Patterns of ChatGPT Engagement and Argumentative Writing Development

Kwak, Subeom

Yonsei University

Graduate School of Education

Assistant Professor

I. Introduction

Students' ways of thinking, reading, and writing are rapidly changing in the classroom thanks to artificial intelligence agents such as ChatGPT. This is particularly noticeable in university writing courses, as generative AI can provide instant feedback, offer sample texts, suggest alternatives for structuring ideas, and even generate entire sections of writing (Wang, 2025).

More specifically, these tools raise key teaching and ethical questions. How do students use AI for their writing? Does AI use help them strengthen argumentative writing skills, or does it get in the way (Levine, Beck, Mah, Phalen, & Pittman, 2024; Wang, 2025)? Are some patterns of AI use helpful for learning to argue well, while others are harmful? How can instructors design writing courses that take advantage of AI's strengths while reducing risks such as over-reliance or shallow, surface-level changes?

These questions are important because academic writing—especially in first-year composition courses—is a turning point where students are expected to learn more complex structures of argument and more advanced rhetorical strategies (Fontenelle-Tereshchuk, 2024). Argumentative writing is a core academic skill. It involves stating

clear claims, supporting them with relevant evidence, explaining how that evidence connects to the claims, and responding thoughtfully to counterarguments. Classic work on argumentation (e.g., Toulmin, 1969/2003) shows that strong arguments depend not only on having content, but also on organizing ideas well, making warrants clear, and taking opposing views seriously.

In many first-year composition courses, however, students find these skills difficult. Their essays often show weak or insufficient evidence, unclear or hidden warrants, and little engagement with counterarguments. Consequently, their writing can be less convincing and coherent than instructors hope. AI tools like ChatGPT might help support these processes, but they can also tempt students to take shortcuts—for example, letting the system generate ideas or rewrite whole sections for them—if they are not taught how to use AI in a thoughtful and strategic way (Tseng & Warschauer, 2023).

Recent studies suggest that what matters is not simply whether students use AI, but how they use it (Steiss et al., 2024). Some students turn to ChatGPT mainly for low-level editing, like fixing grammar or polishing word choice. Others use AI to expand their ideas, reorganize their arguments, or brainstorm new directions. In those cases, they are using AI to support their thinking rather than to replace it (Tate, Harnick-Shapiro, Ritchie, Tseng, Dennin, & Warschauer, 2025).

For example, research shows that students often rely on ChatGPT during planning and drafting to generate ideas, then adapt those suggestions to build their own arguments. They also use AI when revising, treating it somewhat like a handy editing tool (Levine et al., 2024). When used in this more deliberate way, AI can support many parts of the writing process: building an outline, developing content, revising, proofreading, and even reflecting after the draft is done (Vetter, Lucia, Jiang, & Othman, 2024). This may be especially helpful for demanding tasks such as argumentative essays (Steiss et al., 2024).

At the same time, concerns regarding overreliance on AI persist.

If students accept AI-generated text without questioning it, their ability to evaluate information critically and use higher-order thinking skills may weaken, especially when it comes to judging the quality, accuracy, or bias of AI output (Bašić, Banovac, Kružić, & Jerković, 2023). On the other hand, AI can also be used to promote higher-order thinking—such as critical thinking and problem solving—by exposing students to different perspectives and giving them chances to refine their arguments through repeated feedback (Bedington, Halcomb, McKee, Sargent, & Smith, 2024).

Even when they use AI, many students say they are cautious about simply copying AI-generated ideas. They understand that they still need to develop their own critical thinking and writing skills (Wang, 2025). This suggests that instruction around AI must be nuanced: students need to be actively taught how to question, adapt, and build on AI output, rather than treat it as something to accept passively (Deep & Chen, 2025; Ding, Lawson, & Shapira, 2025).

This implies the need for a clear framework that distinguishes between critical use of AI—where AI supports and extends students' own thinking—and naïve dependence, where AI replaces that thinking (Elycheikh, Svetlana, & Magda, 2025). Making this distinction can help instructors design activities and guidelines that encourage students to engage critically with AI-generated text.

The present study examines how first-year university students use ChatGPT during their argumentative writing and how different patterns of use relate to the development of their argumentation skills. I aim to find out which types of AI use are linked to stronger argumentative writing and which types may slow down or block the development of important writing abilities (Wang, 2025). Understanding these patterns is essential for designing teaching strategies that make the most of AI while reducing its disadvantages (Deep & Chen, 2025).

A key part of this work is to examine how students interact with ChatGPT and how well they perform on writing tasks (Kim, Lee, Detrick, Wang, & Li, 2025). More specifically, do students use ChatGPT

in ways that support critical questioning and reflection, or do they mainly use it to offload thinking, leading to superficial analysis (Nasr, Tu, Werner, Bauer, Yen, & Sujo–Montes, 2025)?

Answering these questions can help schools design learning environments that support effective human–AI collaborations. The goal is for students to build both strong critical thinking skills and strong AI literacy, so they can use these tools in responsible, academically meaningful ways (Jin, Cisterna, Owens, & Riordan, 2025). This is especially important because there is still insufficient empirical research that systematically tracks how students integrate AI into their writing processes. Much of this activity is invisible to instructors (Wang, 2025).

To address this gap, this study uses a mixed-methods design. I analyze process data from students' chatbot interactions, gather qualitative reflections from students on how they used ChatGPT, and assess the quality of their argumentative writing quantitatively. This combination allows us to see not only what students write, but also how they work with AI while writing.

