Generative AI Use in College Writing Classes: An Analysis of Student Chat Logs and Writing Projects

Sarah Madsen Hardy, Pary Fassihi, Shuang Geng, Christopher McVey, & Matt Parfitt

Boston University | US

Abstract: This study contributes to the emerging research on generative AI and writing pedagogy by exploring how college writing students make use of GAI when offered instruction in a range of responsible uses and latitude to integrate it into their writing process as they see fit. We analyzed chat log data and papers from participants recruited from six sections in which students were guided in experimenting with ChatGPT Plus and permitted to use it to produce up to 50% of submitted work. Through a combination of AI and human thematic content analysis of student chat logs, we found that in 18.6% of prompts, students asked ChatGPT to write for them. The rest of the prompts involved work leading up to or in support of the writing process. Human thematic content analysis of papers showed that students used ChatGPT to generate 8.2% of the writing they submitted. The most common rhetorical purpose of the AI-generated text they included was discussion/analysis/synthesis. English as a foreign language students (EFLs) in the sample prompted ChatGPT to clarify understanding less often than non-EFLs and integrated less AI-generated text into their papers, with a particularly notable difference in their use of AI-generated summaries. This unexpected finding merits further research, but it suggests that EFLs may use GAI for somewhat different purposes than non-EFL peers.

Keywords: generative AI, writing pedagogy, higher education, English as a foreign language

1. Introduction

The release of ChatGPT 3.5 in late 2022 created what some have described as an existential crisis for higher education, sparking debate about its role in teaching and learning—including its role in writing instruction. Some scholars and educators express strong concerns about the potential erosion of critical thinking, originality, and ethical integrity (Sano-Franchini et al., 2024; Zhai et al., 2024). Others argue that generative AI (GAI) can be a valuable tool for scaffolding student learning (Joakar, 2024; Wang, 2025), leveling the playing field for English language learners (Ghosh, 2024; Selim, 2024), and tailoring learning experiences to student needs (Fu, 2024; Nikolopoulou, 2024). This divide may be explained in part by the fact that public discourse and faculty member conversations about GAI pedagogy and policy often discuss the use of GAI tools in a binary way: a text is either generated by AI or by a human; students either do their own writing or use AI to do writing for them. Students themselves are ambivalent about the value of GAI for learning (Boillos & Idoiaga, 2025; Črček, N., & Patekar, 2023; Wang et al., 2024). When asked about their use of GAI tools, they report a more complex engagement in which AI is leveraged as a search tool, as a forum for brainstorming, and as a personalized tutor (Harvard Business Publishing, 2024). There is a pressing need to understand the varied ways students use GAI for their writing—a granular understanding of how students interact with GAI and make decisions about if and how to use its outputs, whether as a source for information, a tool for revision, or as language to integrate into their texts. By studying student use empirically and analyzing it quantitatively, we can better understand the ways GAI is changing student writing practices. As we do, we can begin to identify the learning conditions that may help students engage with these new technologies effectively and responsibly.

This research is beginning to emerge. By tracking prompts and document histories, Wang and Ren (2024) found that students used GAI most to provide definitions, present arguments, and analyze concepts. In open-ended questions about the experience, students demonstrated critical thinking about GAI outputs and regarded GAI as having had a positive impact on their work. The study authors conclude that these positive outcomes were dependent on the assignment's intentional integration of AI tools. Huston and colleagues (2024) revealed different AI usage patterns, distinguishing between students who effectively leveraged AI tools to refine arguments and those who attempted to use them as shortcuts. They noted students' integration of AI support in iterative writing processes, such as argument refinement and paragraph development, with more experience. Yang and colleagues (2024) also found that students who modified GAI-generated text produced higher quality texts than those who incorporated output without modifying it. Nguyen and colleagues (2024) recorded doctoral students' engagement patterns during a 30-minute writing session, finding that high-performing students used GAI effectively to supplement their writing process, while low-performing students tended to use GAI as a supplementary information source without engaging in deeper revisions, leading to weaker overall writing performance. In another study that directly analyzed prompts, Usher and Amzalag (2025) found that students used GAI in

varied ways, but that lower order uses like content generation and source retrieval were most common, and higher order uses such as brainstorming and synthesis less so. They speculate that this may be due in part to the fact that the subjects, who received no instruction in responsible or effective GAI use, lacked the experience needed to engage GAI in more sophisticated ways: "When students lack a deep understanding of how to optimize chatbot interactions, they may underutilize AI for higher-order cognitive tasks, restricting their engagement with the tool to basic requests rather than leveraging its potential for more complex and strategic writing assistance." Almunen and Jouhar (2025) found that explicit instruction in using ChatGPT for prewriting brainstorming enhanced student performance in persuasive writing. Taken together, these studies suggest that the degree of interactive engagement with AI significantly influences subsequent writing quality and point toward the need to develop writing pedagogies that guide and support iterative and critical interactions with GAI tools.

