



## An Analysis of ChatGPT-Generated Feedback on Argumentative Essays in Korean EFL Classrooms

Hyun-Joo Oh (University of Seoul)  
Shu-Fei Hsieh (National Taiwan Normal University)

Received: March 11 2025  
Revised: April 6 2025  
Accepted: April 14 2025

Oh, Hyun-Joo, & Hsieh, Shu-Fei. (2025). An analysis of ChatGPT-generated feedback on argumentative essays in Korean EFL classrooms. *Modern English Education*, 26, 166-181.

### Keywords

ChatGPT-generated feedback, Artificial Intelligence, argumentative essays, Korean EFL classrooms  
ChatGPT 피드백, 인공지능, 논쟁적 에세이, 한국 EFL 교실

### Hyun-Joo Oh

(First author)  
Visiting Professor  
Institute of Education Innovation  
University of Seoul  
[hyunjoo@uos.ac.kr](mailto:hyunjoo@uos.ac.kr)

### Shu-Fei Hsieh

(Corresponding author)  
Ph.D. Student  
Department of English  
National Taiwan Normal University  
[sandymissesin@gmail.com](mailto:sandymissesin@gmail.com)

### Abstract

In response to the increasing use of ChatGPT, this study aimed to analyze ChatGPT-generated feedback on 126 argumentative essays written by university students in Korean English as a Foreign Language (EFL) classrooms, focusing on content, organization, language and grammar, and mechanics. Results revealed that ChatGPT assigned the highest score to content but the lowest score to language and grammar using a five-point scale. Furthermore, ChatGPT identified recurring writing issues, including a lack of clear thesis statements, underdeveloped arguments, insufficient specific examples or evidence, weak transitions and cohesion, unclear counterarguments, and grammar and syntax problems. These issues need to be reflected in writing instruction for improving argumentative essays. When it comes to the strengths of ChatGPT-generated feedback, it provided immediate and scalable feedback, explicit corrective feedback on language and mechanics, and a standardized overall evaluation that offered a comprehensive view at a glance. However, ChatGPT-generated feedback also had the limitations, including formulaic comments on content and organization, generic and overgeneralized feedback on final suggestions for improvement, and a lack of in-depth analysis of counterarguments and logical flow. The current study has pedagogical implications, offering valuable insights into the interpretation and analysis of ChatGPT-generated feedback as well as recommendations for its optimal integration within EFL contexts.

## INTRODUCTION

With the rise of artificial intelligence (AI) integration in education and various other fields, ChatGPT, one of the AI-powered tools, is increasingly utilized by both students and teachers in English language classes, causing a paradigm shift. ChatGPT is a state-of-the-art language model based on OpenAI's Generative Pretrained Transformer (GPT), designed to generate human-like text drawn from patterns it has acquired through vast amounts of data and interact with users in a seemingly natural and intuitive manner (Rudolph et al., 2023). Since its launch in November 2022, ChatGPT has developed rapidly, with ChatGPT-4 currently outperforming ChatGPT-3 in terms of logical structure, language complexity, vocabulary

diversity, and text linking (Herbold et al., 2023). At this point, ChatGPT is the world's most advanced chatbot (Rudolph et al., 2023), and it is expected to transform learning objectives, teaching methods, materials, assessment, and evaluation (Zhai, 2022), improving how English as a Foreign Language (EFL) is taught and learned (Kim, 2025), thereby revolutionizing EFL education (Koraishi, 2023).

In EFL classrooms, students can use ChatGPT to generate ideas, summarize reading texts, check grammar and difficult words, or assist with assignments. However, there are some concerns about ethical issues such as plagiarism, data privacy, and over-reliance on AI, and it may be difficult for teachers to distinguish ChatGPT-generated writing from the writing produced by students without the help of AI detectors including GPTZero, ZeroGPT, or GPT killer. Considering these issues, Herbold et al. (2023) claimed that as with the use of calculators, it is essential to engage students in critical discussions about the appropriate use of this tool reflecting on when and how to use it. Meanwhile, teachers can utilize ChatGPT to enhance students' English language learning, assist with lesson planning and classroom activities, or provide feedback on students' writing and grade their work.

In the realm of academic writing, argumentative writing—the most common genre in higher education—is important because the ability to compose argumentative essays is regarded as an indicator of students' writing proficiency (Hirvela, 2017) and requires students' problem-solving skills and critical thinking abilities. Argumentative essays ask whether one agrees or disagrees with a topic or issue and appear in various fields, such as university academic writing, standardized tests like TOEIC Writing and TOEFL, as well as research papers and journals. Thus, engaging with argumentative essays can empower students academically by improving their ability to search for reliable sources and effectively summarize and integrate strong evidence to support their arguments (Thompson, 2021). Unlike other language skills, improving writing typically demands consistent practice and thorough feedback (Chen, 2024). Providing feedback is particularly crucial in EFL writing classes, as there are limited opportunities for students to use English outside the classroom in EFL contexts. One form of writing feedback, ChatGPT, can provide immediate suggestions and is user-friendly; however, it remains uncertain whether ChatGPT-generated feedback offers comprehensive guidance or adequately addresses the fundamental elements of argumentative writing. Therefore, it is necessary to examine the analysis of ChatGPT-generated feedback on students' argumentative essays.

Given that more research is needed to investigate whether ChatGPT has the potential to supplement teacher feedback and how teachers can optimize the integration of ChatGPT-generated feedback in EFL writing classes, the present study aimed to categorize and analyze the patterns and tendencies of ChatGPT-generated feedback, provide suggestions for the improvement of argumentative essays, and find out the strengths and weaknesses of ChatGPT-generated feedback. It has implications for offering practical insights into the effective use of ChatGPT-generated feedback in EFL writing instruction. Hence, the following research questions were formulated:

- 1) What are the most frequent writing issues identified by ChatGPT-generated feedback in students' argumentative essays in Korean EFL classrooms?
- 2) What suggestions for improvement does ChatGPT provide, based on students' common errors or repeated mistakes, to teach argumentative essays effectively?
- 3) What are the strengths and limitations of ChatGPT-generated feedback, and how can it be integrated into writing instruction?

## LITERATURE REVIEW

### Key Features of Argumentative Essays

The ability to recognize, evaluate, and construct arguments has been considered a pivotal academic skill, and the development of argumentative writing has been one of the essential objectives of writing instruction (Helms-Park & Stapleton, 2003; Mitchell & Riddle, 2000). Argumentative essays are essays in which students agree or disagree with an issue, using reasons to support their opinions, and the goal of argumentative writing is to convince a reader that one's opinion is valid (Oshima & Hogue, 2020). Therefore, argumentative essays encourage students to think critically about a topic and present their ideas clearly and precisely with logical reasoning. According to Oshima and Hogue (2020), "Argumentation is a popular kind of essay question because it forces students to think on their own: They have to take a stand on the issue,

support their stand with solid reasons, and support their reasons with solid evidence” (p. 153). Primary elements of argumentative essays are “an explanation of the issue, a clear thesis statement, a summary of the opposing arguments, rebuttals to the opposing arguments, and one’s own arguments” (Oshima & Hogue, 2020, p. 156). In argumentative essays, it is crucial to organize one’s arguments effectively and to discuss the other side’s reasons and rebut them, which means pointing out problems with the other side’s reasons in order to prove that they are not good reasons and being reasonable and open-minded, considering all sides of the issue (Oshima & Hogue, 2006).

In order to help students write stronger argumentative essays with well-supported arguments and provide effective feedback, teachers and students need to know about key features of argumentative writing and understand the Toulmin Model of Argumentation (Toulmin, 2003), a framework for constructing and evaluating arguments, making it applicable to argumentative essays. It identifies six essential components: 1) Claim, 2) Grounds (Evidence/Data), 3) Warrant, 4) Backing, 5) Qualifier, and 6) Rebuttal (see Figure 1).



