How Generative AI Affects Chinese College Students' English Academic Writing Skills

Xiaotian Li*

Department of Chinese Study, Faculty of Arts, Chinese University of Hong Kong, Hong Kong, 999077, China

Abstract. Generative AI (e.g., ChatGPT, Grammarly, QuillBot) is rapidly diffusing into English for Academic Purposes writing, yet its relationship with learners' self-efficacy, perceived ability, cognitive load, and dependence remains unclear. This cross-sectional survey ($N = 93$ Chinese undergraduates) quantified one-week AI usage (days/hours, stages, tools), writing self-efficacy (SE), self-rated ability across four EAP dimensions (AB: content/argumentation; structure/coherence; evidence/citation; language/style), cognitive load (CL), feedback-literacy behaviors (editing/verification), and perceived AI dependence (DPN). Results show a significant group difference in AB: high-frequency AI users rated their academic writing ability higher than low-frequency users (one-way ANOVA, $p = .041$), and AI use time correlated positively with AB ($r \approx .30$). Perceived dependence was higher among high-frequency users than low-frequency users (t-test, $p < .01$), and AI use time correlated with dependence ($r = .369$, $p < .001$). A multiple regression predicting DPN from CL, SE, and AB was significant but small in magnitude ($R^2 \approx .08$); cognitive load uniquely and positively predicted dependence ($p = .025$). Non-significant associations (e.g., SE with AI use) are discussed as design targets for future work. The author argues for calibrated, feedback-literate AI use and planned "AI-off" practice to preserve autonomous writing competence while harnessing efficiency gains.

1 Introduction

Generative AI (e.g. ChatGPT, Grammarly, QuillBot) is increasingly embedded in English for Academic Purposes (EAP) writing. Empirical work and reviews report short term gains in grammatical accuracy, coherence, and lexical variety when learners use AI based support during drafting and revision [1–3]. Learners also tend to report lower anxiety and increased autonomy or self-efficacy even after structured exposure to AI-assisted writing tasks [4, 5]. However, both longitudinal and post-use evidence warn that benefits will fade away, and heavy users may become dependent and unable to produce unaided writing once the AI is withdrawn [6–8]. Additional academic integrity concerns surrounding the use of text generation and paraphrase tools further hinder the adoption of these tools for coursework [9].

* Corresponding author: 1155203239@link.cuhk.edu.hk

Against such a convergence of forces, Chinese undergraduates confront high-stakes EAP requirements (genre-appropriate argumentation, coherence, accurate citation, formality). It is important to know how AI will affect their self-efficacy, perceived cognitive load and self-editing behaviour, as well as their attitudes toward fallback during periods of non-use, so that pedagogy can leverage AI productively without undermining unaided writing competence. This paper describes the design of a one-month quantitative study of patterns and outcomes of AI use and the relationship between different variables by survey and data analysis.

2 Literature review

A recent review summarizes burgeoning use of ChatGPT and related systems for brainstorming, outlining, drafting and revision, showing robust near-term improvements in linguistic output with open questions for long-run development [1]. Studies focused on Grammarly report consistent gains on grammar and some lexical measures, with small or mixed effects on fluency and complexity [2]. A randomized controlled trial for eight weeks of AI-mediated feedback found grammatical accuracy and cohesion to be superior in the treatment group compared to controls [3]. A semester-long study showed reductions in anxiety and increases in autonomy and self-efficacy with sustained use of Grammarly [4]. Six-week intervention with QuillBot showed improvements in writing performance and writing self-efficacy [5].

In a randomized design for two semesters, initial gains for a ChatGPT-assisted group narrowed over time as dependence increased, with some learners feeling less capable when the AI was unavailable [6]. While automated tools provide immediate corrective feedback, research suggests that passive reliance on these technologies can reduce learners' cognitive engagement, potentially inhibiting the development of self-regulated editing strategies required for independent writing [7]. Research indicates that while automated writing evaluation tools like Grammarly can enhance immediate text accuracy, they may foster student dependency, as learners often mechanically accept corrections without the deep cognitive processing necessary for long-term language internalization. [8]. The integrity risks from text outsourcing and opaque paraphrase serve to argue for guidance and policy [9].