Through this analysis, we explore the specific prompting strategies and revision cycles students use when interacting with ChatGPT, and we examine which of these strategies seem to help them produce more complex and coherent arguments. We also examine the types of help students seek from AI and what their prompts reveal about their communication style, in order to understand more fully how AI can support academic writing in higher education (Usher & Amzalag, 2025).

This study seeks to provide a more detailed and realistic picture of AI's role in teaching and learning writing. Rather than treating AI only as a threat or as a helpful tool, I show how human thinking and AI assistance interact in complex ways when students learn advanced academic writing (Hutson, Plate, & Berry, 2024; Kim et al., 2025).

This study seeks to promote both theoretical viewpoints and real-world practices in literacy instruction, particularly in the field of writing. This study contributes to the growing body of research on AI-

supported writing education. This study examines university students' use of ChatGPT and distinguishes between practices that enhance learning and those that do not considerably aid students' advancement. This study examines argumentative writing by evaluating its separate components using analytic metrics rather than relying solely on overall ratings. The findings provide research-based advice for instructors seeking to use generative AI in first-year writing classrooms in ways that will benefit students' long-term argumentative writing development.

II. Literature Review

1. Generative AI and the Evolution of Writing Education

Writing instruction has consistently adapted to major technological transformations throughout history. The introduction of typewriters fundamentally altered the writing process, word processors redistributed cognitive demands across planning and revision phases, and the Internet expanded both the scope and complexity of audiences and knowledge sources available to writers (Graham, McKeown, Kiuhara, & Harris, 2012). Each technological shift prompted educational institutions to reconsider how writing instruction should evolve to support learners in leveraging new tools effectively (Karchmer, 2007; Rodriguez & Rodriguez, 1986). Large language models including Chat-GPT represent perhaps the most significant pedagogical challenge to date, as these generative AI systems can produce written text autonomously, raising fundamental questions about the nature of writing as a learning activity (Godwin-Jones, 2021; Ranalli, 2018). The primary challenge for educators is to explore whether and in what ways AI integration can enhance writing skills, such as argumentative reasoning, rhetorical awareness, and critical thinking, rather than merely fa-

cilitating basic text creation or replacing crucial cognitive processes that contribute to meaningful writing development (Rudolph, Tan, & Tan, 2023).

Argumentation is widely used as a major evaluative practice in education. It is also taught intensively in secondary schools and higher education. This prominence underscores the essential function of argumentation in academic discourse (McNeil & Krajcik, 2009). Argumentative writing involves a complex reasoning process that necessitates the simultaneous integration of multiple cognitive and communicative skills (Duschl & Osborne, 2002). Effective argumentative writing requires structural fluency—the capacity to organize claims, evidence, and reasoning into coherent and logically sound arguments while acknowledging and addressing counterarguments (Van Eemeren, Henkemans, & Grootendorst, 2002). Argumentative writing also requires high levels of linguistic control. This is reflected in accurate word choice, appropriate management of tone, and stylistic decisions that strengthen persuasion without obscuring meaning. Argumentation requires dialogic fluency, which involves the ability to recognize social contexts, appreciate diverse perspectives, and craft appeals that resonate with specific audiences while addressing their concerns (Aull & Ross, 2020). These components operate in constant interaction throughout the writing process. Consequently, learners building argumentative skills need to manage structure, linguistic choices, and rhetorical moves simultaneously.

The prospect of integrating ChatGPT into argumentative writing instruction has generated substantial disagreement among researchers and teachers. Some scholars and educators advocate for thoughtful adoption, highlighting ChatGPT's potential as an auxiliary tool that could provide feedback on both surface-level language issues and deeper argumentative structures, or function as a virtual learning partner through iterative dialogue (Fryer et al., 2019; Fuchs, 2023; Kasneci et al., 2023). Others, however, urge caution, arguing that integration may be premature without deeper understanding of the system's sig-

nificant limitations (Godwin-Jones, 2021). Chief among these limitations is ChatGPT's insufficient consideration of social context—the variable cultural norms, audience expectations, historical circumstances, and power dynamics that fundamentally shape how arguments are constructed and received. Owing to its architecture based on pattern recognition within training data, generative AI lacks a genuine understanding of how social context varies across people, regions, cultures, and time periods, yet such context-sensitivity is essential to argumentation (Godwin-Jones, 2021). Moreover, ongoing concerns about the quality of logical reasoning, hallucinations, factual inaccuracies, and the system's propensity to produce content that sounds plausible yet is misleading, raise questions about its reliability as a writing support tool (Su, Lin, & Lai, 2023).

2. Conceptualizing Argumentative Writing

The argumentative process is intrinsically dialogic, requiring authors to rebuild knowledge while actively considering social settings and opposing ideas (Duschl & Osborne, 2002; Van Eemeren et al., 2002). This emphasizes the importance of argumentative writing, as it requires the integration of social intelligence, communicative accuracy, and cognitive thinking (Aull & Ross, 2020). Any auxiliary tool that significantly improves argumentative competence must function on several levels at once: provide insightful feedback on the logical coherence and organization of arguments, help writers improve their language for accuracy and persuasive impact, and critically raise awareness of rhetorical contexts, audience considerations, and the viewpoints of those who might disagree with or challenge the writer's claims.

Generative AI systems lack a genuine understanding of how cultural norms, audience expectations, historical particulars, and power dynamics vary across communities and contexts, and cannot meaningfully guide writers in developing the contextual sensitivity that

effective persuasion demands (Godwin-Jones, 2021). This limitation suggests that while ChatGPT might serve a supplementary role in developing structural and linguistic competencies, educators seeking to develop students' full argumentative sophistication—particularly the dialogic fluency essential to authentic persuasion—must carefully design instruction to compensate for the system's fundamental blindness to social and rhetorical context (Zhang & Li, 2021; Thorp, 2023).

