INNOVATION, TECHNOLOGIES, EDUCATION & COMMUNICATION

INNOVATION, TECHNOLOGIES, EDUCATION ET COMMUNICATION

N° 10

L'INTELLIGENCE ARTIFICIELLE DANS L'ENSEIGNEMENT SUPERIEUR : ENTRE PROMESSES ET MENACES

Coordonnateur : Sabrine Jmad

Responsables :

- Abdessamad Mimoun Elkhir
- Mehdi Kaddouri
- Mohamed Boukare
- Mourad Benali
- Siham Benmessaoud

EDUC
COMMUNICATION, EDUCATION, DIGITAL USAGE AND CREATIVITY LABORATORY
URL CNRST 19

ISSN Numérique : 2737-8195
Novembre 2025

Evaluating English Undergraduate Students' Academic Writing Literacy in Connection with ChatGPT AI Tool: A Pre-, and Post-Assessment Study

Khaoula Fagroud[1], Isam Mrah[2], Mustapha Fagroud[3]

Mohammed the First University, the Faculty of Letters and Human Sciences,

Oujda, Morocco

kfagroud@gmail.com

Abstract: This study investigates the academic writing proficiency of undergraduate English students (Literature major), at the Faculty of Letters and Human Sciences in Oujda, and examines the effect of AI tools, particularly ChatGPT, on their performance. A Pre/Post-Test was administered to those students, with a training session conducted between the tests. The findings indicate that while ChatGPT can positively influence academic writing, its impact is significantly less substantial compared to the training program. The training, with or without ChatGPT, demonstrates long-term benefits, whereas ChatGPT alone showed limited significance. This research demonstrates the critical need for an academic writing training in MA and PhD programs, since it is the ultimate goal of the participants in this research. The study's limitations include the restricted evolution of participants' academic writing skills due to the single training session, which only provided core information and a basic framework for academic writing.

Keywords*:* Academic Writing, AI tools, ChatGPT, Training, Pre-Test, Post-Test.

Introduction

Academic writing literacy has long been a fundamental requirement in higher education. As digital literacy has advanced, particularly with the rise of Artificial Intelligence (AI), there has been a notable intersection between AI tools and the educational field, especially within higher education. According to Zohery [1] (2024), this advancement has brought about significant changes and opportunities that reshape the field

of academic writing and learning processes. AI tools are now increasingly integrated into educational settings, which offer new ways to enhance writing skills and facilitate academic tasks.

Leading this wave of innovation is ChatGPT, an interactive language model with the potential to transform student engagement and academic performance. Launched by OpenAI on November 30, 2022, ChatGPT is a highly advanced pre-trained natural language processing model. It has the ability to understand and interpret the nuances of natural language, generate coherent and contextually appropriate text, answer questions, and participate in conversations based on the given context (Zhai, 2023) [2]. This powerful tool represents a significant leap forward in the integration of AI in education, which in turn offers new possibilities for enhancing learning experiences and academic writing outcomes.

The widespread use of ChatGPT shows the importance of investigating its usefulness in student academic writing. As more students and educators integrate AI tools into learning environments, understanding how ChatGPT can affect writing tasks becomes crucial. Additionally, assessing its impact on student learning outcomes through evaluations becomes crucial for educators to bear in mind.

Similarly, training is another critical factor that declares investigation when assessing its comparative usefulness with ChatGPT. Training serves as a long-term tool to build academic proficiency in writing by providing students with essential skills that contribute to their development over time (Shannon, 2023) [3]. Training often incorporates ongoing feedback mechanisms that allow students to improve their writing skills progressively. Over the long term, consistent training cultivates habits of disciplined writing practice and encourages students to refine their voice and style. This continuous engagement with writing fosters confidence and competence, which prepare students to navigate complex academic writing tasks (Struan, 2021) [4].

Given this background, this thesis posits that undergraduate English Literary students exhibit a low level of academic writing, a challenge that significantly affects their academic performance. This study aims to investigate the potential effects of using ChatGPT on these students' academic writing abilities, as well as the effect of training. By examining the performance of students with and without the intervention of ChatGPT and structured training, this research seeks to provide insights into the effectiveness of AI tools training programs in enhancing academic writing skills. The goal is to determine if these interventions can address the prevalent writing deficiencies observed among these students and foster long-term improvement in their academic performance.

To explore these aspects, the study employs a "test-teach-test" methodology, encompassing a pre-test, training phase, and post-test. This is followed by a targeted training phase. Finally, a post-test evaluates the improvements in their academic writing skills after the training and ChatGPT as interventions. By comparing pre-test and post-test results, the study aims to determine the extent to which structured training and AI tools can mitigate the academic writing level observed in these students. To achieve these objectives, the study addresses the following research questions; Are Undergraduate English students able to write academically and adhere to its rules? And Is there an effect of ChatGPT on students' academic writing skills (RQ2)?

I. Literature Review

Academic writing literacy and digital literacy are integral components of modern education (Strobl and Ali, 2019) [5], with academic writing skills being a paramount for student success. In recent years, the emergence of AI tools like ChatGPT has introduced new dynamics to the academic writing field (Rafi, et al., 2023) [6], which arguably raises questions about their impact on students' writing proficiency and the traditional human-centric approach to academic discourse.

1.1. Defining Academic Writing

According to Irvin (2010) [7], academic writing is a crucial tool for scholars to articulate ideas, analyze research, and construct arguments. In this sense, in a research conducted at George Mason University, Chris Thaiss and Terry Zawacki [8] (2006) identified three key characteristics that define academic writing: Initially, academic writing is characterized by clear evidence of persistence, open-mindedness, and discipline in study. Following this, the dominance of reason over emotions is required. Finally, academic writing is not just about conveying information; it is about engaging with readers who are expected to critically evaluate the arguments presented in the text.

1.1.1. Academic Writing vs. General Writing

This section aims to differentiate between academic writing and general writing to establish clear criteria for each, ensuring that the distinct characteristics of academic writing are well understood.