**FIGURE 1**

*Toulmin Model of Argumentation (Purdue Online Writing Lab, 2025)*

According to the model, every argument begins with a claim, the grounds, and the warrant; a claim is an assertion or the main point, the grounds are the evidence such as reasons, facts or proof that support the argument, and the warrant is the logical reasoning that links the grounds to the argument (Purdue Online Writing Lab, 2025). Following that, backing is the additional support for the warrant, the qualifier describes the strengths of the argument, and rebuttals are counterarguments that show a different perspective on the issue (Toulmin, 2003). These elements can be criteria or standards for analyzing the structure and content of students’ argumentative essays, providing feedback, and evaluating them.

## Feedback in Writing Instruction

Feedback is a powerful tool that can facilitate students’ writing process (Bitchener & Storch, 2016) and improve their writing performance. In addition, feedback helps to heighten students’ self-awareness regarding their strengths and areas for improvement (Ramsden, 2003), and students become more aware of their errors when they recognize a gap between their current language use (interlanguage) and the target language form by receiving feedback (Ataş et al., 2024). According to Zhai and Ma (2023), it becomes important to offer instructional writing feedback not only in the language components but also in dimensions such as content and organization because writing is a multifaceted skill, engaging with complex elements including idea development, unity and coherence, and writing style.

The most commonly used feedback type in the English as a Second Language (ESL) or EFL context is teacher feedback (Hyland & Hyland, 2019), and Kurt and Kurt (2024)’s study showed that students perceived teacher feedback as helpful and motivating and preferred teacher feedback rather than other types of feedback. Teacher feedback can represent the social side of learning with the expert (Kushmar et al., 2022) and provide personalized feedback reflecting an individual student’s writing performance and gradual progress since teachers are most familiar with each student (see Table 1). Furthermore, teacher feedback facilitates active dialogue with students, allowing for clarification

and further improvement, and provides constructive feedback based on students' needs and proficiency levels (Lent et al., 2022). However, the role of teachers as feedback providers has been challenged due to factors such as large class sizes, time constraint, and heavy workloads (EIEbyary & Shabara, 2024; Lee, 2017). In addition, human subjectivity or bias may affect the feedback and scoring process, such as inter-rater reliability stemming from a lack of adherence of standardized criteria, inexperience, or preconceived biases, and intra-rater reliability stemming from unclear criteria, bias toward students, or simple carelessness (Brown, 2010).

In the meantime, peer feedback is another highly valued form of feedback, offering a variety of perspectives on writing (Yao et al., 2021), and it is advantageous for both providers and recipients since it contributes to the improvement of their writing performance (Kurt & Kurt, 2024). Just as young children acquire language more quickly through interactions with their peers, students are able to enhance their writing skills through peer feedback by learning from each other, which leads to collaborative learning that aligns with constructivism. Moreover, peer feedback provides students with the opportunity to analyze and evaluate others' writing from a different viewpoint, engage in self-reflection, and gain a new perspective on their own writing (Kurt & Kurt, 2024). Students also reflect on misunderstandings or misinterpretations in their writing through peer feedback and develop their critical thinking skills by immersing themselves in a dynamic learning environment (Valero Haro et al., 2023). Despite these advantages, there are limitations; peer feedback may not always offer high-quality feedback because peers are not experts and lack experience or a solid understanding of feedback principles (Banihashem et al., 2024; Noroozi et al., 2016). Additionally, peer feedback cannot be consistent or standardized, and it can vary depending on different peers, influenced by close friendships (Sousa et al., 2024).

With the advancement of AI technology, ChatGPT provides automated feedback and assists teachers in the feedback process, thereby reducing their workload (Kurt & Kurt, 2024). ChatGPT can be effectively utilized as a valuable tool to enhance productivity in delivering quality feedback when prompts are presented clearly and precisely (Carlson et al., 2023). The advantages of ChatGPT-generated feedback include offering timely responses on numerous assignments while saving time and effort (EIEbyary & Shabara, 2024; Solak, 2024) and demonstrating medium-high to high effects on learning (Stevenson & Phakiti, 2024). Moreover, Algaraady and Mahyoob (2023) found that ChatGPT-generated feedback can effectively identify surface-level errors such as grammar, spelling, and mechanics. However, there are some drawbacks to ChatGPT-generated feedback; it may hallucinate, producing inaccurate information (Alkaissi & McFarlane, 2023), and it may be less effective regarding cohesion and coherence, lacking practical and actionable steps for improvement (Yoon et al., 2023).

**TABLE 1**  
*Advantages and Disadvantages of Different Feedback Types*

Feedback types	Advantages	Disadvantages
Teacher feedback	<ul style="list-style-type: none"> <li>- is the social side of learning with the expert (Kushmar et al., 2022).</li> <li>- provides personalized guidance that reflects an individual student's writing performance and gradual progress.</li> <li>- facilitates active dialogue with students, allowing for clarification and further improvement, while also providing constructive feedback tailored to their needs and proficiency levels (Lent et al., 2022).</li> </ul>	<ul style="list-style-type: none"> <li>- may impose time and workload constraints on teachers (EIEbyary &amp; Shabara, 2024; Lee, 2017).</li> <li>- may be subjective or biased, leading to inconsistent feedback or evaluations, influenced by inter-rater reliability or intra-rater reliability (Brown, 2010).</li> </ul>
Peer feedback	<ul style="list-style-type: none"> <li>- promotes a collaborative learning, aligning with constructivism.</li> <li>- gives the opportunity to analyze and evaluate others' writing and engage in self-reflection (Kurt &amp; Kurt, 2024).</li> <li>- allows students to identify misunderstandings or misinterpretations in their writing and improve their critical thinking skills by engaging in an interactive learning environment (Valero Haro et al., 2023).</li> </ul>	<ul style="list-style-type: none"> <li>- may not always offer high-quality feedback due to a lack of experience or a solid understanding of feedback principles (Banihashem et al., 2024; Noroozi et al., 2016).</li> <li>- may not be standardized and can vary depending on different peers (Sousa et al., 2024).</li> </ul>

Feedback types	Advantages	Disadvantages
ChatGPT-generated feedback	<ul style="list-style-type: none"> <li>- provides instant and immediate feedback, saving time and effort (EIEbyary &amp; Shabara, 2024; Solak, 2024).</li> <li>- demonstrates medium-high to high effects on learning (Stevenson &amp; Phakiti, 2024).</li> <li>- effectively identifies surface-level errors such as grammar, spelling, and mechanics (Algaraady &amp; Mahyoob, 2023).</li> </ul>	<ul style="list-style-type: none"> <li>- may hallucinate, generating inaccurate information (Alkaissi &amp; McFarlane, 2023).</li> <li>- may be less effective with regard to cohesion and coherence and deficient in actionable steps for improvement (Yoon et al., 2023).</li> </ul>

## Previous Studies on ChatGPT-Generated Feedback

There has been considerable previous research on ChatGPT-generated feedback in writing classes, including comparisons of teacher feedback with ChatGPT-generated feedback (EIEbyary & Shabara, 2024; Guo & Wang, 2023; Solak, 2024), comparisons of peer feedback with ChatGPT-generated feedback (Banihashem et al., 2024), the effectiveness of ChatGPT-generated feedback (Ataş et al., 2024; Kurt & Kurt, 2024), the evaluation of the quality of ChatGPT-generated feedback (Yoon et al., 2023), and the analysis of ChatGPT-assisted teacher feedback's impact on students' writing and motivation (Park & Kim, 2024).