3. ChatGPT as an Auxiliary Tool

Recent studies have shown that students' responses to ChatGPT are not binary. They neither accept nor reject it. Instead, they employ it in various ways that defy simple categorization (Kwak, 2024). Rather than wholesale adoption or avoidance, research on university students shows that they selectively absorb ChatGPT comments during revision processes, exercising critical judgment regarding which suggestions to implement and which to ignore (Kwak, 2024). This evidence shows conditional utility rather than transformative impact, implying that ChatGPT works best when integrated into intentional instructional contexts rather than as a stand-alone writing resource (Fryer, Nakao, & Thompson, 2019; Kasneci et al., 2023). Process-oriented writing instruction frameworks hold particular potential for establishing ChatGPT as a mediating tool during the drafting stages of composition, where exploratory discussion and idea production are key cognitive demands (Kwak, 2024).

4. Research gaps and future directions in AI-supported argumentative writing

A considerable research gap still exists between general awareness of ChatGPT's technological capabilities and the specific uses of generative AI in writing processes (Kwak, 2024). While previous research has examined the quality of AI feedback and documented

broad usage patterns, few studies have delved deeply into the decision-making processes that students use during composition and the actual shifts in argumentation quality that result from AI engagement. This discrepancy is especially significant because generic use patterns conceal crucial compositional behaviors, such as whether the work becomes superficially polished while losing argumentative depth. To understand this, we must extensively examine the decisions students make during the writing process rather than simply judging the finished work after it is completed.

III. Research Methods

1. Participants

Participants were 124 first-year undergraduate students enrolled in a required academic writing course at K University, a private university located in Seoul, South Korea. At K University, the first-year academic writing course is a mandatory component of the general education curriculum for all freshmen and is offered over a 15-week semester. All students in this study were taking a 15-week course at the time of data collection.

Only students who provided informed consent to participate in the study were included in the dataset. Thus, the final sample for analysis consisted of 124 consented students enrolled in the freshman writing course.

The research was conducted under the ethical regulations approved by the institution. The principal investigator completed the university's Institutional Review Board (IRB) research ethics training and obtained IRB approval prior to data collection. At the start of the course, students received an explanation of the study's purpose and procedures and were told that taking part was entirely optional and

that they were free to discontinue their involvement at any point without affecting their academic progress. Students who chose to take part provided their written consent, and only data from these consenting students were used for research purposes. All identifying information was removed or anonymized prior to the analysis.

2. Course context and instructional procedure

The research was embedded within the existing first-year composition curriculum at K University. This course runs for 15 weeks and is designed to introduce first-year students to the core concepts and practices of academic writing, with particular emphasis on argumentative writing.

As part of the regular instruction, all participating students were taught the basic principles of argumentation based on Toulmin's model of argument (Toulmin, 1969/2003). The instructions covered key elements of an argument (e.g., claim, evidence, warrant, counterargument, rebuttal) and how these elements function together in written argumentative texts.

During this course, the students completed a staged argumentative writing task. All research participants were required to complete the following:

1. Develop and submit an argumentative outline, specifying their main claim, supporting reasons, and anticipated counterarguments.

2. Write and submit an initial draft of an argumentative essay based on the outline.

3. Revise and submit a revised draft of the same argumentative essay, incorporating instructor feedback and, where applicable, AI-supported drafting and revision processes.

These tasks were part of the normal course requirements for all the students in the class.

Writing Task

Can the expansion of online learning replace traditional classroom education?

With the recent rapid spread of online learning, the debate over the necessity of traditional classroom-based education has intensified. Online learning offers significant advantages in terms of accessibility and convenience, and some have argued that it can completely replace traditional educational methods. On the other hand, others argue that online learning, lacking face-to-face interaction and emotional exchange, cannot match the quality of traditional education, emphasizing the importance of traditional schooling and its benefits.

Write an essay logically presenting your opinion on whether online learning can replace traditional classroom education.

* The task was originally written in Korean and translated by the author.

This study gathered three categories of data to analyze ChatGPT use and the development of argumentative writing. One category comprised writing produced in class, including outlines in early and revised forms and the first and final versions of argumentative essays. The second category consisted of screenshots or URLs that recorded students' interactions with ChatGPT.

Qualitative coding was performed on the collected ChatGPT conversation logs, dividing them into six categories: Surface Revision (SR), Content Elaboration (CE), Organization/Coherence (OC), Meta-cognitive Use (ME), Dependence (DE), and Collaborative Refinement (CR).

ChatGPT Log Codebook

1. Surface Revision (SR)

Definition: Requests focused on improving the linguistic surface of the text, such as grammar, vocabulary choice, sentence structure, and fluency issues.

Focus: Fixing errors and polishing the style.

Exclusion: Does not include requests to change the logical structure or add new arguments.

2. Content Elaboration (CE)

Definition: Requests to add to or strengthen the substance of the argument. This includes asking for evidence, examples, data, or counterarguments to support the claims.

Focus: Adding key details and evidence to support the claims.

Exclusion: Distinct from simply polishing the tone (SR) or reordering the paragraphs (OC).

3. Organization / Coherence (OC)

Definition: Requests to improve the logical flow, structure, and unity of the content at the paragraph or essay level.

Focus: Rearranging ideas, refining transitions, and double-checking logical connections.

Exclusion: Does not cover the correction of single sentences (SR) or the development of new ideas(CE).

4. Metacognitive Use (ME)

Definition: High-level requests in which students ask ChatGPT to examine, diagnose, or critique their writing.

