ALTRALANG Journal
e-ISSN: 2710-8619 p-ISSN: 2710-7922
Volume 8 Issue 1 / June 2026 pp. 230-246

Assessing the Use and Impact of an AI Digital Assistant Chatbot to Enhance Learners' Writing Abilities: The Case of First Year EFL Students at the University of Oran 02 – Mohamed Ben Ahmed

Amira Nedjar[1*]
[1]*University of Oran 2 Mohamed Ben Ahmed, Algeria*
nedjar.amira@univ-oran2.dz
Fatma Zaghar[2]
[2]*University of Oran 2 Mohamed Ben Ahmed, Algeria*
zaghar.fatma@univ-oran2.dz

Received: 15/11/2025, **Accepted:** 26/05/2026, **Published:** 05/06/2026

ABSTRACT: *Artificial intelligence systems have found their way into education through tools like grammar checkers, automated grading systems, and interactive chatbots. ChatGPT is a widely embraced example, yet its versatility poses challenges due to occasional inaccuracies and misuse. This article investigates the utilization of Poe, a platform enabling the creation of customized chatbots tailored to specific tasks designated through a prompt during their creation. The chatbot was proposed to First-Year English as a Foreign Language (EFL) students to serve as an assistant to improve their writing by providing specialized academic vocabulary, suggestions, and constructive feedback. The results show that the chatbot demonstrated a high degree of usability. It serves as a good aide to students during their writing process.*

KEYWORDS: Academic writing, English as a Foreign Language (EFL), English for Academic Purposes (EAP), explorative study, generative artificial intelligence, task-oriented chatbot

* Corresponding author
ALTRALANG Journal / © 2026 The Author(s). Published by the University of Oran 2 Mohamed Ben Ahmed, Algeria.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Introduction

Artificial Intelligence (AI) technology advanced significantly in the twenty-first century. All the theoretical concepts developed by earlier computer scientists, such as Alan Turing and John McCarthy (McCarthy et al., 2006; Turing, 2004), have been refined into an intelligence that outperforms their quondam predictions. In this time and age, human beings and machines collaborate in different fields as they become mutually complementary; human intelligence helps enhance and hone these machines for people to achieve goals and complete tasks quickly and efficiently. Machines, in return, provide individuals with assistance in different aspects.

In education, the earliest use of AI started with B.F. Skinner's the Intelligent Tutoring System (ITS) (Skinner, 1958), but as technology developed through time and Natural Language Processing Systems (NLP) got enhanced and their reliability rate increased significantly, AI tools started being used in foreign language learning. As a result, a brand-new academic study area called AI in Education (AIED) was created to track the advancement of AI and evaluate its efficacy in both teaching and learning (Holmes et al., 2022). The goal of AIED is to blend formal and informal learning by using AI to help students become more autonomous learners. With adaptive learning AI tools, chatbots, sensors, hardware, virtual platforms, and automatic assessment, the teacher's role in this context is to monitor the process through dashboards (Holmes et al., 2022), as opposed to traditional instruction where the teacher must plan, conduct assessments and learners' analysis, select methods and materials as per the allotted curriculum, evaluate the latter, and make decisions based on the results obtained (Yu & Lu, 2021).
The Fourth Industrial Revolution and the significant effects of globalisation have made AI become a popular tool for English language learning. The increasing demand of the modern world motivates self-conscious English learners to employ cutting-edge tools to master the language by honing their linguistic skills (speaking, writing, listening, and reading), as well as competencies (pragmatic, textual, and intercultural communicative competence) (Nong et al., 2021). Students may now study English using AI-powered tools that offer individualised lessons, tasks, and adaptable materials that are created using algorithms to suit to the learner's objectives.

Literature Review

1. Chatbots

As per Writerbuddy AI's compiled statistics for 2023 and a survey carried out by Statista Consumer Insights, the AI tools that garnered the highest visitor numbers were primarily chatbots (Richter, 2023; Sarkar, 2023). Taking the spotlight as the most visited website is ChatGPT, followed closely by character.ai, and Quillbot, claiming the third position as an AI writing tool. There are, indeed, other tools such as Bard, akin to ChatGPT in its capacity to generate creative text and content, translate, provide information across various subjects, and responding adeptly to diverse users' questions (Rudolph et al., 2023). However, there exist chatbots which are more specific to the academic context in which the bot acts like a personalized tutoring assistant to students. This is the case with Ada (Academic Digital Assistant), which was launched in 2017 (Kabiljo et al., 2020), Replika, which can offer emotional and mental support to students (Pentina et al., 2023), or Habitica, which supports students in forming positive habits by incorporating game elements into their learning approach and making it easier for students to organize their tasks and manage their time, thereby boosting motivation (De Sales & Antunes, 2021).

1.1. Chatbots in Education

Many studies within the field of education have highlighted substantial benefits associated with the integration of chatbots in the learning and teaching processes (Deng & Yu, 2023; N.-Y. Kim et al., 2019; Mushthoza et al., 2023; Wahyuni, 2022; Yang et al., 2022). As instructional approaches nowadays emphasize learner autonomy to promote metacognitive strategies, elevate learning motivation, and expedite the achievement of educational objectives (E. Lee & Hannafin, 2016), educators are increasingly receptive to incorporating AI tools in the educational environment. The study conducted by Huang et al. (2022) revealed that the use of chatbots as assistants to students proved advantageous in terms of timesaving, accessibility, and providing feedback. However, certain limitations were evident, particularly related to technological constraints and the potential for information overload. Moreover, according to Fryer (L. K. Fryer et al., 2020), chatbots have reached a level of development suitable for integration into language learning. In fact, they are deemed more effective and hold a more promising future in language learning due to their current capability to deliver and comprehend natural language, both in oral and textual forms, with minimal technical adjustments.