Our study builds on this body of research by examining student engagement with GAI in six pilot sections of introductory writing courses (i.e., individual classes sharing a common number and title) that were experimenting with AI-intensive instructional strategies. We analyze chat logs and research papers from students enrolled in two different introductory writing classes at a US research university. The piloting instructors offered students lessons in AI literacy—i.e., basic technical understanding of GAI plus an awareness of its ethical implications (Usher & Amzalag, 2025). Throughout the semester, they modeled a variety of possible uses and invited students to engage GAI critically and iteratively during their research and writing process, permitting them to incorporate AI-generated text for up to 50% of their submitted work. The pilot offered a structured, real-world context in which to explore the ramifications of GAI for student learning in the kinds of introductory writing and research classes commonly required at North American universities for students in their first year of study. While the pilot sections all included exercises that integrated GAI, students weren't asked to use the tools in a particular way. The pilot was designed in an intentionally open-ended manner that prioritized iterative experimentation and student agency and was premised on the idea that trust between teachers and students was crucial to developing accessible, ethical GAI-integrated pedagogies. This approach aligns with recent calls to attend to heterogeneity in how students approach GAI, since writers may bring different orientations, levels of awareness, and patterns of use to the same instructional context (Helm & Hesse, 2025).

This study aims to contribute to understanding how student writers make use of GAI tools in a context where they are given both permission and support. Rather than attempting to measure how GAI use affects writing quality, as in the studies above, we are interested in students' purposes as they prompt and re-prompt GAI systems. We are also interested in the rhetorical functions of the AI-generated text they choose to integrate into the writing assignments they submit. Our participants include a substantial proportion of international students who speak a range of home languages other than English. While the emerging empirical research on student GAI often involves multilingual students writing in English for

academic purposes, there is much we do not yet know about how English as a foreign language students (EFLs) interact with and integrate text from GAI when writing in English and how their uses may differ from those of first speakers of English (non-EFLs).

We address the following research questions:

1. What can chat log data tell us about how students choose to use GAI in a structured academic context that encourages critical engagement with GAI in the research and writing process?
2. What differences, if any, are there between the chat log prompting of EFLs and non-EFLs?
3. What is the percentage of AI-generated text that students in this context choose to include in submitted work, and what patterns can we see in how they employ it?
4. What differences, if any, are there between the AI-generated text integration of EFLs and non-EFLs?

Key findings include evidence that when students are provided both instructional support and autonomy around GAI use, they engage with GAI in many ways other than for generating text products wholesale. We also see evidence that while their engagement with GAI was similar in some ways, EFLs and non-EFLs' interactions differed depending on the purpose. EFLs sought understanding or clarification far less frequently in their prompts than non-EFL peers and included fewer AI-generated summaries in their papers, suggesting that students use GAI resources selectively and strategically in order to meet their unique linguistic and academic needs.

2. Methods

2.1 Context and Participants

The context for the study is a large private US research university, where undergraduates can fulfill university-wide requirements for introductory writing in several ways. Half of the study participants were taking an introductory rhetoric class that is part of an interdisciplinary team-taught program for students in their first two years of university study, and the other half were taking the second of two topic-based introductory writing classes open to all undergraduates. Students enrolled in these sections are a roughly representative cross-section of the university's undergraduate demographics, which include 42% people of color, 21% first in their families to attend college, and 23% international students, most of whom are studying English as a foreign language and have advanced English skills. As piloting instructors, we experimented with new ways to reach shared writing and research learning goals by (1) providing students with ChatGPT Plus subscriptions; (2) inviting and supporting students' critical engagement with GAI throughout the semester; and (3) embedding in each pilot section an undergraduate AI affiliate to collaborate with the instructor on GAI learning activities and support students using AI productively and ethically. While our individual classes and lesson plans for GAI integration varied, all piloting instructors took a critical approach to

GAI, framing it as a potentially useful yet fallible interlocutor in an ongoing dialogue rather than as a source of information or finished text, and all of us designed our class assignments and activities to support a common set of learning goals. Examples of learning activities that we designed and tried out in the pilot include AI literacy exercises to help students recognize GAI biases, verify sources, and write effective prompts. They also include scaffolding activities exposing students to options for integrating GAI at each phase of the research and writing process—for example, using GAI to locate and map scholarly conversations, support scholarly reading comprehension, brainstorm and refine research questions, seek feedback, generate relevant media, etc. We frequently asked students to reflect on their writing and research processes, including those that integrated GAI. In these ways, we aimed to learn alongside our students if and how GAI might complement traditional writing and research skills and support their progress toward the courses' common learning outcomes.[1]

As piloting instructors, we created policies (sometimes in conversation with students) that allowed up to 50% of work submitted as major assignments to be AI-generated. We asked all students in pilot sections to self-report their AI use for text generation by visually signaling word-for-word GAI-composed text with blue font. We also asked students to include a brief endnote for each major assignment describing how they used GAI in ways other than text-generation, if any. The pilot's main goals were to help instructors understand how students choose to interact with AI as part of the writing and research process and to encourage instructors to develop new pedagogical approaches to AI literacy. The pilot also offered an excellent opportunity to collect data on how students interacted with GAI tools in a supportive but open-ended context like that offered in the six pilot sections. Our institutional review board determined that our research was exempt from the IRB review process and did not require IRB oversight.

---

[1]Common learning outcomes for introductory writing, research, and inquiry classes:

Students will be able to critically assess both scholarly and public-facing sources, recognizing a variety of ways that sources can be credible; use sources ethically in domains such as attribution of ideas and treatment of human subjects; and interpret and analyze information.

Students will demonstrate understanding of the overall research process and its component parts. As a result, they will be able to formulate good research questions or hypotheses, use disciplinary modes of inquiry, select and deploy sources strategically to address research questions or hypotheses, and contribute to knowledge production.