Academic writing quality may be taken as genre-specific moves and rhetorical organization (Swales) as well as Complexity-Accuracy-Fluency (CAF) dimensions (Housen & Kuiken) [10, 11]. Bandura’s self-efficacy theory explains how mastery experiences (with or without AI) recalibrate confidence for EAP tasks [12]. Carless & Boud’s feedback literacy (seeking, sense making, and acting) illuminates how learners interpret and apply AI suggestions [13]. Technology Acceptance Model constructs (perceived usefulness/ease) account for adoption and intensity, while trust in automation theory differentiates calibrated reliance from over trust and automation bias [14, 15].

3 Methodology

3.1 Participants and sampling

Target N =93 undergraduates enrolled in EAP/academic writing. Inclusion: ≥18 years; at least one academic writing task in the past month. Self-reported AI use in the past 7 days supports usage grouping: Low/None (≤1 day or <0.5 hr), High (≥4 days and ≥2 hrs), with Moderate retained for sensitivity. Background covariates: gender, major, year, recent weekly hours of writing practice, EAP course enrollment, and baseline English proficiency (standardized score or 5-point self-rating).

3.2 Measures and analysis

The questionnaire operationalized seven constructs relevant to AI-assisted EAP writing.

First, AI use over the prior seven days was captured by self-reported days and total hours, disaggregated by stage of use (idea/outline generation, paragraph generation, structural advice, revision/linguistic polishing, evidence/citation support, and formatting), by tool type, and by prompt-craft proficiency (1–5 Likert). From days × hours the author derived an intensity grouping for between-group comparisons: Low/None (≤1 day or <0.5 hr), High (≥4 days and ≥2 hrs), and Moderate (otherwise).

Second, writing self-efficacy (SE) was measured with six 1–5 items tailored to EAP tasks (e.g., organizing an argument under time limits, integrating sources and citing accurately, maintaining coherence, sustaining academic register); their mean served as the SE score (internal consistency checked via Cronbach's α).

Third, self-rated EAP writing ability (AB) was assessed with four 1–5 items aligned to standard rubric dimensions—content/argumentation, structure/coherence, evidence/citation, and language/style—averaged to form AB_mean.

Fourth, cognitive load (CL) for the most recent EAP task was indexed with a single 0–10 mental-effort rating.

Fifth, feedback-literacy/editing behaviors (E) were captured with five 1–5 items covering evaluative adoption (assessing reliability before use), rewriting rather than copying AI output, fact/quotation verification, transforming suggestions into actionable checklists, and genre/register calibration; the item mean was used descriptively (E_mean).

Sixth, perceived dependence (DPN) was measured with five 1–5 items (e.g., difficulty initiating without AI, expected efficiency loss when AI is unavailable, tendency to adopt suggestions with minimal modification), averaged as DPN_mean.

Finally, AI-off expectations were gathered with two 5-point items asking anticipated changes in time-on-task and self-editing frequency during a hypothetical one-week ban; these served to contextualize dependence and usage-intensity effects.

4 Result

4.1 Usage panorama (context for subsequent analyses)

On average, students reported 2.40 days of AI use (SD = 1.72) and 3.18 hours total (SD = 1.54) in the past week. Usage clustered in revision/polishing (≈ 75%) and idea/outline support (≈ 59%), while paragraph generation (≈ 35%) and evidence/citation help (≈ 30%) were less common. Prompt proficiency averaged around "moderate." The baseline percentages set the stage for the subsequent analyses of how AI-use intensity relates to writing performance and cognitive load.

4.2 AI use and self-rated EAP writing ability (AB)

A one-way ANOVA found a significant group difference in AB across usage groups: High users reported higher EAP writing ability than Low/None users ($p = 0.041$). Descriptively, the High AB mean was ≈ 3.47 while the Low/None mean was ≈ 3.01. The Moderate AB mean was in between, at ≈ 3.19. Complementarily, total AI hours were positively associated with AB ($r \approx 0.30$), indicating a small-to-moderate positive association between using AI more and feeling more competent as an EAP writer.

In this sample, those using AI more reported higher self-rated EAP writing ability, supporting a small-to-moderate positive association between AI use and perceived competence, particularly for the composite rubric dimensions (content / structure / evidence / language).

4.3 Perceived dependence (DPN) and its correlates

Perceived dependence overall was small (mean ≈ 2.87/5), but there were large group differences: High users reported significantly higher dependence than Low/None users (independent-samples t-test, $p < .01$). At the continuous level, AI hours correlated positively with DPN ($r = 0.369$, $p < .001$). These two converging signals indicate a dose–response pattern: more time with AI, stronger feelings of reliance.

