HNSJ Humanities & Natural Sciences Journal
ISSN: (e) 2709-0833
www.hnjournal.net

Peer-Reviewed Journal
Indexed at: NSP - SJIF
Impact Factor 2022 = 4.91

Effectiveness of Generative Artificial Intelligence Models in Improving University Students' Academic Writing Skills

Dr. Mohammed Ali Badawi Al-Awad[1]; Khabir Othman Mohamed Badawi[2]; Elromisa Hafez Khalaf Alaah Babikir[1]; Elaf Abdalhadi Abdalla Ebrahim[1]

[1] Assistant Professor , Nile Valley University, Sudan.
[2] Taibah College of Sciences & Technology, Sudan.

HNSJ, 2026, 7(11); https://doi.org/10.53796/hnsj711/7

Received at 05/10/2026 **Accepted at 12/10/2026** **Published at 01/11/2026**

Abstract

Recent years have witnessed rapid advances in generative artificial intelligence (AI) models such as ChatGPT and Gemini, opening new prospects for supporting the educational process. This study aims to measure the effectiveness of generative AI models in improving university students' academic writing skills, using a quasi-experimental design with experimental and control groups (N = 60 second-year students of the Faculty of Education, randomly assigned to the two groups). The experimental group received a six-week training program that integrated generative AI as a writing assistant across brainstorming, drafting, language revision, and referencing, while the control group followed traditional instruction. An academic writing skills test and an observation checklist were used as instruments. The study hypothesizes statistically significant differences ($\alpha = 0.05$) favoring the experimental group in thesis clarity, idea organization, linguistic accuracy, and academic referencing, alongside positive attitudes toward AI as a supportive tool rather than a substitute for thinking. In addition, an exploratory descriptive survey of 87 university students found that 93.1% had previously used generative AI tools for academic writing, 95.4% supported clear university policies governing their use, and 72.4% reported concern about over-reliance on them. The study recommends training faculty members to integrate generative AI tools into academic writing courses and establishing an ethical charter to govern their use.

Key Words: Generative Artificial Intelligence; Academic Writing; University Students; ChatGPT; Adaptive Learning.

فاعلية نماذج الذكاء الاصطناعي التوليدي في تحسين مهارات الكتابة الأكاديمية لدى طلبة الجامعات

د. محمد علي بدوي العوض[1]؛ خبير عثمان محمد بدوي[2]؛ الرميصاء حافظ خلف الله بابكر[1]؛ إيلاف عبدالهادي عبدالله إبراهيم[1]

[1] أستاذ مساعد، جامعة وادي النيل، السودان.
[2] كلية طيبة للعلوم والتكنولوجيا، السودان.

HNSJ, 2026, 7(11); https://doi.org/10.53796/hnsj711/7

تاريخ الاستقبال: 2026/10/05م تاريخ القبول: 2026/10/12م تاريخ النشر: 2026/11/01م

المستخلص

شهدت السنوات الأخيرة تطورًا متسارعًا في نماذج الذكاء الاصطناعي التوليدي، مثل ChatGPT وGemini، مما أتاح آفاقًا جديدة لدعم العملية التعليمية. تهدف هذه الدراسة إلى قياس فاعلية هذه النماذج في تحسين مهارات الكتابة الأكاديمية لدى طلبة الجامعات، باستخدام تصميم شبه تجريبي بمجموعتين تجريبية وضابطة، تشملان ٦٠ طالبًا من السنة الثانية بكلية التربية، وُزّعوا عشوائيًا على المجموعتين. تلقت المجموعة التجريبية برنامجًا تدريبيًا مدته ستة أسابيع، يدمج الذكاء الاصطناعي التوليدي بوصفه مساعدًا للكتابة في مراحل توليد الأفكار، وإعداد المسودة، والمراجعة اللغوية، والتوثيق، بينما اتبعت المجموعة الضابطة التدريس التقليدي. واستُخدم اختبار لمهارات الكتابة الأكاديمية وبطاقة ملاحظة أداتين للدراسة. وتفترض الدراسة وجود فروق دالة إحصائيًا عند مستوى ٠,٠٥ لصالح المجموعة التجريبية في وضوح الأطروحة، وتنظيم الأفكار، والدقة اللغوية، والتوثيق الأكاديمي، إلى جانب اتجاهات إيجابية نحو الذكاء الاصطناعي بوصفه أداة داعمة لا بديلًا عن التفكير. كما أظهرت استبانة وصفية استطلاعية شملت ٨٧ طالبًا جامعيًا أن ٩٣,١٪ سبق لهم استخدام أدوات الذكاء الاصطناعي التوليدي في الكتابة الأكاديمية، وأن ٩٥,٤٪ يؤيدون وضع سياسات جامعية واضحة لتنظيم استخدامها، بينما أبدى ٧٢,٤٪ قلقًا من الإفراط في الاعتماد عليها. وتوصي الدراسة بتدريب أعضاء هيئة التدريس على دمج هذه الأدوات في مقررات الكتابة الأكاديمية، ووضع ميثاق أخلاقي ينظم استخدامها.

