the fact that ChatGPT occasionally offers revisions that, while grammatically accurate and stylistically polished, may

misinterpret context-specific meanings or inadvertently strip away subtle rhetorical and cultural nuances embedded in student writing. ChatGPT relies on statistical patterns in language rather than cultural insight, lived experience, or contextual awareness. For multilingual students—especially those writing in an academic second language—these nuances are not incidental. They are part of how meaning is negotiated across cultural and linguistic boundaries. Take, for instance, the following student sentence:

Original text (Text 4):

> I know that we don't to choose them, but it's very good when you have a good and happy family and you grow in love, just I know children who grows in alcoholic family (parents can hit children, smoke wids and drink alcohol – it's very bad and sad). I'm recognized it in my 19 years.

ChatGPT-improved text:

> While we cannot choose our parents, it is fortunate when a child grows up in a loving, supportive family. Sadly, some children grow up in dysfunctional or abusive households, where parents may drink excessively, use drugs, or even physically abuse their children. This environment can have a lasting impact, often leading to criminal behaviour later in life. (OpenAI, 2024)

While the AI-generated version is more coherent and syntactically refined, it entirely omits the phrase: *"I'm recognized it in my 19 years."* This omission likely stems from the phrase's grammatical inaccuracy and ambiguity, which made it difficult for the model to integrate fluently. However, this erasure is more than a stylistic decision—it is a loss of cultural and rhetorical meaning. The phrase signals a reflective, autobiographical stance in which the student—despite her limited English proficiency—asserts her personal awareness of social issues, such as the link between childhood trauma and future criminal behaviour. The sentence serves not only as a temporal anchor but also as a marker of personal maturity and lived insight.

As a native Romanian speaker and experienced writing instructor, I can decode this non-standard phrase. It is not merely an error; it represents an attempt to express a culturally inflected idea: a sense of precocious awareness, formed through personal and possibly painful observation. In Romanian, expressions of recognition or awakening often use reflexive or self-positioning verbs to describe moral or emotional realizations over time. Transferred into English through interlanguage, *"I'm recognized it in my 19 years"* becomes a hybrid construct—imperfect grammatically, but powerful rhetorically. The student was essentially saying: *"Even though I am only 19, I have already come to realise how much childhood and family life can affect someone's future choices."* This grounds the argument in lived experience and subtly shifts the tone from abstract commentary to personal testimony. By omitting it, ChatGPT not only sanitises the prose but also erases the student's cultural voice, flattening the distinctiveness of her narrative positioning.

This example illustrates why AI cannot replace the role of a culturally literate teacher. ChatGPT can identify and "correct" errors based on mainstream English norms, but it lacks the ability to interpret how non-native constructions might carry cultural, emotional, or rhetorical significance. Teachers, especially those with cross-cultural expertise, are able to preserve and even elevate these moments—offering feedback that improves clarity without silencing identity.

In short, writing is not just about linguistic form; it is a site of cultural negotiation. When AI tools revise student texts without sensitivity to this reality, they risk producing writing that is technically correct but culturally diminished. This is precisely why human oversight remains essential in AI-assisted pedagogy.

5 Conclusions and Recommendations

This study adopts a mixed-methods approach, combining quantitative measures—including CAE assessment scores, word count comparisons, and Word Error Rate—with qualitative analysis of error types and the nature of textual revisions. The aim is to evaluate the effectiveness of ChatGPT in improving the quality of argumentative essays. By examining both numerical gains and qualitative shifts in student writing, the study provides a comprehensive assessment of the tool's pedagogical potential, particularly in the context of preparing future English language teachers.

ChatGPT has the capacity to transform writing instruction by offering immediate, personalised feedback, enhancing linguistic accuracy, and supporting content development. However, its classroom use requires careful management to prevent overreliance and to uphold academic integrity. Educators should guide students in viewing AI not as a substitute for original thinking but as a resource to support the revision process. Clear instructional frameworks and ethical guidelines are essential for ensuring that students use such tools responsibly and effectively.