A multiple regression predicting DPN from cognitive load (CL), self-efficacy (SE), and self-rated ability (AB) was significant but small ($R^2 \approx 0.083$). Crucially, CL uniquely and positively predicted DPN ($p = 0.025$): students who experienced higher mental effort during recent EAP tasks were more likely to feel dependent on AI. The SE coefficient was negative and near-significant ($p \approx 0.055$), consistent with the interpretation that greater confidence may buffer against dependence, whereas AB had no unique effect.

Perceived dependence concentrates among heavy users and among those who feel EAP tasks are mentally effortful. These findings suggest that perceived dependence is higher among heavy users and among students experiencing greater cognitive load, while self-efficacy may modestly buffer against dependence.

4.4 Fallback attitudes under “AI-off” scenarios

When imagining a one-week AI-off scenrio, ≈ 66% anticipated more time on task (approximately 20% reported a substantial increase (‘much more’), and ≈ 50% anticipated more self-editing. High-frequency users were over-represented among those expecting to spend more. Such expectations parallel the forecasted dependence costs, underlining the pedagogical rationale for planned AI-off practice so students can maintain performance without support.

5 Discussion

5.1 Interpretive and pedagogical implications of the significant findings

Two results are striking and cohere with a balanced view of AI-supported EAP writing. First, those using AI more report higher self-rated EAP ability. Classroom reports that AI support for students writing arguments can help them structure their arguments, be coherent, polish their style, and manage citation mechanics—all of which are reflected in our AB composite—suggest that frequent users are (a) better writers to begin with, and (b) are better able to translate AI suggestions into concrete improvements in their text. Another interpretation is that AI scaffolds provide “worked examples” of academic phrasing and cohesion that help students notice and rehearse strategies they can then internalise. Either mechanism (or their interaction) would lead to higher AB ratings.

Second, dependence is greater the more one uses AI and the greater the perceived cognitive load. The positive DPN–usage coefficient is relevant for at least two reasons. First, it supports the claim that dependence reflects a behavioral reinforcement loop: the more students feel they need to offload effort to AI to get by, the more they will feel they need it—especially when a task is cognitively taxing. The CL→DPN pathway provides one actionable suggestion: if teachers reduce the extraneous load students experience when completing tasks (by, for example, scaffold prompts to be less demanding, sequence research steps, or model genre moves), and also encourage feedback-literate editing (evaluate→ verify→ act), then students may feel less need to offload to AI and hence may exhibit weaker dependence.

The near-significant negative coefficient for SE is in the expected direction: students who believe they can complete EAP tasks without help from AI need AI less. Thus, this triangle of load variables (usage, extraneous, and efficacy) provides a pathway to a practical recipe: scaffold cognitive load before digital pedagogy and enhance self-efficacy before sequentially scaffolding further via a structured use of AI tools, not the other way around.

5.2 Pedagogical implications

5.2.1 Calibrated integration rather than blanket adoption

The AB benefit among High users should not be taken as support for always-on AI use. Instead, the results presented here argue for careful staging: teachers can choose where to let students place AI (e.g., giving alternative outlines to stimulate planning, having students simulate reviewer feedback, or showing cohesion problems to promote revision), but can also require independent planning and revision. These findings indicate that while AI can support writing performance, its use should be strategically integrated to foster independent skills. Alternating AI-assisted and AI-off tasks during a course—especially before high-stakes assessments—can help students transfer strategies and verify they can sustain performance without tool support.

5.2.2 Teach feedback literacy as the guardrail

In EAP writing, editing/verification items showed students generally evaluate reliability, rewrite, and sometimes fact-check, but the lowest mean was on turning suggestions into actionable checklists. Instruction should therefore train the full feedback-literacy cycle—seek→ make sense→ act—including source-checking and genre/register calibration. Making this cycle explicit may maintain AB gains while attenuating DPN.

5.2.3 Address cognitive load upstream

Because CL uniquely predicts DPN, reducing avoidable load is a dependence-reduction strategy. In practice: decomposing EAP tasks into clear subgoals (move structures), providing model exemplars with annotations, allowing scaffolded drafting (e.g., topic sentence banks), and offering low-stakes rehearsal. When tasks feel doable, students are less compelled to offload them wholesale to AI.