According to Lee Ann [9] (2002), academic writing goes beyond the surface level of constructing sentences and paragraphs. It involves a higher level of critical thinking, research proficiency, and an understanding of disciplinary concepts, all of which contribute to the development of well-informed and scholarly written work. In contrast, general writing is more versatile in terms of audience, purpose, and style. It may not require the specialized skills and formalities associated with academic writing, which make it accessible to a broader range of readers and purposes.

Apparently, the purpose of academic writing is to eventually target a specific audience and transmit knowledge to a particular group, while general writing lacks a specific target audience and encompasses a wide range of topics that can appeal to anyone. Thus, general writing and academic writing exhibit distinct characteristics and conventions; in fact, academic writing is more demanding and necessitates a higher degree of effort, extensive research, rigorous critical thinking, and adherence to disciplinary conventions compared to general writing.

1.2. Digital Literacies (AI tools such as ChatGPT) in Higher Education

In the context of adult learners, the discussion primarily pertains to higher education, given the academic nature of our discourse. The emergence of AI tools and digital literacy has sparked considerable controversy and unpredictability. While universities have witnessed significant technological advancements and widespread adoption of digital literacies, there is a growing trend wherein AI tools are portrayed as viable replacements for various human roles. These tools are advertised as capable of performing tasks such as searching, correcting mistakes, writing, and problem-solving with ease. For instance, with just a click, one can generate entire presentations on any topic or craft well-structured essays and research papers.

ChatGPT uses deep learning, a subset of machine learning, to produce humanlike text through transformer neural networks. The transformer predicts text -- including the next word, sentence, or paragraph -- based on its training data's typical sequence. However, it also brings some precautions for consideration as AI becomes more ingrained into the university landscape. There exists a potential risk of fostering excessive reliance on them among students, educators, and institutions although AI tools such as ChatGPT detectors and other sophisticated chatbots bring substantial benefits in personalized learning and efficiency.

In the context of higher education, particularly at the university of Applied sciences Kaiserslautern in Germany, ChatGPT emerges as a significant tool with diverse applications. It can be used for text analysis, writing tasks, and tutoring to assist students in understanding complex concepts and skills (Zhai, 2022). Its ability to optimize texts, generate suggestions, summarize longer texts, and even facilitate collaboration among students has revolutionized teaching methodologies and university related services in this context (Cotton et al., 2022) [10].

ChatGPT presents a dichotomy of opportunities and challenges in higher education, which deliver significant advantages to both students and teachers while encounter obstacles in its seamless integration into academic settings. Nevertheless, despite these challenges, ChatGPT maintains its status as a pivotal tool that significantly influences the educational landscape. Looking ahead, our focus will shift to examine its precise impact on academic writing skills within the realm of education.

1.3. ChatGPT and Academic Writing skills in Education

The use of ChatGPT and similar AI tools in academic writing has sparked a viral debate within scholarly circles. On one side, proponents argue that these tools can greatly assist in the writing process, while detractors raise concerns about the potential for plagiarism and the impact on students' development of academic writing skills. This debate investigates the delicate balance between leveraging AI for productivity and creativity and avoiding over-reliance or misuse that may compromise the integrity of academic work.

ChatGPT is widely used in academic writing as a helping tool to increase learning outcomes, to maximize academic success, and to support varied learning preferences and special needs, so as to ensure that every student may grow at their own speed (Lund,2023) [11].

Another study conducted by Jarrah, A. M. et al., [12] (2023) affirmed that supporters argue that using ChatGPT in academic writing is not a form of plagiarism for several reasons. Firstly, ChatGPT and similar AI language models are seen as tools that aid in the writing process rather than replace human creativity entirely. Secondly, proponents emphasize the importance of critically evaluating and rephrasing the content generated by ChatGPT to align with the author's voice and style (Perkins, 2024) [13]. Additionally, they highlight the necessity of proper citation when integrating AI-generated content, viewing it as comparable to cite other sources in academic writing. This perspective posits that with careful use and citation, AI-generated content can be a valuable asset in academic writing without constituting plagiarism.

Conversely, the well-known linguist, Noam Chomsky (2023) [14]perceived ChatGPT as “plagiarism software” that does not create nor does it look like the human mind, it copies existing works from existing artists and adapt it continuously to escape copyrights laws. This viewpoint underscores the importance of originality, human creativity, and the unique insights that can only be originated by real authors in academic writing.

1.4. Recent Studies on the Writing Proficiency of English Undergraduate Students

Students in higher education are expected to develop into proficient academic writers in the course of their studies. In this section the aim is to reveal the potential of whether undergraduate students are proficient in writing academically or their awareness is limited, particularly in putting this latter into practice, in this section we shrink the scope in the Moroccan context of the faculty of letters and human sciences in Oujda.

As stated by Noori (2020) [15], academic writing is crucial for undergraduate students, but they encounter issues with language, structure, and content in order of priority. This finding is similar to Wang and Yu's (2008) [16] culminates that language provides the greatest difficulty, followed by content and structure. The structure presented issues with regard to citations, idea coherence, and content arrangement (Abdulkareem, 2013) [17]. Despite theoretical understanding, a lack of practical experience contributed to these problems (Wallace, Schirato, and Bright, 1999) [18].

On this basis, another study was conducted by Ingunn Ofte (2014) [19], to explore the challenges students face in academic writing. Students struggle due to difficulties in selecting appropriate academic words, as many of the students voice a concern that their academic vocabulary is “limited” and not advanced enough for writing academic texts in English, and that they spend much time on searching for appropriate vocabulary in dictionaries. As one student comments, she spends “a lot of time finding synonyms to words I feel to be a bit easy.” (Ofte, 2014).

In the Moroccan context, The Ministry of Higher Education, Scientific Research, and Innovation in Morocco has recently introduced reforms that include removing the Graduation Research unit from bachelor's degree courses starting in the 2023-2024 academic year. This decision was approved by the National Committee for Coordination of Higher Education on April 14, 2023.