1.2. Chatbots in EFL Contexts

In English as a Foreign Language (EFL) setting, students aspire to achieve proficiency in using the language. In essence, their goal is to master the four primary language skills: listening, speaking, reading, and writing. Several studies have focused on examining the impact of chatbot use across various language skills. Kim's (2018) empirical investigation on the influence of AI chatbots on Korean EFL students' English vocabulary indicated a significant improvement in lexical knowledge among students who interacted with chatbots, as opposed to those in the control group. In another study conducted by Hwang and Lee (2021), pre-service teachers were tasked with creating task-oriented chatbots. The rationale behind this approach was the belief that such chatbots offer more effective assistance by keeping students focused on specific tasks. This stands in contrast to standard multi-context chatbots, which can overwhelm students with excessive information and lead to distraction. Furthermore, teachers can control the content generated by these chatbots, dictating what they should or should not produce. However, this approach has not been thoroughly explored regarding its effect on English language writing skills specifically, much less academic writing, thereby creating a research gap that warrants further investigation.

1.3. Criticism Surrounding Generative Chatbots

Generative AI platforms have faced significant criticism due to their extensive utilization, particularly for creating content on behalf of students (Alharbi, 2023). This raises concerns regarding the potential impact on students' cognitive skills, including originality, critical thinking, and ethical awareness (Ju, 2023; Rahman & Watanobe, 2023; Shidiq, 2023). Therefore, it is advisable for teachers to monitor the use of these AI platforms, guiding and training students on their effective utilization, instead of implementing a complete ban on them in the learning process (Otsuki, 2020). This approach is more feasible, given that such technology has become an integral aspect of our lives and serves as a global indicator of innovation and development.

2. Writing

Writing can be defined as a crucial skill that enables students to expand their lexical competence, acquire accuracy, and strengthen their command of grammar. It is also described as a cognitive as well as a physical process through which human beings express themselves (Hermansson & Lindgren, 2019). It fulfils a social function as people deliver their intentions through structured written language, conforming to linguistic

norms. It is used in various contexts, including the professional and academic settings, serving as a powerful medium that differs from spoken language (Harmer, 2004).

2.1. Issues in Writing Faced by EFL Learners

In the Algerian tertiary context, writing is frequently characterized as a challenging task for students learning English as a foreign language. A study conducted at the University of Biksra, involving 26 third year Licence Master Doctorate (LMD) students majoring in English Language, revealed that 35% of participants rated writing as the second most challenging skill, after speaking which was ranked at 42% (Dernoun, 2015). The same study indicated that the primary challenges in writing for these students stemmed from issues such as vocabulary limitations (29%), grammatical concerns (23%), as well as various other factors including constraints on time, lack of motivation, apprehension regarding negative feedback, and insufficient support from teachers. This perspective was further reinforced by a study undertaken at the University of Ain Témouchent, where PhD students in the Department of Mechanical Engineering received English for Academic Purpose training and were pre-assessed to identify deficiencies in their writing skills. The results of the pre-test revealed that many students encountered difficulties expressing their ideas coherently and methodically, respecting the format, employing appropriate vocabulary, avoiding spelling mistakes, and constructing sentences with proper grammar and punctuation (Belmokhtar & Lamri, 2023).

Another study conducted in Indonesia explored the challenges faced by EFL learners in essay writing. The findings indicated that these students encountered significant obstacles related to grammar, spelling, coherence, cohesion, as well as the organization of words or paragraphs, along with vocabulary issues (Ariyanti & Fitriana, 2017). Similar conclusions were drawn from additional research carried out in countries where English is taught as a foreign language, revealing that most EFL students struggle with the aforementioned difficulties, along with mechanics, syntax, and fundamental academic writing skills such as summarizing, paraphrasing, synthesizing, and so forth (Ahmed, 2019; Lalam, 2018; Maharani et al., 2023; Mohammad Almatarneh et al., 2018; Sarwar et al., 2022; Solikhah, 2017).

Considering the increasing attention in research on the impact of AI on language learning and learners themselves, there remains a notable gap in comprehensive research on tailored interactive chatbots and their effects on students' academic writing. Additionally, addressing the persistent challenges faced by EFL learners in writing remains an ongoing struggle for educators, particularly within the Algerian tertiary education context. Consequently, this paper seeks to investigate the implementation of a specialized chatbot, designed to focus specifically on assisting students with their academic writing process, as opposed to utilizing platforms like ChatGPT. The research questions are outlined as follows:

- What are the benefits and drawbacks associated with the use of a custom-made chatbot for academic writing support?
- To what extent does the implementation of this chatbot contribute to the alleviation of specific writing challenges faced by EFL learners in the Algerian tertiary education context?

The methodological framework of this study comprises 3 phases:

- Pre-assessment phase: This assessment will evaluate the academic writing performance of first-year Licence students in the Written Expression module. To establish a baseline of their overall level in academic wiriting, participants are assigned to write a free-topic short essay which would be evaluated based on rubrics. The second step involves conducting a survey to analyse the needs and preferences of participants. The data collected will then be used to craft a tailored prompt for our custom chatbot that will be used during class sessions.

- Post-assessment phase: The participants will be tasked with writing an academic short essay about the topic of their choice during a class session, using the chatbot for support as needed. The written works will then be evaluated. A final survey will be conducted to gather qualitative data on students' perceptions and attitudes toward Poe's integration into their writing process.
- Performance comparison: The data collected from the previous phases will first be tested for normal distribution. This step is crucial in deciding whether to use a parametric (Paired Students' *t*-Test) or non-parametric statistical test (Related-Samples Wilcoxon Signed Rank Test) to compare the performances between the two phases. After completing the comparison, the effect size will be calculated to measure the magnitude of difference. Further analysis, including the difference between the average pre-assessment and post-assessment scores, will be conducted to determine whether the effect is positive or negative. A second survey is also conducted to evaluate the participants' experience as well as gather feedback on the custom-chatbot utilization as a writing assistant.