5.3 Why some plausible links were not significant—and how to detect them better

This paper deliberately de-emphasized non-significant associations (for example, between AI use and self-efficacy, or small differences in editing behaviors), yet these null findings are theoretically meaningful and likely reflect design or measurement limits rather than the absence of an effect. With $N \approx 93$, effects smaller than roughly $r \approx 0.25$ are may not reach statistical significance; increasing the sample (e.g., $N \geq 200$), using more discriminating exposure groupings based on hours and paragraph-generation frequency, and indexing cumulative exposure across months rather than a single week would stabilize estimates and increase statistical power.

A second improvement is to pair self-reports with objective writing outcomes. Double-blind rubric ratings on a brief standardized EAP task (with inter-rater reliability, ICC ≥ 0.75) would reduce shared-method bias and test whether the self-rated ability (AB) advantage among heavy users translates into externally judged quality. Beyond linear models, it is also worth testing non-linearities and usage profiles. An inverted-U pattern is plausible if moderate AI use balances efficiency with learning; profile-by-outcome analyses contrasting "revision-only," "paragraph-generation heavy," or "structure-oriented" users may reveal structure obscured in aggregate analyses.

Finally, dependence as measured here must be instrumented behaviorally as well as through attitude. Time-to-first-keystroke, self-editing counts, and brief "AI-off" microtasks (e.g., a five-minute outline with no assistance) would be sensitive indicators, and forced-choice items in vignette-based multiple-choice questions can outperform agreement scales. Analytically, adding tests for mediation and moderation would clarify which processes are operating—e.g., whether feedback literacy mediates AI use→AB or whether self-efficacy weakens the dependence on usage→dependence link; bootstrap CIs (PROCESS-style) would improve inferential precision. Methodologically, quasi-experimental and longitudinal designs would provide within-person contrasts that are sensitive even when samples are not sufficiently large to provide robust between-person comparisons and speak directly to persistence and transfer once assistance is removed—a short crossover assignment in which a matched set of prompts is used (AI-assisted one week, AI-off the next) or a pre/post track with an AI-off follow-up.

5.4 Limitations

This study's cross-sectional study design means that we cannot infer causality: it is equally likely that students who feel more competent choose to use AI more as it is that AI use increases competence, and heavier users may simultaneously be facing more difficult tasks that lead both to use and outcome. Measures relied largely on self-report (AB, SE, DPN, usage), which introduces common-method variance; future work should incorporate objective writing tasks, system logs, and revision traces. In this study, the sample size was relatively small (N=93), and the majority of participants were drawn from a single institution. Consequently, the findings have limited generalizability to programs with different EAP curriculum frameworks, assessment systems, or AI usage policies. To prioritize feasibility, several constructs were brief, including a single-item cognitive load index and compact SE items; multi-item load scales (e.g., short-form NASA-TLX) and richer self-efficacy batteries would likely improve reliability and sensitivity. Finally, although qualitative triangulation was planned, no interviews were conducted, so mechanisms—for example, how students decide to adopt versus rewrite AI suggestions—remain inferred rather than observed.

5.5 Practical summary

Taken together, the most robust patterns are that heavier AI use is associated with higher self-rated EAP ability and higher perceived dependence, and that cognitive load is a unique positive correlate of dependence. Pedagogically, the present work argues for balanced use of AI in instruction: teach genre moves and feedback literacy first and foremost, place AI only where instructional need exists, and institutionalize periodic AI-off practice such that autonomous competence is exercised and maintained. Methodologically, the next steps should explore whether moderate (as opposed to minimal or heavy) AI use produces the greatest objective writing gains with the least dependence and whether self-efficacy-building interventions reduce the usage-dependence linkage-preferably within longitudinal or crossover designs that can advance the field from association to credible causal knowledge.

6 Conclusion

This study described how Chinese undergraduates currently use generative AI for English for Academic Purposes (EAP) writing and how their engagement with AI relates to key learner outcomes. Through a cross-sectional survey (N = 93) that quantified recent AI use alongside self-efficacy (SE), self-rated EAP writing ability (AB), cognitive load (CL), feedback-literacy behaviors, perceived dependence (DPN), and AI-off expectations, we found two robust and complementary patterns. First, heavier AI use aligned with higher self-rated EAP ability: high-frequency users reported significantly higher AB than low-frequency users, and weekly AI hours correlated positively with AB. Second, perceived dependence increased with usage, showing a clear dose–response pattern; moreover, cognitive load uniquely and positively predicted dependence in a multivariable model. Taken together, these results suggest that, in real classroom conditions, students who use AI more tend to feel more capable as academic writers, yet this perceived competence can coexist with a stronger sense of reliance on the tool—especially when writing feels mentally effortful.