Focus: Self-regulation and identifying weaknesses rather than direct editing.

Exclusion: Asking ChatGPT to "fix it" without explanation falls under other categories.

5. Dependence (DE)

Definition: Instances in which the student relies on ChatGPT to write or completely rewrite the text, replacing their own cognitive effort.

Focus: Passive acceptance of AI generation.

Exclusion: If the student selectively uses parts of the AI's output while maintaining control, it is not considered dependence.

6. Collaborative Refinement (CR)
Definition: A sophisticated, student-led partnership in which students edit their own work, ask for feedback, and critically select specific AI suggestions.
Focus: A back-and-forth process of writing, obtaining AI advice, and selecting what to use.
Exclusion: Mere copy-pasting or asking for simple grammar fixes does not constitute high-level collaboration.

3. Statistical analyses

All statistical analyses were conducted to examine how different patterns of ChatGPT use related to students' learning gains in argumentative writing. Analyses were performed using Python 3.12.7, with a significance level set at $\alpha = .05$.

First, descriptive statistics were computed for all focal variables, including pre- and post-test scores for overall argumentative writing performance and for key components (evidence and counterargument), as well as the six ChatGPT-use subscales: Surface Revision (SR), Content Elaboration (CE), Organization/Coherence (OC), Metacognitive Use (ME), Dependence (DE), and Collaborative Refinement (CR). The four subscales associated with higher-order and constructive uses of ChatGPT (CE, OC, ME, CR) were combined to form a Positive_Use_Index, and the two subscales reflecting more surface-level and potentially problematic uses (SR, DE) were combined to form a Risk_Use_Index. A Balance_Index was then computed by subtracting the Risk_Use_Index from Positive_Use_Index, such that higher scores indicated that constructive uses of ChatGPT outweighed surface-level and dependency-oriented uses.

To explore the broad patterns of ChatGPT use, a cluster analysis

was conducted on the six ChatGPT-use subscales (SR, CE, OC, ME, DE, and CR). Standardized scores for these subscales were used as input variables. The resulting clusters were interpreted as qualitatively distinct profiles of ChatGPT use (e.g., relatively positive, mixed, low use). These clusters were then compared on learning gains in argumentative writing, operationalized as gain scores (post–pre) for overall writing quality (Gain_Total) and for specific components such as Evidence (Gain_Evidence) and Counterargument (Gain_Counterargument).

Next, to focus more specifically on the role of constructive ChatGPT use, the Positive_Use_Index was used to form three groups based on tertile splits (low, mid, and high). A one-way ANOVA was conducted with the Positive_Use_Index group (low/mid/high) as the between-subjects factor and Gain_Evidence (post–pre Evidence score) as the dependent variable. This analysis tested whether students who made more frequent and constructive use of ChatGPT for content elaboration, organization, metacognitive engagement, and collaborative refinement showed greater gains in their use of evidence. The omnibus F-test and corresponding effect size were used to evaluate group differences, and post hoc comparisons were used to interpret the significant effects.

Finally, to examine the unique contribution of positive ChatGPT use to gains in addressing counterarguments while controlling for initial ability, multiple regression analysis was conducted. Gain_Counterargument (post–pre counterargument scores) served as the dependent variable. The predictors were Pre_Counterargument, Positive_Use_Index, and their interaction terms (Pre_Counterargument × Positive_Use_Index). The model fit (R^2), significance and direction of each regression coefficient, and interaction term were examined.

The analyses resulted in a unified group of assessment tools. These tools linked broad patterns of ChatGPT use to both helpful and potentially risky practices. They also included detailed indicators of effective use, such as how students worked with the evidence and counterarguments.

IV. Results and Discussion

Students' use of ChatGPT for argumentative writing was assessed using six subscales that captured qualitatively distinct functionalities of AI help. To summarize these patterns, a Positive_Use_Index was created by combining four subscales related to the higher-order and constructive applications of ChatGPT: CE, OC, ME, and CR. The Risk_Use_Index was created by combining two subscales that reflect superficial surface-level and possibly hazardous use: SR and DE. The Balance_Index was calculated by subtracting the Risk_Use_Index from the Positive_Use_Index, with higher values indicating that students' constructive, higher-order uses of ChatGPT exceeded their surface-level and dependency-oriented usage.[1]

1. Descriptive statistics and preliminary analyses

Table 1. Descriptive statistics for pre-test, post-test, and gain scores

Scale	n	Pre-test M (SD)	Post-test M (SD)	Gain M (SD)
Total	89	3.84(2.12)	9.20(3.42)	5.36(3.23)
Evidence	120	1.14(0.49)	2.47(0.59)	1.33(0.76)
Counterargument	120	0.37(0.61)	1.73(1.08)	1.36(1.17)

Note. Pre-test and post-test scores are based on analytic rubric ratings for argumentative writing. Gain scores were computed as post–pre.

---

1   The grading criteria and survey for argumentative writing were based on the Argumentative Writing Education Project (Newell, Bloome, & Hirvela, 2015). Considering that they emphasize Toulmin's argumentation model, which is also heavily relied upon in Korean argumentative writing education; that they are designed to measure only 'argumentation' elements, resulting in a relatively simpler form compared to comprehensive writing scoring rubrics; and that I personally participated as a co-researcher in the argumentative writing project in the United States, directly experiencing the rater training process and actual scoring, these criteria were applied to the present study.