Students will be able to craft responsible, considered, and well-structured written arguments, using media and modes of expression appropriate to the situation.

Students will be able to read with understanding, engagement, appreciation, and critical judgment.

Students will be able to write clearly and coherently in a range of genres and styles, integrating graphic and multimedia elements as appropriate.

2.2   Data Collection

When the semester ended, we recruited students from the pilot to participate in the study. Of 89 students enrolled in the pilot's six sections, 50 students gave us permission to analyze their work after the semester ended, so our dataset includes 50 papers that responded to a range of assignment prompts from four different piloting instructors. Twenty-six percent of these students (13) identified themselves as English as a foreign language students (EFLs). We also invited participants to submit chat logs related to their work on a major class project, offering us an additional kind of data. Our dataset includes 44 chat logs from 34 students (26% EFLs), with many capturing extensive exchanges with ChatGPT.

2.3   Data Analysis

AI-Assisted Content Analysis

We conducted content analysis on the student-submitted writing samples and chat logs to discern major patterns in the rhetorical purposes of students' AI use. In addition to traditional coding approaches involving human raters, we utilized a Large Language Model (LLM) to enhance the efficiency of the coding process. Emerging studies in the social sciences have shown that, given a structured codebook, LLM models used to code textual data have reached substantial agreement with human experts while enhancing efficiency (Chew, 2023; Pattyn et al., 2024; Prescott et al., 2024; Xiao et al., 2023). We developed the LLM-assisted content analysis method as described in the following two subsections.

Codebook Development

The members of the research team who taught participating courses reviewed a randomly selected sample of the full dataset (10%) and collaboratively developed an initial codebook. Categories were developed using a combination of deductive and inductive coding around research questions while narrowing the focus based on the predominant themes that emerged from the sample. This codebook was used to design the initial prompt to the LLM (ChatGPT 4o model). The LLM served as an additional rater to human raters and coded through the selected samples. We then compared coding results of human raters and the LLM to identify repetitive false coding patterns and refined the codebook and LLM prompt based on the findings. Tables 1 and 2 show the codebooks.

Coding and Validation

The instructors and LLM coded another randomly selected sample (20%). We conducted interrater reliability using Cohen's Kappa to compare human and LLM coding results. Cohen's Kappa was used to decide whether the LLM has reached desirable performance and could be used to code through the rest of the dataset.

Table 1. Codebook for Chat Logs

Code	Full Name	Description
P	Planning/ Brainstorming	Requests to generate ideas, explore topics, or structure a project. Includes brainstorming examples or potential research directions, asking for outlines, plans, or topic explorations for writing, generating ideas for research questions or arguments
S	Sourcing	Requests for assistance in locating, identifying, or verifying sources. Includes research advice on how to find credible sources, direct requests to have the AI identify or list relevant sources
U	Understanding/ Clarification	Requests aimed at clarifying, explaining, or comprehending information. Includes asking for definitions of terms or synonyms for words, requests for background or general information on a topic without requiring source attribution, inquiries related to understanding an attached or linked source overall (general comprehension) or seeking answers to specific aspects of the text, including just providing an attachment or link
F	Feedback	Requests for evaluative commentary on text quality without directly altering the content. Includes critiques on text effectiveness or adherence to genre conventions, general suggestions for improvement or writing advice
R	Revising	Requests that involve directly altering existing text. Includes changing style, tone, or length of text. Includes adding to, finishing, improving, or editing drafted text, specific instructions for revision or rewriting.
W	Writing	Requests for generating new or original text. Includes composing sections or complete pieces of text, creative or academic writing tasks, where text is produced from scratch

For the chat log dataset, the LLM's performance was substantial (Cohen's Kappas by category are between 76.2% and 93.7%). Therefore, the LLM was used to code the rest of the dataset alongside one instructor, who took into consideration its coding results.. The final coding is a collaborative effort between a human rater and the LLM.

Table 2. Codebook for Writing Samples

Code	Full Name	Description
EP	Entire paragraph	Entire paragraph (or paragraphs) formatted with blue font
TI	Title	Titles and subtitles
C	Citation	Bibliographic information about a book (author, title, press, DOI, etc.) in any academic format
BI	Background information	Factual information that is necessary to understand the research question and/or topic of the paper
Q	Question	Research question or more general questioning (revisiting or refining a question or part of a question, musing); description of what is not yet known and why it matters in the context of what experts have already found out
TH	Thesis	Main claim as asserted provisionally early in the paper; what the writer thinks they can show through the evidence they have found and analyzed
S	Summary	Short description of the main ideas in a single source (summary) or in a specific part of a single source (paraphrase). Note that the code Summary is used only for another writer's text. This code is not used when a writer is summing up their own text, as in an academic abstract or as part of a conclusion section
DAS	Discussion/analysis/synthesis	Interpretation of texts or data that are relevant to the research question, discussion of general conclusions that can be drawn from an example; language that brings together a number of discrete ideas, points, findings, concepts, sources, etc. DAS (not Summary) is used when a writer sums up their own ideas, as in an academic abstract or a conclusion section

O	Other	Anything that does not fit into any of the other codes (except entire paragraph)

For the writing samples dataset, agreement between the LLM and human raters was moderate (Cohen's Kappas by category are between 40% and 66.3%). Therefore, we switched back to a traditional coding approach, with each of the samples in the validation phase coded by two raters. Cohen's Kappa by category between human raters ranged from 85.6% to 100%. Given the substantial interrater reliability, each of the rest of the writing samples were coded by a single rater.

