

# Examining the Effects of Structured ChatGPT Use on First-Year ESL Writing Development Over Two Semesters

Omar Serwor Massoud, Jing Zhang,  
and Paul McQuilling

## Abstract

As generative AI enters university writing classrooms, its impact on EFL writing and academic integrity remains unclear. This two-semester quasi-experimental study examined structured ChatGPT use in first-year Japanese university classes. Two intact course sections served as naturally occurring groups: an experimental section ( $n = 18$ ) that used ChatGPT as a scaffold for idea generation, organization, grammar clarification, and citation support, and a control section ( $n = 12$ ) that completed the same curriculum without AI. Pre- and post-course timed writings were scored by two trained raters using a five-category analytic rubric. Independent-samples t-tests on gain scores (one-tailed) showed a significant advantage for the ChatGPT group in Mechanics ( $p = .036$ , Hedges'  $g = 0.678$ ) and a marginal advantage in Organization and Coherence ( $p = .056$ ), with no group difference in Language Use and Grammar. No between-group differences were observed in Content or Vocabulary. Qualitative responses indicated that students found paragraph organization easier and became more aware of academic integrity, though many remained unsure about grammar accuracy, vocabulary choice, and appropriate paraphrasing. Taken together, structured ChatGPT use, positioned as a tutor rather than a text generator, appears to support Mechanics and aspects of organization and ethical awareness. Key limitations include the small, non-equivalent sample and genre differences between pre- and posttests.

## Introduction

Generative AI tools such as ChatGPT are rapidly entering university writing classrooms, offering feedback, idea support, and language scaffolding. In English as a foreign language (EFL) context, they may strengthen organization, clarity, and adherence to academic conventions, yet raise concerns about over-reliance and academic integrity. Much of the existing research is short term or perception based and tends to emphasize sentence-level accuracy over higher-order outcomes (e.g., Song & Song, 2023;

Kim et al., 2024), with reviews noting the need for more robust classroom studies on learning effectiveness and ethical use (Li, Tan, Wang, & Lowell, 2025; Lee et al., 2025; Bittle & El-Gayar, 2025). From an applied-linguistics lens, carefully constrained AI can function as scaffolded support for writing, a point developed further in the theoretical framework below, but classroom evidence on sustained learning and integrity remains limited, particularly outside Western settings (Cotton et al., 2023).

This study reports a two-semester quasi-experimental investigation with first-year Japanese university students (CEFR A2-B1), extending a six-week pilot on self-efficacy and academic-integrity awareness (Massoud & Zhang, 2025). Two intact classes followed the same writing curriculum; the experimental class used ChatGPT as a tutor for brainstorming, organization, grammar clarification, and citation (copy-pasting AI text was prohibited), while the control class completed identical tasks without AI. Pre- and post-course timed writings were rated using an analytic rubric; bilingual surveys and brief reflections captured confidence and ethical-use awareness, and a teacher journal<sup>1</sup> provided classroom context. The study offers outcome-focused evidence on what sustained, scaffolded AI improves in EFL writing and what it does not, and it informs guidelines for pairing AI with explicit language instruction, integrity routines, and teacher monitoring.

## **1. Literature Review**

### **AI in Language Education**

AI has reshaped language learning by supporting practice, feedback, and adaptive instruction. Early Computer-Assisted Language Learning (CALL) and later Mobile-Assisted Language Learning (MALL) introduced programmed drills, feedback systems, and portable practice opportunities (Liang et al., 2021; Rüschoff & Ritter, 2001; Kukulska-Hulme, 2013). More recently, Intelligent CALL has used natural language processing and machine learning to provide individualized tutoring and automated assessment (Heift, 2021; Khatun & Miwa, 2016). These tools can personalize materials, offer automated writing feedback, and adjust to learners' proficiency (Pokrivcakova, 2019; Huang et al., 2023). At the same time, concerns remain about confusing or inconsistent AI responses, teacher readiness, and the reliability of feedback on complex skills such as writing (Roscoe & McNamara, 2013; Grimes & Warschauer, 2010; Marr, 2024).

## **Theoretical Aspect**

Theoretical perspectives on learning help explain how AI tools such as ChatGPT might support writing development and confidence. Bandura's (1997) concept of self-efficacy highlights that beliefs about one's capability influence motivation, persistence, and achievement in writing (Pajares, 2003). From a sociocultural perspective, Vygotsky's (1978) Zone of Proximal Development (ZPD) and Wood et al.'s (1976) notion of scaffolding describe how learners progress when they receive guidance from more knowledgeable others, while Krashen's (1985) *i+1* Input Hypothesis suggests that exposure to slightly challenging but comprehensible language supports growth. Together, these theories suggest that learners benefit when they receive support, face manageable challenges, and develop confidence. In this view, generative AI can function as a guided assistant: it can provide feedback, model texts, and suggest revisions in ways that resemble tutoring. Stojanov (2023) shows how ChatGPT 3.5 can act as a "more knowledgeable other," helping learners extend their abilities through scaffolded interaction. This suggests that immediate, adaptive feedback from ChatGPT may enhance writing self-efficacy when it is used as a support rather than a replacement for students' own work.

## **AI and Academic Integrity**

Academic integrity has become a critical issue in the context of AI use, especially with tools like ChatGPT, because they can both support and potentially threaten ethical writing practices. This study conceptualizes academic integrity as students' commitment to ethical writing, including accurate citation, proper referencing, and avoidance of plagiarism. Although ChatGPT, introduced in November 2022, has proven effective in supporting writing tasks, it has also raised concerns about undermining academic standards (Bin-Nashwan et al., 2023; Cotton et al., 2023). Reflecting these concerns, several universities, including Stanford and Oxford, specify that generative AI tools may only be used in summative assessments when explicitly permitted, treating unapproved use as academic misconduct (Stanford University, 2024; University of Oxford, 2025).