Guo and Wang (2023) compared ChatGPT-generated feedback with teacher feedback from five teachers on 50 university students' argumentative essays in Chinese EFL classes, analyzing them in terms of content, organization, and language. The results showed that ChatGPT generated more feedback comments compared to teacher feedback, and teachers mainly focused on content and language, whereas ChatGPT delivered more balanced feedback across content, organization, and language aspects (Guo & Wang, 2023). Similarly, EIEbyary and Shabara (2024) compared the corrective feedback generated by teachers and ChatGPT on 32 cause-and-effect essays, focusing on the adequacy of feedback strategies and guidance as well as the accuracy and polarity (positive or negative feedback). The study revealed that teachers were skilled at providing direct error correction, while ChatGPT tended to offer more metalinguistic feedback, rewording correct sentences and delivering guidance in a less direct manner (EIEbyary & Shabara, 2024). Likewise, Solak (2024) examined feedback from both teachers and ChatGPT on students' assignments concerning content, vocabulary, grammar, cohesion, coherence, and overall evaluation, and found that teachers were valued for their empathetic and personalized feedback as well as their emotional insight. However, ChatGPT-generated feedback was regarded as more thorough and detailed, stressing the importance of integrating both human and AI feedback to enhance overall effectiveness (Solak, 2024). A similar but slightly different study was conducted by Banihashem et al. (2024), who compared peer feedback with ChatGPT-generated feedback. The results showed that peer feedback and ChatGPT feedback could play complementary roles in the feedback process, with peer feedback concentrating on identifying issues and encouraging critical analysis, while ChatGPT offered more detailed, descriptive comments on essay composition (Banihashem et al., 2024).

In relation to the effectiveness of ChatGPT-generated feedback in EFL classrooms, Kurt and Kurt (2024) concluded that ChatGPT has the potential to be a useful feedback tool based on affordances such as practicality, interactivity, and adaptability. Also, although teachers and peers may not always be available to provide immediate and extensive feedback, ChatGPT can fill this gap and function as a facilitator owing to its interactive and adaptable characteristics (Kurt & Kurt, 2024). In the same vein, Ataş et al. (2024) analyzed the effectiveness of ChatGPT-generated feedback on EFL compare-and-contrast essays, and the results showed that ChatGPT provided exceedingly precise feedback on both content and organizational aspects of writing. In spite of the high accuracy, ChatGPT was unable to provide feedback in all necessary areas or offer a holistic view of the entire essay, suggesting that a collaborative approach between teachers and AI is essential for delivering the optimal feedback to students (Ataş et al., 2024). Similarly, Yoon et al. (2023) evaluated the quality of ChatGPT-generated feedback regarding coherence and cohesion of argumentative essays written by English language learners, using English Language Learners' Integrated Performance and Scoring Evaluation (ELLIPS) rubric. The findings of their study revealed that ChatGPT-generated feedback was mostly generic and general, relying on superficial linguistic elements, which was not effective for adjustments and meaningful revisions of coherence and cohesion (Yoon et al., 2023). Finally, Park and Kim's (2024) study examined the effects of ChatGPT-assisted teacher feedback on the writing performance and motivation of Korean EFL students in comparison with teacher feedback in control and experimental groups. According to the research findings, combining the ChatGPT tool with teacher feedback was found to improve students' writing skills including accuracy, fluency, and complexity, and to boost their motivation (Park & Kim, 2024).

## METHOD

### Participants

The participants in this study consist of 126 freshmen (76 male students & 50 female students) who took *College English (Writing)* at S University which is located in Seoul, Korea (see Table 2). At S University, first-year students are required to take *College English (Speaking)* for one semester and *College English (Writing)* for the other semester as part of their required electives, depending on their major. One of the two researchers was the instructor and taught three *College English (Writing)* classes in the spring semester and three in the fall semester in 2024. The instructor earned her master's and doctoral degrees in the United States, specializing in foreign and second language education and TESOL, and has been teaching at universities since 2006.

**TABLE 2**  
*Number of Student Participants*

	Spring semester	Fall semester
Male	44	32
Female	22	28
Total	66	60

For the present research, argumentative essays written by 66 students majoring in STEM fields in the spring semester and 60 students majoring in humanities and social sciences in the fall semester of 2024 (see Table 3) were collected. Students took a *pre-test* (G-TELP and an essay writing test) before the semester began for the placement test, and the average G-TELP score of student participants in the spring semester was 180.85 out of 300, and that of the fall semester students was 194.45 out of 300. Their English proficiency levels ranged from intermediate to upper intermediate.

**TABLE 3**  
*Students' Majors in the Spring & Fall Semesters*

Spring semester (Science, technology, engineering, and mathematics, $N=66$ )	$N$	Fall semester (Humanities and social sciences, $N=60$ )	$N$
Department of Architecture	13	Department of Economics	14
Department of Computer Science	11	Department of Public Administration	10
Department of Chemical Engineering	7	Department of Social Welfare	6
Department of Civil Engineering	7	Department of International Relations	6
Department of Statistics	6	Department of English Language and Literature	5
Department of Mechanical and Information Engineering	6	College of Convergence Studies	5
Department of Geospatial Information Engineering	5	Department of Taxation	4
Department of Artificial Intelligence	4	Department of Korean History	3
Department of Physics	2	Department of Philosophy	3
Department of Applied Chemistry	2	Department of Korean Language and Literature	2
College of Convergence Studies	1	Department of Chinese Language and Culture	1
Department of Advanced Convergence Studies	1	College of Liberal Studies	1
Department of Environmental Horticulture	1		

### Data Collection

In the *College English (Writing)* class, academic English writing was taught to freshmen using a step-by-step approach (writing a paragraph, writing an essay, and understanding sentence structure) with the textbook *Longman Academic Writing*

*Series 4.* The objective of the course was that students would be able to write an argumentative essay consisting of 4-5 paragraphs by the end of the semester. For this study, a total of 126 argumentative essays were collected from the spring and fall semesters of 2024. Over the course of 12 weeks, the classes were conducted as follows each week (see Table 4).

**TABLE 4**  
*Weekly Course Schedule*

Week	Course content
Week 1	Introduction
Week 2	Writing a Paragraph - Paragraph Structure (1)
Week 3	Writing a Paragraph - Paragraph Structure (2)
Week 4	Writing a Paragraph - Unity and Coherence
Week 5	Writing an Essay - From Paragraph to Essay (1)
Week 6	Writing an Essay - From Paragraph to Essay (2)
Week 7	Writing an Essay - Argumentative Essays
Week 8	Midterm Exam
Week 9	Sentence Structure - Types of Sentences
Week 10	Sentence Structure - Parallelism and Sentence Problems
Week 11	Sentence Structure - Adverb Clauses
Week 12	Sentence Structure - Adjective Clauses

After learning about argumentative essays, in the 13th week of the semesters, during class, students from six classes (three classes per semester) were instructed to write an argumentative essay composed of 4-5 paragraphs as a one-hour assignment. They were permitted to use two computers at the front of the classroom when needed, with only a Korean-English or English-Korean dictionary allowed. Students were not permitted to use their cell phones to secure a fair testing environment, where they could write based on their writing skills without the assistance of translation tools or AI. All the essays were written on the same topic to guarantee consistency in the data analysis: ‘The country is a better place for children to grow up than a large city is. Do you agree or disagree with that statement? Use specific examples and reasons to explain your answer.’