The new pedagogical strategy focuses on developing students' digital literacy, self-advocacy skills, and national identity awareness. A plethora of studies were conducted to investigate the academic writing proficiency of undergraduate students, and the impact of this decision on their future studies as researchers who contribute in academia. As claimed by Jamal Labradhi [20] (2023) Monographs are submitted in partial fulfillment of the requirements for a certain degree. That is, students do not conduct research for the sake of research per se; rather, they administer it in order to obtain the degree they study for.

These reforms were implemented as a response to students' passivity towards monographs and due to the aforementioned reasons. However, these reforms may pose challenges for students pursuing MA and Ph.D. studies in terms of academic writing skills. With the reduction in emphasis on research and analysis at the undergraduate level, students may face difficulties in conducting rigorous research, critically evaluating scholarly literature, and producing high-quality monographs or dissertations required for advanced degrees.

1.5. Gaps in the Literature about ChatGPT uses

Academic Writing is a core subject that undergraduate students take during their three years of BA studies, which may continue into MA and PhD studies. Several studies have been conducted to explore the difficulties students face globally and locally, in Morocco, yet it seems that there is no actionable solutions after exhibiting what are the obstacles and the challenges, to improve students' academic writing level. A significant gap remains in the absence of empirical experiments or practical actions that demonstrate how and to what extent such training is truly effective.

Moreover, the literature does not provide a clear and detailed explanation of when ChatGPT might be unnecessary, beneficial, or even misleading in the context of academic writing. The nuanced circumstances under which ChatGPT can either support or hinder students' writing development are not well-documented.

II. Methodology

This section is the starting of the empirical part of the current study. The following sections of this part are mainly the core of the methodology and data collection phases being employed.

This research aims to investigate the academic writing levels of S6 English students in FLSH Oujda (literature major), with a specific focus on the relationship between their proficiency and the use of ChatGPT, as one of the most used AI tool. The study seeks to investigate undergraduate students' academic writing proficiency and to determine whether there is a cause-effect relationship between students' use of ChatGPT and their academic writing skills or if these variables are separate. This investigation is crucial in highlighting a gap in the academic writing proficiency of S6 students, which can inform the development of continuous approaches in MA programs.

The study involves a purposeful sample of undergraduate students from the English department, literature major, who are willing to pursue MA and PhD studies. This focused group from FLSH Oujda is convenient for examining the specific research questions posed. The primary research questions are: "Are S6 English students able to write academically and respect its rules?" and "Is there any effect of ChatGPT on students' academic writing skills?"

Moreover, the methodology employs a Test-Teach-Test approach due to the nature of the thesis statement. This approach is chosen because neither questionnaires nor interviews would provide concrete and credible data regarding the students' academic writing skills. By testing the students' writing before and after a targeted teaching intervention, the study aims to gather reliable data on their writing abilities and the potential influence of ChatGPT.

2.1. Design

The initial measurement (pretest) is conducted to establish a baseline of the participants' academic writing skills. In the training phase, which lasted for two hours, Participants are randomly divided into two groups: Group A, which receives training to improve academic writing skills, and Group B, which does not receive any training. In the post-test phase, each of the initial groups (Group A and Group B) is further divided into two subgroups, resulting in four groups. Group 1 (trained and uses ChatGPT), Group 2 (trained and does not use ChatGPT), Group 3 (not trained and uses ChatGPT), and Group 4 (not trained and does not use ChatGPT). In this Phase, all four groups pass a post-test to assess their academic writing skills.

II.2. Sampling

Simple random sampling enhances the representativeness of the sample. By giving each S6 literature student an equal opportunity to participate, the sample is more likely to accurately reflect the diverse abilities and characteristics of the entire population. This representativeness is crucial for generalizing the findings to all S6 literature students, thereby increasing the external validity of this study.

II.3. Instruments

In this study, the testing instrument is particularly significant as it offers quantifiable data on students' academic writing skills. The use of the MEAL framework and relevance to the question ensures that the assessments are systematic and standardized, thus enhancing the validity of the findings.

II.4. Data Collection Procedures

The data collection procedures were systematically structured into three phases: the pre-test to establish baseline measurements, a training phase to enhance academic writing skills, and a post-test to evaluate the impact of the training and ChatGPT usage. The pre-test involved two initial questions and a writing task. The training comprised practical workshops, followed by a post-test immediately after the training, with groups differentiated by training and ChatGPT use. Finally, ethical considerations were diligently addressed, ensuring informed consent, confidentiality, and the fair treatment of all participants throughout the study.

III. Findings

In this section, we start by descriptions of the four groups in the pre-test and the post-test, progressing by independent comparisons, and culminating by the evolution of paired groups in different phases.

3.1. Description of Results

3.1.1. The Pre-Test

3.1.1.1. Group 1: No Training/ No ChatGPT (TN-CN)

The pre-test data for the TN_CN group, based on seven valid responses, shows an average grade of 2.43, a median of 2.00, and a mode of 0, which means that the scores are very low while the test was scored out of 20 (Table 1).

Statistics		
Pre Grades		
N	Valide	7
Mean		2,43
Median		2,00
Mode		0
Standard Deviation		2,225
Minimum		0
Maximum		5
Quartiles	25	,00
	50	2,00
	75	5,00
a. Group = TN_CN		

Table 1: Statistical measures of TN_CN Group

The standard deviation is 2.225, indicating significant variation. Scores ranged from 0 to 5, with the 25th percentile at 0, the median at 2.00, and the 75th percentile at 5.00. This suggests a relative wide range of abilities, with a concentration of lower scores.

3.1.1.2. Group 2: No-Training – Yes-ChatGPT (TN_CY)

The pre-test data for the group TN_CY reveals that there were eight valid responses. The results are similar to the first group with some differences (Table 2).