Historical accounts of China's seventh-century civil service exams describe ghostwriting and harsh penalties (Lang, 2013; Iqbal, 2023), indicating that academic dishonesty is longstanding, even if its forms change. In the contemporary context, Niloy et al. (2024) report that frequent use of AI tools such as ChatGPT can weaken students' understanding of ethical writing and increase the risk of plagiarism. Recent classroom work, including Massoud and Zhang (2025), has begun to examine how ChatGPT influences students' writing self-efficacy and awareness of academic integrity,

but further research is needed to understand its long-term impacts and broader ethical implications.

## Research Gap

Existing scholarship shows that ChatGPT can enhance sentence-level accuracy in areas such as grammar, vocabulary, and phrasing, as demonstrated in recent ESL writing studies (Mahapatra, 2024; Deep et al., 2025). What remains less clear is how sustained classroom use relates to broader aspects of writing quality such as organization, coherence, and mechanics over time. A 12-week intervention reported gains in writing performance and motivation (Song & Song, 2023), but most generative AI (GenAI) studies in language education are short term, focus on sentence-level accuracy, and rarely use controlled designs (Li et al., 2025; Lee et al., 2025). Comparative evidence further suggests that AI-only support is strongest for lower-level features, while human-AI hybrids yield larger gains in higher-order qualities such as organization and reasoning (Zhang, Aubrey, Huang, & Chiu, 2025).

Motivational benefits are frequently reported, yet we still know little about how extended, guided use shapes students' confidence and self-belief in academic writing. Short-term studies link AI feedback with increased motivation and confidence (Song & Song, 2023), but have not tracked longer-term change. Reviews call for attention to psychological and cognitive dimensions, including self-efficacy and autonomy (Li et al., 2025; Lee et al., 2025), and learner-focused research on metacognition emphasizes plan-monitor-evaluate processes as central to responsible ChatGPT use (Teng, 2025). Early classroom work also points to perceived benefits while urging a shift beyond perceptions toward measured change (Massoud & Zhang, 2025).

Ethical considerations likewise require further study. Much of the current conversation centers on institutional responses to plagiarism rather than on how students develop ethical awareness while using AI in their courses. A recent review highlights gaps in understanding generative AI's effects on academic integrity and calls for policy, training, and assessment redesign, particularly outside Western contexts (Bittle & El-Gayar, 2025). Media-based analysis reaches similar conclusions, noting that institutional responses dominate coverage while student ethical development receives less attention (Sullivan, Kelly, & McLaughlan, 2023). Mali (2025) and Massoud and Zhang (2025) similarly emphasize the need to track how integrity awareness develops under sustained AI use. Together, these works point to the need for longer-term, classroom-based studies that link GenAI use with measurable outcomes in organization,

mechanics, confidence, and ethical awareness gaps that the present two-semester trial seeks to address.

## 2. Research Questions

This study investigates the effects of structured ChatGPT use on first-year Japanese university students' English (EFL) academic writing development and ethical awareness. Building on prior evidence that ChatGPT enhances students' self-efficacy in writing (Massoud & Zhang, 2025), the research addresses the following questions:

1. Does regular, guided use of ChatGPT lead to greater improvement in Japanese EFL students' English writing quality particularly in terms of organization, coherence, mechanics, and grammar accuracy compared with instruction without AI?
2. How does extended use of ChatGPT influence students' confidence in academic writing tasks?
3. How does sustained ChatGPT use affect students' understanding and awareness of academic integrity, including ethical paraphrasing and citation practices?
4. What benefits and challenges do students report when using ChatGPT as a writing support tool?

## 3. Methodology

### Participants

Thirty first-year Japanese university students enrolled in a required academic writing course at a private university in Tokyo participated in the study in 2024. All were English majors or in closely related programs, aged 18-20, with a mix of genders. Institutional placement testing placed students within the CEFR A2-B1 range, and the two class sections showed no meaningful baseline differences (pre-course level-test comparison,  $p > .05$ ).

Two intact class sections were used as naturally occurring groups: the Tuesday section ( $n = 18$ ) served as the experimental group, and the Thursday section ( $n = 12$ ) served as the control group. Because students registered for sections based on scheduling availability, the groups were non-equivalent but comparable rather than randomly assigned. None of the students had received formal instruction in AI-assisted writing prior to the study.

The course met once weekly for 90 minutes across two semesters (approximately 28 weeks). All participants provided informed consent and were briefed on responsible AI

use and academic-integrity expectations.

## Research Design

This study employed a two-semester quasi-experimental, non-equivalent groups design using two intact first-year writing classes at a Japanese university. Across the two semesters, instruction progressed from paragraph-level writing to short essays, with an emphasis on organization, coherence, and sentence-level accuracy. Because students registered for either the Tuesday or Thursday section based on scheduling availability, the classes functioned as naturally occurring groups rather than randomly assigned clusters. Both sections were taught by the same instructor, followed an identical syllabus and textbook sequence (*Great Writing 2*), met once per week for 90 minutes, and completed the same writing tasks and feedback cycles. The only planned instructional difference between the two sections was the in-class, structured use of ChatGPT in the experimental section.

In the experimental section (Tuesday), students received structured opportunities to use ChatGPT as a writing support tool. Students used ChatGPT (OpenAI; GPT-3.5) via the freely available web interface, which was familiar to students and required no cost or special access. ChatGPT was positioned as a tutor rather than a text generator: students used short English prompts appropriate for A2-B1 proficiency levels to support brainstorming, outlining, paragraph organization, sentence-level clarification (grammar and mechanics), and citation formatting. For paragraph organization, students commonly asked about opening or concluding sentences, adding supporting examples, or linking ideas (e.g., “*Help me start a strong first sentence about [topic]*” or “*How can I link my ideas in this paragraph?*”). ChatGPT’s responses focused on broad aspects of paragraph structure, such as topic sentences, supporting details, and transitions, rather than rewriting or reordering student text.

Students also used ChatGPT for help with paraphrasing or summarizing ideas and for guidance on MLA citation format. Citation-related use was limited to formatting support (e.g., in-text citations and Works Cited entries). Decisions about when a citation was required, whether quotation was necessary, and how sources were integrated into student writing were addressed through classroom instruction and teacher feedback rather than delegated to ChatGPT. Japanese translations were allowed for comprehension, but all composing and prompts remained in English. Students were not permitted to copy or paste generated text into drafts, and the instructor monitored in-class use and checked drafts for originality.