## Data Analysis

After data collection, the researchers entered the following prompt in the same manner and received ChatGPT-generated feedback for each essay, respectively, using ChatGPT-4:

*The topic of argumentative essay is ‘The country is a better place for children to grow up than a large city is. Do you agree or disagree with that statement? Use specific examples and reasons to explain your answer.’ Evaluate the following argumentative essay and provide detailed feedback on its **content, organization, language, and mechanics**. Highlight strengths, identify weaknesses, and offer specific suggestions for improvement.*

Since the research findings could vary depending on the wording, sentence structure, and content of the prompt, the two researchers thought deeply about the prompt and had extensive discussions while thoroughly examining six essential components (claim, grounds (evidence/data), warrant, backing, qualifier, and rebuttal) identified by the Toulmin Model of Argumentation (Toulmin, 2003). The researchers then experimented with various prompts reflecting Lo (2023)’s CLEAR (“Concise, Logical, Explicit, Adaptive, and Reflective” (p. 1)) framework as a standard method for creating prompts. The content of the prompt includes a thesis statement, arguments and evidence, counterarguments and rebuttals, and relevance and depth. The organization of the prompt consists of an introduction, body paragraphs, a conclusion, along with unity and coherence. The language component involves grammar, sentence structure, word choice, clarity, and academic tone, while the mechanics encompass spelling, punctuation, capitalization, and formatting. In order to analyze the common patterns and tendencies in ChatGPT-generated feedback on students’ argumentative essays as well as the most frequent writing issues across multiple essays, “Pattern Coding” was employed to condense segments of data and group those summaries into

categories, themes, or concepts (Saldaña, 2016). “Pattern Coding” can be an effective method for organizing qualitative data and drawing significant insights from complex data sets. This method enabled the researchers to identify common writing issues, such as weaknesses in argument development, coherence, or language usage, and analyze how these challenges appeared consistently in different student essays. Accordingly, the researchers were able to recognize recurring patterns and gain a deeper understanding of prevalent writing difficulties.

## FINDINGS AND DISCUSSION

### Most Common Writing Problems Identified by ChatGPT-Generated Feedback

When each argumentative essay was entered with the same prompt, ChatGPT provided feedback as follows, in terms of strengths, areas for improvement, and suggestions regarding content, organization, language and mechanics, including an overall evaluation (see Figure 2):

<p><b>Content</b></p> <p><b>Strengths:</b></p> <ul style="list-style-type: none"> <li>The essay presents a clear stance in favor of cities as a better place for children to grow up.</li> <li>Three main arguments are provided: <b>lack of diversity in the countryside, inconvenient transportation, and better education in cities.</b></li> <li>Some real-world examples (e.g., public transportation limitations) help support the argument.</li> </ul> <p><b>Areas for Improvement:</b></p> <ul style="list-style-type: none"> <li>The thesis statement is somewhat unclear and contains grammatical errors that affect readability. For example, “<i>I degree that the country is a better place for children to grow up than a large city.</i>” seems to contradict the position taken.</li> <li>The explanations for each point lack depth and specific examples. For instance, when discussing education, mentioning <b>specific programs, school rankings, or extracurricular opportunities</b> in cities would strengthen the argument.</li> <li>The conclusion is missing. A strong conclusion should summarize key points and reinforce the argument.</li> </ul> <p><b>Suggestions:</b></p> <ul style="list-style-type: none"> <li>Revise the thesis statement for clarity. For example: <i>Although some people believe that the countryside is a good place to raise children, I argue that a city provides better opportunities due to its diversity, convenient transportation, and superior education system.</i></li> <li>Expand on the reasons with more examples or statistical evidence where possible.</li> <li>Add a <b>clear concluding paragraph</b> that restates the argument and offers a final thought.</li> </ul> <hr/> <p><b>Organization</b></p> <p><b>Strengths:</b></p> <ul style="list-style-type: none"> <li>The essay follows a <b>logical structure</b> with an introduction, body paragraphs, and an implied conclusion.</li> <li>Each body paragraph introduces a main idea and follows a general pattern of explanation.</li> </ul> <p><b>Areas for Improvement:</b></p> <ul style="list-style-type: none"> <li>The introduction could be <b>more engaging</b>. The phrase “<i>To persuade this, I will explain...</i>” is unnatural and does not effectively introduce the argument.</li> <li>The transition between ideas could be smoother. For example, the second paragraph begins with “<i>First, the countryside lacks diversity.</i>” While this is clear, using a transition like “<i>One major drawback of living in the countryside is its lack of diversity.</i>” would improve flow.</li> </ul> <p><b>Suggestions:</b></p> <ul style="list-style-type: none"> <li>Improve the introduction with a <b>hook</b> (a general statement about the topic) before stating the thesis.</li> <li>Use <b>better transitions</b> between ideas, such as “<i>Additionally,</i>” or “<i>Another reason why cities are better is that...</i>”.</li> <li>Add a <b>conclusion</b> that restates key arguments and leaves a strong final impression.</li> </ul>	<p><b>Language and Mechanics</b></p> <p><b>Strengths:</b></p> <ul style="list-style-type: none"> <li>The essay uses <b>simple and understandable vocabulary.</b></li> <li>The overall meaning of the essay is clear despite some grammatical mistakes.</li> </ul> <p><b>Areas for Improvement:</b></p> <ul style="list-style-type: none"> <li>Many <b>grammar mistakes</b> and awkward phrases affect clarity. Examples:             <ul style="list-style-type: none"> <li>“<i>To persuade this</i>” → “<i>To support my argument</i>”</li> <li>“<i>The city have much higher education system than country.</i>” → “<i>Cities have a much better education system than the countryside.</i>”</li> <li>“<i>This can intend children to look at the worlds in narrow sight.</i>” → “<i>This may cause children to develop a narrow perspective of the world.</i>”</li> </ul> </li> <li>Word choice is sometimes awkward. “<i>Homogeneous in terms of socioeconomic status</i>” is too formal compared to the rest of the essay.</li> <li>Incorrect pluralization and verb agreement (e.g., “<i>The city have</i>”, “<i>the number of bus station is only one</i>”) should be corrected.</li> </ul> <p><b>Suggestions:</b></p> <ul style="list-style-type: none"> <li>Proofread carefully or use <b>grammar-checking tools</b> to correct errors.</li> <li>Pay attention to <b>verb agreement</b> (e.g., “<i>The city has</i>”, not “<i>The city have</i>”).</li> <li>Replace unnatural phrases with more natural alternatives.</li> </ul> <hr/> <p><b>Overall Evaluation</b></p> <table border="1"> <thead> <tr> <th>Category</th> <th>Score (Out of 5)</th> <th>Comments</th> </tr> </thead> <tbody> <tr> <td>Content</td> <td>★★★☆☆ (3/5)</td> <td>Needs more development and specific examples. A conclusion is missing.</td> </tr> <tr> <td>Organization</td> <td>★★★☆☆ (3/5)</td> <td>Logical structure, but transitions and introduction need improvement.</td> </tr> <tr> <td>Language &amp; Grammar</td> <td>★★☆☆☆ (2/5)</td> <td>Many grammar errors and awkward phrasing reduce clarity.</td> </tr> <tr> <td>Mechanics</td> <td>★★☆☆☆ (2/5)</td> <td>Incorrect pluralization, verb agreement, and sentence structure issues.</td> </tr> </tbody> </table> <p><b>Final Suggestions</b></p> <ul style="list-style-type: none"> <li><b>Revise the thesis statement</b> to make it clearer.</li> <li><b>Strengthen arguments</b> with more specific details and examples.</li> <li><b>Improve transitions</b> to make the essay flow more smoothly.</li> <li><b>Fix grammar and sentence structure errors</b> to improve readability.</li> <li><b>Add a conclusion</b> to reinforce the argument</li> </ul>	Category	Score (Out of 5)	Comments	Content	★★★☆☆ (3/5)	Needs more development and specific examples. A conclusion is missing.	Organization	★★★☆☆ (3/5)	Logical structure, but transitions and introduction need improvement.	Language & Grammar	★★☆☆☆ (2/5)	Many grammar errors and awkward phrasing reduce clarity.	Mechanics	★★☆☆☆ (2/5)	Incorrect pluralization, verb agreement, and sentence structure issues.
Category	Score (Out of 5)	Comments														
Content	★★★☆☆ (3/5)	Needs more development and specific examples. A conclusion is missing.														
Organization	★★★☆☆ (3/5)	Logical structure, but transitions and introduction need improvement.														
Language & Grammar	★★☆☆☆ (2/5)	Many grammar errors and awkward phrasing reduce clarity.														
Mechanics	★★☆☆☆ (2/5)	Incorrect pluralization, verb agreement, and sentence structure issues.														