Due to missing pre-test scores for the total scale, analyses involving pre–post changes in total scores were based on 89 students with complete data for both time points. In contrast, pre- and post-test scores for the evidence and counterargument components were available for 120 participants.
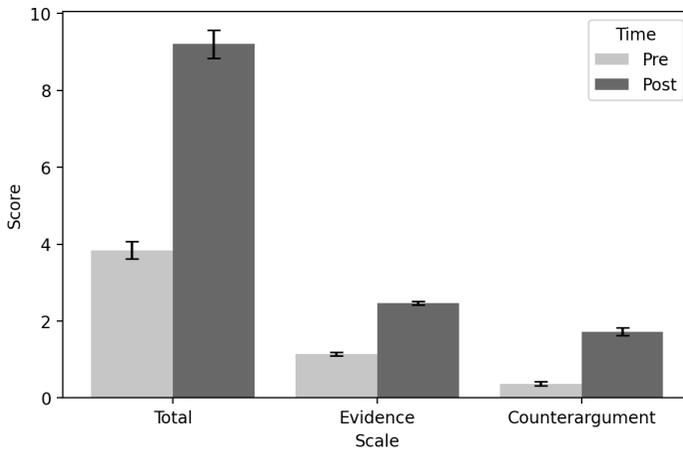

Figure 1. Pre-Post test results

The students showed a noticeable but limited improvement in their argumentative writing skills. Most of their pre-test scores fell within the lower to mid-range, suggesting considerable potential for further development. Overall, post-test scores were higher. The score differences between the post- and pre-tests were mostly positive. The same trend was observed for the major components of argumentative writing. Students showed notable mean increases in their evidence and counterargument skills. By the end of the semester, they were able to provide stronger evidence and address opposing views more effectively.

Students' use of ChatGPT varied in terms of frequency and quality. Scores on the positive-use subscales—Content Elaboration (CE), Organization and Coherence (OC), Meta-cognitive Use (ME), and Col-

laborative Refinement (CR)—ranged from very low to relatively high, indicating that some students used ChatGPT only sparingly for higher-order writing support, while others used it extensively to generate ideas, structure arguments, and revise their reasoning. The average scores on this index were in the low-to-moderate range. However, the scores varied widely, indicating substantial individual differences in constructive AI-supported writing practices.

In contrast, the risk-oriented subscales, Surface Revision(SR) and Dependence (DE), tended to be somewhat lower on average, although again with notable variability. Some students appeared to use ChatGPT primarily for quick surface-level fixes or direct copying, whereas others reported a more strategic use of ChatGPT. Descriptively, most students had slightly higher positive than risky use scores.

As a whole, the students improved their argumentative writing outcomes. Improvements were also found in sub factors of argumentation such as evidence and counterarguments. Yet, the pre- and post-test data demonstrated marked differences across individual learners. This variation provides a solid platform for further research in this field. This enables academics to investigate how different patterns of ChatGPT use are related to growth in critical parts of argumentative writing.

2. Cluster-based comparisons of learning gains

To explore naturally the occurring patterns of ChatGPT use, a k-means clustering analysis was conducted using the six subscale scores ('SR', 'CE', 'OC', 'ME', 'DE', and 'CR'). A three-cluster solution was chosen to represent roughly "low", "moderate", and "high" levels of ChatGPT use and sophistication, based on the relative profiles of the clusters.

I then compared clusters on pre-test total scores ('Pre_Total') and total gain scores ('Gain_Total') using one-way ANOVAs. Although there were small numerical differences in means across clusters, nei-

ther the difference in 'Pre_Total' nor in 'Gain_Total' reached statistical significance (both $p > .05$). In other words, students' overall writing gains did not differ reliably as a function of these broad usage patterns.

This suggests that simply belonging to a certain cluster of Chat-GPT usage (e.g., more vs. less use overall) does not, by itself, strongly differentiate students' overall writing improvement. These findings motivated a shift from unsupervised clustering toward more theory-driven groupings based on the quality and balance of ChatGPT use.

3. Balance of positive vs. risky ChatGPT use: extreme-group comparison

I focused on the Balance_Index—Positive_Use_Index-Risk_Use_Index—which captures the extent to which students' use of ChatGPT is dominated by constructive, higher-order strategies versus more superficial or potentially problematic ones.

To amplify contrasts, I selected the bottom 25% and top 25% of students on the Balance_Index as "Low_Balance" and "High_Balance" groups, respectively. These two extreme groups were compared on: total gain score ('Gain_Total') and evidence gain score ('Gain_Evidence').

Table 2. Descriptive statistics for gain scores (post–pre) by balance group

Scale	Group	n	Gain M (SD)
Total	Low_Balance	28	5.21(3.17)
Total	High_Balance	8	6.25(3.92)
Evidence	Low_Balance	30	1.20(0.89)
Evidence	High_Balance	23	1.52(0.79)

Note. Gain scores were computed as post-test minus pre-test. High_Balance and Low_Balance groups were defined using the upper and lower quartiles of the Balance_Index.