The control section (Thursday) completed the same writing curriculum without AI assistance, relying instead on peer review, dictionaries, and teacher conferencing.

Timed writing tasks were evaluated using a five-category analytic rubric (Content and Development, Organization and Coherence, Language Use and Grammar, Vocabulary, Mechanics). All five categories were scored. Analyses focused on Organization and Coherence, Language Use and Grammar, and Mechanics because these areas aligned most closely with the instructional focus of the course and the research questions. Content and Development and Vocabulary were examined descriptively to provide contextual information.

### **Data Collection**

This mixed-methods study gathered quantitative and qualitative data across two semesters (April 2024-January 2025). Instruments used in class are reproduced in the appendices, including the pretest and post test surveys (Appendices A-C).

1. The Week 1 writing task required students to produce a timed, single-paragraph opinion composition, while the end-of-semester task required a timed short multi-paragraph opinion essay with multiple paragraphs, greater length, and more sustained argumentation. The posttest task therefore placed higher demands on organization, coherence across paragraphs, and language control. Two trained raters independently scored all scripts using a five-category analytic rubric (Content and Development, Organization and Coherence, Grammar, Vocabulary, Mechanics), and averaged sub-scores were used for analysis. The raters were blind to group assignment and test timing (pretest vs. posttest) during scoring.
2. Alongside the baseline writing, students completed a bilingual Japanese-English pretest survey on demographics, familiarity with the university's academic-integrity policy, prior AI use, and attitudes toward ethical AI use. With the posttest timed short multi-paragraph opinion essay, they completed group-specific surveys: the control version asked about confidence while writing, aspects of writing that felt easier, areas still needing improvement, and learning about academic integrity; the experimental version added items on AI use, its perceived effects on writing, confidence in using AI ethically, and views on classroom guidelines. Both versions included short-answer items that yielded brief qualitative comments.
3. A brief teacher's journal recorded lesson focus, permitted ChatGPT prompt functions in the experimental section, and parallel non-AI activities in the control

section; these entries were later used to contextualize the quantitative results.

All instruments were administered during regular class time. Students had already been placed in the course level through the university's routine placement process, and no additional CEFR test was administered. Informed consent was obtained prior to data collection, and all surveys and writing scripts were de-identified before scoring and analysis.

## Data Analysis

Two raters independently scored each script on five rubric categories (1-5 scale). If their scores differed by more than one point on any category, the raters briefly discussed the script and reached a consensus score; otherwise, the two scores were averaged.

Because the study involved two intact first-year classes, the experimental and control groups were defined at the class level rather than through random assignment. However, all analyses were conducted at the student level. For each focal rubric category, Organization and Coherence, Language Use and Grammar, and Mechanics, we calculated an individual gain score (posttest minus pretest) based on the average of the two raters' scores. These gain scores were compared between the two groups using independent-samples *t*-tests. Given prior evidence that ChatGPT improves students' self-efficacy in organization, paragraphing, and citation practices (Massoud & Zhang, 2025), the primary hypothesis for the present study was directional: students with structured access to ChatGPT were expected to show greater gains in writing quality than those in the non-AI condition. Accordingly, gain score comparisons for the three pre-specified primary outcomes Organization and Coherence, Language Use and Grammar, and Mechanics, were evaluated using one-tailed independent-samples *t*-tests ( $\alpha = .05$ ). Prior to the main analyses, Shapiro-Wilk tests were conducted to examine the assumption of normality. Additionally, baseline equivalence was examined by comparing the two groups' Week 1 writing scores with a *t*-test. All statistical analyses were conducted in IBM SPSS 29.

Survey items were summarized using counts and percentages. Short-answer responses and weekly teacher observation notes were coded inductively into a small set of recurring themes to illustrate students' experiences with ChatGPT, including aspects they found helpful or challenging.

## 4. Results

### 4.1 Quantitative Data Analysis

Independent-samples *t*-tests were conducted on gain scores (posttest minus pretest) for the three pre-specified primary outcomes: Organization and Coherence, Language Use and Grammar, and Mechanics. Content and Development and Vocabulary were examined descriptively because they were not part of the study's confirmatory hypotheses.

Across the two semesters, instruction progressed from paragraph-level writing in Semester 1 to a timed short multi-paragraph opinion essay in Semester 2, increasing both organizational and linguistic demands; as a result, decreases in posttest scores were expected, and gain scores were used to compare relative change between groups under identical task conditions.

As shown in Table 1, students in the ChatGPT-supported section showed significantly smaller decreases in Mechanics scores ( $M = -0.806$ ,  $SD = 1.273$ ) than those in the control section ( $M = -1.667$ ,  $SD = 1.174$ ),  $t(28) = 1.870$ ,  $p$  (one-tailed) = .036. This indicates that students with structured access to ChatGPT demonstrated greater improvement (or less decline) in spelling, punctuation, and general mechanics compared with peers whose course did not include AI support during the two semesters. For Organization and Coherence, the experimental group again showed a smaller decline ( $M = -1.028$ ,  $SD = 1.169$ ) than the control group ( $M = -1.792$ ,  $SD = 1.356$ ). Although the effect was in the predicted direction, the difference did not reach statistical significance,  $t(28) = 1.645$ ,  $p$  (one-tailed) = .056. For Language Use and Grammar, differences between the experimental ( $M = -0.917$ ,  $SD = 1.047$ ) and control groups ( $M = -1.167$ ,  $SD = 1.094$ ) were not significant,  $t(28) = 0.630$ ,  $p$  (one-tailed) = .267. The two groups showed comparable declines in Content and Development (experimental:  $M = -1.472$ ,  $SD = 1.277$ ; control:  $M = -1.708$ ,  $SD = 1.698$ ), with no reliable group differences. Vocabulary gains were likewise similar (experimental:  $M = -1.167$ ,  $SD = 1.200$ ; control:  $M = -1.208$ ,  $SD = 1.356$ ),  $t(28) = 0.088$ ,  $p$  (one-tailed) = .465.