الكلمات المفتاحية: الذكاء الاصطناعي التوليدي؛ الكتابة الأكاديمية؛ طلبة الجامعات؛ ChatGPT؛ التعلم التكيفي.

1. Introduction

Academic writing is among the most important skills expected of a university student, as it translates the student's capacity for critical thinking, analysis, and idea organization into scholarly texts that are methodologically and linguistically disciplined. Nevertheless, Arab and international educational studies alike continue to report a marked decline in this skill among a large proportion of university students, whether in the clarity of the thesis statement and the coherence of the argument, or in linguistic accuracy and adherence to recognized standards of scholarly referencing.

At the same time, recent years—particularly since the wide commercial release of large language models at the end of 2022—have seen an unprecedented spread of generative AI tools among university students, who now use them spontaneously across a variety of writing tasks, without structured pedagogical guidance that would ensure their sound use or limit their potential risks to intellectual independence and academic integrity.

Hence the importance of moving from ad hoc use of these tools to deploying them within a pedagogically designed training program that integrates generative AI into the successive stages of the writing process—brainstorming, drafting, revision, and referencing—as a cognitive assistant that develops the skill rather than a substitute that replaces it. This study responds to that challenge through a quasi-experimental design that measures the actual effect of such deployment on the academic writing skills of a sample of Faculty of Education students.

2. Theoretical Framework and Previous Studies

2.1 The Concept of Generative Artificial Intelligence

Generative AI refers to the branch of artificial intelligence based on large language models (LLMs), which are trained on vast amounts of text so that they can generate new content—textual or conversational—in response to user input (prompts). Models such as ChatGPT and Gemini are the most prominent examples of such systems, which for the first time offered the general public, including university students, near-human conversational text interaction.

2.2 Theoretical Grounding for Integrating AI into Writing Instruction

This study rests on a composite theoretical framework that combines social constructivism, which views learning as the product of the learner's interaction with tools and cognitive mediators that can serve as temporary "scaffolding" until the learner attains independence, and the adaptive learning model, which holds that tailoring support to the learner's level and pace enhances learning effectiveness. Within this framework, generative AI models can be viewed as a "digital scaffold" that offers immediate, personalized feedback at every stage of writing, provided that the student remains the decision-maker in the thinking and in the final wording of the text.

2.3 Previous Studies

A growing body of literature—particularly between 2023 and 2025—has examined the relationship between generative AI and higher education from multiple angles. Kasneci et al. (2023) discussed the educational opportunities and challenges of using large language models in education, pointing to their potential to support personalized learning and to generate explanations of concepts, as against risks related to academic integrity, bias, and the quality of outputs; they concluded that teachers and students need to be prepared for the responsible use of these tools.

Cotton et al. (2024) focused on the problem of academic integrity in the era of ChatGPT,

showing that the educational benefit of the tool—in stimulating critical thinking and supporting writing—depends on the existence of clear institutional policies that regulate its use and balance its benefits against the need to preserve the originality of students' academic work.

In a study more specific to writing, AlAfnan (2023) examined the impact of ChatGPT on academic writing and concluded that the tool can assist with the linguistic and structural organization of academic texts, while stressing that students must distinguish between using the tool as an aid to editing and revision and relying on it to generate the original ideas of a piece of research.

At the Arab level, Al-Zahrani (2024) addressed the use of generative AI applications in teaching Arabic and noted that the gap between the actual spread of these tools among Arab students and the existence of institutional pedagogical frameworks for using them remains wide, calling for their integration into controlled teaching plans rather than leaving them to ad hoc individual use.

In the same vein, UNESCO (2024) issued guidance on generative AI in education and research that stressed a human-centred approach and the importance of building institutional policies that safeguard privacy and digital equity and that clearly define the limits of acceptable use for both students and teachers.