Introducing Poe

This research focuses on the utilization of the platform *Poe*, a service created by Quora and introduced in 2022. The platform is built on Large Language Models (LLMs), which allows users to interact with chatbots in a human-like manner.

Poe offers a variety of 'official bots' developed by different companies, including the standard assistant powered by GPT-3.5, or GPT-4, Claude, DALL-E 3, and others. Additionally, the platform provides AI-powered bots designed for tasks beyond text generation, such as image generation.

Users have a range of options depending on their specific needs. They can initiate conversations with bots directly upon opening the application on their phone, tablet, or computer, or they can create a customized bot.

When users opt to create a bot, they are required to provide essential details related to the chatbot. These details encompass the bot's name, the base system it operates on, the prompt, the greeting message, and the bot's profile description:

1- The bot's name should be distinctive, to ensure it is not already in use by other platform users.

2- The base systems are AI tools developed by other companies that have their respective data sources. These systems include a range of options such "ChatGPT, GPT-4 and DALL-E 3 from OpenAI, Claude Instant and Claude 2 from Anthropic, StableDiffusionXL from Stability AI, PaLM and Gemini-Pro from Google, Llama 2 from Meta, Playground-v2 by Playground, Mistral-Medium by Mistral, and many other bots created by the community" (Poe, 2023).

3- The prompt, which is an integral component of generative AI, functions as a guiding instruction for the bot to develop content produced through the large model. It serves to define the scope of the bot by specifying the necessary context, information, or task that the bot is expected to undertake. The specificity of the prompt significantly influences the nature of the generated text. As per *Poe*'s developer website (*Poe for Developers*, 2023), there are recommended points to take into consideration to write an effective prompt:

- Using the second person to address the bot.
- Clarity
- Using brackets and / or markdowns for enhanced specificity

4- The greeting message is used to initiate an engagement with the users as would a real human being. It also sets the tone of the conversation and builds rapport with the person. This message can also contain

essential information such as instructions on how to use the bot effectively as well as expressions to avoid while interacting with the bot to prevent a potential ban (profanities, racial affronts, etc.).

5- The bot's profile description serves to introduce the functionalities of the bot if the latter is shared with other users. The creator of the bot can decide whether to make the bot publicly accessible or keep it private. Private bots can only be shared with selected individuals through a provided link.

Upon completion of these requisite details, the bot will be ready to use. Subsequently, the creator may also share the link of the chatbot to enable other people to utilize it.

Writing Skills

While the definition of writing as an ability may vary depending on the approach used, its fundamental role remains consistent: that of a "social practice" (Coffin et al., 2003). In Academia, writing plays a crucial role in communicating research on various topics and new discoveries which are then shared to an international community of researchers-like. This, consequently, promotes critical thinking and analysis among students and researchers as they seek professional development and intellectual exchange. Moreover, it serves as a social indicator of an academic's affiliation within a specific scientific community, reflecting the conventions and practices that define the appropriate written communication and identity within that domain (2003). Nevertheless, assessing academic writing remains a topic of debate due to the diverse range of teaching approaches employed.

In applied linguistics, several scholars have been debating whether to adapt a prescriptive or a descriptive approach to teach and assess language, especially in written form (Hinkel, 2018). While the prescriptive approach is often described as 'discriminatory', it is nonetheless the standardized version of language that serves as the basis for EFL learners, especially in the academic context where correct reproduction of written language is often considered to avoid any miscommunication. Therefore, during their learning process, EFL learners can be exposed to different approaches to writing instruction, and each approach emphasizes specific aspects of effective writing (Yi, 2009).

The various approaches and their distinctive features are outlined in Figure 01. The visual also highlights the differences between each approach, as each one of them focuses on a particular aspect of writing. However, given the complex nature of learning and teaching, which is based on assumptions about the intricacy of human behaviour, it is best to take a holistic approach while evaluating learners and consider every angle that could contribute to a comprehensive understanding of writing as a skill, especially in EAP context.