**FIGURE 2**  
ChatGPT-Generated Feedback on an Argumentative Essay

The results of the analysis revealed that the average word count of argumentative essays in STEM fields was 274.85, whereas that of humanities and social sciences was 323.55 (see Table 5). ChatGPT-generated feedback presented an overall evaluation for each essay, assigning a score based on a five-point scale. According to ChatGPT-generated feedback, for the 66 argumentative essays written by students in STEM fields, the average score was 2.89 out of 5 for content, 2.87 for organization, 2.25 for language and grammar, and 2.99 for mechanics. Meanwhile, for the 60 argumentative essays written by students in humanities and social sciences, the average score was 2.78 for content, 2.75 for organization, 2.01 for language and grammar, and 2.44 for mechanics.

**TABLE 5***Overall Evaluation of Argumentative Essays*

Category	STEM fields ( <i>N</i> = 66)	Humanities and social science ( <i>N</i> = 60)	Average
Word count	274.85	323.55	298.04
Content	2.89	2.78	2.83
Organization	2.87	2.75	2.81
Language & grammar	2.25	2.01	2.13
Mechanics	2.99	2.44	2.73

Research findings showed that students majoring in humanities and social sciences wrote longer essays (48.7 more words) than STEM students but had slightly lower scores in content, organization, language and grammar, and mechanics, suggesting that while they elaborated more on their ideas, their writing contained more errors and lacked clear structure. Additionally, the content and organization scores of STEM students and humanities students were similar, yet the largest difference lay in mechanics, which implied that humanities students made more typos, punctuation mistakes or errors, with inconsistent formatting. ChatGPT assigned both groups a higher mean score for content (2.83 out of 5) than for other categories, whereas language and grammar were rated the lowest (2.13 out of 5), suggesting that the content of the essays was generally well-developed, while the use of language and grammar might require improvement to enhance overall writing proficiency.

According to the interpretation of ChatGPT-generated feedback, although there were slight differences between STEM students and humanities students, recurring feedback themes revealed more common patterns across multiple essays, as outlined below:

- 1) Lack of clear thesis statements: In some essays, the thesis statement was not explicitly stated, making it difficult to identify the writer's position on the topic. As a clear thesis is considered a key element of argumentative essays, its absence can contribute to an unclear main argument from the beginning. The following is a representative example of ChatGPT-generated feedback on essays that lacked a clear thesis statement, taken from W13:  
*"The introduction is brief and does not provide enough background information or a clear thesis statement. It would be better if the introduction expanded on the issue, explaining why the question of where children should grow up is important."*
- 2) Underdeveloped arguments: In some essays, arguments were introduced but lacked further development through detailed-explanations, relevant examples, or supporting evidence. Consequently, the arguments were not fully explored, which could have impacted the overall depth and persuasiveness of the essays. Below is a representative example of feedback provided by ChatGPT on W29:  
*"The argument about school commute time is valid but not sufficiently explained. The essay mentions that children in the city have shorter commute times, but it does not explain how this affects their overall well-being or growth. It could discuss how less commuting time allows children more time for leisure or study."*
- 3) Lack of specific examples or insufficient use of evidence: While the essays presented claims, they often failed to support their arguments with concrete examples or strong evidence. In some essays, references to studies or statistics were vague and lacked full citations. Providing clear examples or evidence enhances clarity and strengthens logical reasoning. The following is a representative example of ChatGPT-generated feedback on essays that lacked specific examples or sufficient evidence, taken from W27:  
*"The arguments are not fully developed or supported with concrete examples. For instance, the claim that "children grow up in a large city can have diverse experience" is vague. It would be more compelling if specific examples were provided, such as access to cultural events, extracurricular activities, or diverse social interactions."*
- 4) Weak transitions and cohesion: Some essays had issues with smooth transitions between sentences and paragraphs. This interrupted the logical flow and cohesion of the essays and made it difficult for the reader to follow the arguments or understand them explicitly. Below is a representative example of feedback given by ChatGPT on W81:  
*"The transitions between ideas are abrupt. For example, the jump from discussing the financial advantages of rural areas to the argument that cities have better infrastructure is jarring. Transitions should be smoother to help the reader follow the logic of the essay."*

- 5) Unclear counterarguments: Whereas some essays mentioned the benefits of the city or the countryside, the opposing arguments and rebuttals to counterarguments were not fully addressed or explored in depth. These elements are essential for strengthening the persuasiveness of the writer's arguments. The following is a representative example of ChatGPT-generated feedback on essays that lacked clear thesis counterarguments, taken from W68:  
*"The essay tries to refute city life's benefits, but the counterarguments are confusing. For example, linking cybercrime to city life is not well explained."*
- 6) Grammar and syntax problems: Common issues included subject-verb agreement, article usage, awkward sentence structure, and incorrect word choices, which affected the clarity of the essays and reduced their overall quality. Below is a representative example of feedback provided by ChatGPT on W93:  
*"There are several grammatical issues, such as incorrect use of articles and awkward phrasing. For example: 'children may take peaceful and eco-friendly childhood' → 'children may have a peaceful and eco-friendly childhood' 'help them to become health and relaxed adults' → 'help them become healthy and relaxed adults' 'helps they grow up safely' → 'helps them grow up safely' Sentence structures can be more varied to avoid monotony and add sophistication to the writing."*

In addition to these writing issues, one more frequently observed feedback for humanities students was redundancy and repetition. This suggested that unnecessary repetition of words and ideas often occurred within the same essay. Even though humanities students wrote more, it was found that similar points were restated multiple times without offering new insights.

## Suggestions for the Improvement of Argumentative Essays

Regardless of students' majors, recurring writing issues were identified in argumentative essays based on the interpretation and analysis of ChatGPT-generated feedback. Accordingly, the following areas for improvement were recommended to enhance argumentative writing:

- 1) Write clearer thesis statements: An argumentative essay should have a clear and specific thesis statement that explicitly states which side the writer is for or whether the writer agrees or disagrees with the issue, directly addressing the direction of the essay. Teachers can teach students how to write a focused thesis statement that connects the opposing point of view to the writer's own argument using transitions of contrast. It is important that the writer's argument is expressed in the main clause, while the opposing point of view is placed in a subordinate structure (Oshima & Hogue, 2006), as shown in the example below.  
 Example of areas for improvement in W78 – *"Revise the introduction to immediately address the main argument and provide a clear thesis statement. For example: Although many parents move to cities for educational opportunities, I believe the country offers a better environment for children to grow up in."*
- 2) Clarify and develop arguments: After writing a clear thesis statement in the introductory paragraph, each body paragraph must focus on one main point that connects this thesis clearly, making the argument stronger and more cohesive. It is significant to fully explain each main point in detail. Therefore, teachers are expected to help students practice writing deeper and more detailed arguments instead of surface-level arguments, as shown in the following example.  
 Example of areas for improvement in W59 – *"Develop arguments more fully by explaining the positive aspects of city life (e.g., education, social experiences, career opportunities)."*
- 3) Incorporate specific examples or supporting evidence: Each argument can be supported with specific examples, experiences, facts, studies, or statistics, which makes the argument more reliable and persuasive. Thus, teachers need to teach students how to use relevant examples to back up their arguments by practicing writing two or three supporting main points for various thesis statements. Moreover, having students read an essay with support and an essay without support for comparison or providing writing models that integrate specific supporting details and reliable outside sources may be useful for learning about logical reasoning in argumentative essays.  
 Example of areas for improvement in W19 – *"Specific examples or studies to support the claims about education, healthcare, and safety would make the essay more convincing. For example, referencing statistics about healthcare outcomes in rural vs. urban areas or discussing the types of extracurricular activities available in cities would add depth."*

