https://doi.org/10.29140/97817637116240-26

Using Large Language Models to Enhance Data-driven Learning in Academic Writing

Xuhua Zhang[1] and Jianbin Zeng[1]
[1]Fudan University, China

Introduction

Academic writing instruction in English for Specific Purposes (ESP), particularly for students in science and engineering disciplines, increasingly incorporates genre-based approaches to help learners understand the rhetorical structure and discourse conventions of research writing (Swales, 1990; Hyland, 2004). To support learner autonomy and disciplinary appropriateness, many instructors supplement genre-based instruction with corpus-informed methods such as Data-Driven Learning (DDL) (e.g., Charles, 2007), which encourages learners to discover authentic usage patterns from language corpora (Johns, 1991; Boulton & Cobb, 2017).

While the effectiveness of DDL has been widely acknowledged across educational levels (O'Keeffe, 2021), its classroom implementation often faces practical challenges. These include the cognitive and technical demands of corpus tools, learners' limited proficiency in interpreting concordance data, and the time constraints inherent in classroom settings (Ädel, 2010). To respond to these challenges, recent advances in artificial intelligence—especially large language models (LLMs) such as DeepSeek, ChatGPT and KIMI—offer new opportunities for supporting writing instruction.

LLMs can provide context-sensitive feedback, diagnose language errors, and assist with paraphrasing, citation checking, and sentence-level revision (Fathi & Rahimi, 2024; Liu et al., 2024; Wang et al. 2025). These capabilities are not merely theoretical; recent classroom-based research demonstrates their practical benefits and highlights the need for AI literacy. A randomized controlled trial involving 259 Chinese undergraduates compared LLM-generated feedback with traditional instructor feedback and found that

E-mail: zhangxuhua@fudan.edu.cn & jbzeng@fudan.edu.cn

students receiving AI feedback showed significantly greater improvements in organization and content development (Zhang, 2025). Another classroom study used structural equation modelling to show that AI literacy and self-regulated learning positively predicted writing performance and digital well-being among university students in generative-AI-supported courses (Shi, Liu & Hu, 2025). These findings, alongside earlier CALL-focused trials (e.g., Fathi & Rahimi, 2024; Liu et al., 2024), underscore both the potential of LLMs to enhance revision and the need for deliberate instruction in AI literacy. AI literacy is now regarded as an extension of digital literacy, encompassing the ability to understand, apply, evaluate and ethically engage with AI technologies (Mansoor et al., 2024). These perspectives informed our integration of LLMs into a genre-based, DDL-supported curriculum. Moreover, there remains a need for empirical classroom-based research to evaluate how LLMs can be meaningfully integrated into genre-based and corpus-assisted academic writing instruction.

This study reports on a two-year teaching practice in an "Academic English for Science and Engineering" course that primarily adopted a genre-based instructional approach supplemented by DDL activities. In a subset of class sections, this combined approach was further enhanced with the integration of LLMs. Focusing on synthesis writing tasks, this study compares student performance between classes using corpus-assisted instruction alone and those additionally supported by LLMs, examining the impact on learners' ability to identify and revise linguistic and citation-related errors.

Methods

This study was conducted over five consecutive semesters (Spring 2023 to Spring 2025) in an "Academic English for Science Engineering" course offered to undergraduate and graduate students majoring in science and engineering at a leading Chinese university. A total of 673 students from 23 classes participated in the study, all following a genre-based curriculum supplemented by Data-Driven Learning (DDL) methods. Thirteen classes (384 students) used DDL techniques such as analyzing collocations and citation patterns with corpus tools to improve academic writing. The remaining ten classes (289 students), including two at the graduate level, followed the same instructional sequence but also incorporated LLMs, such as DeepSeek, ChatGPT, and KIMI, to support language analysis, error detection, and revision tasks.

In both groups, synthesis writing was introduced in the middle of the semester through model analysis and guided exercises. A key component of the instruction involved corpus consultation using the Corpus of Contemporary

American English (COCA)(Davies, 2009). Students were guided to use COCA's online interface to investigate collocational patterns, phrase structures, and citation-related usage (e.g., reporting verbs, hedging expressions). Specific tasks included searching for high-frequency academic collocations, identifying field-appropriate verbs for integrating sources, and comparing grammatical patterns across registers. These hands-on DDL activities aimed to raise students' awareness of academic language and inform their revisions during writing tasks.

In the LLM-enhanced classes, students completed the same DDL tasks but were also encouraged to use AI tools to diagnose and revise language and citation issues in model synthesis texts. Instructors provided guidance on prompt formulation, AI feedback evaluation, and cross-referencing with corpus findings to validate or challenge AI suggestions.

To evaluate the impact of integrating LLMs into DDL academic writing instruction, we analyzed student performance on a synthesis writing quiz administered one week after classroom practice. The quiz was designed to assess students' ability to identify and correct pre-embedded errors in a synthesis text. It embedded thirteen deliberate errors covering four categories: (1) language mistakes (e.g., grammatical and lexical errors), (2) citation-related issues—including redundant or misformatted reference information, (3) logical inconsistencies in the argument or structure, and (4) mechanical errors such as improper punctuation and spacing. Students received one point for each correctly identified error; identifying ten or more earned full credit.

Data were thus collected to compare performance between the DDL-only and AI-supported groups. Scoring was based on the number of correct identifications. A score of 10 indicated full marks.