Welch's t-tests showed that the High_Balance group had higher mean gains in both total score and evidence than the Low_Balance group. However, these differences were not statistically significant ($p > .05$). In other words, there was a consistent trend in the expected direction—students whose positive use of ChatGPT outweighed risky use tended to improve more—but the magnitude of the difference, given the sample sizes (n ≈ 22–29 per group) and variability, was insufficient to reach conventional significance.
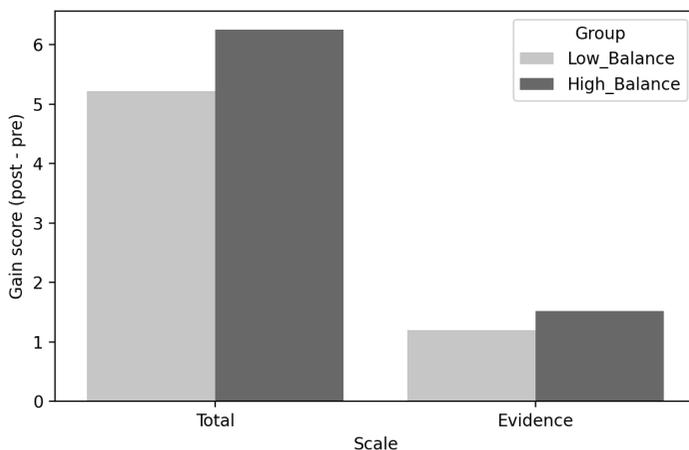

Figure 2. Mean gain scores (post–pre) in Total and Evidence by Balance group

When considered alongside the cluster results, the findings imply that rough classifications do not produce substantial differences in the overall progress. This prompted further analyses that focused on specific argumentative writing components and more precise indicators of constructive use.

4. Positive ChatGPT use and gains in evidence

Constructive uses of ChatGPT, such as expanding reasons, identifying counterpoints, and refining claims, are closely related to the

development of argumentative evidence. Based on this, I evaluated whether the Positive_Use_Index predicted improvements in the evidence components.

Students were divided into three groups based on tertiles of Positive_Use_Index: low, mid, and high positive use. A one-way ANOVA was conducted with 'Gain_Evidence' as the dependent variable and Positive_Use tertile as the independent variable. The analysis revealed a significant effect of Positive_Use_Index on Evidence gains the omnibus F-test was significant ($F = 3.27$, $p < .05$). Visually, the distribution of 'Gain_Evidence' across the three groups is illustrated in this boxplot.
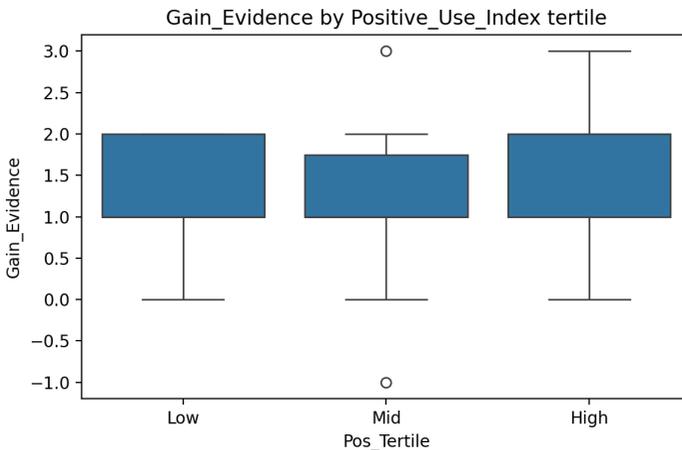

Figure 3. Distribution of gain_evidence

As shown, students in the high Positive_Use group achieved the largest average gains in evidence, followed by the low and mid groups. While the exact pattern across the low and mid groups is modest, the overall ANOVA result indicates that students who engaged in more extensive and constructive use of ChatGPT exhibited significantly greater improvement in providing evidence in their argumentative writing.

These results suggest that students' patterns of ChatGPT use shape the development of particular argumentative writing. This implies that higher-level generative or reflective practices are more strongly associated with these improvements than with overall writing performances. A one-way ANOVA indicated a significant effect of Positive_Use_Index tertiles on evidence, with students in the high tertile demonstrating the largest improvement in evidence. The descriptive statistics and ANOVA results are shown in Table 3.

Table 3. Descriptive statistics and one-way ANOVA

Tertile of Positive_Use_Index	n	M Gain_Evidence	SD Gain_Evidence
Low	41	1.29	1.10
Mid	42	1.14	1.05
High	37	1.57	1.15

Note: Gain_Evidence = post-test Evidence score – pre-test Evidence score.
Positive_Use_Index was computed as the sum of CE, OC, META, and CR scores and then divided into tertiles (Low, Mid, High) using sample-based quantiles.

Although the overall pattern suggested that a higher Positive_Use_Index was associated with greater gains, the mid group did not outperform the low group in evidence. This non-linear pattern indicates that more positive use is not automatically better and that only consistently high levels of elaborative and reflective engagement may translate into robust improvements. One plausible explanation is that mid group students sometimes adopted ChatGPT's suggestions without sufficient critical evaluation, leading to revisions that were more fragmented or even confusing at the level of evidence construction. By contrast, some low group students may have relied more on their own reasoning or course materials, achieving comparable gains despite lower reported AI use. These interpretations remain speculative, but they highlight the need for future research that triangulates survey-based indices with log data and qualitative analyses of drafts

to understand how different patterns of AI engagement actually play out in students' argumentative writing processes.

Table 4. ANOVA Summary

Source	SS	df	MS	F	p
Between groups	5.40	2	2.70	3.27	.042
Within groups (error)	89.90	109	0.82		
Total	95.30	111			

A one-way ANOVA revealed a significant effect of Positive_Use_Index tertiles on evidence gain, $F(2, 109) = 3.27$, $p = .042$. Students in the High Positive_Use_Index group showed the largest improvement in evidence (see Table 3).

5. Regression analysis: Positive use and gains in Counterargument

To further investigate the component-level effects, I examined the gains in counterargument (`Gain_Counterargument`) using a multiple regression model that included both initial performance and positive ChatGPT use, as well as their interaction. The model was:

Gain_Counterargument = β_0 + β_1 * Pre_Counterargument + β_2 * Positive_Use_Index + β_3 * (Pre_Counterargument × Positive_Use_Index) + error

The key findings from this model are Pre_Counterargument had a significant negative coefficient ($\beta_1 < 0$, $t \approx -3.07$). This indicates a clear ceiling effect: students who started with higher counterargument scores tended to show smaller subsequent gains, presumably because there was less room for improvement in their scores.