It is important to note that there may be differences in language used while prompting, leading to those prompts being categorized as different codes. For instance, while we would code "Explain the most important concepts in this text" (referring to an uploaded source) as *understanding*, the student might not use the output to shore up comprehension, but instead for the purpose of directly copying the output into their own paper. Alternatively, if the student had prompted GAI to "Write a summary of this text for inclusion in my paper," we would code it as *writing*. In situations such as these, the context of the prompt in the longer chat significantly helped to increase coding accuracy during our code validation process.

3.   Results

3.1   What Can Chat Log Data Tell Us About How Students Use GAI in a Structured Academic Context that Encourages Critical Engagement with GAI in the Research and Writing Process? What Differences, If Any, Are There Between Chat Log Prompting of EFLs and non-EFLs?

Our analysis of chat log data shows that 18.6% of 290 student prompts asked ChatGPT to directly write for them (i.e., "Can you generate an introduction for my 2500–3000-word research essay?"), with the rest of the prompts involving work leading up to or in support of the writing process. Prompts most frequently concern revision (24.5%, i.e., "Less fancy word and shorter"), followed closely by prompts asking for explanation of class materials or concepts (23.8%, i.e., "Explain how theory is used in this PDF"). Writing ranks third (18.6%), with lower percentages for planning (13.1%, i.e., "What would be the best way to break this essay down for a 5 minute video?"), sourcing (10.7%, i.e., "Find me an article discussing how AI is portrayed in the media") and feedback (10%, i.e., "This is my research paper draft. Is the paper well-organized, with clear headings, sections, and a logical progression of ideas?").

Our analysis also reveals some intriguing contrasts between EFL and non-EFL GAI interactions, as shown in Figure 1. While both used GAI to plan/brainstorm and ask for feedback in relatively similar—and low—proportions, there were notable differences in how these two populations interacted with GAI in the chat logs we collected. EFLs used GAI about twice as much for revision as their non-EFL peers (27% vs 13.1% of prompts), and only about half as much for help locating sources (7.9% vs. 14.2% prompts). The most striking contrast is EFLs' comparatively low use of GAI for understanding/clarification (4.8% of prompts compared to 34.1% for non-EFLs), a sevenfold difference. Understanding/clarification was the least common kind of chat log prompts submitted by EFLs, and it was the most common one for their non-EFL peers.

Figure 1: Purpose of Student Chat Log Prompts

Grouping our six chat log codes into two broader categories allowed us to see other patterns in student prompting. By grouping planning, sourcing, understanding, and feedback under the larger category of helping students with *learning,* and revision and writing under the larger category of helping students with *writing*, we see that students interacted with ChatGPT more often to ask it to give them advice, leads, resources, and explanations (57.6%) than to ask it to produce *or* revise text for them (43%).

These broader categories of help with learning vs. help with writing highlight further contrasts between EFL and non-EFL peers in our sample, as shown in Figure 2. EFL student prompts more often sought help with writing than non-EFL students (52.4% and 29%, respectively); non-EFLs sought learning support more than twice as often as they sought writing support (71.4 and 39.7%, respectively).
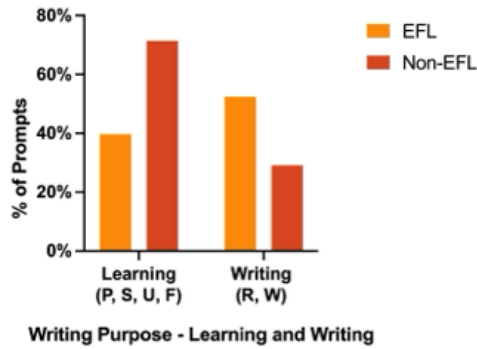
Figure 2: Learning and Writing in Student Chat Log Prompts