Pedagogically, the findings argue for calibrated integration of AI rather than blanket adoption. Instructors can position AI where it adds demonstrable value (e.g., outlining alternatives, simulating reviewer feedback, highlighting cohesion issues) while deliberately preserving AI-off phases for independent planning and revision. Parallel instruction in feedback literacy—evaluating reliability, verifying sources, rewriting instead of copying, and translating suggestions into actionable checklists—can help students convert AI outputs into learning rather than shortcutting effort. Because cognitive load is a unique correlate of dependence, upstream measures to reduce extraneous load (clearer move structures, annotated exemplars, scaffolded drafting, low-stakes rehearsal) are likely to lower reliance and build durable self-efficacy.

At the same time, the study's design limits interpretation. Because it was cross-sectional in nature, we cannot infer causality; self-report measures are subject to common-method bias; and the fact that the sample was drawn from a single institution with limited variation in sample size preclude making strong generalizations. Here, some links that were plausible on theoretical grounds were small and not significant, which may reflect low power, the short instruments, or the need for objective measures. Future research should thus complement self-reports with double-blind rubric scoring of standardized writing tasks, model non-linear usage profiles (e.g., revision-only vs. paragraph-generation heavy), test mediation/moderation pathways (e.g., whether feedback literacy mediates AI use→ AB; whether SE buffers usage→ DPN), and use longitudinal or crossover designs that alternate AI-assisted and AI-off assignments to examine persistence and transfer.

This survey shows that more frequent use of generative AI is linked to higher perceived EAP writing ability, but also to greater perceived dependence—especially when cognitive load is high. Practically, programs should prioritize genre skills, including feedback literacy, first and then integrate AI purposefully while maintaining regular AI-off practice to sustain autonomous competence. Methodologically, to move beyond associations, programs should use objective writing evaluations and stronger causal designs to explore mechanisms and effects.

References

1. C.K. Lo, P.L.H. Yu, S. Xu, D.T.K. Ng, M.S.Y. Jong, Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learn. Environ. **11**, 50 (2024).
2. G. Dizon, J.M. Gayed, A systematic review of Grammarly in L2 English writing contexts. Cogent Educ. **11**, 2397882 (2024).

3. L. Kohnke, D. Zou, R. Zhang, Evaluating the impact of AI mediated feedback on academic writing: A randomized controlled trial. System **108**, 102897 (2022).
4. G. Dizon, Exploring the effects of Grammarly on EFL students' foreign language anxiety and learner autonomy. JALT CALL J. **19**, 299–316 (2023).
5. M.S.M. ElGarawany, The effects of a QuillBot based intervention on English language majors' EFL writing performance, apprehension, and self efficacy. Lang. Teach. Res. Q. **43**, 167–189 (2024).
6. S. Park, M. Song, Longitudinal effects of ChatGPT assisted writing on EAP students: A two semester randomized controlled study. J. Second. Lang. Writ. **65**, 101055 (2024).
7. L. Tian, Y. Zhou, Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. System, **91**, 102247 (2020). https://doi.org/10.1016/j.system.2020.102247
8. R. O'Neill, A. Russell, Grammarly: Help or hindrance? Academic learning support professionals' perceptions. Journal of Academic Language and Learning, **13**, A88–A107 (2019).

   https://journal.aall.org.au/index.php/jall/article/view/591
9. D.R. Cotton, P.A. Cotton, J.R. Shipway, Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innov. Educ. Teach. Int. **61**, 228–239 (2024).
10. J.M. Swales, Research Genres: Explorations and Applications (Cambridge University Press, 2004).
11. A. Housen, F. Kuiken, Complexity, accuracy and fluency in second language acquisition. Appl. Linguist. **30**, 461–473 (2009).
12. A. Bandura, Self Efficacy: The Exercise of Control (W. H. Freeman, 1997).
13. D. Carless, D. Boud, The development of student feedback literacy: Enabling uptake of feedback. Assess. Eval. High. Educ. **43**, 1315–1325 (2018).
14. F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**, 319–340 (1989).
15. J.D. Lee, K.A. See, Trust in automation: Designing for appropriate reliance. Hum. Factors **46**, 50–80 (2004).