Statistics		
Pre_Grades		
N	Valid	8
Average		3,25
Median		3,50
Mode		0
Standard deviation		3,059
Minimum		0
Maximum		8
Quartiles	25	,00
	50	3,50
	75	5,75
a. Group = TN_CY		

Table 2: Statistical measures of TN_CY Group

The mean grade was 3.25, which indicates a moderate performance level among students. The median grade was 3.50, which suggests that half of the students scored above and half below this value. The mode was 0, which shows that the most frequently occurring grade was 0. The standard deviation was 3.059, which indicates a substantial variation in the grades for this group compared to the previous one. The minimum score was 0, while the maximum score reached 8. Additionally, the 25th percentile was at 0, the 50th percentile (median) at 3.50, and the 75th percentile at 5.75. This distribution highlights a wide range of abilities within the group, with a concentration of both lower and higher scores.

3.1.1.3. Group 3: Training/ No ChatGPT (TY_CN)

The pre-test data for the TY_CN group, comprising seven valid responses, unveils noteworthy aspects of participants' achievement (table 3).

Pre_Grades		
N	Valide	7
Average		3,14
Median		2,00
Mode		0
Standard deviation		3,024
Minimum		0
Maximum		8
Quartiles	25	,00
	50	2,00
	75	6,00
a. Group = TY_CN		

Table 3: Statistical measures of TY_CN Group

The mean grade, relatively close to the second group, which was 3.14, indicates a moderate level of overall performance. The median grade, at 2.00, shows that half of the students scored below and half above this value. The data shows that the scores ranged from a minimum of 0 to a maximum of 8. The mode was 0, though multiple modes were present. The standard deviation was 3.024, indicating considerable variability in the grades. The 25th percentile was 0, the median (50th percentile) was 2.00, and the 75th percentile was 6.00.

3.1.1.4. Group 4: Training – ChatGPT (TY_CY)

Evidently, in the pre-test data for the TY_CY group, the mean grade of 4.25 indicates a moderate to good overall performance among the students (table 4).

Pre_Grades		
N	Valid	8
Average		4,25
Median		4,00
Mode		3
Standard deviation		2,712
Minimum		0

Maximum		8
Quartiles	25	2,25
	50	4,00
	75	6,75
a. Group = TY_CY		

Table 6: Statistical Measures of TY_CY Group

The median grade of 4.00 shows that half of the students scored below and half scored above this value, while the mode of 3 reveals that the most frequently occurring grade was 3. The grades ranged from a minimum of 0 to a maximum of 8, highlighting a wide range of abilities within the group. The 25th percentile was at 2.25, the median (50th percentile) at 4.00, and the 75th percentile at 6.75, indicating that while some students scored low, a significant number scored quite high.

However, it should be noted that this difference is not that significant for this group compared to the other groups. Consequently, we can say that the level of all the students in the different groups had a very low level of competence for the theme studied. This observation confirms our choice of students and their similarities. Thus, if there were differences in the post-test, they would mainly be linked to the tools tested, namely the training carried out and the use of digital tools such as the ChatGPT.

3.1.2. The Post Test

3.1.2.1. Group 1: No Training/ No ChatGPT (TN_CN)

Moving to The post-test data, the first group TN_CN exhibits seven valid responses with an average grade of 2.29. The mode, which was 4 and the most frequently occurring grade, matched the median grade of 4.00 (table 5).

Statistics[a]		
Post_Grades		
N	Valid	7
Average/ Mean		2,29
Median		4,00
Mode		4
Standard deviation		2,138
Minimum		0
Maximum		4
quartiles	25	,00
	50	4,00
	75	4,00
a. Group = TN_CN		

Table 5: Statistical Measures of TN_CN Group

The standard deviation of 2.138 indicated significant variation in grades. Scores ranged from a minimum of 0 to a maximum of 4. The 25th percentile was 0, the median (50th percentile) was 4.00, and the 75th percentile was also 4.00. This suggests that while the average performance was low, a significant portion of students achieved a grade of 4.00 or above.

We can notice that this group has not improved since the Average and all other parameters indicate that their performance is almost equivalent to the pre-test. We will return to this comparison in the next section.

3.1.2.2. Group 2: No Training/ ChatGPT (TN_CY)

Regarding the second group TN_CY (table 8), it shows that there were eight valid responses. The mean grade was 9.25, with a median and mode of 8.

Statistics[a]		
Post_Grades		
N	Valid	8
Average		9,25
Median		8,00
Mode		8
Standard deviation		3,196
Minimum		4
Maximum		14
Quartiles	25	8,00
	50	8,00
	75	12,00
a. Group = TN_CY		

Table 6: Statistical Measures of TN_CY Group

The standard deviation was 3.196, indicates variability in the scores. Grades ranged from a minimum of 4 to a maximum of 14. The 25th and 50th percentiles were both at 8, while the 75th percentile was at 12. Overall, this suggests a generally high performance with some variation among students.

3.1.2.3. Group 3: Training No ChatGPT (TY_CN)

For the group labeled TY_CN, there were seven valid responses; the mean grade was 16.29, which indicates a high overall performance (table 7).

Statistics[a]		
Post_Grades		
N	Valid	7
Average/Mean		16,29
Median		16,00
Mode		16
Standard deviation		2,138
Minimum		14
Maximum		20
Quartiles	25	14,00
	50	16,00
	75	18,00
a. Group = TY_CN		

Table 7: Statistical Measures of TY_CN Group

The median grade was 16.00, and the mode was also 16, suggesting this was the most frequently occurring score. The standard deviation was 2.138, which shows some variability in the grades. The minimum score was 14, and the maximum was 20. Additionally, the 25th percentile was at 14.00, the 50th percentile (median) at 16.00, and the 75th percentile at 18.00. This distribution indicates that the majority of students performed well, with scores clustering around the higher end of the scale.

The improvement in the score is very clear for this group which went from an Average of 3.14 to 16.29. This confirms that the training session carried out achieved its objectives and allowed the students to acquire the targeted skills and also to obtain a more satisfactory result.

3.1.2.4. Group 4: Training/ChatGPT (TY_CY)

The last post-test group, TY_CY, reveals that the mean grade was 17.00, which demonstrates that, on average, students in this group performed very well (table 8).