We also wanted to understand how GAI interactions evolved over the course of longer chat sessions. By cross-referencing the coded prompts with their location in the chat (as a cumulative percentage showing the prompting progress of the chat, with the first prompt occurring at 0% of the chat and the last prompt at 100%), we can visually map the evolution of student prompts, as shown in Figures 3 and 4. This data reinforces some of the general findings of the previous tables; compared with their counterparts, EFL students more often begin chats with prompts related to feedback or writing (25% each as early prompts, Figure 3) as compared with non-EFL counterparts, (14.3% each, Figure 4). However, even for EFL students, first prompts were most often related to planning (33.3%, Figure 3), while middle prompts were dominated by revision; prompts that ask GAI to produce writing dominated in the final 25% of EFL chat sessions (at 30%, Figure 3). This shows that even when students use GAI for the purposes of composing writing, such interactions often only occur after a longer conversation that still engages in many traditional parts of the writing process. This is especially the case for non-EFL students, who seldom begin chats by asking GAI to produce writing; the first 25% of the chat is instead dominated by planning (20%), sourcing (20%), and understanding (31.4%) (Figure 4).
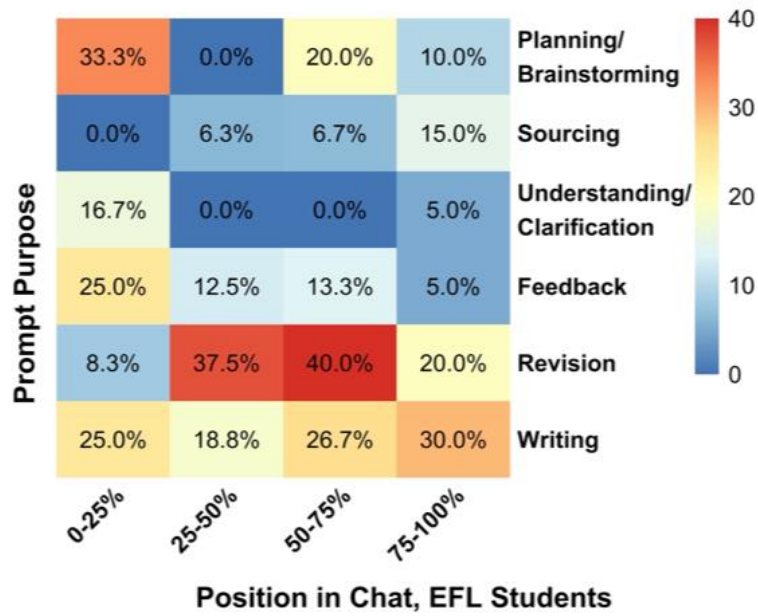
Figure 3: Frequency of Prompt Purpose in Relation to Position in Chat, EFLs

Note: Color intensity reflects the percentage of prompts in each chat position quartile that exhibit the specified purpose. Because a single prompt can demonstrate multiple purposes, the percentages in each column do not add up to 100%.

Figure 4: Frequency of Prompt Purpose in Relation to Position in Chat, non-EFLs

Note: Color intensity reflects the percentage of prompts in each chat position quartile that exhibit the specified purpose. Because a single prompt can demonstrate multiple purposes, the percentages in each column do not add up to 100%.
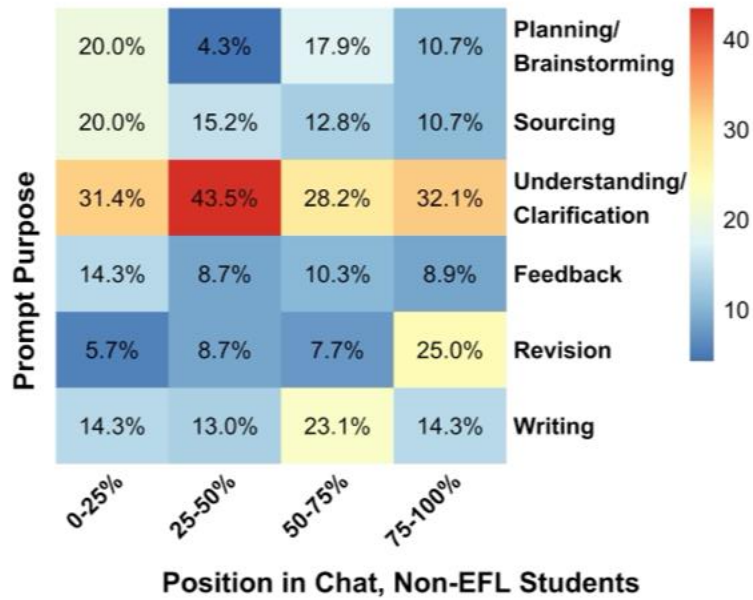
3.2 What Is the Percentage of AI-generated Text That Students Include in Submitted Work, and What Patterns Can We See in How They Employ It? What Differences, If Any, Exist Between the AI-Generated Text Integration of EFLs and non-EFLs?

The chat log analysis gave us a useful snapshot of the range, length, and kind of interactions with ChatGPT that students initiated, offering a behind-the-scenes look at the variety of ways students find GAI useful that would otherwise remain invisible. Direct analysis of writing samples offers a different source of information about some of the ways students find ChatGPT useful. As we mentioned above, in the spirit of experimentation, we all had open AI policies in which students could include up to 50% of AI-generated language in their papers. We asked, by the honor system, for students to format AI-generated text in blue font. We did not ask them to use blue font for minor stylistic or grammatical changes suggested by ChatGPT that are similar to those available through familiar tools like Grammarly and the proofreading features built into Microsoft Word and Google docs. We only asked them to indicate more substantive language that ChatGPT had generated, as when they cut and pasted from a chat

(see Figure 5). We acknowledged that there is some gray area in these distinctions and simply asked students to do their best in determining whether ChatGPT had generated text directly (blue font required) or whether it has helped inspire or revise text that they themselves had composed (no blue font necessary).