- 4) Enhance stronger transitions and coherence: As in the example below, strengthening transition signals such as subordinators (although, because, since), coordinators (but, and, so), and prepositions (despite, because of) gives an essay logical coherence by connecting ideas across sentences and paragraphs, which can guide the reader smoothly from one idea to the next. There should be no sudden jumps between ideas for the logical flow of the essay. Transition signals are like turning on an indicator before making a left turn while driving, as they inform the reader in advance about what is coming next, helping them follow the flow of the essay more easily. For this reason, teachers should teach different transitions to students depending on their functions when introducing an additional idea, an opposing idea, an example, or a conclusion and result, and encourage students to practice using transitions in context so that they can apply them in their writing for logical organization.  
Example of areas for improvement in W115 – *“Transitions between paragraphs are abrupt. Each body paragraph introduces a new reason, but the ideas don’t flow smoothly from one to the next. More transitional phrases would improve the flow.”*
- 5) Engage more with counterarguments: When writing an argumentative essay, instead of providing only reasons to support the writer’s claims or arguments, it is essential to discuss the other side’s reasons and rebut them, thoroughly considering both perspectives of the issue. Accordingly, teachers can guide students in understanding how to make their argumentation more persuasive by presenting both sides of the issue and effectively rebutting opposing points of view with a variety of expressions that convey contrast, which enhances credibility, as demonstrated in the following example.  
Example of areas for improvement in W72 – *“The counterargument about the countryside offering a peaceful, eco-friendly environment is only briefly acknowledged and not fully addressed. A more balanced counterargument would improve credibility.”*
- 6) Improve grammar and syntax: As part of the revision process, students should focus on correcting common grammatical errors and syntax problems, improving clarity and readability as well as avoiding awkward expressions. Furthermore, careful proofreading of an essay focusing on subject-verb agreement, articles, verb tenses, singular and plural forms, and sentence structure is crucial for its improvement. Since language and grammar are often challenging aspects for students in EFL classes, teachers may allocate more time to teaching different types of sentences (simple, compound, complex, and compound-complex sentences) and addressing common grammatical errors through repeated practice.  
Example of areas for improvement in W85 – *“Review subject-verb agreement (e.g., “institutions are” vs. “institutions is”). Proofread for clarity and rephrase awkward sentences for better readability. Pay attention to punctuation and consistency in tense to improve the essay’s overall grammar.”*

These suggestions may offer helpful pedagogical guidance for teaching argumentative essays in an effective manner and can be integrated into writing instruction both in Korean and other EFL classrooms.

## Strengths and Limitations of ChatGPT-Generated Feedback

Based on the analysis of ChatGPT-generated feedback patterns in students’ argumentative essays, the strengths and limitations of AI feedback were categorized as follows:

### *Strengths of ChatGPT-Generated Feedback*

- 1) Immediate and scalable feedback: When each argumentative essay was entered into ChatGPT with the same prompt, feedback was generated within seconds. In EFL writing classes, offering timely and constructive feedback to students is vital for developing their writing skills; however, offering detailed and individualized feedback can be time-consuming for teachers, particularly in large class settings. ChatGPT-generated feedback addresses this challenge by providing immediate and scalable responses, reducing teachers’ workload and ensuring equal feedback opportunities for all students.
- 2) Explicit corrective feedback on language and mechanics: Another strength of ChatGPT-generated feedback is that it offers specific and personalized suggestions on language (grammar, sentence structure, word choice, clarity, and academic tone) and mechanics (spelling, punctuation, capitalization, and formatting) in essays along with more detailed comments that provide guidance to enhance clarity and accuracy. As seen in Figure 3,

ChatGPT directly presented sentence-level suggestions, offered grammatically correct alternatives or rewrote whole sentences, clearly indicating what was incorrect and how it could be improved. Although this type of direct and explicit corrective feedback is not always beneficial for all learners, it can be helpful for students with lower proficiency levels who struggle to identify their errors in EFL classrooms, as it allows them to compare and better understand the corrections.

Language and Mechanics	
<b>Strengths:</b>	
<ul style="list-style-type: none"> <li>✓ The essay attempts to use comparative structures (e.g., "better place for children to grow up").</li> <li>✓ Some key ideas are communicated despite grammar issues.</li> </ul>	
<b>Areas for Improvement:</b>	
<ul style="list-style-type: none"> <li>• <b>Grammar Errors:</b> The essay contains many <b>grammatical mistakes</b> that affect readability. Examples:               <ul style="list-style-type: none"> <li>○ "Living in the country is inconvenience." → "Living in the countryside is inconvenient."</li> <li>○ "Because the country can't well build infrastructure." → "Because the countryside does not have well-developed infrastructure."</li> <li>○ "Last, many shopping mall, hospital, and museum etc. make convenience to life." → "Lastly, shopping malls, hospitals, and museums provide convenience."</li> </ul> </li> <li>• <b>Awkward or Unclear Phrasing:</b> <ul style="list-style-type: none"> <li>○ "So, children learn missing things to busy in the nature." → This sentence does not make sense. It could be rewritten as:  <b>"In the countryside, children can learn valuable lessons from nature that they might miss in a busy city."</b></li> <li>○ "Because it know oneseft." → "Because it helps children understand themselves better."</li> </ul> </li> <li>• <b>Incorrect Article and Plural Usage:</b> <ul style="list-style-type: none"> <li>○ "Many advantage." → "Many advantages."</li> <li>○ "The large city can various experiences." → "Large cities provide various experiences."</li> </ul> </li> </ul>	
<b>Suggestions:</b>	
<ul style="list-style-type: none"> <li>• <b>Proofread carefully</b> or use grammar-checking tools to correct mistakes.</li> <li>• <b>Rewrite unclear phrases</b> in a more natural way.</li> <li>• <b>Pay attention to verb agreement</b> (e.g., "the city has," not "the city have").</li> </ul>	

**FIGURE 3**

*ChatGPT-Generated Feedback on Language and Mechanics*

- 3) Standardized overall evaluation: At the end of feedback, ChatGPT provided a rubric evaluating content, organization, language and grammar, and mechanics, with scores out of five, which was presented in a visual format to provide a summary overview (see Figure 4). Even though comments were generic, lacking personalization, the standardized overall evaluation might allow both teachers and students to more easily identify the strengths and weaknesses of essays and gain a comprehensive view at a glance. Moreover, by analyzing this visual data, teachers can recognize students' weaknesses in their writing, address areas in which students have problems, and incorporate them in argumentative essay writing instruction.

Category	Score (Out of 5)	Comments
Content	★★★☆☆ (3/5)	Needs stronger rebuttals, specific evidence, and better explanation of key points.
Organization	★★★☆☆ (3/5)	Logical structure but requires smoother transitions and a clearer thesis statement.
Language & grammar	★★☆☆☆ (2/5)	Frequent grammar errors and unnatural phrasing affect readability.
Mechanics	★★☆☆☆ (2/5)	Issues with article usage, sentence structure, and word choice.

**FIGURE 4**

*ChatGPT-Generated Feedback on Overall Evaluation*

### Limitations of ChatGPT-Generated Feedback

Although there were advantages of ChatGPT-generated feedback discussed above, there were also the limitations as described below.