Considering this literature as a whole, the researchers note a relative scarcity of Arab experimental studies that quantitatively measure the effect of a structured training program—rather than mere free use—of generative AI tools on academic writing skills specifically. This is the research gap that the present study seeks to help fill.

3. Problem, Objectives, and Research Questions

3.1 Statement of the Problem

Despite the wide spread of generative AI tools among university students, their use generally remains ad hoc and pedagogically unguided, amid evident weakness in academic writing skills among a considerable share of students and an absence of sufficient Arab experimental studies evaluating the effect of integrating these tools within a structured training program on improving those skills. The problem of the study is therefore captured in the following main question:

What is the effectiveness of using generative AI models, within a structured training program, in improving university students' academic writing skills?

3.2 Research Questions

1. Are there statistically significant differences at the 0.05 level between the mean scores of the experimental and control groups on the post-test of academic writing skills as a whole, attributable to the use of generative AI models?
2. Are there statistically significant differences between the mean scores of the two groups on each sub-skill separately: thesis clarity, idea organization, linguistic accuracy, and adherence to academic referencing?
3. What are the attitudes of the experimental-group students toward using generative AI as a supporting tool in academic writing?
4. What is the reality of university students' use of generative AI tools in academic writing, how do they self-assess their effect on their writing skills, and what are their concerns and attitudes toward regulating their use institutionally?

3.3 Objectives

The study seeks to: (1) measure the effectiveness of a training program based on the use of generative AI models in improving university students' academic writing skills; (2) identify the program's effect on each sub-skill of academic writing separately; (3) explore students' attitudes toward using these tools as an assistant rather than a substitute for thinking; (4) document the reality of university students' use of these tools in academic writing, and examine their self-assessment of the tools' effect and their concerns and attitudes toward institutional regulation, through an exploratory survey; and (5) offer practical recommendations for faculty members and university decision-makers on the optimal way to integrate these tools.

3.4 Significance of the Study

The study derives its significance from moving the discussion of generative AI in education from general theoretical argument to quantitative empirical verification, and from offering Arab university institutions a replicable procedural model for integrating these tools into academic writing courses, thereby serving the conference track "Artificial Intelligence in Education."

4. Methodology and Procedures

4.1 Research Design

The study adopted a quasi-experimental design with two equivalent groups (experimental and control) and pre- and post-test administration of the measurement instrument. This is the most suitable design for examining the effect of an independent variable that can be implemented within a natural educational setting (existing class sections) in which full random assignment of individuals is difficult. The design was complemented by an exploratory descriptive survey of university students' views (see Section 4.6).

4.2 Population and Sample

The study population consisted of students of Wadi Al-Nil University (Sheikh Abdallah Al-Badri University). A purposive sample of (60) second-year students of the Faculty of Education was drawn for the quasi-experimental study and randomly assigned to two groups equivalent in academic level, with their equivalence verified through the pre-test: an experimental group (n = [number]) that underwent the training program based on generative AI, and a control group (n = [number]) that studied the same topics in the traditional way without the use of AI tools. The exploratory descriptive survey (Section 4.6), by contrast, has an independent sample of (87) participants, who were self-selected from diverse universities and disciplines and are unrelated to the two experimental groups.

4.3 Research Instruments

The study relied on two main instruments: (a) an academic writing skills test, a performance-based achievement test administered before and after the program, which measures four sub-skills: thesis clarity, idea organization, linguistic accuracy, and adherence to academic referencing according to the APA system; and (b) an observation checklist used during the training sessions to record students' behavior while interacting with generative AI models at each stage of writing. Both instruments are to be submitted to a panel of specialized referees to verify face validity and content validity, and their reliability is to be computed using Cronbach's alpha and inter-rater agreement.

4.4 The Training Program

The training program delivered to the experimental group extended over six weeks, with two

sessions per week. During it, generative AI models were integrated as an educational assistant—not as a substitute tool for completing the work—in four successive stages of the writing process:

- **Brainstorming and idea generation:** using the tool to pose probing questions and generate multiple perspectives on the topic.
- **Drafting the first version:** using the tool to suggest organizational structures for the text without writing the text on the student's behalf.
- **Language and style revision:** using the tool to detect linguistic errors and propose stylistic alternatives, with the student critically reviewing each suggestion.
- **Scholarly referencing:** training students to use the tool to check the formulation of citations and references according to APA (7th edition) without blindly accepting its outputs.