We also used the writing pedagogy best practice of meeting with students individually to discuss their drafts-in-progress and give them formative feedback, which offered us a good sense of student ownership of the papers they were in the midst of writing. Because of this opportunity to discuss drafts face to face, and because students knew we were open to and interested in how they would choose to use GAI, we have no reason to believe that their writing samples significantly underrepresent the amount of AI-generated text they included. However, we acknowledge the possibility that students were not entirely transparent or consistent in how they disclosed their use.
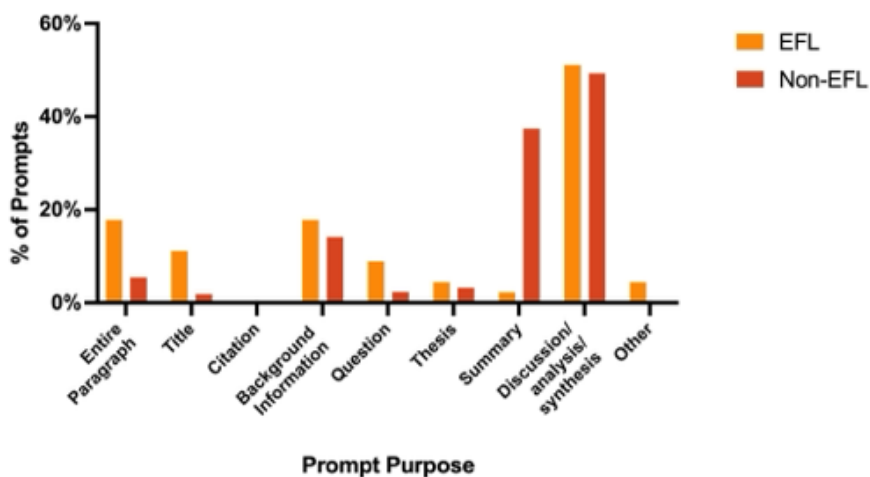

Figure 5: AI-Generated Text in Writing Samples, by Purpose

Our analysis revealed that about half of the students (26/50) who engaged with ChatGPT throughout the term as part of the pilot chose not to use it to generate any text in the papers we collected. EFL (9/13) and non-EFL (18/37) students made this choice in similar proportions. Papers that did not include any blue font accounted for 38% of the words in the writing we sampled. Overall, 8.2% of the words in the writing samples we collected were blue (GAI-composed). If we leave aside the samples from students who did not integrate GAI text at all

and consider only the papers that included at least some blue font, 13.2% of the words submitted were AI generated—still far less than class policies allowed.

We coded a total of 296 blue passages by their rhetorical purpose and also noted if the passage was an entire paragraph. (While we were most interested in the uses to which students put GAI output their own texts, we included a code for "entire paragraph" as a potential indication that students were using GAI in a passive way.) Overall, 6.8% of the blue words were copied and pasted as entire paragraphs, including 17.8% of blue words in EFL writing samples and 5.5% for non-EFLs (see breakdown in Figure 5). However, EFLs integrated only 45 passages of blue text, which is 15.2% of total AI-generated passages. Since 26% of students were EFLs, they seemed to use AI-generated text less often than their counterparts. When we count number of words instead of number of passages, we find the same thing: 5.2% of words were blue in the papers EFLs submitted, compared to 9.8% for non-EFLs. If we leave aside papers with no GAI text included, 10.8% of words were blue in papers EFLs submitted, compared to 14% for non-EFLs.

The most frequent rhetorical purpose of the blue text overall was discussion/analysis/synthesis (DAS) at 50%; this finding is unsurprising, since these purposes are central to academic writing. Numbers for DAS were similar across EFL and non-EFL groups. However, blue text used to summarize a single text, with the next highest percentage at 32.4%, revealed a striking contrast between EFLs and their non-EFL peers. For first speakers of English, summary made up 37.4% of the blue text in their papers, while for EFLs, summary was just 2.2%. This result lines up with our finding that in the chat logs: EFLs relied less on GAI for understanding/clarification than their non-EFL peers.

4. Discussion

Our results suggest that if students have the opportunity to engage GAI frequently and are encouraged to critically reflect on its strengths and weaknesses, they will use it selectively. We see clear evidence that when students participating in the pilot classes employed GAI to produce writing, they did so by making a complex series of decisions about how to weave its language into their own. The desire to succeed in the class clearly had an effect on these decisions: We assume that—as in any of our classes—students were motivated to submit good work, both intrinsically (through their interest in the material and their desire to learn) and extrinsically (through their desire to meet our assessment criteria, which differed by section and assignment). Those students who judged that the best way for them to submit good work was to use AI-generated text did so by integrating it in limited and specific ways. The papers in our sample integrated AI-generated passages were hybrid texts that reflected significant student agency; the 93% of blue passages that were not entire paragraphs were, by definition, actively selected, revised, and woven into the students' own language. Furthermore, while inserting an entire paragraph of GAI output might suggest a more passive, less critical engagement than weaving in output more selectively (see Huston et al., 2024; Yang et al., 2024), it's also possible that an entirely AI-generated paragraph is the result of a long series of prompts and revisions that could signal a different kind of student agency. Further

investigation would need to be done on the nature of the paragraphs before we could draw any conclusions.

In the context of a pilot designed to explore a range of different approaches to teaching writing and research with AI rather than to test one particular approach or assignment, we found students making varying judgements about the value of a wide range of potential uses. Doubtless, the circumstances of the pilot sections—which students registered for knowing that GAI experimentation would be encouraged—and piloting instructor pedagogy—one that emphasized process over product and devoted class time to critically exploring GAI output— affected our results. It is notable that half of the students—even in the context of regular exposure to prompting strategies in class and a course policy explicitly allowing liberal use of AI-generated prose in their papers—determined that it was not in their interest (in relation to their learning and/or their grades) to do so. These same students may well have used GAI to research, brainstorm, seek revision ideas, etc., and future research might analyze individual students' chat logs alongside their submitted work to tease apart such use patterns. But, based on the analysis we conducted, we can say that students who chose to incorporate GAI-generated text used it quite differently from one another. These findings confound any binary understanding of student AI use. They encourage us to think of GAI as part of a continuum of options for technology-integrated composing that includes many shades of gray. And they support an approach Gegg-Harrison and Shapiro (among others) describe as critical AI literacy—a pedagogy that aims to "help students make informed, confident choices regarding whether and how they use these technologies" (39).