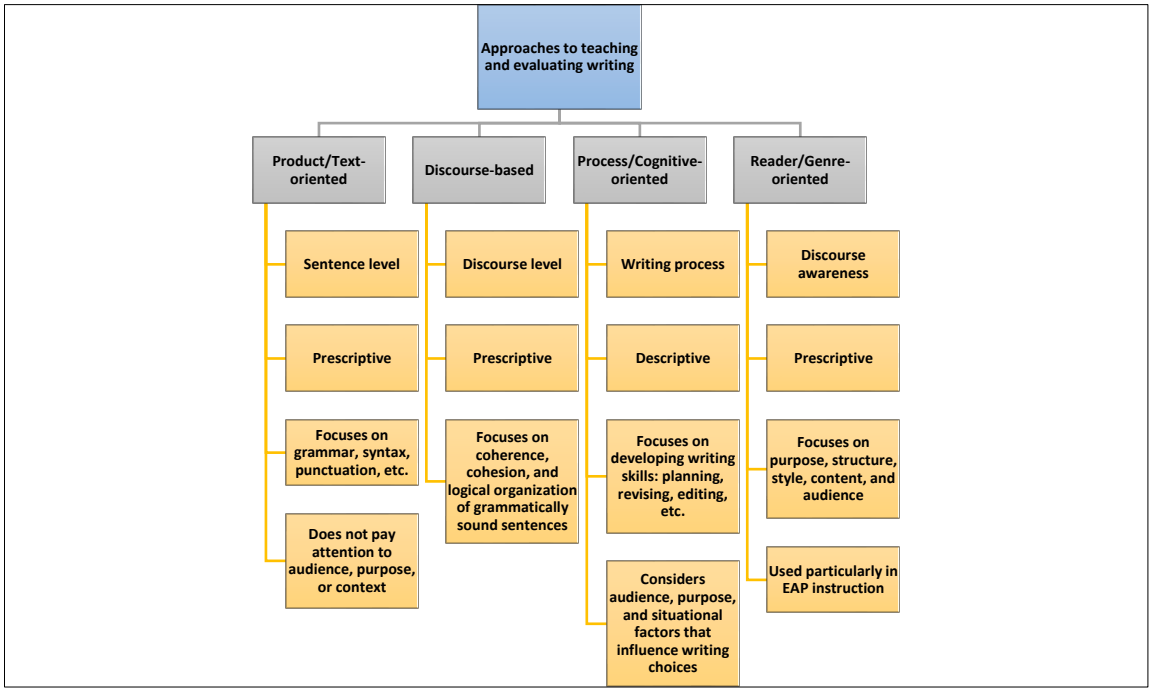


Figure 1 ***Approaches to teaching writing based on Jyi-yeon Yi's paper (2009)***

Academic writing differs from general writing in several key aspects. Academic writers require not only subject matter expertise (schemata) but also contextual understanding, effective writing strategies, and proficiency in language mechanics and grammar (Tribble, 2006). Furthermore, effective writing requires consideration of the purpose, audience, and genre (Nation, 2008). The purpose could be exploring the field of study, communicating information objectively, persuading readers, or storing information. According to Langan's (1993) intellectual/rhetorical approach, students may employ various rhetorical tasks and patterns of exposition to achieve these purposes, such as exemplification, process description, cause-and-effect analysis, comparison and contrast, definition, classification, argumentation, and narration (Cushing Weigle, 2002). The audience of academic writing is typically a specific discourse community with the necessary background knowledge to understand domain-specific registers. The genre of academic writing is determined by the purpose and audience and can range from paragraphs and essays to reports, research proposals, dissertations, and research papers. Additionally, academics must possess proficiency in research, summarization, paraphrasing, outlining, referencing, reading comprehension, and proofreading (Coffin et al., 2003; Nation, 2008).

In academia, writing is evaluated based on various criteria, which may differ depending on the field of study and the genre of writing. Instructors typically define specific criteria for assessing students' writing compositions, considering the purpose, audience, and conventions of the discipline.

Methodology

1. Participants

We combined two class groups of first-year students (L1) from the Department of English at the Faculty of Foreign Languages, University of Oran 02 – Mohamed Ben Ahmed, into one experimental group (*n=22*). It is noteworthy to know that these students underwent an identical and mandatory educational curriculum provided by the Ministry of Education, with all of them studying English as a Foreign Language (EFL) as

part of this curriculum. Their initial academic exposure to the language began in lower secondary school, where they received four years of instruction in the fundamentals of the English language. However, some students may have had the opportunity to study English at an early age in kindergartens or undergo language training in private or public institutions. This educational process is then extended in upper secondary school, involving an additional three years of intermediate to upper-intermediate courses in English.
The participants in both groups were selected randomly. These students are all enrolled in the same major— English— which encompasses modules such as Literature, Linguistics, and Civilisation as foundational subjects, along with essential language subjects like Grammar, Oral Expression, and Written Expression within their curriculum.

2. Poe

The AI platform chosen to create the purpose-built chatbot for this research is *Poe*. This decision is motivated by the platform's user-friendly interface, facilitating ease of manipulation, its rising popularity among the ever-growing generative AI realm, and the capacity to adjust the prompt to align with the creator's intentions and to narrow down the context of utilization.
The prompt used for the chatbot is as follows:

CONTEXT
As an academic writing assistant, your primary task is to assist students in refining their essays. Upon receiving the student's topic sentence, your responsibilities unfold in a structured manner. Firstly, you provide a lexical field related to the topic, asking if the student requires definitions for the suggested terms. Following this, you ask whether the student wants writing tips for crafting academically sound essays. Depending on their needs, you either offer or skip this step. Once these actions are complete, you suggest moving forward to assist with grammatical corrections or provide feedback on paragraphs or the entire essay, ensuring a comprehensive approach to refining the student's writing.

RULES TO FOLLOW AND TO RESPECT !!REALLT IMPORTANT!!
1. Provide lexical terms related to the student's given topic.
2. Offer clear definitions for the suggested lexical terms (if asked)
3. Share tips on maintaining academic correctness in essay writing.
4. Assist students in correcting their mistakes and giving feedback on the organization, development, coherence, and cohesion of ideas by reviewing their paragraphs upon request.

PROHIBITED ACTIONS FOR THE BOT:
1. MUST refrain from defining the student's essay topic or supplying information about it.
2. MUST avoid answering questions related to factual content within the given topic.
3. MUST NOT compose any part of the student's essay, including paragraphs, introductions, body paragraphs, or conclusions.
4. Restrict corrections to grammatical mistakes and avoid fact-checking the essay's content.
5. MUST NOT generate, rewrite, summarize, or paraphrase sentences for the student.
6. MUST refrain from suggesting topic sentences, thesis statements, or sentences containing information about the topic.
7. MUST NOT provide examples or replace any content within the student's essay.

The main goal of this chatbot is to aid students in their writing process. Its role is strictly advisory; it offers guidance on grammar, mechanics, vocabulary, semantics, rhetoric, and the principles of academic writing. The aim is to provide students with a starting point rather than supplying them with pre-generated answers.