4.5 Statistical Methods

Data will be analyzed using the Statistical Package for the Social Sciences (SPSS) by computing means and standard deviations, an independent-samples t-test to compare the two groups, and a paired-samples t-test to measure differences between the pre- and post-test within each group, in addition to computing the effect size (Cohen's d) to determine the practical—and not merely statistical—significance of the differences.

4.6 The Exploratory Descriptive Survey

In addition to the quasi-experimental design, an exploratory descriptive survey was conducted to document the reality of using generative AI models in academic writing among university students and to examine their self-assessment of the tools' effect and their concerns and attitudes toward institutional regulation. The survey comprised (87) responses across (40) columns, including descriptive data, multiple-choice questions, (21) items on a five-point Likert scale, and two open-ended questions. It should be noted that the survey sample differs from the planned quasi-experimental sample (n = 60), that it is self-selected, and that it includes students from various universities and disciplines; its results therefore cannot be generalized.

The data underwent preliminary cleaning that included removing extra spaces and unifying the labels of clearly equivalent answers. Record identifiers, timestamps, e-mail addresses, and any identifying fields were excluded from the analysis and presentation to protect participants' privacy. Likert items were coded from (1) strongly disagree to (5) strongly agree, and the "agreement" percentage was computed by summing the "agree" and "strongly agree" responses over the valid denominator for each item. The analysis was limited to frequencies, percentages, and descriptive means, with no tests of differences or causal inference; the means here are approximate descriptive indicators for ordinal items, not a psychometrically validated scale. That the percentages of multiple-choice questions sum to more than 100% is expected, because participants could select more than one answer.

As for data quality, Likert-item responses were complete for all participants (87), while (85) answered the questions on tools and stages of use, (57) answered the first open-ended question, and (62) answered the challenges question. Inconsistencies were observed between the prior-use question and the questions on frequency and tools, which are detailed in Section 5.1. Open-ended responses were coded thematically in a preliminary way using keywords, allowing a single response to fall under several themes, without independent coders adjudicating.

5. Results

5.1 Results of the Exploratory Survey (Actual Data)

This section presents the results of the exploratory survey described in Section 4.6, which are based on actual data collected from (87) participants. These results are to be read as participants' subjective impressions, not as a direct measure of improvement in writing performance.

First: Sample description. The sample consisted of (87) students, of whom females made up (75.9%) and males (24.1%); age concentrated in two equal groups, 18–22 and 23–27 years, at (47.1%) each (Figure 1, Table 1). Fourth-year students and above made up (56.3%), and (49.4%) reported a "very good" grade-point average. University and major names varied and were entered in different forms; differences by major or university were therefore not interpreted without broader coding clean-up.

Figure 1: Distribution of the survey sample by gender and age (percentages calculated from the total of 87)

Table 1: Characteristics of the exploratory survey sample (n = 87)

Indicator	Count	Percentage
Female	66	75.9%
Male	21	24.1%
Aged 18–22	41	47.1%
Aged 23–27	41	47.1%
Fourth year or above	49	56.3%
"Very good" GPA	43	49.4%

Second: Patterns of use. A total of (81) participants (93.1%) reported having previously used a generative AI tool in academic writing. Reported frequency of use, from (84) valid responses, was distributed among daily use (30 responses; 35.7%), several times a week (30; 35.7%), once a week (3; 3.6%), and rarely (21; 25.0%) (Figure 2). ChatGPT led the tools named (79 of 85 responses; 92.9%), followed by Gemini (60; 70.6%), then Claude (17; 20.0%) and Copilot (15; 17.6%) (Figure 3). As for the writing stages in which the tools are used, the most prominent were generating ideas and themes (53 of 85; 62.4%), organizing and structuring the text (52; 61.2%), and translation (49; 57.6%), while use for referencing and citations was lower (25; 29.4%).