**Table 1**

*Independent-Samples t-Tests, Descriptive Statistics, and Tests of Normality for Gain Scores*

| Variable           | Group | Descriptive Statistics |       | t-test for Equality of Means |                |              | Tests of Normality |                |
|--------------------|-------|------------------------|-------|------------------------------|----------------|--------------|--------------------|----------------|
|                    |       | Mean                   | SD    | t(df)                        | p (one-tailed) | 95% CI Lower | 95% CI Upper       | Shapiro-Wilk p |
| Gain_<br>cont&dev  | Exp   | -1.472                 | 1.277 | 0.435                        | .334           | -.876        | 1.349              | .418           |
|                    | Ctrl  | -1.708                 | 1.698 | (28)                         |                |              |                    |                |
| Gain_<br>org&coh   | Exp   | -1.028                 | 1.169 | 1.645                        | .056           | -.187        | 1.715              | .444           |
|                    | Ctrl  | -1.792                 | 1.356 | (28)                         |                |              |                    |                |
| Gain_<br>luse&gram | Exp   | -.917                  | 1.047 | 0.630                        | .267           | -.563        | 1.063              | .052           |
|                    | Ctrl  | -1.167                 | 1.094 | (28)                         |                |              |                    |                |
| Gain_<br>voc       | Exp   | -1.167                 | 1.200 | 0.088                        | .465           | -.923        | 1.007              | .050           |
|                    | Ctrl  | -1.208                 | 1.356 | (28)                         |                |              |                    |                |
| Gain_<br>mech      | Exp   | -.806                  | 1.273 | 1.870                        | .036*          | -.082        | 1.804              | .497           |
|                    | Ctrl  | -1.667                 | 1.174 | (28)                         |                |              |                    |                |

*Note.* cont&dev = content and development; org&coh = organization and coherence; luse&gram = language use and grammar; voc = vocabulary and word choice; mech = mechanics; Exp = Experimental; Ctrl = Control

Effect sizes (Hedges'  $g$ ) were also calculated. Mechanics showed a medium effect ( $g = 0.678$ ), indicating a meaningfully smaller decline in the experimental group than in the control group. Organization and Coherence showed a small-to-medium effect ( $g = 0.597$ ), consistent with its marginal  $p$ -value. The effects for Content and Development ( $g = 0.158$ ), Grammar ( $g = 0.228$ ), and Vocabulary ( $g = 0.033$ ) were small, suggesting minimal group differences for these dimensions. These effect sizes reinforce the quantitative pattern: structured ChatGPT use most strongly supported surface-level accuracy and modestly improved text flow, with limited effects on content development, grammatical accuracy, or vocabulary growth.

## 4.2 Qualitative Data Analysis

Survey and open-ended responses from the experimental and control groups were summarized to describe students' perceptions of confidence, ease of writing, areas needing improvement, academic integrity awareness, and experiences with AI tools. The main descriptive patterns are presented below.

#### **4.2.1 Confidence in writing**

In the experimental group (n = 18), most students reported feeling “a little confident” (50.0%) or “neutral” (38.9%), with very few selecting “confident” (5.6%) or “not confident” (5.6%). The control group (n = 12) showed a broadly similar distribution, with most students selecting “a little confident” or “neutral”; however, a higher proportion of control-group students (75.0%) reported being “confident” or “a little confident” compared with the experimental group (55.6%).

#### **4.2.2 Perceived ease in writing**

Experimental-group responses most frequently mentioned organizing or structuring essays (44.4%), followed by grammar or sentence construction (22.2%) and idea generation (16.7%). In the control group, students most commonly cited organizing or structuring their writing (41.7%) or explaining examples (33.3%). A small number indicated no specific improvement (16.7%).

#### **4.2.3 Areas needing improvement**

Among experimental-group students, vocabulary and grammar were the most frequently reported areas needing further improvement (77.8%). Control-group students mentioned a wider range of needs, including vocabulary improvement (33.3%), organization or structure (33.3%), and essay writing technique (16.7%).

#### **4.2.4 Awareness of academic integrity**

Most students in both groups reported learning more about academic integrity since the pretest (experimental: 61.1% “yes,” 38.9% “a little”; control: 66.7% “yes,” 33.3% “a little”). No students in either group selected “no.”

#### **4.2.5 Use of AI tools and perceived impact (experimental group only)**

When asked how they used AI tools, students most commonly reported idea generation (44.4%) and grammar or spelling checks (38.9%). Regarding perceived impact, students identified grammar/spelling correction (33.3%), idea generation or feedback (27.8%), and vocabulary improvement (16.7%). Four students (22.2%) indicated AI tools did not help improve their writing skills.

#### **4.2.6 Suggested classroom guidelines and additional comments**

Experimental-group students most frequently suggested guidelines related to

plagiarism prevention (38.9%), learning independence or reducing over-reliance on AI (22.2%), idea generation (22.2%), and acceptable grammar support (11.1%). Most students provided no additional comments (83.3%). A small number suggested banning ChatGPT, sharing brainstorm ideas, or clarifying allowed vs. not-allowed AI usage. Control-group comments were limited, consisting mainly of general expressions of appreciation (41.7%) or no additional remarks (58.3%).

## 5. Findings

### 5.1 Writing performance (RQ1)

Independent-samples *t*-tests on gain scores revealed a partial advantage for the ChatGPT-supported section. Mechanics showed a statistically significant difference favoring the experimental group,  $t(28) = 1.870$ ,  $p = .036$  (one-tailed), with students in the ChatGPT section demonstrating a larger reduction in mechanical errors ( $M = -0.806$ ) than those in the control group ( $M = -1.667$ ). Organization and Coherence also showed a near-significant trend,  $t(28) = 1.645$ ,  $p = .056$ . Differences in Content and Development, Grammar, and Vocabulary were small and non-significant (all  $ps > .267$ ). Although the end-of-semester writing task placed greater organizational and linguistic demands on students, the use of gain scores allowed comparison of relative change between groups under identical task conditions, making it possible to examine whether structured ChatGPT use mitigated performance decline.