Statistics[a]		
Post_Grades		
N	Valid	8
Average		17,00
Median		16,00
Mode		16
Standard deviation		2,138
Minimum		14
Maximum		20
Quartiles	25	16,00
	50	16,00
	75	19,50
a. Group = TY_CY		

Table 8: Statistical Measures of TY_CY Group

Second, the close alignment of the median and mode at 16.00 suggests a consistency in the performance, with a significant number of students achieving this score. Additionally, the standard deviation of 2.138 shows that while there is some variability in the scores, most students performed within a narrow range around the mean. Last but not least, the 75th percentile at 19.50 indicates that the top 25% of students scored very high, with grades approaching the maximum of 20. Overall, the data suggests that the majority of students in the TY_CY group achieved high grades.

3.2. Groups Comparison

3.2.1. The Pre-test

3.2.1.1. Training – No Training

To verify the hypothesis that the students had the same level before the experiment, we considered that the group was made up of two subgroups and proceeded to compare their average (table 9).

Group Statistics				
	Training	N	average	Standard deviation
Pre-Grades	Yes	15	3,73	2,815
	No	15	2,87	2,642

Table 9: Group Statistics for Pre-Grades with and without Training

The group with training had an average pre-test score of 3.73, while the group without training had an average score of 2.87. The standard deviations were 2.815 and 2.642, respectively, indicating some variability in scores within each group. The difference between the two is very slight but to decide, we applied a test of equality of two averages (table 12), this is the "independent samples test". This test is crucial as it helps determine whether the assumption of equal variances is valid (training variance of scores= No training variance of scores), we called this the null hypothesis; otherwise we accept the alternative hypothesis that they are not equal.

The result shows that the difference is not statistically significant since the significance value = 0.392 which is well above 0.05 (the minimum from which a significant difference must be declared).

Independent Sample Test

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	t	ddl	Sig. (bilateral)
Pre Grades	,052	,821	,869	28	0,392

Table 10: Independent Samples Test for Pre-Grades with and without Training

This result confirms that the students had the same level before the application of the different experimental treatments, even if very small differences may appear but they are only due to the randomness of the sampling. Thus, the validity of the results of our experiment would be increased and any differences that might appear at post-test could be attributed with high confidence to the effect of the training.

3.2.1.2. ChatGPT – No ChatGPT

We carried out the same test, this time considering that the students were made up of two subgroups in relation to the use of ChatGPT (table 12).

The group that used ChatGPT had an average pre-test score of 3.75, while the group without ChatGPT scored an average of 2.79. Both groups show some variability in their scores, with standard deviations of 2.840 and 2.577, respectively (table 11).

Group statistics				
	ChatGPT	N	Average	Standard deviation
Grades_Pre	Yes	16	3,75	2,840
	No	14	2,79	2,577

Table 11: Group Statistics for Pre-Grades with and without ChatGPT

To see if this difference in scores is significant, an "independent samples test" Is conducted (table 12). The test results indicated that the difference wasn't statistically significant, as the significance value was 0.341, much higher than the 0.05 value needed to declare a significant difference.

Independent Sample Test

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	t	ddl	Sig. (bilateral)
Pre_Grades	,152	0,700	0,968	28	0,341

Table 12: Independent Samples Test for Pre-Grades with and without ChatGPT

3.2.2. The Post-test

3.2.2.1. No Training/ Training

Since the average post-test of the two groups are extremely different (table 15), we can see that the training had a significant positive effect on the post-test scores. The group that received training scored, on average, much higher (16.67) compared to the group that did not receive training (6.00). The training appears to have been effective in improving the participants' performance on the post-test, as evidenced by the higher mean score and more consistent results in the training group compared to the no training group.

Group Statistics

	Training	N	Mean	Standard Deviation

Post_Grades	Yes	15	16,67	2,093
	No	15	6,00	4,472

Table 13: Group Statistics for Post-Grades with and without Training

Following by Levene test to measure the significance of the difference (table 14)

Independent Sample Test

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	T	ddl	Sig. (bilateral)
Post_Grades	9,809	0,004	8,367	28	0,000

Table 14: Independent Sample Test for Post-Grades with and without Training

The difference between those who received training and those who did not is very highly significant (sig. = 0.000) and cannot be ignored in any way. In fact, we went from a level well below the Average which is 6.00 to a very remarkable level of 16.7.

3.2.2.2. ChatGPT/ No ChatGPT

The table below (table 57) presents data from a post-test comparing two groups: one that used ChatGPT and one that did not. The moderate difference in the two groups' means indicates a noticeable difference in performance between the two groups, with the ChatGPT group performing better on average. The use of ChatGPT appears to have been beneficial in improving the participants' performance on the post-test, as evidenced by the higher mean score and more consistent results in the ChatGPT group compared to the no ChatGPT group.

Group Statistics

	ChatGPT	N	Mean	Standard deviation
Post_Grades	Yes	16	13,13	4,787
	No	14	9,29	7,549

Table 15: Group Statistics for Post-Grades with and without ChatGPT

To test the significance of the difference in post-test scores between the group that used ChatGPT and the group that did not, we will first conduct Levene's test for equality of variances (table 16).

Independent Sample Test

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	T	ddl	Sig. (bilateral)
Post_Grades	13,148	0,001	1,686	28	0,023

Table 16: Independent Sample Test for Post-Grades with and without ChatGPT

The difference between those who use ChatGPT and those who have not had access to it is highly significant (sig.=0.023 which is well below 0.05). This means that ChatGPT helps improve student performance even if this improvement is not as significant as for training. In fact, the difference is only around 4 points compared to the improvement linked to training which exceeds 10 points.

3.2.2.3. No Training – ChatGPT Yes Vs No

When we see in more detail what happened within each group, we end up with the comparison of several groups and we start with the contribution of ChatGPT for students who have not followed training. The results show that in terms of score, as shown in the table, the students who used ChatGPT had an Average of 9.25 which is much higher than that of those who did not work with ChatGPT who did not have than 2.29 (table 17).