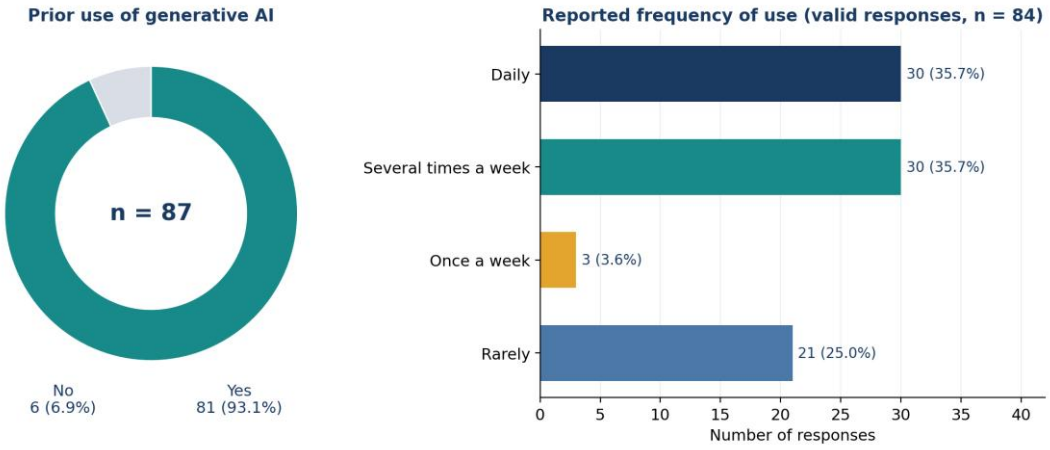

Figure 2: Prior use and reported frequency of use (frequency distribution calculated from 84 valid responses)

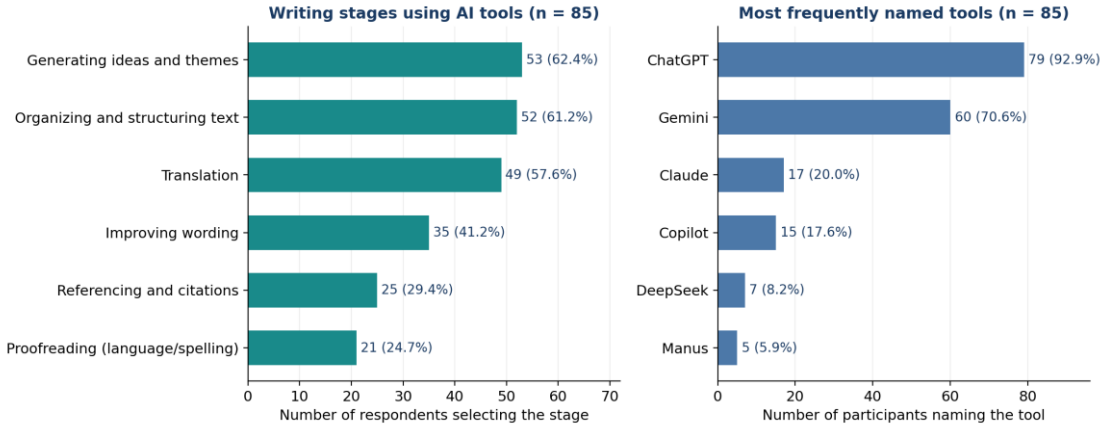

Figure 3: Stages of using the tools and most frequently named tools (multiple-choice question; percentages from valid responses)

It should be noted that the responses are inconsistent across questions: four participants who answered "no" to the prior-use question nevertheless named tools and answered the frequency and stages questions, and one participant answered "yes" without specifying frequency. The frequency distribution was therefore presented on the basis of valid responses as they were given, and it should not be taken as a confirmed distribution among previous users alone.

Regarding motives for use, saving time topped the purposes named (63 of 85; 74.1%), followed by overcoming difficulties with academic language (39; 45.9%) and then improving the quality of writing (34; 40.0%) (Table 2).

Table 2: Reported purposes of using the tools (valid responses, n = 85)

Reported purpose	Named it / valid responses	Percentage
Saving time	63/85	74.1%
Overcoming difficulties with academic language	39/85	45.9%
Improving the quality of writing	34/85	40.0%

Third: Self-assessment of writing skills. Self-assessments were positive in all three domains: the mean for the domain of organizing ideas and building the text and arguments was (3.89 out of 5), for language and style (3.74), and for referencing and citation (3.64) (Figure 4). At the item level, structuring the paper (86.2%), logical organization of ideas

(85.1%), and precision of terminology (82.8%) received the highest agreement among the improvement items, while distinguishing direct from indirect quotation (59.8%) was relatively the weakest (Table 3). These results reflect a perceived benefit in structural aspects more than in the finer aspects of referencing; they are participants' impressions, not a direct measure of improved performance.