These patterns partly align with prior research showing that GenAI tools tend to improve lower-level accuracy and text flow. Mahapatra (2024) reported gains in phrasing and sentence-level correctness after integrating ChatGPT, and Deep et al. (2025) found improvements in performance and engagement in ESL writing courses. Marzuki et al. (2023) likewise noted that AI tools often support idea expansion and basic organizational clarity. In the present study, these benefits were most visible in Mechanics, consistent with ChatGPT's use as a quick-check tutor for punctuation and error detection.

Effect-size estimates strengthen this interpretation. Mechanics showed a medium effect (Hedges'  $g = 0.678$ ), indicating a meaningful educational impact. Organization and Coherence demonstrated a small-to-medium effect ( $g = 0.597$ ), suggesting moderate support for structuring and transitions, which aligns with students' reflections that planning felt easier. In contrast, effect sizes for Grammar ( $g = 0.228$ ), Vocabulary ( $g = 0.033$ ), and Content and Development ( $g = 0.158$ ) were small, suggesting limited influence on higher-order linguistic development when ChatGPT is used in a constrained, tutor-style manner. This pattern also mirrors findings from hybrid human-

AI feedback models (e.g., Zhang et al., 2025), which report stronger gains in reasoning and organization when human feedback remains central.

Taken together, the combined p-values and effect sizes provide partial support for RQ1. Structured ChatGPT use improved mechanical accuracy and moderately supported text flow, but did not lead to notable gains in grammatical control, vocabulary development, or content generation. These results are consistent with the view that ChatGPT functions most effectively as a scaffolding tool for surface-level accuracy rather than as a comprehensive writing enhancer.

## 5.2 Writing confidence (RQ2)

Although students in both sections reported modest confidence at the end of the course, the overall pattern indicates that ChatGPT made writing tasks feel somewhat more manageable without substantially increasing self-efficacy. In the experimental group, most students selected “a little confident” or “neutral”, and only one reported feeling “confident”. While both groups most frequently selected “a little confident” or “neutral,” the control group reported a noticeably higher proportion of confident responses, suggesting that access to ChatGPT did not inflate students’ perceived abilities or lead to overconfidence.

Students’ explanations provide additional insight. Experimental students most frequently reported that organizing and structuring their writing felt easier, followed by producing sentences and checking grammar, whereas control students similarly highlighted organizing ideas and explaining examples. These perceptions echo findings from prior studies suggesting that AI assistance can make writing tasks feel less intimidating or more manageable, without necessarily producing stronger gains in writing self-efficacy. For example, Song & Song (2023) and Deep et al. (2025) report increases in motivation and perceived ease of writing when AI tools are incorporated, yet Teng (2025) emphasizes that durable growth in self-efficacy requires active metacognitive engagement rather than passive reliance on AI feedback.

Our results align with this view. While ChatGPT helped students generate ideas, clarify sentences, and organize their writing, many still felt they needed to strengthen vocabulary and grammar. This suggests that the tool reduced cognitive load and provided reassurance, but did not yet cultivate deeper confidence in writing ability. Overall, the qualitative evidence provides partial support for RQ2: ChatGPT was perceived as making the writing process feel more manageable for students in the experimental group, but it did not substantially elevate students’ self-reported

confidence.

### **5.3 Academic-integrity awareness (RQ3)**

Both groups reported increased awareness of academic integrity by the end of the course, with every student in the experimental group and all students in the control group indicating that they had learned “yes” or “a little” since the pretest. This widespread growth suggests that integrity-related mini-lessons, classroom reminders, and assignment policies had a consistent influence across sections.

However, there were notable differences in how students in the ChatGPT group interpreted and evaluated AI-related ethical issues. Most experimental group students reported only neutral confidence in using AI ethically, and the remainder selected a little confident. None reported feeling fully confident. Their accompanying comments frequently requested clearer boundaries regarding plagiarism prevention, acceptable levels of AI assistance, and limits on over-reliance. Several explicitly mentioned uncertainty about appropriate paraphrasing or how much correction support from ChatGPT was permissible.

This pattern of higher awareness paired with cautious self-assessment, aligns with recent concerns in the literature. Cotton et al. (2023), Niloy et al. (2024), and Bittle and El-Gayar (2025) emphasize that generative AI can blur the boundary between assistance and inappropriate substitution, especially when students are unsure how much rephrasing or idea generation constitutes unethical use. At the same time, our findings resonate with classroom-based recommendations, such as Massoud & Zhang (2025), which argue that AI should be framed as a supporting tool rather than a source of text, and that explicit guidelines are essential to prevent misuse.

Importantly, no students in the experimental group reported engaging in inappropriate copying or deliberate plagiarism. Their concerns focused not on temptation, but on ambiguity: they worried about inadvertently crossing ethical boundaries. This distinction is meaningful. It suggests that the study’s instructional approach requiring original wording, prohibiting copy-paste use, and checking drafts for authenticity was effective in promoting responsible attitudes, yet more detailed guidance is necessary for students to feel confident in their ethical decision-making.

Therefore, the results provide partial support for RQ3, indicating increased awareness of academic integrity across both sections rather than a ChatGPT-specific effect. Students demonstrated improved understanding of academic-integrity expectations, but also articulated ongoing uncertainty about ethical AI use, highlighting

the need for explicit, practical AI-literacy instruction linked to paraphrasing, attribution, and acceptable forms of AI assistance.

#### **5.4 Reported benefits and challenges of ChatGPT (RQ4)**

Students in the experimental section reported several practical benefits of ChatGPT during their writing process, most commonly using it for idea generation and quick grammar or spelling checks. No student selected “structuring my essay” as a direct use category, but qualitative comments suggest that structural support may have been experienced indirectly for example, through clarifying model sentences, better transitions, or suggestions that improved paragraph flow. A small number of students indicated that they did not use AI tools at all, typically because they felt unsure about how to use them effectively or preferred to rely on familiar methods.