Group Statistics[a]

	ChatGPT	N	Mean	Standard deviation
Post_Grades	Yes	8	9,25	3,196
	No	7	2,29	2,138

a. Training = No

Table 17: Comparison of Post- No Training Grades with and without ChatGPT

Since Levene's Test was conducted first to check for equality of variances and indicated equal variances. The t-test for equality of means shows a significant difference between the groups (table 200. (sig.=0.000). This confirms the usefulness of integrating digital tools, especially ChatGPT in learning and preparing for academic writing.

Independent Sample Test[a]

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	T	ddl	Sig. (bilateral)
Post_Grades	,905	0,359	4,878	13	0,000

a. Training = No

Table 18: Independent Sample Test for Post- No Training Grades with and without ChatGPT

3.2.2.4. Training – ChatGPT Yes Vs No

The data suggests that there is a slight improvement in post-test scores for the group that used ChatGPT compared to the group that did not use ChatGPT (table 19), although both groups received training. The mean score for the ChatGPT group (17.00) is slightly higher than that of the no ChatGPT group (16.29), but the difference is relatively small. The identical standard deviation of 2.138 for both groups indicates that the variability in scores is the same for both groups.

Group Statistics

	ChatGPT	N	Mean	Standard Deviation
Post_Grades	Yes	8	17,00	2,138
	No	7	16,29	2,138

Table 19: Comparison of Post- Training Grades with and without ChatGPT

Both groups exhibit the same variability in their scores, and the precision of the mean estimates is high. To determine if this slight difference is statistically significant, further statistical tests, such as a t-test, would be needed (table 20).

Independent Sample Test[a]

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	t	ddl	Sig. (bilateral)
Post_Grades	0,106	0,750	0,645	13	0,530

Table 22: Independent Sample Test for Post- Training Grades with and without ChatGPT

The t-test evaluates whether there is a statistically significant difference in the mean post-training grades between the two groups. Both the equal variances and unequal variances assumptions yield p-

values of 0.530, which are greater than 0.05. This indicates that there is no significant difference in the means of the two groups. Comparing to the previous group (table 20), the current analysis Shows no significant difference in post-training grades between participants who used ChatGPT and who didn't.

3.2.2.5. No ChatGPT Training Yes Vs No

Two groups are compared in this section (table 21); participants who did not use ChatGPT but received training, and ones who did not use ChatGPT as well and also didn't receive any training.

Group Statistics[a]

	Training	N	Mean	Standard Deviation
Grades_Post	Yes	7	16,29	2,138
	No	7	2,29	2,138

Table 21: Comparison of Post- no ChatGPT Grades with and without Training

The table below (table 22) presents the Independent Sample Test with Levene's Test and t-test for Equality of Means, which is to analyze and present statistical evidence regarding the difference in post-training grades between two groups where ChatGPT was not used ("No").

Independent Sample Test

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	t	ddl	Sig. (bilateral)
Post_Grades	0,628	0,444	12,250	12	0,000

Table 22: Independent Sample Test for Post- No ChatGPT Grades with and without Training

Both assumptions (equal and unequal variances) yield very low p-values (0.000), indicating a highly significant difference in mean post-training grades between the two groups (ChatGPT = No). This means that participants who did not use ChatGPT exhibit significantly different post-training grades compared to each other, regardless of variance assumptions.

3.2.2.6. ChatGPT Training Yes Vs No

Concerning the group of students who used ChatGPT, we note that the students who followed the training obtained an average score of 17.00 which is much higher than those who did not follow it with an average score of 9.25 (Table 23). The difference of 7.75 is very large and is statistically significant (t=5.701 and

sig=0.000 which is far less than 0.05) (Table 24). Therefore, we cannot in any way ignore this difference and attribute it to chance. It is directly due to the learning generated by training.

Group statistics[a]

	Training	N	Mean	Standard Deviation
Post_Grades	Yes	8	17,00	2,138
	No	8	9,25	3,196

Table 23: Comparison of Post- ChatGPT Grades with and without Training

Independent Sample Test

	Levene's test for equality of variances		t-test for equality of means		
	F	Sig.	t	ddl	Sig. (bilateral)
Post_Grades	1,397	0,257	5,701	14	0,000

Table 24: Independent Sample Test for Post- ChatGPT Grades with and without Training

We can consider, then, that the training was beneficial to the students and allowed them to improve their performance in terms of knowledge and skills.

3.3. Groups Evolution

In this section, we will present the results relating to the evolution of each of the four experimental groups between the pre-test and the post-test. We will have an idea, for each group, if the level has changed positive or negative or even if the level has remained unchanged. Thus, when the level improves, this means that the corresponding treatment is beneficial. For groups that do not change level, this means that the treatment in question does not improve the students' skills in academic writing.

Unlike previous comparisons, in this study of evolution, we will use the test for comparing two means in the case of paired samples. Indeed, the same students were tested twice, once during the pre-test and a second time for the post-test.

3.3.1. No Training – No ChatGPT

We start with the group that has not received anything since the start of the training. For this group, table 25 shows that the score did not really change and only increased by 0.14 out of 20, going from 2.29 to

2.43 respectively for the pre-test and post-test. The variability is identical for both tests and remains relatively high compared to the Average.

Statistics for paired samples a

		Mean	N	Standard Deviation
Pair 1	Post Grades	2,29	7	2,138
	Pre Grades	2,43	7	2,225

a. Group = TN_CN

Table 25: Paired Sample Statistics for Pre-Test and Post-Test Grades (Group = TN_CN)

We can then say that this group did not change its performance and remained at a low level. These students who did not benefit from training or access to ChatGPT were not able to improve their capacity. This confirms the fact that if we continue to provide no assistance to our students in Academic Writing, they risk having the same gaps and having the same difficulties in Writing.

3.3.2. No Training –ChatGPT

For the group of students who did not attend the training and who had access to ChatGPT, table 26 shows a relatively interesting improvement going from 3.25 in the pre-test to 9.25 in the post-test but which is not very satisfactory, since their performance did not reach the average level which would have allowed the validation of the test.