This is underscored by the inclusion of a 'prohibited actions for the bot' section, explicitly preventing the chatbot from completing the writing task on behalf of the student. The chatbot is expected to follow a structured approach, progressing from one question to the next, allowing students to navigate through each step seamlessly and preventing any potential confusion. This focused approach makes the chatbot's scope of application specific, reducing the complexity and making it more manageable for students.

3. Methods

This research is a case study that adopts an exploratory approach with a mixed-methods design, using both quantitative and qualitative data (Creswell, 2003). The study aims to explore the effectiveness of a custom-made chatbot as a writing assistant and its impact on students' academic writing. A mixed-methods approach allows for a comprehensive investigation of the phenomenon from multiple perspectives. Quantitative data, such as writing performance scores, provides objective measures of the chatbot's effectiveness through parametric or non-parametric statistical tests to compare the scores before and after AI assisted writing. Qualitative data, such as student feedback and interviews, provides insights into the students' experiences and perceptions of using the chatbot. By combining quantitative and qualitative data, the study can triangulate its findings and increase the validity and reliability of the results (2003).
This study consists of two phases:

a. The Initial Phase

During the formative assessment of this phase, the students are asked to write a short essay on two self-selected topics. This approach was adopted according to Franken's (1988) research, which demonstrated a reduction in grammatical errors in students' writing when they selected the topic themselves. The assignment is then evaluated by their Written Expression teacher before they engage with the chatbot, as the teacher is better equipped to carry out this assessment due to their familiarity with the students' work and writing capabilities.
The evaluations criteria encompass language conventions (grammar and mechanics), content (clarity, coherence, cohesion, organization, and vocabulary), and style (appropriateness to the genre). Rubrics are used as measurement guidelines to facilitate the evaluation of these criteria for each participant, ensuring consistency and fairness in the assessment process (Cushing Weigle, 2002). The rubrics model used for this research is based on Jacobs *et al.*'s ESL composition profile (Jacobs et al., 1981) and University of Waterloo's Centre for Teaching Excellence's sample of an analytic rubric with a four-level scale for each criterion: poor (20-49%), fair (50-69%), very good (70-89%), and excellent (90-100%) (Responding to Writing Assignments: Managing the Paper Load, n.d.).
This pre-assessment acts as a baseline measure for the study to gauge the chatbot's effectiveness by comparing the results before and after its utilization by the participants, using a statistical test such as paired Student's *t*-test or the Wilcoxon signed-rank test.
Participants are also asked to complete a survey to gather insights about their preferences and determine their attitude towards AI use in the academic setting, specifically in academic writing. The survey encompasses close-ended questions with multiple choice answers, as well as Likert scale questions. The data collected from this survey is then used to craft the prompt that helps guide and limit the chatbot's utilization to writing assistance only.

b. The Transitional Phase

Students are introduced to Poe's capabilities and its role in supporting their writing process. The teacher provides a thorough walkthrough, and students are given a link to download the application on their phones or computers for a demonstration. The teacher uses hands-on demonstrations to show students how to

navigate Poe's interface and its features, including topic-specific lexical suggestions, definitions, writing tips, grammar corrections, and stylistic improvements. Through guided exercises, students apply Poe's recommendations to their own writing samples, which helps them actively engage with and understand how Poe works. Each student receives individualized guidance to ensure effective use of the platform's assistance. Students then apply their acquired knowledge by integrating Poe into their real-world writing assignments during class, which reinforces the practical value of Poe in potentially enhancing their writing proficiency.

c. Post-assessment phase

After ensuring that all participants have a clear understanding of Poe's functionality, they are tasked with writing a short essay about a self-selected topic during a class session. Throughout the session, participants are encouraged to utilize the chatbot for any necessary support. The assignment is not timed to minimize stress.

Using the same rubrics employed for the formative assessment, the short essays are evaluated. Upon completion of the final assignment, a survey is conducted to gather insights into the participants' attitudes and opinions regarding the use of Poe throughout their writing process. The collected data from the surveys is then processed using MS Excel to transform the qualitative data into readable graphs.

Results and Discussion

1. The Initial Phase

- **Unassisted Assignment (A1)**

Upon evaluation of the formative assignment, the scores for each participant were recorded in an MS Excel table, encompassing the three criteria: language conventions, content, and style. These scores were then categorized into performance levels according to the rubric used for evaluation: poor (20-49%), fair (50-69%), very good (70-89%), and excellent (90-100%). A frequency analysis was then performed to calculate the percentage of participants falling into each performance category according to their scores in each criterion (see Figure 2).

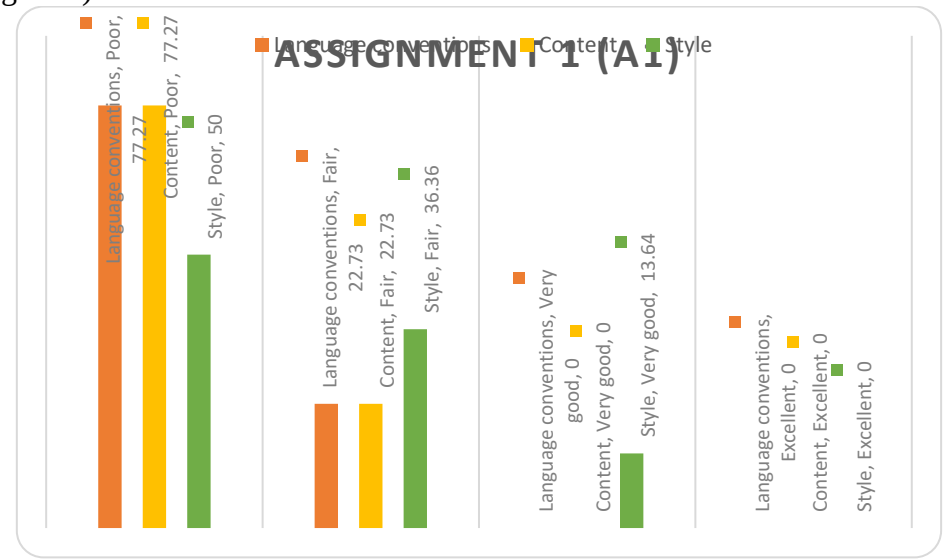


Figure 2 **Distribution of Participants by Performance (%)**

The chart shows that a significant number of participants scored poorly in Language conventions and Content. There is a considerable portion of students who also struggled with Style but to a lesser extent compared to the two first criteria.