When asked whether ChatGPT improved their writing skills, students most frequently highlighted rapid corrective feedback, support in generating ideas, and vocabulary learning. These perceived benefits closely align with earlier reports that EFL learners value AI tools for brainstorming and error checking (Mahapatra, 2024; Kim et al., 2024; Burkhard, 2022). However, a notable minority of students felt that ChatGPT did not improve their writing, either because they used it sparingly or because the feedback felt too general to contribute to meaningful progress.

The challenges identified by students mirror concerns in previous research as well. Several experimental students reported uncertainty about grammar accuracy, vocabulary choice, and paraphrasing, indicating that ChatGPT’s suggestions were sometimes helpful but not fully trustworthy. This mixed perception is consistent with Mali’s (2025) systematic review, which notes that while GenAI tools often enhance fluency and confidence, learners remain cautious about over-reliance and the quality of AI-generated corrections. In our study, a few students explicitly suggested establishing clearer boundaries between acceptable and prohibited uses of AI, reflecting broader concerns about ethical use and academic integrity highlighted in Cotton et al. (2023) and Bittle & El-Gayar (2025).

Taken together, the findings provide support for RQ4. Students perceived ChatGPT as a useful but imperfect tutor that is helpful for brainstorming and surface-level accuracy, yet insufficient on its own for developing more advanced grammatical control or nuanced lexical choice. These perceptions reinforce the principle that AI-assisted writing is most effective when paired with explicit instruction, careful task design, and clear classroom guidelines to help students understand how to use AI meaningfully and

ethically.

## **6. Discussion**

### **6.1 Interpretation of writing outcomes**

The performance data suggest that structured access to ChatGPT improved surface accuracy and in part improved organization. The Mechanics indicates that students used the tool effectively for punctuation, and error spotting, which aligns with how ChatGPT was positioned in class (as a tutor for quick checks and clarifications rather than a text generator). The near-significant advantage in Organization and Coherence points to better flow, linking, and paragraph order in the AI-supported section, consistent with students' reports that structuring felt easier. Unlike Mechanics, Grammar did not differ significantly between groups, which implies that more advanced sentence accuracy still requires explicit teaching and practice rather than relying mainly on AI. Content and Vocabulary gains were similar across groups, further underlining that higher-level development and lexical growth did not automatically follow from AI access. Taken together, the pattern is consistent with a scaffolding effect: ChatGPT appears most helpful for mechanical correctness and text flow, with limited impact on deeper grammatical control unless this is directly taught.

### **6.2 Confidence and learner experience**

Confidence levels were moderate in both groups, with “a little confident” the most common response. The absence of “very confident” ratings in the experimental section suggests that AI access did not inflate students' perceived ability. Students appreciated idea generation and error feedback, yet still felt they needed more vocabulary and grammar, a perception that mirrors the quantitative results and reinforces the need for continued language-form instruction alongside AI use. In terms of Organization and Coherence many students reported that structure felt easier, no one selected “structure support” as a specific AI use category. This suggests that students may experience structural help indirectly, for example, through model sentences, transitions, or clarifying questions without explicitly labeling this as “structure.” Making these links more visible in instruction could help students better understand what aspects of their writing are being supported. One possible explanation for the control group's reported ease with organizing or structuring their writing and explaining examples is the use of peer feedback. Through reviewing classmates' drafts and receiving comments on their own writing, students may have become more aware of what readers found unclear or

underdeveloped, which could have supported improvements in idea organization and elaboration. This reader-focused awareness, developed through peer interaction, may account for some of the perceived ease reported by control-group students despite the absence of AI support.

### **6.3 Academic integrity and AI literacy**

Both groups reported learning more about academic integrity, and experimental students in particular expressed a desire for clear guidelines on plagiarism prevention and acceptable AI use. At the same time, most experimental students reported neutral confidence in using AI ethically, which indicates caution rather than overconfidence. This combination of heightened awareness and cautious self-assessment supports the approach taken in class: ChatGPT was framed as a tutor, original wording was required, and explicit classroom rules were in place. These findings also point to the value of continued AI-literacy activities that move beyond general warnings. Tasks such as paraphrasing practice with source checks, side-by-side comparisons of citations, and explicit reflection on how AI output was used could help students develop justified confidence in their ethical use of AI tools.

### **6.4 Pedagogical implications**

1. The advantages in Mechanics and Organization emerged under clear constraints (no copy-paste of AI text, English prompts, teacher monitoring). Maintaining these norms seems important if ChatGPT is to function as a scaffold rather than a text-replacement tool.
2. Students' self-identified needs and the grammar results suggest that explicit instruction and practice in grammar and vocabulary should continue. AI can provide feedback and examples, but it should not replace form-focused teaching.
3. Because students did not label "structure" as an AI use category, instructors might model prompts that target outlining, topic sentences, and paragraph logic. Short reflection logs could ask students to write which prompts helped their structure, making the link between AI use and organization more visible.
4. Given neutral confidence in ethical AI use and repeated requests for guidance, courses should include brief, recurring activities focused on paraphrase quality, citation checking, proper acknowledgment of AI assistance, and discussion of acceptable versus unacceptable support.

## 6.5 Limitations and directions for future research

This study has several limitations that should be considered when interpreting the findings. First, the sample size was small and drawn from a single institution, which limits the generalizability of the results. The pretest and posttest tasks also differed in genre (a paragraph in Semester 1 and a timed short multi-paragraph opinion essay in Semester 2). Although we attempted to mitigate this issue by focusing on rubric dimensions that apply across genres (Organization and Coherence, Grammar, and Mechanics) and by analyzing gain scores, differences in task demands may still have influenced performance.

Second, the class sections were assigned at the class level rather than through individual randomization, and the analyses used individual students as the unit of analysis. Because only two intact classes were available, it was not possible to adjust for potential class-level clustering, and the effect sizes should therefore be interpreted with caution.