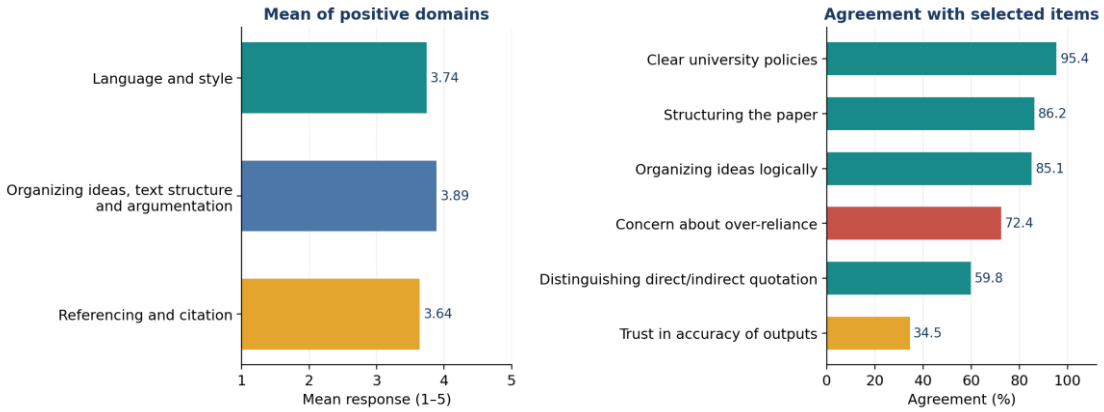

Figure 4: Means of the positive domains and agreement percentages for selected items (concern and trust are separate items, distinct from the domain means)

Table 3: Breakdown of Likert-scale items (n = 87)

Item	Mean (1–5)	Agree / count	Agreement
I. Language and style			
Improved grammatical accuracy	3.69	58/87	66.7%
Precision of terminology	3.91	72/87	82.8%
Clarity and directness of sentences	3.84	66/87	75.9%
Broader vocabulary	3.64	60/87	69.0%
Variety and appropriateness of style	3.74	65/87	74.7%
Reduced lexical repetition	3.63	58/87	66.7%
II. Organizing ideas, text structure, and argumentation			
Logical organization of ideas	4.00	74/87	85.1%
Coherence between paragraphs	3.84	65/87	74.7%
Structuring the paper	4.02	75/87	86.2%
Building academic arguments	3.79	64/87	73.6%
Accuracy of evidence and proof	3.80	67/87	77.0%
Discussing different viewpoints	3.86	68/87	78.2%
III. Referencing and citation			
Citing sources	3.76	66/87	75.9%
Distinguishing direct from indirect quotation	3.56	52/87	59.8%
Fewer citation errors and unintentional plagiarism	3.60	58/87	66.7%
IV. Confidence, concerns, and institutional attitudes			

Item	Mean (1–5)	Agree / count	Agreement
Confidence in the accuracy and quality of outputs	3.11	30/87	34.5%
Concern about over-reliance	3.97	63/87	72.4%
Use raises ethical concerns	3.17	31/87	35.6%
Including the tools in curricula	3.86	66/87	75.9%
Clarity and necessity of university policies	4.26	83/87	95.4%
Feeling that writing skills are declining	3.40	46/87	52.9%

Fourth: Concerns and institutional attitudes. A total of (63 of 87; 72.4%) admitted concern about relying on the tools at the expense of developing their own skills, and (46 of 87; 52.9%) reported feeling that their writing skills had declined, which is a self-report rather than an objective measure of decline. Confidence in the accuracy and quality of outputs was limited (30 of 87; 34.5%), and (31 of 87; 35.6%) agreed that use raises ethical concerns, with a considerable proportion of neutral responses. By contrast, (95.4%) supported establishing clear university policies and (75.9%) supported formally including the tools in curricula.

Fifth: Open-ended responses. The first open-ended question received (57) responses, and the challenges question (62). In the challenges question, the theme of over-reliance, laziness, and weakened thinking or independence emerged in (29) responses (46.8%), followed by the accuracy and reliability of outputs (10; 16.1%), then sources and citation (4; 6.5%) and privacy and security (2; 3.2%), while (6) responses mentioned no concerns or did not know (9.7%) (Figure 5). In the first open-ended question, responses repeatedly referred to organizing ideas and writing (15; 26.3%) and to time, speed, and effort (14; 24.6%). However, the wording of this question in the survey instrument asks about the benefit of writing "without using" AI tools, whereas many responses speak of advantages usually attributed to using them, which suggests ambiguity or a wording error. The coding of these responses is therefore exploratory and should not be taken as specific evidence of the effect of use or non-use.