Table 5. Multiple regression predicting Gain_Counterargument from Pre_Counterargument, Positive_Use_Index, and their interaction

Predictor	B	SE B	t	p
Intercept	1.39	0.15	9.07	< .001
Pre_Counterargument	−0.63	0.21	−3.07	.003
Positive_Use_Index	0.05	0.02	2.53	.013
Pre_Counterargument × Positive_Use_Index	−0.03	0.04	−0.77	.444

The multiple regression model predicting Gain_Counterargument from Pre_Counterargument, Positive_Use_Index, and their interaction was significant, $F(3, 108) = 7.10$, $p < .001$, explaining approximately 23% of the variance ($R^2 = .23$). Pre_Counterargument negatively predicted gain ($B = -0.63$, $p = .003$), whereas Positive_Use_Index positively predicted it ($B = 0.05$, $p = .013$). The interaction between Pre_Counterargument and Positive_Use_Index was not significant ($B = -0.03$, $p = .444$; see Table 5).

The Positive_Use_Index had a significant positive coefficient ($\beta_2 > 0$, $t \approx 2.53$). Even after controlling for initial counterargument ability, students who made more extensive and constructive use of ChatGPT experienced greater gains in their counterarguments. This suggests that positive engagement with ChatGPT uniquely contributed to improvements in students' ability to recognize, generate, or respond to counterarguments in their writing.

Third, the interaction term between Pre_Counterargument and Positive_Use_Index (β_3) was insignificant ($t \approx -0.77$). This implies that the beneficial effect of positive ChatGPT use on counterargument gains did not differ systematically across levels of initial ability. In other words, both lower- and higher-performing students appeared to benefit from positive ChatGPT use, although high initial performers still showed smaller gains overall due to ceiling effects.

From a pedagogical perspective, this pattern suggests that con-

structive uses of ChatGPT can support the development of complex argumentative moves such as addressing counterarguments, and that this support is not limited to only the weakest or strongest writers.

V. Conclusion

Across analyses, a consistent picture emerges: broad usage patterns (e.g., global clusters of ChatGPT use or simple high vs. low balance indices) did not yield strong or consistent differences in overall writing gains. This suggests that merely using ChatGPT frequently or belonging to a particular usage cluster is not, by itself, closely associated with general writing improvement.

More differentiated patterns appeared when the focus shifted to the Positive_Use_Index, which reflects the qualitatively productive uses of ChatGPT, and to specific argumentative subskills. Students with a higher Positive_Use_Index tended to show larger gains in evidence, and regression analyses indicated a positive association between the Positive_Use_Index and gains in counterarguments. This contribution persisted when pre-test scores and ceiling effects were statistically controlled. These results are correlational and do not, by themselves, establish that the positive use of AI caused the observed improvements; for example, students who were already more proficient or motivated may also have been more likely to engage in constructive AI-supported strategies.

The non-significant interaction between pre-test scores and Positive_Use_Index suggests that constructive engagement with ChatGPT is linked to gains for students at multiple proficiency levels, rather than being confined to a single subgroup. Simultaneously, the strong negative effect of pre-test scores on subsequent improvement underscores the importance of taking initial abilities into account when interpreting the outcomes of AI-supported writing instruction.

Overall, these findings highlight the potential importance of purposeful and well-structured ways in which students engage with ChatGPT. Activities involving elaboration, organization of reasoning, and claim revision were statistically associated with gains on rubric-based measures of evidence and counterarguments. However, the present study focused on quantitative scores rather than qualitative analyses of argument content; thus, I cannot fully disentangle whether higher scores primarily reflect deeper argumentative development or more successful fulfillment of formal rubric criteria.

These conclusions should be interpreted as context-specific, reflecting the particular institutional setting, course design, and student population examined in this study, rather than as universal principles for all forms of AI-supported writing instruction. Accordingly, my findings are best understood as implications for AI-supported argumentative writing in this particular first-year composition course context, and should not be taken as general prescriptions for AI use in writing instruction across all educational settings.

Future instructional designs and research on AI-supported writing should therefore prioritize scaffolded and reflective uses of generative AI that center on argumentation, while also incorporating qualitative analyses of students' texts to capture how AI-supported practices shape the substance of structure of their arguments.

REFERENCES

Aull, L. L., & Ross, V. (2020). From cow paths to conversation: Rethinking the argumentative essay. *Pedagogy: Critical Approaches to Teaching Literature, Language, Composition, and Culture 20*(1), 21-34.

Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Humanities and Social Sciences Communications 10*(1), 1-5.

Bedington, A., Halcomb, E. F., McKee, H. A., Sargent, T., & Smith, A. (2024). Writing with generative AI and human-machine teaming: Insights and recommendations from faculty and students. *Computers and Composition 71*, 102833.

Deep, P. D., & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies 15*(9), 247.. https://doi.org/10.3390/soc15090247

Ding, L., Lawson, C., & Shapira, P. (2025). Rise of generative artificial intelligence in science. *Scientometrics 130*(9), 5093-5114.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education 38*(1), 39-72.

Elycheikh, A., Svetlana, M., & Magda, P. (2025). Critical Integration of Generative AI in Higher Education: Cognitive, Pedagogical and Ethical Perspectives. *London Journal of Research in Humanities & Social Science 25*(13), 1-12.