- 1) **Formulaic feedback on content and organization:** One of the limitations was providing template-like feedback on content (a thesis statement, arguments and evidence, counterarguments and rebuttals, and relevance and depth) and organization (an introduction, body paragraphs, a conclusion, and unity and coherence). In other words, ChatGPT offered similar and predictable comments on strengths, areas for improvement, and suggestions to the majority of essays. For example, in terms of content, ChatGPT repeatedly mentioned “Clarify the thesis in the introduction,” “Expand on key points by providing specific examples,” or “The counterarguments are mentioned but not thoroughly addressed, leading to a weak rebuttal” across multiple essays without presenting a definite correction or direct explanation. It seemed to provide general and formulaic feedback by making slight changes to the words and expressions while maintaining the same meaning. Likewise, in terms of organization, ChatGPT recurrently delivered similar feedback, such as the following: “Rewrite the introduction to clearly introduce the main points,” “Use stronger topic sentences to guide the reader,” “The conclusion is underdeveloped and does not effectively reinforce the main argument,” or “Transitions between paragraphs and ideas could be smoother.” These comments suggest that there is a limitation in generating specific and concrete feedback tailored to the content and structure of each student’s essay. As ChatGPT places a strong focus on grammar-related feedback, teachers can supplement it with more specified feedback on content, organization, and analytical writing skills such as expanding the arguments, to provide more balanced and constructive feedback to students.
- 2) **Generic and overgeneralized feedback on final suggestions for improvement:** In the final part of the feedback, with an overall evaluation, ChatGPT provided the following final suggestions for improvement (see Figure 5).

**Final Suggestions for Improvement:**

- ✓ *Revise the thesis statement to be clearer and grammatically correct.*
- ✓ *Expand the arguments by adding specific examples and explanations.*
- ✓ *Improve transitions between paragraphs for better flow.*
- ✓ *Fix grammar errors and reword awkward sentences for clarity.*
- ✓ *Strengthen the conclusion by summarizing key points effectively.*

**FIGURE 5**

*ChatGPT-Generated Feedback on Final Suggestions for Improvement*

This was nearly identical across the essays and lacked personalization, which did not provide contextually relevant feedback or actionable feedback with clear suggestions for improvement specifically. When researchers looked at only this part, it was not clear which essay the feedback was for. Since the feedback was too generic and overgeneralized, rather than being tailored to different essays, it might be difficult for students to identify which specific areas need improvement and how to rectify them on their own. Considering this, teachers need to elaborate on how to write a clear thesis statement, develop arguments with supporting evidence, and improve transitions through writing instruction or more detailed feedback.

- 3) **Lack of in-depth analysis of counterarguments and logical flow:** Another limitation of ChatGPT-generated feedback was that it lacked in-depth analysis of counterarguments and logical flow, and rarely suggested additional justification to strengthen arguments. ChatGPT consistently mentioned “counterarguments” and “logical” in almost all the essays as suggestions or areas for improvement, but it did not provide elaborated feedback, as seen below:

*Introduce a counterargument and refute it effectively (W4).*

*Develop the counterarguments more fully and explain why they do not outweigh the benefits of city life (W8).*

*Address potential counterarguments to show a more balanced understanding of the topic (W9).*

*Improve the logical flow between paragraphs for clearer connections between ideas (W12).*

*Rewrite unclear sentences to improve readability and logical flow (W55).*

*Reorganize the body paragraphs to improve the logical flow of ideas (W120).*

As demonstrated above, even though ChatGPT effectively identified missing counterarguments, it did not propose relevant counterarguments with depth and meaningful revisions based on the context of each essay. From students' perspectives, it may be challenging to understand how to incorporate counterarguments into their argumentative essays, refute them, and enhance the logical flow. Rather than providing superficial and surface-level feedback, it would have been more beneficial if ChatGPT had offered examples of counterargument sentences or explained in detail how to develop logical flow and structure. Therefore, teachers are expected to offer additional clarification or feedback on this aspect when necessary. These limitations reveal gaps that necessitate supplementary instructional support.

## CONCLUSION AND IMPLICATIONS

This study analyzed the common patterns in ChatGPT-generated feedback on 126 argumentative essays written by students from STEM fields and humanities and social sciences in Korean EFL writing classes, identified the most frequent writing issues across essays, and offered suggestions for improvement, thereby highlighting the strengths and limitations of the feedback. The present study has implications for classroom pedagogy, given that argumentative essays play an important role in writing classes by helping students enhance their critical thinking and logical reasoning skills (Sharadgah et al., 2019), and ChatGPT has the potential to be a useful tool to support teachers in providing feedback and to assist students throughout the writing process, particularly in revising and improving their drafts based on specific suggestions.

For the first research question, when ChatGPT offered an overall evaluation with scores based on a five-point scale in terms of content, organization, language and grammar, and mechanics, it was found that ChatGPT assigned the highest score to content and the lowest to language and grammar in students' argumentative essays. This suggests that students demonstrated greater difficulties in linguistic aspects of their writing, especially in relation to grammar, sentence structure, word choice, and clarity. The relatively low scores in this area, along with recurrent comments in the feedback, underscored the pedagogical need to emphasize instruction on key areas such as subject-verb agreement, appropriate article usage, varied sentence structure, and consistent verb tense. In addition, based on the analysis of recurring feedback themes, common writing problems were identified by ChatGPT: a lack of clear thesis statements, underdeveloped arguments, insufficient specific examples or evidence, weak transitions and cohesion, unclear counterarguments, and grammar and syntax problems. These findings show that many essays lacked a well-defined thesis statement, failed to develop a well-supported argument, and did not adequately address opposing viewpoints, which are essential elements of effective argumentative writing. These issues align with the difficulties faced by Iranian EFL learners when writing argumentative essays in previous research: limited language skills, challenges in applying Western argument and persuasive techniques, difficulties in formulating a clear thesis, addressing counterarguments in an effective manner, and ensuring coherence in an essay, limited exposure to argumentative writing in their native language, and insufficient practice in English (Mallahi, 2024).

For the second research question on suggestions for improving students' argumentative essays, teachers need to encourage students to write a clear and specific thesis statement that directly guides the direction of an essay and to develop their arguments in detail. It is also important to incorporate specific supporting details to strengthen the argument and to use clearer transitions to ensure coherence in an essay. Finally, students are required to engage more with counterarguments and rebut them in order to demonstrate consideration of both sides of the issue, and to improve grammar and syntax issues through the revision process. These recommendations can serve as valuable instructional insights for teaching argumentative essays effectively in EFL writing classes.

For the third research question on the strengths and limitations of ChatGPT-generated feedback, the analysis of feedback patterns shows that ChatGPT can provide immediate and scalable feedback, which may reduce the time and workload of teachers, especially when they are burdened by the abundance of students' errors and feel overwhelmed by the task of giving comprehensive feedback to numerous students (Lee, 2017) or when they are not available to offer extensive feedback in a timely manner. Furthermore, ChatGPT presented tailored and individualized feedback on grammar and mechanics for each essay, with grammatically correct alternatives, although some of them were metalinguistic. This explicit corrective feedback on language errors can be particularly beneficial for non-native English speaking teachers without requiring considerable time and effort as well as for students with lower proficiency levels. As another strength, at the end of feedback, ChatGPT provided a standardized overall evaluation through a rubric covering content, organization, language and grammar, and mechanics, which was presented in a visual format to offer a comprehensive view at a glance. On the other hand, there were also the limitations to ChatGPT-generated feedback. ChatGPT gave formulaic responses on content and organization,

delivering template-like comments. It also provided generic and overgeneralized feedback in its final suggestions for improvement, lacking deeper engagement with the content and the context of each essay, which made the suggestions vague and failed to offer strong alternatives. Another limitation was that ChatGPT-generated feedback lacked in-depth analysis of counterarguments and logical structure.