Results and Discussion

Across the five semesters, 13 classes (384 students) received genre-based writing instruction with corpus-assisted DDL activities (Referred to as DDL only in Table 1), while 10 classes (289 students) received the same instruction with LLM integration (Referred to as DDL+LLMs in Table 1). The average scores for the corpus-assisted only group ranged from 4.91 to 6.17, with a mean around 5.57. In contrast, the LLM-enhanced group scored between 6.12 and 7.89, with a higher mean of approximately 6.91 (Table 1).

These results suggest that LLM-enhanced instruction led to significantly better performance in error identification tasks, particularly in recognizing subtle citation formatting issues and phraseological inconsistencies that novice

Table 1. Overview of classes, enrollment, and quiz performance by instructional condition

Semester	Instructional Condition	No. of Classes	No. of Students	Mean Quiz Scores
Spring 2023	DDL only	3	83	6.12, 5.16, 5.3
Spring 2023	DDL+LLMs	2	59	6.32, 7.2
Fall 2023	DDL only	3	80	5.12, 5.41, 6.02
Fall 2023	DDL+LLMs	2	60	6.89, 7.43
Spring 2024	DDL only	3	95	6.17, 5.44, 6.0
Spring 2024	DDL+LLMs	2	65	7.14, 7.89
Fall 2024	DDL only	2	60	5.19, 4.91
Fall 2024	DDL+LLMs	2	50	6.54, 6.12
Spring 2025	DDL only	2	66	5.71, 6.01
Spring 2025	DDL+LLMs	2	55	7.21, 7.4
Total	DDL only	13	384	5.57 (avg. across classes)
Total	DDL+LLMs	10	289	6.91 (avg. across classes)

writers often overlook. Students using LLMs could access immediate, context-sensitive feedback, allowing them to test hypotheses, revise iteratively, and internalize accurate academic language patterns more efficiently.

From an instructional perspective, the integration of LLMs addressed two major limitations of traditional DDL: the cognitive load of corpus navigation and the time constraints of in-class feedback. While DDL fosters analytical thinking and pattern discovery, it can be overwhelming for students with limited corpus literacy. LLMs acted as scaffolding agents, enabling broader student engagement with complex tasks such as synthesis writing and source integration.

Importantly, the LLMs did not replace DDL but complemented it. Students were still encouraged to use corpora for verification and to compare with AI output. This hybrid approach maintained the pedagogical rigor of DDL while improving efficiency and inclusivity.

These findings support the use of LLMs as pedagogically meaningful enhancements to corpus-informed writing instruction, particularly for synthesis writing tasks that demand accuracy, integration, and stylistic control.

Conclusion

This study demonstrates that augmenting a genre-based, corpus-assisted academic writing curriculum with LLMs can substantially improve students' ability to identify and correct linguistic and citation-related errors. Over five semesters and 673 participants, classes that used LLM support scored appreciably higher on a synthesis-writing quiz than those relying solely on corpus-based activities. Learners benefitted from immediate, context-sensitive feedback, which enabled iterative revision and deeper engagement with complex academic language patterns. Crucially, LLMs did not replace traditional DDL practices but complemented them, reducing the cognitive load of corpus navigation and providing scaffolding for learners at varying proficiency levels. These findings underscore the promise of hybrid AI-assisted approaches for more inclusive and efficient writing instruction.

However, several limitations must be acknowledged. First, data were collected across five consecutive semesters; although the curriculum was standardized, variations in teaching style, digital proficiency and feedback practices could have influenced student performance. Second, effective use of COCA and LLMs presupposes a level of digital and AI literacy that is unevenly distributed among learners. AI literacy—viewed as an extension of digital literacy—includes understanding, applying and critically evaluating AI technologies. Variations in students' familiarity with prompt engineering or corpus navigation may therefore have contributed to performance differences. Third, group assignment was based on intact class sections rather than randomization, leaving open the possibility of pre-existing differences in language proficiency, writing experience or motivation. Fourth, the synthesis quiz focused on error identification in a single task and may not fully capture sustained improvements in rhetorical control or synthesis skills. Future research should employ randomized class assignments, incorporate longitudinal writing portfolios, and use more comprehensive assessments to evaluate writing development.

References

Ädel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. In O'Keeffe, A., & McCarthy, M. (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371–382). Routledge.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. https://doi.org/10.1111/lang.12224

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, *6*(4), 289–302. https://doi.org/10.1016/j.jeap.2007.09.009

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: a qualitative study. *Computer Assisted Language Learning*, 1–40. https://doi.org/10.1080/09588221.2024.2374772

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

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning materials. *ELR Journal*, *4*, 1–16.

Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2024). Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1–25. https://doi.org/10.1080/09588221.2024.2389923

O'Keeffe, A. (2021). Data-driven learning–a call for a broader research gaze. *Language Teaching*, *54*(2), 259–272. https://doi.org/10.1017/S0261444820000523

Mansoor, H. M., Bawazir, A., Alsabri, M. A., Alharbi, A., & Okela, A. H. (2024). Artificial intelligence literacy among university students—a comparative transnational survey. *Frontiers in Communication*, *9*, 1478476.

Shi, J., Liu, W., & Hu, K. (2025). Exploring how AI literacy and self-regulated learning relate to student writing performance and well-being in generative AI-supported higher education. *Behavioral Sciences*, *15*(5), 705.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*, 1–28.

Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behavioral Sciences*, *15*(5), 600.