Finally, the study did not track how individual students actually used ChatGPT outside of class, nor did it examine long-term transfer to other courses or genres. Future research could incorporate prompt logs, compare AI-only feedback with combined human-AI feedback, or conduct longitudinal qualitative studies across multiple institutions to better understand how sustained, guided use of generative AI shapes writing development, confidence, and ethical awareness over time.

## Conclusion

This two-semester classroom-based study found that structured access to ChatGPT supported measurable writing gains in the areas where the tool was most actively used. The experimental section outperformed the control group on Mechanics and showed a marginal advantage in Organization and Coherence, while grammar gains were small and similar across groups. Student reflections echoed these results: many experimental students reported that planning and structuring felt easier, yet they continued to perceive vocabulary and grammar as areas requiring improvement. Both groups reported learning more about academic integrity. In the experimental section, most students expressed neutral confidence in using AI ethically and requested clearer guidelines for appropriate use.

Overall, the findings suggest that positioning ChatGPT as a tutor can help improve surface accuracy and text flow over time, particularly for planning, transitions, and quick error checks, but it does not automatically strengthen deeper grammatical

control or vocabulary development. Effective classroom practice may therefore combine short, level-appropriate AI prompts with explicit language-form instruction, regular paraphrasing and citation checks, and teacher monitoring to ensure that final wording remains in the student's own voice. This study is limited by its small sample size, single institutional context, and the shift from paragraph to timed short multi-paragraph opinion essay writing across semesters. Future research should involve larger samples, delayed posttests, and broader outcome measures (e.g., paraphrasing, summarizing, citation), and should compare AI-only feedback with blended human-AI feedback.

## Notes

- <sup>1</sup> The teacher journal was a brief weekly log kept by the instructor, noting the lesson focus, permitted ChatGPT prompts in the experimental section, and parallel non-AI activities in the control section. It was used to contextualize the quantitative findings but was not subjected to formal qualitative analysis.

## References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.
- Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, 75, 102370. <https://doi.org/10.1016/j.techsoc.2023.102370>
- Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. <https://doi.org/10.3390/info16040296>
- Burkhard, M. (2022). Student perceptions of AI-powered writing tools: Towards individualized teaching strategies. Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Educational Technologies, Lisbon, Portugal. <https://eric.ed.gov/?id=ED626893>
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2023.2190148>
- Deep, P. D., Martirosyan, N., Ghosh, N., & Rahaman, M. S. (2025). ChatGPT in ESL higher education: Enhancing writing, engagement, and learning outcomes. *Information*, 16(4), 316. <https://doi.org/10.3390/info16040316>
- Folse, K. S., Muchmore-Vokoun, A., & Solomon, E. V. (2025). *Great writing 2* (5th ed.). National Geographic Learning.

- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A Multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, 8(6), 1-44. <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- Heift, T. (2021). Intelligent computer-assisted language learning. In H. Mohebbi & C. Coombe (Eds.), *Research questions in language education and applied linguistics* (pp. 655-658). Springer. [https://doi.org/10.1007/978-3-030-79143-8\\_114](https://doi.org/10.1007/978-3-030-79143-8_114)
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112-131. [https://doi.org/10.30191/ETS.202301\\_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Iqbal, N. (2023). *A comparison of the linguistic encoding of artificial-intelligence-generated academic essays and academic essays written by MSc ALSLA 2022-23 students* [Master's thesis, University of Oxford]. [https://ora.ox.ac.uk/objects/uuid:ba64e3ad\\_a522-429d8578-a3e32ae79b37/files/dm039k554z](https://ora.ox.ac.uk/objects/uuid:ba64e3ad_a522-429d8578-a3e32ae79b37/files/dm039k554z)
- Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12878-7>
- Khatun, N., & Miwa, J. (2016). An autonomous learning system of Bengali characters using web-based intelligent handwriting recognition. *Journal of Education and Learning*, 5(3), 122-138.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Kukulska-Hulme, A. (2013). Mobile-Assisted Language Learning. *The Encyclopedia of Applied Linguistics* (C. A. Chapelle, Ed.). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405198431.wbeal0768>
- Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2025.2498537>
- Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445. <https://doi.org/10.1016/j.caeai.2025.100445>
- Lang, J. M. (2013). *Cheating lessons*. Harvard University Press.
- Liang, J., Hwang, G., Chen, M. A., & Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2021.1958348>

- Mali, Y. C. G. (2025). Exploring the use of ChatGPT in EFL/ESL writing classrooms: A systematic review. *Journal of Language and Education, 11*(2), 137-156. <https://doi.org/10.17323/jle.2025.21793>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(9). <https://doi.org/10.1186/s40561-024-00295-9>
- Massoud, O. S., & Zhang, J. (2025). Evaluating the impact of ChatGPT on ESL students' perceptions of writing skills and academic integrity. *Sophia University Junior College Division Faculty Journal, 46*, 33-54.
- Marr, B. (2024, April 3). The top 10 limitations of ChatGPT. *Forbes*. <https://www.forbes.com/sites/bernardmarr/2023/03/03/the-top-10-limitations-of-chatgpt/?sh=54b3c5ea8f35>
- Marzuki, S., Widiati, U., Rusdin, D., Darwin, R., & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- Niloy, A. C., Hafiz, R., Hossain, B. M. T., Gulmeher, F., Sultana, N., Islam, K. F., Bushra, F., Islam, S., Hoque, S. I., Rahman, M. A., & Kabir, S. (2024). AI chatbots: A disguised enemy for academic integrity? *International Journal of Educational Research Open, 7*, 100396. <https://doi.org/10.1016/j.ijedro.2024.100396>
- Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly, 19*(2), 139-158. <https://doi.org/10.1080/10573560390143085>
- Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education, 7*(3), 135-153.
- Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology, 105*(4), 1010-1025. <https://doi.org/10.1037/a0032340>
- Rüschhoff, B., & Ritter, M. (2001). Technology-enhanced language learning: Construction of knowledge and template-based learning in the foreign language classroom. *Computer Assisted Language Learning, 14*(3-4), 219-232. <https://doi.org/10.1076/call.14.3.219.5789>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843. <https://doi.org/10.3389/>

fpsyg.2023.1260843

- Stanford University. (2024, March 1). *Generative AI policy guidance*. <https://communitystandards.stanford.edu/generative-ai-policy-guidance>.
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20, 35. <https://doi.org/10.1186/s41239-023-00404-7>
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), <https://doi.org/10.37074/jalt.2023.6.1.17>
- Teng, M. F. (2025). Understanding EFL student writers' metacognitive awareness in utilizing ChatGPT. *System*, 135, 103848. <https://doi.org/10.1016/j.system.2025.103848>
- University of Oxford. (2025, September 22). *Guidance on safe and responsible use of Gen AI tools*. <https://www.ox.ac.uk/students/life/it/guidance-safe-and-responsible-use-gen-ai-tools>.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*. Advance online publication. <https://doi.org/10.1080/17501229.2025.2503890>