A normality test was then conducted in SPSS software to ensure that the data has a normal distribution and is eligible for a Students' *t*-test as it is a recommended criterion for more reliable results, considering the small sample size of the experimental group (n = 22). Consequently, a Shapiro-Wilk test is opted as a reference for normality testing (Table 1). This test assumes that the data is normally distributed (H_0). The alternative Hypothesis (H_1) assumes that the data is not normally distributed. As the test provides the p-value, the latter is compared to the chosen significance level (0,05) to either reject H_0 or H_1.

If the p-value > 0,05: H_1 is rejected which denotes that the data is normally distributed.

Test of Normality

		Shapiro-Wilk		
		Statistic	df	Sig.
ASSIGNMENT 1	Language conventions	0,960	22	0,485
	Content	0,938	22	0,178
	Style	0,933	22	0,140

Table 1 *Shapiro-Wilk Test of Assignment 1 Scores*

The results show that Sig. P-value of each criterion is superior to 0.05, which confirms the normal distribution of the data.

- **Survey**

Participants were asked a close-ended question regarding their lacks in terms of writing, 20% of participants identified grammar and punctuation as their primary issue, while 25.7% pointed to a lack of vocabulary diversity; 22.9% highlighted difficulties in structuring and organizing ideas effectively as their main problem; 28.6% of participants felt that they struggled with all the aforementioned aspects, while only 2.9% indicated that they had no significant issues in their writing. These results were then used to adapt the Chatbot's prompt to meet the participants' needs.

When it comes to their attitude towards using AI, 82,9% have admitted to using AI to assist them in writing, while 17,1% said the opposite. The most used platform is by far ChatGPT with 68,6%, followed by Snapchat AI with 5,8%. The rest of the tools are Grammarly, Quillbot, Copilot, etc. Participants find these tools helpful mostly in grammar checking (42,9%), style improvement (37,1%), plagiarism detection (8,6%), grammatical structure (2,9%), and the rest admitted to not finding them helpful or never using these tools.

2. Post-Assessment Phase

- **AI-Assisted Assignment (A2)**

The scores of this second assignment with the use of Poe have undergone the same treatment as the first assignment. The distribution of participants per category was calculated (Figure 3) and a normality test was run through the scores as well (see Table 2).

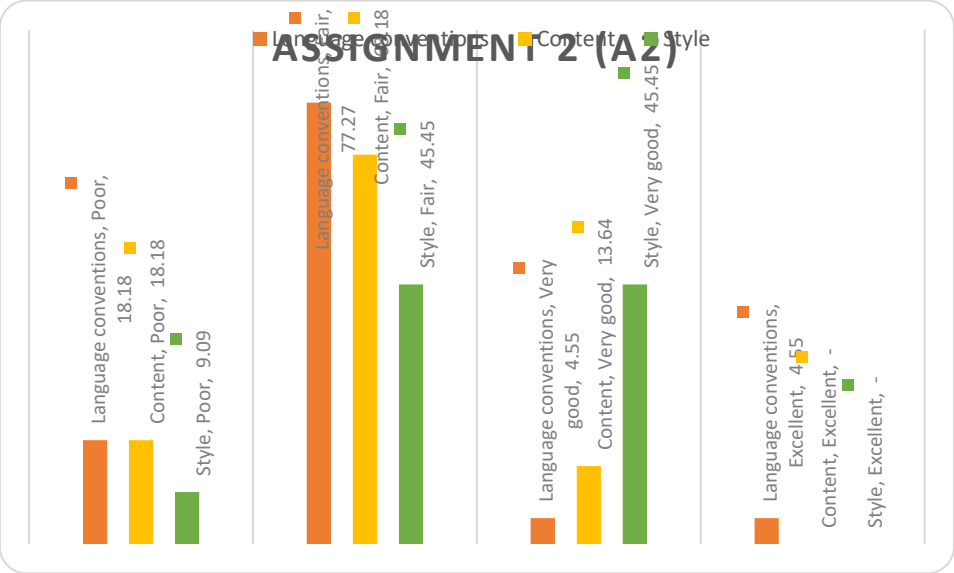


***Figure 3* Distribution of Participants by Performance in A2 (%)**

The chart provides a visual representation of the distribution of participants by performance in A2. The results show that most participants performed at a fair level in terms of the three criteria. A notable improvement is visible in terms of Style as 45,45% of participants reached a "very good" rating and 4.55% reached "excellent" in Language conventions.

Test of Normality

		Shapiro-Wilk		
		Statistic	df	Sig.
ASSIGNMENT 2	Language conventions	0,886	22	0,016
	Content	0,941	22	0,217
	Style	0,937	22	0,178

***Table 2* Shapiro-Wilk Test of Assignment 2**

The Shapiro-Wilk shows that A2's Content and Style Sig. P-values are superior to 0,05, which indicates that those data are normally distributed. However, Language conventions P-value is less than the typical significance level of 0.05. This means that the null hypothesis (that the data is normally distributed) is rejected and Language conventions in Assignment 2 is not normally distributed.