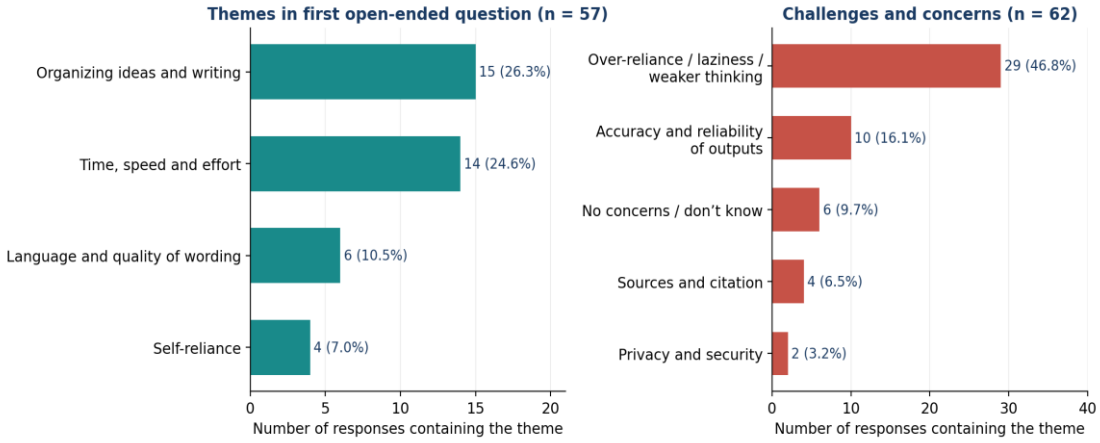

Figure 5: Preliminary thematic coding of open-ended responses (a single response may contain more than one theme)

5.2 Results of the Quasi-Experimental Study (In Progress)

Post-test data for the quasi-experimental study have not yet been collected, so no statistical results are presented in this paper for Questions 1–3. Answers to them are deferred until implementation is complete and the data are analyzed using the methods described in Section 4.5. Question 4 is answered by the survey results in Section 5.1.

6. Discussion of the Results

6.1 Discussion of the Exploratory Survey Results (Actual Data)

The survey results show that using generative AI tools in academic writing has become a common practice among the sample (93.1%), and that a large proportion of respondents use them daily or several times a week, which is consistent with what the study noted in its introduction about the spontaneous spread of these tools. The gap between the prevalence of use and the overwhelming demand for clear policies (95.4%) is consistent with Cotton et al. (2024), who held that the educational benefit of the tool depends on institutional policies that regulate its use; with Al-Zahrani's (2024) observation of the widening gap between actual diffusion and institutional pedagogical frameworks; and with UNESCO's (2024) approach, which rests on clear policies defining acceptable use.

In terms of self-assessment, the highest levels of agreement were for structural tasks (structuring the paper and organizing ideas), and the writing stages most often reported were generating ideas and organizing the text, while the presence of referencing and the distinction between types of quotation were weaker. This accords with AlAfnan's (2023) conclusion that the principal benefit of these tools lies in organizational and structural support for the text. This agreement, however, rests on subjective impressions and does not demonstrate an actual improvement in the quality of writing.

Participants' stance appears dual in practice: they recognize the tools' usefulness in organization, wording, and time-saving, yet they worry about over-reliance (72.4%), more than half (52.9%) feel that their writing skills have declined, and only (34.5%) agreed with the item on trusting the accuracy of outputs. This is consistent with Kasneci et al.'s (2023) warning about risks to the quality of outputs and the need to prepare students for responsible use, and with AlAfnan's (2023) distinction between using the tool as an aid to editing and relying on it to generate original ideas. These results support the need for a training program that makes the tool an assistant rather than a substitute. However, the survey captures free, unguided use; it therefore cannot serve as evidence of the effectiveness of the program itself, a question that only the experiment can settle.

It is noteworthy that the reported motives for use concentrate on saving time (74.1%) more than on improving the quality of writing (40.0%), and that translation was among the most prominent forms of use (57.6%). This suggests that current use leans toward efficiency and quick completion more than toward skill development, which strengthens the need to frame it pedagogically so that the student remains the decision-maker in writing rather than a recipient of ready-made outputs. This is consistent with the open-ended responses, in which the theme of over-reliance, laziness, and weakened thinking or independence predominated (46.8% of responses to the challenges question).

A paradox in the results deserves attention: participants rate the improvement in their skills highly (domain means between 3.64 and 3.89 out of 5) and, at the same time, more than half of them report feeling that those skills have declined (52.9%). The survey data do not allow a definitive interpretation of this paradox; it may reflect a perceived benefit in immediate task completion alongside concern about the longer-term effect of reliance. Verifying it requires comparative analysis across user groups (for example, by frequency of use) or an objective measurement of performance.