Our results also suggest that EFLs and non-EFLs use GAI in some distinctive ways. Some of the differences we found are unsurprising: It makes sense that students writing in English as a foreign language would use entire paragraphs of GAI text more often than non-EFL peers (17.8% vs. 5.5%; see Figure 5) because of the extra cognitive load imposed by writing in English. It also makes sense that EFLs prompt GAI to help them with writing more than to help them with learning (see Figure 2). Both of these results suggest that EFLs choose to use GAI in ways that help offset the difficulty of working in a non-native language. More EFL prompts asking for help with writing and could also be explained by EFL's previous experiences with feedback and their impression that they can best use GAI to improve the quality of their work through help with wording and phrasing.

However, other results suggest more complex and unexpected differences between how EFLs and non-EFLs interact with GAI. EFLs in our sample incorporated fewer GAI-generated passages and fewer GAI-generated words overall than non-EFLs. And EFLs prompted GAI for understanding/clarification at a much lower rate than non-EFLs (4.8% of EFL prompts compared to 34.1% of non-EFL prompts; see Figure 1). They incorporated GAI-generated summary in the papers they submitted at a rate far lower than non-EFL peers (2.2% of their blue text summarized a source compared to 37.4% for non-EFLs; see Figure 5). Prompting for understanding and incorporating AI-generated summary are the two areas that show the strongest contrast between EFL and non-EFL GAI uses in the data we collected.

These results highlight the difficulty of distinguishing between *writing* and *learning* in the analysis of prompting behavior. For example, a student might ask ChatGPT to explain the main concepts in a scholarly article, which we would code as *understanding/clarification* and categorize as prompting for *help with learning*. That same student might then cut and paste the output of this query into the literature review section of their own research paper, using it in a way that we would code as *summary*. Our results indicate that this general pattern of use was far more common among non-EFLs than EFLs. More research is needed to determine whether a student's understanding of the content of the scholarly article was either enhanced or compromised by their choice to use GAI in this way.

Our preliminary finding of lower rates of prompting for understanding and clarification among EFLs should not be mistaken for a lack of initiative or skill. Instead, it may indicate intentional, strategic choices informed by EFL students' self-assessment of their language capabilities and their comfort level with traditional linguistic supports. Future studies might more deeply explore the qualitative dimensions of ELL students' interactions with GAI, such as through interviews or reflective journals, to better understand the motivations behind their specific patterns of GAI engagement. Understanding the root cause of their selective usage of GAI tools could inform instructional practices tailored to EFL students.

Any conclusions about the differences we found between EFL and non-EFL GAI use should be drawn with considerable caution. Though our sample size is comparable to that of some similar studies, it is possible that the students who opted to participate were atypical. There are many reasons why students may have opted not to participate, including not being familiar with the methods of scholarly research and not seeing it as a priority at the busy end of the semester. The lower number of students who opted into sharing chat log data might be explained by the fact that students presume their instructors will read their papers, while sharing chat log data likely feels more vulnerable. Sharing chat log data also entailed cutting and pasting links to chats on the survey form, whereas we already had access to papers submitted. Regardless, more research with larger samples that are more evenly balanced between EFL and non-EFL students is necessary to understand whether the patterns in our data represent a larger trend. If it proves to be, instructors might consider providing EFLs targeted instruction on how GAI could be used effectively in comprehension tasks. Explicit instructional materials tied to GAI prompting strategies specifically tailored to the needs of EFLs may encourage them to use GAI tools more fully (Warschauer et al., 2023).

5. Conclusions

Despite the questions that remain, our findings suggest that we should not assume EFL and non-EFL students use GAI tools in the same way. Furthermore, we should not assume that EFLs depend on GAI more heavily. We underscore the latter point to counteract the misconception that EFLs are more prone to overuse GAI—one reinforced by AI detector biases against EFLs (Liang et al., 2023). The fact that that among our participants, EFLs used GAI less frequently to clarify understanding or generate summaries of texts compared to their non-EFL peers also supports a broader conclusion gleaned from working with this data: There are as

many reasons and ways to use GAI in the writing process as there are students. Our findings show that students who are offered sustained instruction are capable of engaging GAI robustly and adopting its outputs selectively, in ways that suggest both judgement about its effects and investment in their own learning. We didn't consider the quality of writing samples, so we cannot say that students used GAI well or ethically, only that they used it relatively sparingly and in ways consistent with effective GAI use in earlier studies (Nguyen, 2024; Wang & Ren, 2024; Yang, 2024). We also know that the choices students made about GAI in this study were highly variable, suggesting that instructional approaches that offer multiple opportunities and options for experimenting with GAI may help students make autonomous decisions that preserve their agency and serve their own learning. Working with both chat logs and student writing confirmed our experience as instructors in the pilot: that if we model and talk about what responsible GAI use should be, students will adopt responsible practices, using it to augment rather than bypass their engagement in the writing process.