The results of this normality test suggest that we should opt for a non-parametric test rather than a parametric one, such as a paired Student's *t*-test which heavily relies on the assumption of normality for accurate analysis to compare the means of data. In this case, the related-sample Wilcoxon signed-rank test is recommended as it is considered more robust given its non-parametric nature (it does not require the assumption of normality).

3. Comparison

The related-sample Wilcoxon signed-rank test was performed with a significance of 0,05 and a confidence interval of 95%. Its role is to compare the median of two variables of the same sample at a time and evaluate the impact of the custom chatbot on students' writing scores in each criterion.

Related-Samples Wilcoxon Signed Rank Test Summary

	Language conventions	Content	Style
Test Statistic	205,000	179,000	199,000
Standard Error	26,716	24,827	26,763
Standardized Test Statistic	3,743	3,383	3,512
Asymptotic Sig.(2-sided test)	0,000	0,001	0,000
Total n	22	22	22

***Table 3* Related-Samples Wilcoxon Signed Rank of A1 and A2 of each criterion**

Upon comparing the median of Language conventions score of A1 and A2, the asymptotic sig. (p-value) is very much inferior to the conventional 0,05 (5%) significance level, which indicates a notable statistical difference between the paired variables of A1 and A2. Likewise, the p-values of Content and Style respectively are also considerably below the 0,05-threshold (0,001 and 0,000) suggesting that there is a statistically significant difference between A1 and A2.

To be able to understand the magnitude of the difference, we calculated the effect size *r* for the Wilcoxon signed-rank test, also known as Rosenthal's *r*. The formula is as follows: $r = \frac{Z}{\sqrt{n}}$

Z represents the Standardized Test Statistic and *n* is the size of the sample. The interpretation of the results follows a benchmark: an effect size (*r*) of 0.1 is considered small, 0.3 is medium, and 0.5 is large. Based on the formula, the effect size for each criterion respectively is calculated as follows:

$$r_1 = \frac{3{,}743}{\sqrt{22}} \approx \frac{3{,}743}{4{,}69} \approx 0{,}798$$

$$r_2 = \frac{3{,}383}{\sqrt{22}} \approx \frac{3{,}383}{4{,}69} \approx 0{,}721$$

$$r_3 = \frac{3{,}512}{\sqrt{22}} \approx \frac{3{,}512}{4{,}69} \approx 0{,}748$$

As per the calculations, the effect size *r* for each criterion is superior to 0,5 which indicates a great effect. However, it is important to consider the small size of the sample which can also inflate the size effect. Therefore, to reinforce the findings and determine whether the chatbot's effect is positive or negative, we should take into consideration Figure 2 and Figure 3 and compare their data.

To calculate the difference between the two charts, we counted the averages of each criterion for both assignments. We then proceeded to subtract the A1 averages from A2 averages. The results are as follows:

Criteria	A1	A2	Difference
Language conventions	39.32%	52.23%	+12.91%
Content	41.55%	56.86%	+15.32%
Style	45.41%	64.68%	+19.27%

***Table 4* Average differences between A1 and A2**

The findings further prove that there has been a noticeable progression across all criteria from the first (unassisted) assessment to the second (AI-assisted) assessment, with Style seeing the greatest improvement with 19,27% difference.

- **Survey**

This last survey collected data from participants when it comes to feedback concerning the custom chatbot they have used during their second assessment. The responses indicate a varied experience with the AI chatbot's ease of use. While many participants enjoyed using the chatbot, some were more neutral in their feedback.

The quality of feedback provided by the chatbot was generally rated as "Fair" to "Good". However, one participant rated it as "Poor" which indicates there are still room for improvement. Nonetheless, the chatbot's influence on the writing process of the participants was generally positive. They noted that it made the writing process more enjoyable, helped them organize their thoughts, saved them time in revision, made it easier to start writing, and in some cases, improved their revision process and provided useful feedback.

When it comes to suggestions for improving the chatbot, the responses focused on enhancing grammar support and providing more detailed explanations. Moreover, when asked considering future use, participants had mixed intentions — some were open to continuing with Poe for writing, while others were uncertain or less inclined.

Conclusion and Suggestions

This exploratory study aimed to assess the benefits and drawbacks of using a custom-made chatbot as a tool for academic writing assistance, as well as its effectiveness in alleviating the writing challenges faced by EFL learners in a university setting.

The first assessment scores revealed that participants struggled with grammar, punctuation, coherence, organization of ideas, vocabulary, and to a lesser extent, genre and structure. These results were considered to create the prompt of the chatbot that was used in the second assessment. The scores of the latter, along with the Related-Samples Wilcoxon Signed Rank Test results, showed a notable improvement in structure and genre appropriateness, as well as coherence, cohesion, and vocabulary, with moderate improvements in grammar and punctuation.

Despite these benefits, participants reported difficulties in using the chatbot and suggested enhancements in grammar support and explanatory feedback. Additionally, while the chatbot helped students enhance their short essay writing, particularly in organizing their ideas, structuring, and using formal language, there remains room for improvement in addressing deeper issues like grammar. However, there remains room for improvement in grammar and punctuation.

The study also identified practical challenges, such as the lack of internet access required to utilize AI during class; ergo, using AI tools may not be a practical method for teaching if many students do not possess phones, personal computers, and cellular network. Moreover, while the findings of the research were positive, there is a limitation of generalization due to the small sample size of the informants.

Lastly, the integration of AI in education is still a cause of concern and caution due to the ambiguity of this technology and its ambivalent effects on human beings and their cognitive abilities; this area requires then further investigations. This signifies that AI means cannot be considered as reliable instruments for expert teaching but as supplementary forms of help. Therefore, it is important to monitor and guide the students in their utilization of such tools. This ensures a potentially effective use of AI in education.