## Appendix A: Posttest Control Group Survey

### Background items (short answer)

1. 年齢 / Age
2. 性別 / Gender
3. 専攻・研究分野 / Major or field of study
4. 学年 / Year in school

### Essay prompt shown with the survey

Do you think technology has improved the way people communicate? Why or why not?

### Perception items

5. このエッセイを書くとき、最初のエッセイと比べてどのくらい自信がありましたか？ / How confident did you feel while writing this essay compared to the first one?  
Options: Not confident / A little confident / Neutral / Confident / Very confident
6. 今回の方が前より簡単だと感じたことは何ですか（例：アイデアの整理、文法の使用、例の説明など）？ / What did you find easier this time (e.g., organizing ideas, using grammar, explaining examples)?  
Options: Short answer
7. 自分のライティングで、まだ改善が必要だと思う点は何ですか？ / What do you think you still need to improve in your writing?  
Options: Short answer
8. プレテスト以降、学術的誠実性についてより多く学びましたか？ / Have you learned more about academic integrity since the pretest?  
Options: Yes / No / A little
9. 追加のコメントはありますか？ / Do you have any additional comments?  
Options: Short answer

## Appendix B: Posttest Experimental Group Survey

### Background (short answer)

1. 年齢 / Age
2. 性別 / Gender
3. 専攻・研究分野 / Major or field of study
4. 学年 / Year in school

### Essay Prompt

Do you think technology has improved the way people communicate? Why or why not?

### Items and answer options

1. このエッセイを書くとき、最初のエッセイと比べてどのくらい自信がありましたか？ / How confident did you feel while writing this essay compared to the first one?  
Options: Not confident / A little confident / Neutral / Confident / Very confident
2. 今回の方が簡単だと感じたことは何ですか（例：アイデアの整理、文法の使用、例の説明など）？ / What did you find easier this time (e.g., organizing ideas, using grammar, explaining examples)?  
Options: Short answer
3. 自分のライティングで、まだ改善が必要だと思う点は何ですか？ / What do you think you still need to improve in your writing?  
Options: Short answer
4. プレテスト以降、学術的誠実性についてより多く学びましたか？ / Have you learned more about academic integrity since the pretest?  
Options: Yes / No / A little
5. ライティングで AI ツールを倫理的に使用することに自信がありますか？ / Do you feel more confident using AI tools ethically in your writing?  
Options: Not confident / A little confident / Neutral / Confident / Very confident
6. ライティング練習で AI ツールをどのように使用しましたか（使用していれば）？ / How did you use AI tools (if at all) in your writing practice?  
Options: Check all that apply (To generate ideas / For grammar and spelling checks / To structure my essay / I did not use AI tools / Other)
7. AI ツールはライティングスキルの向上に役立ったと思いますか？その理由は？ / Do you think AI tools helped you improve your writing skills? Why or why not?  
Options: Short answer
8. 教師は教室での ChatGPT の使用に関するガイドラインを設定すべきだと思いますか？その場合、どの分野を扱うべきだと思いますか？ / Do you think teachers should set guidelines for using ChatGPT in the classroom? If so, which areas should be addressed?  
Options: Yes / No / Not sure + Short answer (areas)
9. ChatGPT や他の AI ツールの教室での使用方法に関する追加のコメントや提案はありますか？ / Do you have any additional comments or suggestions about how ChatGPT or other AI tools should be used in the classroom?  
Options: Short answer

## Appendix C: Pretest Survey- Experimental and Controlled Group Survey

### Demographics (短答 / Short answer)

1. 年齢 / Age
2. 性別 / Gender
3. 専攻・研究分野 / Major or field of study
4. 学年 / Year in school

### Academic integrity & AI (bilingual items with options)

6. あなたの所属する機関の学術的誠実性ポリシーにどの程度詳しいですか？ / How familiar are you with your institution's academic-integrity policy?  
(Not familiar / A little familiar / Somewhat familiar / Familiar / Very familiar)
7. あなたの教育において、学術的誠実性はどれほど重要だと思いますか？ / How important is academic integrity in your education?  
(Not important / A little important / Neutral / Important / Very important)
8. ライティングの課題に AI ツールを使用したことがありますか？ / Have you used AI tools for writing assignments?  
(Yes / No)
9. (7が Yes の場合) AI ツールの使用目的 (複数選択可) / Purposes used (check all that apply)  
(Brainstorming / Outlining / Grammar & spelling / Paraphrasing & summarizing / Vocabulary & phrasing / Translation & clarification / Citation & formatting / Other)
10. ライティングで AI を使うことは違反につながると感じますか？ / Can using AI for writing lead to violations?  
(Yes / No / It depends)
11. 倫理的に AI を使える自信は？ / Confidence using AI ethically  
(Not confident / A little confident / Neutral / Confident / Very confident)
12. AI 使用はポリシーで規制すべきか？ / Should policy regulate AI use?  
(Yes / No / Not sure)
13. 倫理的使用の指導を受けたことがあるか？ / Received instruction on ethical AI use  
(Yes / No)
14. ガイドラインやワークショップは役立つか？ / Would guidelines or workshops help?  
(Yes / No / Maybe)

### Pretest Essay Prompt

15. *Should people spend more time outdoors instead of on computers and TVs? Why or why not?*  
(英語で回答 / Answer in English)