The findings of this study offer practical insights for integrating ChatGPT into academic writing instruction. While the data underscore the tool's potential to improve grammar, vocabulary, and clarity, they also highlight the importance of pedagogical strategies that foster critical engagement with AI-generated feedback. ChatGPT should be positioned as a complement to, rather than a replacement for, traditional instruction—enhancing students' ability to revise, reflect, and take ownership of their writing.

One of the most productive uses of ChatGPT in the writing classroom lies in supporting iterative drafting. Students can be encouraged to compose initial drafts independently and then use AI to explore alternative phrasing, clarify arguments, or expand underdeveloped sections. To avoid passive acceptance of suggestions, educators should scaffold the revision process by asking students to annotate the changes they make and explain their reasoning. This practice promotes metacognitive awareness and reinforces a sense of authorial agency.

ChatGPT also serves as a differentiated support mechanism, particularly valuable in mixed-proficiency classrooms. For less confident writers, it can provide

scaffolding in grammar and vocabulary, enabling clearer expression of ideas. For more advanced students, it offers opportunities to explore stylistic refinement and discourse conventions. To encourage meaningful engagement, teachers can design targeted prompts that align with specific instructional goals, such as eliminating redundancy or enhancing argumentative cohesion.

Despite these advantages, the study identifies textual coherence as a persistent challenge. While ChatGPT effectively addresses micro-level issues such as grammar and word choice, organisational aspects—such as logical flow and paragraph structure—often remain underdeveloped. This suggests that coherence must be taught explicitly. Educators should continue to focus on text structure through instruction on discourse markers, logical sequencing, and thematic progression. AI-generated outputs can serve as instructional materials for this purpose, with students analysing and revising them to improve structural clarity.

An essential component of AI-assisted writing instruction is the development of students' critical literacy and ethical awareness. Classroom integration of ChatGPT should include discussions about the role and limitations of AI in the writing process. Students need to understand that while AI can improve clarity and correctness, it should not replace their original ideas or rhetorical intentions. Teachers can support this by encouraging students to evaluate AI-generated content critically and reflect on the effects of their revisions. Assignments that include revision logs, reflective commentaries, or comparisons of different feedback sources (peer, teacher, AI) can help students adopt a more discerning and reflective approach.

Assessment practices in AI-integrated classrooms should also evolve. Rather than focusing solely on final drafts, educators should consider how students engage with feedback, revise their texts, and demonstrate understanding of genre and communicative purpose. Reflective writing and documentation of revision processes can offer deeper insights into students' development and provide evidence of learning beyond surface-level correctness.

In conclusion, ChatGPT offers significant potential as a pedagogical tool for enhancing student writing, especially in terms of grammatical accuracy and lexical development. However, its effective classroom integration requires thoughtful instructional design that promotes critical engagement, structural awareness, and responsible authorship. Teachers play a crucial role in mediating students' interactions with AI, ensuring it acts not as a shortcut, but as a means of fostering deeper learning, revision skills, and greater independence in academic writing.

References

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting Ai-generated writing in second language education in higher education. Teaching English with Technology. doi: https://doi.org/10.56297/buka4060/xhld5365

Al-Obaydi, L. H., Pikhart, M., & Klimova, B. (2023). ChatGPT and the general concepts of education: Can artificial intelligence-driven chatbots support the process of language learning? *International Journal of Emerging Technologies in Learning (iJET)*, 18(21), 39–50. doi: https://doi.org/10.3991/ijet.v18i21.42593

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. doi: https://doi.org/10.1016/j.asw.2023.100745

Bellman, R. (1978). *An introduction to artificial intelligence: can computers think?*. Boyd & Fraser.