Paired Sample Statistics

		Mean	N	Standard deviation
Paire 1	Post_Grades	9,25	8	3,196
	Pre_Grades	3,25	8	3,059

a. Group = TN_CY

Table 26: Paired Sample Statistics for Pre-Test and Post-Test Grades (Group = TN_CY)

The comparison of the scores obtained in the two tests (post-test and pre-test) confirms the increase in the Mean which is 6.00 and statistically significant (Sig=0.005 which is less than 0.05) (Table 27).

Paired Sample Test

		Paired differences		t	ddl	Sig. (bilateral)
		Mean	Standard Deviation			
Pair 1	Grades_Post - Grades_Pre	6,000	4,036	4,205	7	0,004

a. Group = TN_CY

Table 27: Paired Sample Test for Pre-Test and Post-Test Grade Differences (Group = TN_CY)

We can therefore say that the use of Chat GPT was beneficial overall for all individuals because the average score was higher during the post-test but did not improve performance individually since the correlation is not important. In addition, on the one hand some students benefited a lot from ChatGPT and their score improved significantly and on the other hand some did not improve their performance and remained at a low level even with the use of ChatGPT. It is possible that these students do not master the use of such digital tools and are not familiar with ChatGPT and therefore we could plan for an introduction of these ones if we ever opt for their integration into the training.

3.3.3. Training – No ChatGPT

We then move on to the group who followed the training session and were not allowed to use ChatGPT during the post-test. The results from table 28 show that the improvement in the performance of the participants in this group is significant. In fact, they went from an average score of 3.14 in the pre-test to an average score of 16.29 in the post-test. Better yet, the variability within this group is very low compared to the other groups, which indicates that the entire group improved and produced quality reports that resulted in such a high score.

Paired Sample Statistics

		Mean	N	Standard Deviation
Pair 1	Grades_Post	16,29	7	2,138
	Grades_Pre	3,14	7	3,024

a. Group = TY_CN

Table 28: Paired Sample Statistics for Pre-Test and Post-Test Grades (Group = TY_CN)

Furthermore, table 29 confirms what has been said since the difference between the pre-test score and that of the post-test is 13.14 and cannot be attributed to chance and is therefore statistically very highly significant. (sig=0.000 which is much less than 0.05).

Paired samples test

	Paired Differences		t	ddl	Sig. (bilateral)
	Mean	Standard Deviation			
Pair 1 Post_Grades & Pre_Grades	13,143	3,625	9,592	6	0,000

a. Group = TY_CN

Table 29: Paired Sample Test for Pre-Test and Post-Test Grade Differences (Group TY_CN)

These results show the importance and usefulness of additional training in scientific writing techniques. The training was very beneficial to the students.

3.3.4. Training - ChatGPT

The last group is that of students who followed the training and had access to ChatGPT during the post-test. Table 30 displays a performance which is the most important of all the groups with an average score of 17.00. The improvement linked to ChatGPT is not as important as that linked to training since we went from 16.29 (Training-No ChatGPT) to 17.00 (Training – ChatGPT) or an improvement of only 0.81.

As for the previous group, the correlation is very weak and does not exceed 0.05 (Table 37)

Paired Sample Statistics

		Mean	N	Standard Deviation
Pair 1	Post_Grades	17,00	8	2,138
	Pre_Grades	4,25	8	2,712

a. Group = TY_CY

Table 30: Paired Sample Statistics for Pre-Test and Post-Test Grades (Group = TY_CY)

Finally, table 31 confirms the improvement in performance for this group which gained 12.75 points between the pre-test and the post-test. This gain is statistically very significant (Sig=0.000).

Paired Sample Test

		Différences appariées		t	ddl	Sig. (bilateral)
		Mean	Standard Deviation			
Pair 1	Post_Grades & Pre_Grades	12,750	3,370	10,701	7	0,000

a. Group = TY_CY

Table 31: Paired Sample Test for Pre-Test and Post-Test Grade Differences (Group TY_CY)

We can then say that the training considerably improves the performance of students and much more than what ChatGPT can do.

IV. Discussion

The findings suggest that students are not yet fully proficient in meeting these expectations. They exhibit significant challenges with key aspects of academic writing, such as constructing coherent arguments, These struggles are consistent with Bouziane and Abouabdelkader's (2016) observation that "writing in Morocco needs more improvements" (p. 41).

In the post-test phase, we observe significant differences in the performance of the four groups, which provide further insights into the impact of training and the use of ChatGPT on students' academic writing abilities. Students who utilized ChatGPT without training exhibited significant improvements in their post-test academic writing scores. The use of ChatGPT appears to have positively influenced their ability to produce higher quality written work. Similar to the findings of the study conducted by Jarrah et al. (2023), ChatGPT is perceived as a tool that supports the writing process, enhancing it rather than replacing human creativity entirely. However, this positive effect is unreliable, further interpretations will guide us to measure its effect.

The post-test results provide compelling evidence supporting RH1, indicating that academic writing assessment is crucial in MA programs. The findings suggest that structured training and digital support play pivotal roles in improving academic writing proficiency. The post-test outcomes also validate RH2 by demonstrating varying impacts of ChatGPT usage on students' writing abilities.

As stated previously by Mohamed Khalifa et al., in 2024, the findings also suggest that pedagogical strategies in higher education should incorporate a balanced approach that includes both constant training and digital tools. While ChatGPT showed promise in enhancing academic writing skills, its long-term efficacy remains questionable without formal training. Educators should consider integrating AI tools as supplementary resources rather than primary teaching methods.

The comparison of different groups underlines the importance of structured training and the potential supplementary role of digital tools like ChatGPT in enhancing students' academic writing skills. However, to gain a deeper understanding of how these interventions impact student performance over time, it is essential to examine the evolution of the same groups across different phases of the study.