Fontenelle-Tereshchuk, D. (2024). Academic writing and ChatGPT: Students transitioning into college in the shadow of the COVID-19 pandemic. *Discover Education 3*(1), 6. https://doi.org/10.1007/s44217-023-00076-5

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior 93*, 279-289.

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is Chat GPT a blessing or a curse? *Frontiers in Education 8*, 1166682.

Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology 25*(1), 4-19.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology 104*(4), 879-896.

Hutson, J., Plate, D., & Berry, K. (2024). Embracing AI in English composition: Insights and innovations in hybrid pedagogical practices. *International Journal of Chang-*

es in Education 1(1), 19‑31.

Jin, H., Cisterna, D., Owens, D., & Riordan, B. (2025). Collaboration with ChatGPT: preservice elementary teachers analyzing and interpreting student responses. *Disciplinary and Interdisciplinary Science Education Research 7*(1), 27.

Karchmer, R. (2007). Best practices in using the Internet to support writing. In C. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (pp. 222‑241). Guilford.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., & Krusche, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences 103*, 102274.

Kim, J., Lee, S.‑S., Detrick, R., Wang, J., & Li, N. (2025). Students‑Generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528‑025‑09444‑6

Kwak, S. (2024). Generative AI: A Writing Companion or a Hindrance? *Journal of Reading Research*, (73), 45‑80.

Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2024). How do students use ChatGPT as a writing support? *Journal of Adolescent & Adult Literacy 68*(5), 445‑457. https://doi.org/10.1002/jaal.1373

McNeil, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain‑specific and domain‑general knowledge in writing arguments to explain phenomena. *Journal of the Learning Sciences 18*(3), 416‑460.

Nasr, N., Tu, C., Werner, J. S., Bauer, T., Yen, C., & Sujo‑Montes, L. (2025). Exploring the Impact of Generative AI ChatGPT on Critical Thinking in Higher Education: Passive AI‑Directed Use or Human‑AI Supported Collaboration? *Education Sciences 15*(9), 1198. https://doi.org/10.3390/educsci15091198

Newell, G. E., Bloome, D., & Hirvela, A. (2015). *Teaching and learning argumentative writing in high school English language arts classrooms.* Routledge.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it. *Computer Assisted Language Learning 31*(7), 653‑674.

Rodriguez, D. & Rodriguez, J.J. (1986). *Teaching writing with a word processor, grades 7Y13.* Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills and National Council of Teachers of English.

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), 342‑363.

Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. (2024). Comparing the quality of human and ChatGPT feedback on students' writing. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2024.101894

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing 57*, 100752.

Tate, T., Harnick-Shapiro, B., Ritchie, D., Tseng, W., Dennin, M., & Warschauer, M. (2025). Incorporating generative AI into a writing-intensive undergraduate course without off-loading learning. *Discover Computing 28*(1), 72. https://doi.org/10.1007/s10791-025-09563-9

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science 379*(6630), 313-313.

Toulmin, S. E. (2003). *The Uses of Argument*. Cambridge University Press.

Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning 3*(2), 258-262.

Usher, M., & Amzalag, M. (2025). From Prompt to Polished: Exploring Student–Chatbot Interactions for Academic Writing Assistance. *Education Sciences 15*(3), 329. https://doi.org/10.3390/educsci15030329

Van Eemeren, F. H., Henkemans, A. F. S., & Grootendorst, R. (2002). *Argumentation: Analysis, Evaluation, Presentation*. Routledge.

Vetter, M. A., Lucia, B., Jiang, J., & Othman, M. (2024). Towards a framework for local interrogation of AI ethics: A case study on text generators, academic integrity, and composing with ChatGPT. *Computers and Composition 71*, 102831.

Wang, C. (2025). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology Knowledge and Learning 30*(3), 1825-1846.

Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research 1*(6), 831-833.

ChatGPT 활용 패턴과 논증적 글쓰기 발달 양상

곽수범

본 연구는 한 학기 내 진행한 신입생 대학생 글쓰기 수업 중 논증적 글쓰기 과정에서 ChatGPT를 어떻게 활용했는지 탐구하였다. 연구 참여자 124명의 학생은 툴민 기반 논증 모형을 학습한 후 논증적 글쓰기를 하였다. 쓰기 과정에서의 ChatGPT 사용은 6개 척도(표면적 수정, 내용 강화, 조직/통일성, 메타인지 활용, 과의존도, 학생 주도성 조절)로 측정하였다. 분석적 평가 결과 점수는 전반적으로 소폭 향상되었으며, 특히 근거 제시와 반론 부분에서 점수 상승이 나타났다. 이 연구는 논증적 글쓰기 결과물의 질적 향상에 있어 생성형 AI의 피상적인 사용은 유의미한 결과로 이어지지 않는다는 사실을 재확인하였다.

핵심어 인공지능, 챗GPT, 논증적 글쓰기, 쓰기 교육, 혼합 연구

Patterns of ChatGPT Engagement and Argumentative Writing Development

Kwak, Subeom

This study explored how first-year university students used ChatGPT during a semester-long academic writing course on argumentative writing. One hundred and twenty-four students enrolled in a required freshman writing course at a Korean university received explicit instruction in Toulmin-based argumentation and wrote argumentative essays. ChatGPT use for writing was measured across six subscales (surface revision, content elaboration, organization/coherence, metacognitive use, dependence, and collaborative refinement). Analytic rubric scores showed modest overall gains, with particularly strong improvements in evidence and counterarguments. The findings highlight that the quality, rather than mere presence, of AI use matters in strengthening the quality of argumentative writing.

KEYWORDS Artificial intelligence, ChatGPT, argumentative writing, writing instruction, mixed-methods research