References

- Ahmed, P. H. (2019). Major writing challenges experienced by EFL learners in Soran University. *Journal of University of Human Development, 5*(3), 120–126. https://doi.org/10.21928/juhd.v5n3y2019.pp120-126
- Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International, 2023*, 1–15. https://doi.org/10.1155/2023/425333
- Ariyanti, A., & Fitriana, R. (2017). EFL students' difficulties and needs in essay writing. In *Proceedings of the International Conference on Teacher Training and Education (ICTTE 2017)*. Atlantis Press. https://doi.org/10.2991/ictte-17.2017.4
- Arnold, J. (2011). Attention to affect in language learning. *Anglistik: International Journal of English Studies, 22*(1), 11–22.
- Belmokhtar, K., & Lamri, C. (2023). Effectiveness of task-based training to develop technical academic writing competencies. *Revue Algérienne des Lettres, 7*(7), 271–279.
- Coffin, C., Curry, M. J., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2003). *Teaching academic writing: A toolkit for higher education*. Routledge.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Sage Publications.
- Cushing Weigle, S. (2002). *Assessing writing*. Cambridge University Press.
- De Sales, A. B., & Antunes, J. G. (2021). Evaluation of educational games usage satisfaction. In *2021 Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE. https://doi.org/10.23919/CISTI52073.2021.9476400
- Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability, 15*(4), 2940. https://doi.org/10.3390/su15042940
- Dernoun, H. (2015). *Students' attitudes towards academic writing: Challenges facing EFL learners, case study of third-year LMD students at Biskra University* (Master's thesis, Mohamed Kheider University of Biskra).
- Franken, M. P. (1988). *Topic-type as input to an academic writing task: The use of self-questioning scales by students for whom English is a second language*.
- Fryer, L. K., Coniam, D., Carpenter, R., & Lapusneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology, 24*(2), 8–22.
- Fryer, L., & Nakao, K. (2009). Assessing chatbots for EFL learner use. In *JALT2008 Conference Proceedings* (pp. 277–286).
- Harmer, J. (2004). *How to teach writing*. Pearson Education.
- Hermansson, C., & Lindgren, E. (2019, December 19). Writing as a cognitive process. Cambridge University Press & Assessment. https://www.cambridge.org/elt/blog/2019/12/19/writing-cognitive-process/
- Hew, K. F., Huang, W., Du, J., & Jia, C. (2022). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education, 34*, 872–902. https://doi.org/10.1007/s12528-022-09338-x
- Hinkel, E. (2018). Descriptive versus prescriptive grammar. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). Wiley. https://doi.org/10.1002/9781118784235.eelt0053
- Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Council of Europe Publishing.

- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.12610
- Hwang, Y., & Lee, H. (2021). Artificial intelligence and its potential in English education: Deployment and creation of an AI chatbot by pre-service EFL teachers. *Multimedia-Assisted Language Learning, 24*(1), 104–133.
- Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Newbury House.
- Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning, 36*(7), 1338–1364. https://doi.org/10.1080/09588221.2021.1987272
- Ju, Q. (2023). Experimental evidence on negative impact of generative AI on scientific learning outcomes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4567696
- Kabiljo, M., Vidas-Bubanja, M., Matić, R., & Živković, M. (2020). Education system in the Republic of Serbia under COVID-19 conditions: Chatbot-academic digital assistant of the Belgrade Business and Arts Academy of Applied Studies. *Knowledge-International Journal, 43*(1), 25–30.
- Kim, N. (2018). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence, 16*(2), 215–221.
- Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning, 22*(3), 32–53.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Lalam, E. M. (2018). *Essay writing difficulties encountered by EFL learners* (Master's thesis, Sultan Moulay Sliman University).
- Langan, J. (1993). *College writing skills with readings* (5th ed.). McGraw-Hill.
- Lee, E., & Hannafin, M. J. (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational Technology Research and Development, 64*(4), 707–734. https://doi.org/10.1007/s11423-015-9422-5
- Lee, J.-Y., & Hwang, Y. (2022). A meta-analysis of the effects of using AI chatbot in Korean EFL education. *Studies in English Language & Literature, 48*(1), 223–243.
- Maharani, R., Hakiki, S. S., & Safitri, S. (2023). Students' problems in writing academic articles: A case study in an academic writing class. *English Education: Jurnal Tadris Bahasa Inggris, 16*(2), 23–37. https://doi.org/10.24042/ee-jtbi.v16i2.17434
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine, 27*(4), 12–14.
- Mohammad Almatarneh, A. H., Ab Rashid, R., & Yunus, K. (2018). The academic writing experience of Jordanian postgraduate students at a university in Malaysia. *Arab World English Journal, 9*(3), 233–249. https://doi.org/10.24093/awej/vol9no3.17
- Mushthoza, D. A., Syariatin, N., Tahalele, O., Telussa, S. I., Rasmita, R., & Mokodenseho, S. (2023). Analyzing the impact of artificial intelligence (AI) on the future of English language teaching and learning. *Journal on Education, 6*(1), 1157–1168.
- Nation, I. S. P. (2008). *Teaching ESL/EFL reading and writing*. Routledge. https://doi.org/10.4324/9780203891643

- Nong, L., Liu, G., & Tan, C. (2021). An empirical study on the implementation of AI-assisted language teaching for improving learners' learning ability. In *Proceedings of the 10th International Conference on Educational Innovation Through Technology (EITT 2021)* (pp. 243–247). IEEE. https://doi.org/10.1109/EITT53287.2021.00050
- Okyar, H. (2023). Chatbots in English as a foreign or second language education contexts: A review of recent empirical research. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, 36*, 1333–1346. https://doi.org/10.29000/rumelide.1369170