References

Almumen, H. ., & Jouhar, M. (2025). The Impact of ChatGPT as a Brainstorming Tool on Gifted Students' Persuasive Writing. *Journal of Writing Research*, *17*(2), 181-207. https://doi.org/10.17239/jowr-2025.17.02.01

Boillos, M.M., & Idoiaga, N. (2025) Student perspectives on the use of AI-based language tools in academic writing. *Journal of Writing Research*. https://doi.org/10.17239/jowr-2025.17.01.06

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv preprint arXiv:2306.14924*. https://doi.org/10.48550/arXiv.2306.14924.

Črček, N., & Patekar J. (2023). Writing with AI: University Students' Use of ChatGPT. *Journal of Language and Education*, *9*(4), 128-138. https://doi.org/10.17323/jle.2023.17379

Fu, Y. (2024). A Research of the Impact of ChatGPT on Education. *Applied and Computational Engineering*, 35, 26-31. https://doi.org/10.54254/2755-2721/35/20230354

Gegg-Harrison, W., & Shapiro, S. (2025). From policing to empowerment: Promoting student agency in the context of AI text-generators and AI-detection tools. In C. Wang & Z. Tian (Eds.), *Rethinking writing education in the age of generative AI* (pp. 26–41). Routledge. https://doi.org/10.4324/9781003426936-4

Ghosh, S. (2024, June), *ChatGPT as a Tool for Equitable Education in Engineering Classes.* Paper presented at 2024 ASEE Annual Conference & Exposition, Portland, Oregon. https://doi.org/10.18260/1-2--48458

Harvard Business Publishing. (2024). How students are actually using generative AI. https://hbsp.harvard.edu/inspiring-minds/how-students-are-actually-using-generative-ai

Helm, G., & Hesse, F. (2025). Training programmes on writing with AI – but for whom? Identifying students' writer profiles through two-step cluster analysis. *Journal of Writing Research*, *17*(1), 1-21. https://doi.org/10.17239/jowr-2025.17.01.01

Hutson, J., Plate, D. ., & Berry, K. (2024). Embracing AI in English Composition: Insights and Innovations in Hybrid Pedagogical Practices. *International Journal of Changes in Education*, *1*(1), 19-31.

Joakar, Ajit. (2024, October 5) A scaffolding based pedagogy for AI assisted education. https://www.linkedin.com/pulse/scaffolding-based-pedagogy-ai-assisted-education-ajit-jaokar-yfnle/

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7). https://doi.org/10.1016/j.patter.2023.100779

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education, 49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Nikolopoulou, K. (2024). Generative Artificial Intelligence in Higher Education: Exploring Ways of Harnessing Pedagogical Practices with the Assistance of ChatGPT. *International Journal of Changes in Education*, *1*(2), 103-111. https://doi.org/10.47852/bonviewIJCE42022489

Pattyn, F. (2024). The value of generative AI for qualitative research: A pilot study. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS4202964

Prescott, M. R., Yeager, S., Ham, L., Rivera Saldana, C. D., Serrano, V., Narez, J., ... & Montoya, J. (2024). Comparing the efficacy and efficiency of human and generative AI: Qualitative thematic analyses. *JMIR AI*, *3*, e54482. https://doi.org/10.2196/54482

Sano-Franchini, J., McIntyre, M., & Fernandes, M. (2024, November). Refusing GenAI in writing studies: A quickstart guide. *Refusing Generative AI in Writing Studies*. https://refusinggenai.wordpress.com

Selim, A. S. M. (2024). The Transformative Impact of AI-Powered Tools on Academic Writing: Perspectives of EFL University Students. *International Journal of English Linguistics*, *14*(1), 14-29. https://doi.org/10.5539/ijel.v14n1p14

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

Wang, L., & Ren, B. (2024). Enhancing academic writing in a linguistics course with generative AI: An empirical study in a higher education institution in Hong Kong. *Education Sciences, 14*(12), Article 1329. https://doi.org/10.3390/educsci14121329

Wang, C., Aguilar, S. J., Bankard, J. S., Bui, E., & Nye, B. (2024). Writing with AI: What College Students Learned from Utilizing ChatGPT for a Writing Assignment. *Education Sciences*, *14*(9), 976. https://doi.org/10.3390/educsci14090976

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*, 101071. https://doi.org/10.1016/j.jslw.2023.101071

Yang, K., Raković, M., Liang, Z., Yan, L., Zeng, Z., Fan, Y., Gašević, D., & Chen, G. (2024). Modifying AI, enhancing essays: How active engagement with generative AI boosts writing quality. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2412.07200

Yang, M., Jiang, S., Li, B., Herman, K., Luo, T., Moots, S. C., & Lovett, N. (2025). Analysing nontraditional students' ChatGPT interaction, engagement, self-efficacy and performance: A mixed-methods approach. *British Journal of Educational Technology*, 00, 1–28. https://doi.org/10.1111/bjet.13588

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. Y. (2023, March). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In *Companion proceedings of the 28th international conference on intelligent user interfaces* (pp. 75-78). https://doi.org/10.1145/3581754.3584136

Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2022). Examining the Relative Effects of Task Complexity and Cognitive Demands on Students' Writing in a Second Language. *Studies in Second Language Acquisition*, *44*(2), 483–506. https://doi.org/10.1017/S0272263121000310

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(28). https://doi.org/10.1186/s40561-024-00316-7