Based in the evolution of the four groups, the use of ChatGPT may conceal students' weaknesses in academic writing and create a false impression that they are proficient writers. By relying on AI-generated content, students might appear to have stronger writing skills than they actually possess, which can be misleading both for the students themselves and for their educators. Second, the long-term effectiveness of ChatGPT in developing genuine writing skills is questionable. While ChatGPT can provide short-term assistance, it may ultimately prove to be deceptive, as it does not foster the deep understanding and skill development required for sustained academic success. These findings are in line with the study by Ismail Dergaa (2023) presents a similar perspective, noting that while ChatGPT can enhance academic writing skills, it may inadvertently lead students to believe that the improvements made are solely the result of their own efforts.

Additionally, as suggested in the literature, it is difficult to forbid, resist, or reject revolutionary tools like ChatGPT (Eke, 2023), which is not the objective here. Instead, the aim is to integrate these tools effectively while ensuring that students still develop their own skills. Training, as a form of a well-structured syllabus, will provide students with the confidence that they can perform better on their own.

Conclusion

The study investigated the level of academic writing skills for undergraduate English students, literature major at the FLSHO and focused on the effect of ChatGPT and structured training among these students. Key findings indicate that initial academic writing skills were notably low across all groups in the pre-test phase. Without training, the use of ChatGPT only showed minimal improvement, which suggests its limited efficacy in enhancing writing proficiency independently. Conversely, groups that received structured training exhibited significant improvements in post-test scores, which indicate the critical role of formal instruction in fostering academic writing skills. Furthermore, the combination of training and ChatGPT demonstrated relevant effects, which led to the highest levels of improvement in writing abilities.

The present research endeavor represents an effort aimed at filling the existing gap in the literature. Specifically, this study aims to investigate the challenges encountered by English undergraduate, literature major in academic writing and their relationship with the use of the ChatGPT tool. By addressing these issues, the study seeks to contribute to the ongoing discourse on enhancing academic writing skills and integrating AI tools in higher education.

However, this study's methodology revealed several inherent limitations. On the one hand, one significant limitation of the study was the limited duration of the training sessions. Although the initial aim was to conduct sessions lasting more than two hours, the actual duration had to be reduced due to the students' lack of motivation. On the other hand, another limitation was the inaccessibility of participants from other majors, such as linguistics, since the inclusion of those students could have provided a more diverse sample and led to more generalizable findings.

In light of the acknowledged limitations, there is a room for future studies to concentrate on the numerous critical challenges that students and educators in this matter. Design comprehensive training programs that extend beyond a two-hour session to ensure thorough understanding and practice. As

previously highlighted, a recommendation concerning inclusive participant selection is needed. To ensure that the findings are relevant and applicable across different fields, future research and training programs should include participants from other majors and disciplines, since academic writing is a core requirements in dissertations even if it is about scientific research papers, In addition to those from English departments.

References

1. Zohery, M. (2024). Chapter 2: ChatGPT in academic writing and publishing: A comprehensive guide. *ResearchGate*. https://doi.org/10.5281/zenodo.7803702

2. Zhai, X. (2023). ChatGPT and AI: the game changer for education. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389098

3. Shannon. (2023). 7 ChatGPT prompts to improve your writing. *Forbes*. https://www.forbes.com/sites/aytekintank/2024/06/25/7-chatgpt-prompts-to-improve-your-writing/

4. Sturuan, A. (2021). Improving academic writing skills to boost student confidence and resilience. *Times Higher Education*. https://www.timeshighereducation.com/campus/improving-academic-writing-skills-boost-student-confidence-and-resilience

5. Strobl, C., & Ail, E. (2019) Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*. https://www.sciencedirect.com/science/article/abs/pii/S036013151830318X

6. Rafi, M., JianMing, Z., & Ahmad, K. (2023). Technology integration for students' information and digital literacy education in academic libraries. Information Discovery and Delivery. https://www.emerald.com/insight/content/doi/10.1108/IDD-07-2019-0049/full/html

7. Irvin, L. (2017). *What Is Academic Writing. Parlor Press*. https://pressbooks.ulib.csuohio.edu/eng-102/chapter/what-is-academic-writing-by-lennie-irvin/

8. Thaiss, C., & Zawacki, T. M. (2006). Engaged writers and dynamic disciplines: Research on the academic writing life*. Boynton/Cook Publishers*.

9. Lee. A. (2002). Rehearsing new roles: How college students develop as writers. *Southern Illinois University Press*

10. Cotton, D., & Cotton, P. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. https://pearl.plymouth.ac.uk/handle/10026.1/20577

11. Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? Library Hi Tech News.

12. Jarrah, A. M., Wardat, Y., & Fidalgo, P. (2023). Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say?. *Online Journal of Communication and Media Technologies*, 13(4), e202346. https://doi.org/10.30935/ojcmt/13572

13. Perkins, M. (2024). Academic publisher guidelines on AI usage: A ChatGPT-supported thematic analysis. *F1000Research, 13*, Article 142411. https://doi.org/10.12688/f1000research.142411.2

14. Chomsky, N. (2023, March 8). The false promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

15. Noori, A. (2020). An Investigation of Afghan Undergraduate English Major Students Academic Writing Difficulties. Journal of Foreign Language Teaching & Learning, 5(2), 99-114. https://doi.org/10.18196/ftl.5249

16. Wang, W. and Yu, X. (2008). A survey of learners' self-perceived difficulties in L2 writing. Foreign Language Learning Theory and Practice, (1), 31-35.

17. Wallace, A., Schirato, T., & Bright, P. (1999). Beginning university: Thinking, researching, and writing for success. Crows Nest, NSW: Allen & Unwin

18. Wallace, A., Schirato, T., & Bright, P. (1999). Beginning university: Thinking, researching, and writing for success. Crows Nest, NSW: Allen & Unwin

19. Ofte, I. (2014). English academic writing proficiency in higher education: Facilitating the transition from metalinguistic awareness to metalinguistic competence. *ADNO, 8*(2), Article 17.

20. Jarrah, A. M., Wardat, Y., & Fidalgo, P. (2023). Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say?. *Online Journal of Communication and Media Technologies*, 13(4), e202346. https://doi.org/10.30935/ojcmt/13572