Boskabadi, E.; Nikfar, M.; Ugwuoke, J.; Ali, H. (2024). ChatGPT in Teaching and Learning: A Systematic Review. *Education Sciences, 14*, 643. doi: https://doi.org/10.3390/educsci14060643

Burkhardt, J., MacDonald, M., & Rathemarcher, A. (2010). *Teaching information literacy*. American Library Association.

Cambridge University Press & Assessment. (2024). *Teacher guide for writing: C1 Advanced*. https://www.cambridgeenglish.org/images/600976-teacher-guide-for-writing-c1-advanced.pdf

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, *27*(3), 419–436. doi: https://doi.org/10.1177/0265532210364391

De Laat, M., Joksimovic, S., & Ifenthaler, D. (2020). Artificial intelligence, real-time feedback and workplace learning analytics to support in situ complex problem-solving: A commentary. *International Journal of Information and Learning Technology, 37*(5), 267–277. doi: https://doi.org/10.1108/IJILT-03-2020-0026

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 498–506. doi: https://doi.org/10.1016/j.dss.2010.06.003.

Ferlazzo, L. (January 18, 2023). 19 Ways to Use ChatGPT in Your Classroom. *Education Week.* https://www.edweek.org/teaching-learning/opinion-19-ways-to-use-chatgptin-your-classroom/2023/01

Frankfurt, H. (2005). *On Bullshit*, Princeton.

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., Uhlig, S., & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems, 4*, 19–23. doi: https://doi.org/10.1016/j.iotcps.2023.06.002

Gonda, D. E., & Chu, B. (2019). Chatbot as a learning resource? Creating conversational bots as a supplement for teaching assistant training course. In *Proceedings of the IEEE International Conference on Engineering, Technology and Education (TALE)*, Yogyakarta, Indonesia, 10–13 December 2019 (pp. 1–5). IEEE. doi: https://doi.org/10.1109/TALE48000.2019.9225974

Gottschalk, K., & Hjortshoj, K. (2004). *The elements of teaching writing: A resource for instructors in all disciplines*. Bedford/St. Martin's.

Habeb Al-Obaydi, L., Pikhart, M., & Klimova, B. (2023). ChatGPT and the General Concepts of Education: Can Artificial Intelligence-Driven Chatbots Support the Process of Language Learning?. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(21), pp. 39–50. doi: https://doi.org/10.3991/ijet.v18i21.42593

Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology, 26*, Article 38. doi: https://doi.org/10.1007/s10676-024-09775-5

Jasin, J., Ng, H. T., Atmosukarto, I., Iyer, P., Osman, F., Wong, P. Y. K., Pua, C. Y., & Cheow, W. S. (2023). The implementation of chatbot-mediated immediacy for synchronous communication in an online chemistry course. *Education and Information Technologies*, *28*, 10665–10690. doi: https://doi.org/10.1007/s10639-023-11602-1

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed - an argument for AI in education. Retrieved from http://discovery.ucl.ac.uk/1475756/

OpenAI. (2024). *ChatGPT* (May 16 version) [Large language model]. https://chat.openai.com/

Pecorari, D. (2008). *Academic writing and plagiarism*. Norfolk.

Popescu, T. (2013). A corpus-based approach to translation error analysis: A case-study of Romanian EFL learners. *Procedia - Social and Behavioral Sciences, 83*, 242–247. doi: https://doi.org/10.1016/j.sbspro.2013.06.048

Ramos, J. (2023). *ChatGPT for education: A practical guide*. Berlin: Verlag GD Publishing Ltd. & Co KG.

Skrabut, S. (2023). 80 Ways to use ChatGPT in the classroom using AI to enhance teaching and learning.

Strzelecki, A., Cicha, K., Rizun, M. et al. (2024). Acceptance and use of ChatGPT in the academic community. *Education and Information Technologies, 29*, 22943–22968. doi: https://doi.org/10.1007/s10639-024-12765-1

Zawacki-Richter, O., Marín, V. I., Bond, M., et al. (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*(39). doi: https://doi.org/10.1186/s41239-019-0171-0