Considering the strengths and limitations of ChatGPT-generated feedback, it is necessary to explore how teachers and students can use ChatGPT effectively to improve writing skills, integrating the research findings into recommendations for its implementation in writing classes. The results of this study showed that ChatGPT seemed to provide comprehensive feedback based on the prompt, but not balanced feedback; it offered more specific and detailed feedback on language and mechanics, whereas it gave mechanical feedback on content and organization of argumentative essays. Accordingly, this suggests implementing a complementary approach that combines teacher feedback or peer feedback with ChatGPT-generated feedback, by taking advantage of only the strengths of each type of feedback. Even though ChatGPT is a useful tool in that it can provide prompt and broadly applicable feedback, taking its weaknesses into account, it cannot replace teacher feedback as a comprehensive writing evaluator or a primary feedback provider. It can be utilized as an initial feedback tool such as identifying grammatical errors before in-depth revision stages, but teachers need to fill in the gaps by providing additional explanations or guidance on content and organization as well as counterarguments and logical reasoning, which can result in a synergistic effect. In addition, ChatGPT cannot engage in meaningful and evolving discussions like teachers and may provide incorrect information. In such cases, teachers are able to offer additional personalized formative feedback, enabling continuous revision and progress through the stages of process writing. For students, it is important to develop a critical perspective to determine which information from ChatGPT is accurate and which feedback is valuable and relevant without becoming overly reliant on the tool.

This study contributes to the relatively understudied area of research on the analysis of ChatGPT-generated feedback on argumentative essays and sheds light on its strengths, limitations, and potential for effective integration into EFL writing instruction. However, the research findings should be interpreted with caution because the study is limited to ChatGPT-generated feedback on a single type of writing prompt and one essay type. Such constraints may affect the consistency and reliability of the output, especially given ChatGPT's tendency to generate slightly different responses depending on timing and phrasing of the input. Additionally, the study did not include expert or instructor evaluations to cross-check the accuracy or appropriateness of the feedback, making it difficult to determine the degree to which ChatGPT's feedback can be trusted as instructional guidance. To extend this line of research, future studies could explore how ChatGPT's feedback patterns differ across different essay types, examine inter-rater agreement between ChatGPT and human raters across proficiency levels, and investigate how its feedback evolves when the same text is resubmitted at different intervals. Further research is also needed to understand students' perceptions and attitudes toward ChatGPT-assisted teacher feedback or ChatGPT-assisted peer feedback reflecting the dynamic nature of AI.

## References

- Algaraady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journals, Special Issue on CALL 9*, 3–17.
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*(2), e35179. <https://doi.org/10.7759/cureus.35179>
- Ataş, A. H., Cengiz, B. C., & Çelik, B. (2024). Evaluating the ChatGPT in generating feedback on content and organization components of EFL compare and contrast essays. *Sakarya University Journal of Education, 14*(3), 515–538.
- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachslar, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(23), 1–15.
- Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Multilingual Matters.
- Brown, H. D. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson.
- Carlson, M., Pack, A., & Escalante, J. (2023). Utilizing OpenAI's GPT-4 for written feedback. *TESOL Journal, 15*(2), e759. <https://doi.org/10.1002/tesj.759>
- Chen, Y. (2024). Is ChatGPT any good: Evaluating the potential of ChatGPT-reformulated essays as written feedback in L2 writing. *SSRN*. <https://doi.org/10.2139/ssrn.4963755>
- EIlebyary, K., & Shabara, R. (2024). ChatGPT-generated corrective feedback: Does it what it says on the tin? *Teaching English with Technology, 24*(3), 68–89.
- Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*, 8435–8463.
- Helms-Park, R., & Stapleton, P. (2003). Questioning the importance of individualized voice in undergraduate L2 argumentative writing:

- An empirical study with pedagogical implications. *Journal of Second Language Writing*, 12(3), 245–265.
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays. *arXiv*. <https://doi.org/10.48550/arXiv.2304.14276>
- Hirvela, A. (2017). Argumentation & second language writing: Are we missing the boat? *Journal of Second Language Writing*, 100(36), 69–74.
- Hyland, K., & Hyland, F. (2019). Contexts and issues in feedback on L2 writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 1–22). Cambridge University Press.
- Kim, Rakhun. (2025). Generational gap in accepting AI integration in Korean EFL classrooms: Comparing pre-service and in-service teachers within technology acceptance model. *Modern English Education*, 26, 113–129.
- Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology*, 3(1), 55–72.
- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23(24), 1–17.
- Kushmar, V., Vornachev, O., Korobova, O., & Kaida, O. (2022). Artificial Intelligence in language learning: What are we afraid of? *Arab World English Journal, Special Issue on CALL 8*, 262–273.
- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer.
- Lent, N., Harrison, T., & Rolle, S. (2022, October 13). Assessment and feedback as an active dialogue between tutors and students. *THE Campus: Learn, Share, Connect*. <https://www.timeshighereducation.com/campus/assessment-and-feedback-active-dialogue-between-tutors-and-students>
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4), 102720.
- Mallahi, O. (2024). Exploring the status of argumentative essay writing strategies and problems of Iranian EFL learners. *Asian-Pacific Journal of Second and Foreign Language Education*, 9(19), 1–26.
- Mitchell, S., & Riddle, A. (2000). *Learning to argue in higher education*. Heinemann.
- Noroozi, O., Biemans, H., & Mulder, M. (2016). Relations between scripted online peer feedback processes and quality of written argumentative essay. *The Internet and Higher Education*, 31, 20–31.
- Oshima, A., & Hogue, A. (2006). *Writing academic English 4* (4th ed.). Pearson.
- Oshima, A., & Hogue, A. (2020). *Longman academic writing series 4* (5th ed.). Pearson.
- Park, Pu-Nahm, & Kim, Tai-Won. (2024). The effects of ChatGPT-assisted teacher feedback on Korean college students' English writing. *The Journal of Yeolin Education*, 32(6), 43–65.
- Purdue Online Writing Lab. (2025, February 14). *Toulmin argument*. [https://owl.purdue.edu/owl/general\\_writing/academic\\_writing/historical\\_perspectives\\_on\\_argumentation/toulmin\\_argument.html](https://owl.purdue.edu/owl/general_writing/academic_writing/historical_perspectives_on_argumentation/toulmin_argument.html)
- Ramsden, P. (2003). *Learning to teach in higher education*. Routledge.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 1–22.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.
- Sharadgah, T. A., Sa'di, R. A., & Ahmad, H. H. (2019). Promoting and assessing EFL college students' critical thinking skills through argumentative essay writing. *Arab World English Journal*, 10(4), 133–150.
- Solak, E. (2024). Examining writing feedback dynamics from ChatGPT AI and human educators: A comparative study. *Pedagogika-Pedagogy*, 96(7), 955–969.
- Sousa, F., Alves, T., Gama, S., Jorge, J., & Gonçalves, D. (2024). Impact of social relationships on peer assessment in e-learning. *arXiv*. <https://arxiv.org/abs/2407.01077>
- Stevenson, M., & Phakiti, A. (2024). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65.
- Thompson, V. E. (2021). Integrating global Englishes into literature and writing units: Advice for secondary teachers. In M. D. Devereaux & C. C. Palmer (Eds.), *Teaching English language variation in the global classroom* (pp. 82–91). Routledge.
- Toulmin, S. E. (2003). *The uses of argument* (2nd ed.). Cambridge University Press.
- Valero Haro, A., Noroozi, O., Biemans, H. J., Mulder, M., & Banihashem, S. K. (2023). How does the type of online peer feedback influence feedback quality, argumentative essay writing quality, and domain-specific learning? *Interactive Learning Environments*, 32(9), 5459–5478.
- Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, 50, 100568.
- Yoon, Su-Youn, Miszoglád, E., & Pierce, L. R. (2023). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion. *arXiv*. <https://doi.org/10.48550/arXiv.2310.06505>
- Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, 61(4), 875–900.
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN*. <https://doi.org/10.2139/ssrn.4312418>