These results call for caution in interpretation for several reasons. The sample is self-selected and concentrated in particular groups (females 75.9%, fourth-year students and above 56.3%), with variation in universities and majors, so it cannot be considered representative of all university students. The data are cross-sectional and self-reported, and they establish neither

causality nor actual improvement. The Likert scale is ordinal and its psychometric properties were not tested. Inconsistencies were noted between the prior-use question and the questions on frequency and tools; the forms in which university and major names were written varied; the wording of the first open-ended question is potentially ambiguous; and the coding of open-ended responses is automated and preliminary, and may fail to capture the meaning of some colloquial or ambiguous responses. Future measurement of effect should therefore use a pre-/post-test design or a comparison group, with standardized writing tasks assessed against a published scoring rubric by independent raters.

6.2 Theoretical Interpretation of the Expected Results of the Quasi-Experimental Study (Hypothetical)

Data for the quasi-experimental study have not yet been collected (see Section 5.2). This section therefore offers a theoretical account of what is expected to contribute to improved performance in the experimental group, and it remains a hypothesis to be tested with data later. The expected educational factors are as follows. First, the immediate feedback that generative AI models provide shortens the time gap between the student's production of a text and receipt of improvement-oriented comments, unlike the traditional pattern in which the teacher's marking is delayed by days or weeks. Second, the tool's provision of multiple structural suggestions at the brainstorming stage eases the "blank page anxiety" that hinders many students from starting to write. Third, direct support in scholarly referencing addresses one of the most burdensome aspects of academic writing for novice students, namely mastery of the details of the APA system.

These hypothetical interpretations are consistent with AlAfnan's (2023) conclusion that the principal benefit of generative AI tools lies in editorial and structural support for the text rather than in generating original intellectual content, which explains the importance of the pedagogical design that the present study took care to adopt so that the student remains the decision-maker at every stage. The expected results also intersect with the warnings of Cotton et al. (2024) about the dangers of total reliance on the tool; the expected positive attitude among students toward using the tool as an assistant rather than a substitute is anticipated to stem from the design of the training program, which balances benefiting from the tool with preserving independence of thought, an expectation that needs to be verified with data.

Nevertheless, it should be noted that the limited sample size (60 students) and the short duration of the program (six weeks) impose methodological caution in generalizing the results, which calls for later studies with larger samples and longer training periods to verify whether the effect persists after the program ends.

7. Recommendations and Proposals

In light of the study's theoretical and methodological framework and the results of the exploratory survey, the following practical recommendations can be offered:

5. Train faculty members in methods of integrating generative AI models into academic writing courses, including designing learning tasks that use the tool as an assistant rather than a substitute.
6. Establish a clear university ethical charter regulating students' use of generative AI tools in research and academic work, specifying the limits of acceptable use and the disclosure required.
7. Incorporate training units into university study-skills courses that specifically address the critical, responsible use of AI tools in writing and research, in keeping with the (75.9%) of participants who supported formally including the tools in curricula.

8. Conduct longitudinal studies that measure the long-term effect of such programs and examine their cumulative effect across multiple academic levels and on broader samples from different disciplines.
9. Encourage universities to develop alternative assessment tools that take into account the widespread use of generative AI tools (93.1%) and that focus on the writing process and its stages rather than on the final product alone, instead of settling for traditional prohibition policies, in light of the concentration of motives for use on saving time (74.1%).
10. Train students to verify the accuracy of tools' outputs and to manually check facts, sources, and quotations before adopting them, in light of the limited trust in outputs (34.5%) and the relative weakness in distinguishing quotations (59.8%).
11. Improve survey instruments in later studies by removing the ambiguity in the open-ended benefit question, making the options of the "prior use" and "frequency" questions consistent, specifying a reference time period, and standardizing the names of majors and universities.

8. References (APA 7th Edition)

AlAfnan, M. A. (2023). The impact of ChatGPT on academic writing: Opportunities and challenges. *Journal of Artificial Intelligence and Technology, 3*(2), 60–68.

Al-Zahrani, S. (2024). Employing generative artificial intelligence applications in teaching the Arabic language [in Arabic]. *Journal of Educational and Psychological Sciences*, [volume]([issue]), [pages].

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

UNESCO. (2024). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization.