International Journal of Basic and Applied Sciences, 14 (7) (2025) 19-32

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/537pxe76
Research paper



Leveraging A Large Language Model for Precision Enhancement in EFL Email Writing: A Quasi-Experimental Study

Mohd Nazim [1], Taj Mohammad [2], Ali Abbas Falah Alzubi [3] *, Soada Idris Khan [4]

[1] *Associate Professor, Department of English, College of Languages and Translation, Najran University, Saudi Arabia*
[2] *Associate Professor, English Skills Department, Preparatory Year, Najran University, Najran, Saudi Arabia*
[3] *Assistant Professor of Applied Linguistics, Department of English, College of Languages and Translation, Najran University, Saudi Arabia*
[4] *Teacher and trainer, Sanabel Al Noor International School Najran, Saudi Arabia*
**Corresponding author E-mail: aliyarmouk2004@gmail.com*

Received: October 13, 2025, Accepted: October 22, 2025, Published: November 2, 2025

Abstract

Writing formal emails requires clarity, conciseness, and accuracy. Many English as a Foreign Language (EFL) learners find these skills challenging. Large Language Models (LLMs), such as ChatGPT, represent a significant advancement in natural language processing. However, their effectiveness in structured, task-specific language tasks for education is not well studied. This study investigates whether the GPT-3.5 model can assist EFL students in writing formal emails with greater precision and structural accuracy. The research treats the classroom as a real-world test case for the model's abilities. It applied a quasi-experimental method. Data were collected from control and experimental groups (N = 60). The study used a test and semi-structured interviews. The results show ChatGPT improved formal email writing among EFL students. Qualitative findings highlight the tool's ease of use and time-saving nature. ChatGPT also boosts clarity and confidence. Some concerns about overreliance on ChatGPT emerge. The tool sometimes struggled with subject lines, sign-offs, and signatures. This suggests that human intervention is needed. Contributing to AI-assisted pedagogy, this study shows how ChatGPT can improve EFL students' formal email writing skills.

Keywords: *AI-Powered Precision; ChatGPT; EFL students; Formal email writing skills; Large Language Models.*

1. Introduction

Writing is an important talent for academic advancement, professional communication, and personal expression [1]. As a reflective process, it promotes critical thinking and creativity while requiring organized thought, clarity of communication, and a solid command of language aspects to attain effective learning outcomes [2]. The incorporation of information technology (IT) into writing instruction has recently transformed teaching approaches by increasing engagement and accessibility with technologies such as ChatGPT, facilitating personalized learning and collaboration [3]. However, effective implementation of ChatGPT-mediated instruction necessitates careful planning to ensure that technology enhances rather than substitutes essential writing skills [2].

Among the various forms of writing, formal email communication has been a vital component of formal interaction in the digital age. Establishing and maintaining professional relationships, exchanging critical information, promoting organizations, and improving performance all rely on effective formal email writing skills. Despite its central role, EFL students often struggle to achieve clarity, conciseness, and precision in formal email writing [4]. This poses a particular challenge for EFL learners, especially in the context of Saudi Arabia. Email has long been regarded as the most useful and favored tool for formal communication due to its ability to be corrected, detailed, organized, and allow for extended phrases. As part of a syllabus, email offers numerous advantages for L2/FL learners, particularly in the Najran university setting with facilities of internet and electronic gadgets; e-mail provides rapid feedback and allows students to discuss and communicate directly, inexpensively, swiftly, and consistently. In addition, students have greater influence over message planning, composition, editing, and delivery than in face-to-face communication [4].

According to previous studies, email can help with communication, social learning facilitation, and writing ability improvement [5]. The integration of AI-generated content, including emails, offers significant potential for both businesses and educational institutions. An increasing number of professionals and students are turning to artificial intelligence to compose their electronic correspondence [6]. Email offers L2/FL learners a real audience, extending language learning time and space, facilitating real-world communication and fast feedback

Copyright © Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[4]. It encourages them by bringing the content and subject matter to life, allowing them to draw critical links between the classroom and the real world [4]. To accomplish intended outcomes and increase collaboration, EFL learners as future professionals must be able to compose emails that are clear, concise, and suitable.

The incorporation of AI-generated content, including emails, holds immense promise for both enterprises and educational institutions. A rising number of professionals and students use artificial intelligence to write their electronic letters [7]. LLMs, such as Google's Gemini and OpenAI's ChatGPT, have grown in popularity in recent years, owing to technological maturation that has made them viable for general use. One of the most practical applications of these tools has been in business communication, where LLMs assist professionals in drafting clear, concise, and contextually appropriate emails. Although others existed before it, ChatGPT was the game-changer for users and AI enthusiasts by allowing anyone to sign up for free, open a chat dialogue, and begin receiving legible LLM output without having to understand the concept or prompt engineering [8].

From a technical perspective, ChatGPT is built upon the Generative Pre-trained Transformer (GPT) architecture, a class of LLMs known for their deep learning capabilities. These models are trained on vast corpora of text data and utilize a transformer-based neural network to predict and generate human-like text sequences. The model's ability to assist in formal email writing can be understood as a function of its natural language generation (NLG) and in-context learning capacities. In this study, we investigate the potential of these capabilities in a constrained, generous, genre-specific task, i.e., email writing, assessing not only student improvement but also the model's proficiency in generating contextually appropriate, structurally sound, and precise linguistic outputs. Several corporate sectors have already begun to investigate the possible impact of AI on their respective disciplines. It is critical that communication educators, particularly business educators, understand how these new technologies will affect corporate communication and teach students about the uses, functions, and ethics of AI in the workplace [9].

Despite the numerous benefits, critics warn of potential drawbacks such as students abusing ChatGPT for assignments, ethical concerns about plagiarism, and the risk of AI replacing essential cognitive skills such as critical thinking and problem-solving [10]. Nonetheless, while ChatGPT has been employed for a variety of text processing tasks, little research has investigated its role in enhancing formal email writing skills among EFL students, particularly at Najran University in Saudi Arabia.

To achieve the objectives of this research, the researchers examine three key aspects of how effectively ChatGPT enhances EFL students' formal email writing skills. First, it compares the post-test scores of the experimental and control groups to determine the impact of ChatGPT on EFL students' business formal email writing skills. Second, it seeks to establish baseline equivalence between the two groups by examining their pre-test scores before the intervention, thereby confirming the comparison's validity. Finally, the study investigates students' experiences and perceptions of utilizing ChatGPT as a learning tool for writing formal emails, with a focus on subject lines, salutations, content development, and closings.

1.1 Research questions

1. How effective is the use of an AI-powered tool for improving EFL students' business writing skills?
2. What are students' experiences regarding the integration of ChatGPT in the process of formal email writing?

2. Theoretical Framework

This study is based on Computer-Assisted Language Learning (CALL) theory, which provides a conceptual framework for investigating how technological tools can help second and foreign language acquisition. CALL argues for the use of digital technologies in language education to provide students with authentic input, timely feedback, meaningful engagement, and chances for self-directed learning [11]. Within this paradigm, the usage of ChatGPT as an AI-driven writing aid is extremely important for improving EFL students' writing skills. ChatGPT, as an interactive platform, allows learners to create, modify, and improve their written work, which aligns with CALL's basic concepts of enhancing linguistic clarity, fluency, and learner independence [12].

ChatGPT supports CALL's emphasis on situated learning and task authenticity, two essential elements for transferring skills outside of the classroom by assigning students writing assignments that mimic real-world situations [12]. Additionally, CALL emphasizes how crucial it is to empower students at every stage of the writing process [5]. By enabling students to plan, draft, edit, and receive feedback at their own pace, ChatGPT helps students become more competent and confident in their ability to communicate professionally [13]. In this sense, CALL not only guides the pedagogical application of ChatGPT but also emphasizes how it connects traditional EFL instruction with digitally enhanced learning settings. This study shows how students can develop precision, professionalism, and clarity of their paragraph writing, skills, necessary for academic success, by strategically utilizing AI tools.

While the CALL framework provides pedagogical context, the mechanism of interaction is fundamentally driven by advances in computational linguistics and NLP. The students' engagement with ChatGPT is, in essence, a human-computer interaction (HCI) loop involving prompt engineering, model inference, and output refinement. The model's transformer architecture allows it to process complex student prompts and generate coherent, multi-part texts (emails) that adhere to genre conventions. Therefore, this study sits at the intersection of applied linguistics and the applied science of NLP, using a language learning task to evaluate the practical utility of a specific AI model's architectural strengths and limitations.

2.1 Literature review

Writing is a focused activity in which the author uses alphabetic letters, symbols, and signs to convey information to an audience. It is one of the four fundamental English language skills, along with speaking, listening, and reading [14]. It is also one of the most difficult activities as Elliot and Williamson [15] emphasize the difficulty of writing by stating, "writing is among the most complex of human behaviors" (p. 2). It is important to develop writing skills for EFL students. The primary goal of writers should be to maintain a good academic writing style, which includes a careful selection of language, logical progression from one idea to another, and a clear, honest, and objective presentation of ideas [14]. A good command of writing skills assists students in developing academic writing.

Email writing is one of the examples of academic writing that EFL students need to master, as it is a very important part of academic writing courses. Nashruddin et al. [16] conducted a study to know the perceptions of email writing, which reflects its usefulness. The findings revealed the teacher's perspectives about the use of e-mail in learning: email is an effective device, e-mail can be used to communicate various documents, e-mail can be utilized to meet learning objectives, and e-mail facilitates learning. Students' impressions of the use of e-mail include that it may be utilized to obtain learning materials, it is simple to use, and students can access many assignments

via email. However, a tiny minority of students had problems using email as a learning medium. Therefore, simplifying writing, no matter academic or professional, has always been challenging for EFL students. To facilitate students, many techniques have been introduced from time to time.

Integration of technology and AI-mediated tools in the EFL writing classroom is the best example of it. The pedagogical applications of tools like ChatGPT are underpinned by significant advancements in the field of NLP and the development of LLMs. The paradigm shift was largely catalyzed by the introduction of the transformer architecture [17], which moved away from previous recurrent models by using a self-attention mechanism. This allows the model to weigh the importance of all words in a sentence simultaneously, enabling more efficient training on vast datasets and a superior grasp of long-range contextual dependencies.

This architectural innovation paved the way for generative pre-trained models. Generative Pre-trained Transformers (GPTs) are a class of LLMs that utilize a decoder-only transformer architecture and are trained using a self-supervised, autoregressive next-word prediction objective on massive, diverse text corpora [18]. This pre-training phase equips the model with a broad, albeit shallow, understanding of grammar, facts, reasoning patterns, and stylistic conventions present in the data. The model used in this study, ChatGPT, is an application of the GPT-3.5 series, fine-tuned using Reinforcement Learning from Human Feedback (RLHF) to better align its outputs with human instructions and preferences [20]. This technical foundation is critical for understanding its capabilities: the model's ability to assist in formal email writing is a direct application of its in-context learning ability, where it uses the prompt as a context to generate a coherent, structured, and contextually appropriate text sequence, mimicking the genre conventions it learned during training [19].

An AI-assisted language learning application was created by Gayed et al. [21] to help EFL students overcome cognitive obstacles and enhance their writing abilities. The results showed that the AI-powered language learning tool successfully improved students' writing abilities and decreased the cognitive obstacles they faced when completing writing assignments. Additionally, Liu et al. [22] investigated the effect of AI on the writing abilities of EFL learners using a quasi-experimental research approach. The results showed significant gains in writing abilities as compared to the traditional curriculum. Effective writing performance was also facilitated by the AI-supported language learning technique, which also improved learners' self-efficacy, self-regulated learning, and decreased cognitive load.

Using AI-mediated tools in the writing classroom has been very rewarding for students. Ng et al. [23] conducted a study that investigated the impact of Gemini on the writing abilities and motivation of EFL learners. 60 Omani learners were divided into treatment and control groups, each with 30 learners. The treatment group received extra feedback from Gemini, leading to improved performance. The study highlights the importance of Gemini in enhancing EFL learners' motivation and writing abilities. Likewise, Wordcraft, an AI-powered text editor, was used by Ippolito et al. [24] to assist experienced authors with creative writing activities, specifically brainstorming. The outcomes showed that the AI-powered tool had a beneficial effect on creative writing.

Lam and Moorhouse [25] discovered that WordTune efficiently helped students identify their writing deficiencies, boosting self-assessment and learning. This tool goes beyond mere grammatical correction and explores the stylistic aspects of writing. Jenni is another AI-powered writing helper that makes predictive text suggestions to help users write emails, reports, articles, and other documents more effectively.

Liu et al. [22] investigated user perceptions of AI-Mediated Communication (AI-MC), with a particular focus on AI-generated messages. Through large-scale surveys and in-depth interviews, their studies examine how various factors, especially interpersonal emphasis, shape trust in AI-mediated email exchanges. Findings indicate that awareness of AI involvement generally decreases trust, but surprisingly, trust increases when AI is used for more interpersonal emails compared to transactional ones.

Marzuki et al. [26] conducted research using a qualitative methodology and organized it as a case study. Data on the variety of AI writing tools and their effects on Indonesian university students' writing quality were gathered through semi-structured interviews. According to the study, EFL teachers found that AI technology helped students manage emails and other business correspondence, which increased professionalism and clarity. According to Kohnke et al. [13, p.3], it can "identify the meaning of a word in context, correct and explain language mistakes, create texts in various genres (e.g., emails, stories, recipes), develop quizzes, annotate texts, and offer dictionary definitions, example sentences, and translations."

ChatGPT is one of the AI tools that has been very effective in improving the writing skills of EFL students, which are crucial in the development of academic and professional writing. Nguyen T.C. [10] evaluated Vietnamese university teachers' grasp of ChatGPT, as well as its perceived benefits and drawbacks. The study employed an online questionnaire to collect responses from English teachers representing various Vietnamese universities. The results demonstrated that ChatGPT is beneficial in improving students' writing skills, although concerns such as academic dishonesty and student laziness were discovered.

Romadhon [29] conducted a quantitative survey of ESP students and discovered positive reviews of ChatGPT for formal email writing, with the majority indicating an increased, fun, and engaging learning experience. The study proposes incorporating ChatGPT into curricula to improve business writing, as well as recommending that teachers control AI overreliance through balanced usage and human contact, and suggests additional research into optimizing its educational effects.

Although recent studies highlight the potential of AI to support EFL writing development, researchers such as Touvron et al. [28] and Al-Afnan et al. [27] emphasize notable limitations. AI tools like ChatGPT often struggle with conveying emotional nuance, pragmatic appropriateness, and culturally embedded communication norms—particularly in professional contexts. In business and intercultural correspondence, AI-generated texts may appear impersonal, lacking the warmth and contextual sensitivity expected in human exchanges. These challenges suggest that, despite linguistic accuracy, AI outputs may fall short of meeting socio-pragmatic expectations across diverse settings. Furthermore, these findings inform ongoing debates about the ethical, pedagogical, and cognitive dimensions of AI integration in education. While AI offers efficiency and creative support, concerns persist about overreliance, reduced critical thinking, and the authenticity of learner-generated content. The current literature thus reflects both the promise and complexity of incorporating AI into EFL writing instruction. Quantitative studies report gains in accuracy, organization, and motivation, yet critical perspectives advocate for pedagogical approaches that are culturally responsive and ethically grounded. This duality points to the need for research that not only tracks performance outcomes but also explores the deeper cognitive and communicative dimensions of AI-assisted writing.

ChatGPT assisting EFL learners in honing their writing abilities, artificial intelligence (AI) has completely transformed language learning and writing teaching. Several studies have been conducted to investigate various features of ChatGPT. Nevertheless, there has been little research on how ChatGPT can assist students in improving their formal email writing skills, notably at Najran University in Saudi Arabia. To address this gap through a context-sensitive lens, the present study explores how ChatGPT-mediated instruction can support EFL learners in developing email writing skills, while critically engaging with the cultural and communicative limitations inherent in AI-generated discourse.

3. Methodology

3.1 Research design

This study used a quasi-experimental mixed-method design to examine the impact of ChatGPT on EFL students' formal email writing skills at Najran University. A total of 60 undergraduates were divided equally into two groups: an experimental group that received ChatGPT-mediated education and a control group that was taught using traditional methods. Both groups took a pre-test to ensure baseline equality, followed by a four-week intervention and a post-test to assess outcomes. The formal email writing exam, with eight components (subject line, salutation, writer information, content, matter of concern, closing line, sign-off, and signature), was validated through expert review and item analysis, demonstrating good internal reliability ($\alpha = 0.94$). Independent sample t-tests and effect size estimates were used to assess quantitative data, while semi-structured interviews were conducted to gather students' experiences of ChatGPT's usefulness.

To ensure the internal validity of the quasi-experimental design and isolate the effect of the ChatGPT intervention, several potential external variables were addressed. First, to control for pre-existing differences in prior AI experience and familiarity, a brief requirements analysis survey was administered to all participants during the diagnostic and baseline phase. This survey assessed their device access, confidence in formal email writing, and previous exposure to AI tools. The results confirmed that both groups had statistically comparable and generally low levels of prior experience with advanced AI writing assistants like ChatGPT, minimizing its influence as a confounding variable. Second, the assignment of students to the control and experimental groups was managed at the class-section level to maintain ecological validity, but efforts were made to ensure baseline equivalence. As demonstrated by the pre-test results (see Table 7), no statistically significant differences existed between the two groups across all eight domains of formal email writing at the outset of the study ($p > 0.05$ for all domains). This confirms the groups' equivalence in terms of initial writing proficiency. Third, to mitigate the Hawthorne Effect (where participants change behavior because they are being studied), both groups were informed they were participating in a study to improve email writing skills, without emphasizing the AI tool as the primary variable of interest for the control group. Furthermore, both groups received the same amount of instructor attention, used the same curriculum and learning objectives, and were assessed using identical rubrics and tests. The only systematic difference was the use of ChatGPT for guided practice in the experimental group, thereby increasing the likelihood that the observed post-test differences are attributable to the intervention itself.

3.2 Population and sample

The study included 90 undergraduate students (aged 18-22) who were enrolled in the Reading and Writing course and were chosen through purposeful sampling from PY in the second semester of the 2025 academic year. Sixty students were selected from this population. Participants were selected from the same cohort and academic level to minimize variance in academic background and language proficiency. To meet the study's aims, one treatment session was selected via purposeful sampling. These students are Arabic-speaking learners as well as prospective engineers, doctors, and computer science experts who need to pass PY to continue their studies in their respective disciplines. They come to PY after finishing upper secondary school and fulfilling the enrollment requirements for Najran University. They completed their upper secondary schooling at Najran. They studied English as a foreign language and were classified as PY level between elementary and upper intermediate according to a diagnostic test.

The study's ethical approval number is 0076-00076-DS. Before providing consent, the student participants were fully informed about the research process. If they agreed, participants might withdraw or skip any questions at any time. Furthermore, participants were free to ask questions about the research. They were advised that participating in the study would provide no direct or indirect benefits. Participants were assured that all information obtained would be kept strictly confidential and used only for the study. Furthermore, they were given the research team's contact information in case they needed any additional information or clarification.

Moreover, this research utilized Generative AI to assist with the initial analysis of qualitative data, search for authentic academic sources, proofreading, and editing. It was also used to rephrase and organize ideas, and draft and format sections like results and references.

3.3 Instruments

To answer the research questions, the researchers administered a test and conducted a semi-structured interview. The test was developed based on a review of the literature and the researchers' own teaching experience [25], [29], [31]. It was designed to assess the efficacy of the intervention program by comparing pre- and post-test outcomes. The test focused on eight key themes: subject line, salutation, information about the writer, appropriate content, matter of concern, closing line, sign-off, and signature. Using the provided rubric, students were asked to write an email to their friend enquiring about a project or write an email to their friend requesting him to provide notes of the last English class. These emails were then evaluated according to the rubric. On average, the test took about twenty minutes to complete. The researchers administered the test in the PY building to the same group of participants.

A semi-structured interview was conducted to obtain EFL students' experiences and opinions about the effectiveness of ChatGPT-mediated training in improving formal email writing skills. The interviews were conducted at Najran University during the second semester of the 2024-2025 academic year, in a designated lecture room located in the Deanship of Preparatory Year building, Language Skills Unit, which provided a comfortable, professional, and distraction-free setting.

Participants ($N = 15$) were recruited from the experimental group to guarantee a range of proficiency levels. Despite the relatively small sample size, participants were purposefully selected to enable rich exploration of individual experiences and nuanced qualitative insights. The sample encompassed diverse language proficiency levels and academic backgrounds, ensuring a wider spectrum of perspectives within the study's scope.

The researcher questioned each participant individually, with another colleague acting as an observer, recording nonverbal reactions such as hesitancy, confidence, and enthusiasm. The interviews lasted around 10-15 minutes per student and were conducted over two weeks following the post-test.

The interview focused on difficult concepts such as ChatGPT's function in enhancing writing across specific components, including subject line, salutation, writer information, appropriate content, matter of concern, closing line, sign off, and signature. The semi-structured questions were divided into two categories: experiences and opinions. The interview questions were prepared based on the researchers' teaching skills and a review of past research [27], [9], [22]. The interview protocol was structured around two important categories: experiences and opinions.

To address potential biases, particularly social desirability bias, several measures were implemented to ensure the authenticity and reliability of participant responses:
(1) Participants were assured complete anonymity and confidentiality, with no identifying information collected.
(2) The interviewer emphasized that there were no "right" or "wrong" answers, and that both positive and negative feedback were equally valuable.
(3) An independent observer was present to objectively record students' verbal and nonverbal cues.
(4) Responses were triangulated with quantitative performance data and classroom observations to enhance trustworthiness and reduce subjective bias.
These measures were implemented to ensure that students' responses authentically reflected their experiences rather than perceived expectations from the researcher.
Moreover, the researchers employed the following themes to elicit the participants' experiences and opinions.
Experience: Students were asked to describe how ChatGPT helped them compose an email. Prompts included the following:

- ChatGPT to improve various aspects of an email, such as the subject line, closing, or signature?
- ChatGPT to write better email content and express information about the writer?

Opinion: Students were invited to consider how ChatGPT helped them reform formal email writing. These included the following questions:

- Your overall experience using ChatGPT in terms of simplicity of use and user friendliness.
- ChatGPT is a time-saving tool that effectively solves frequent formal email writing problems. Why, or why not?
- ChatGPT to impact your confidence and ability to tackle formal email writing challenges.

3.4 Intervention

The intervention focused on whether ChatGPT-enhanced writing instruction improved Preparatory Year (PY) students' formal email writing as compared to traditional instruction. For four months (February–May 2025), the experimental group (N = 30) and control group (N = 30) met twice a week for 200 minutes (100 minutes per class) in the same university language laboratory. Both groups were taught by the same instructor, used the same physical language laboratory, and followed the same schedule to control for instructor style and environmental factors. In addition, both groups used the identical curriculum, class objectives, activities, and evaluation rubrics; the only difference was that the experimental group used ChatGPT for 50 minutes each session.
Instruction in both groups was based on the program's prescribed formal email writing units. It focuses on:

1. Subject line
2. Salutation
3. Information about the writer
4. Appropriate contents
5. Matter of concern
6. Closing line
7. Sign off
8. Signature.

The AI intervention utilized the OpenAI GPT-3.5 model (the engine behind the freely available ChatGPT at the time of the study). The model's parameters were accessed via the standard chat interface. The intervention was designed to test the model's in-context learning ability, where students provided task-specific prompts (e.g., 'Generate a subject line for an email to a professor requesting an appointment') and refined the outputs based on rubric guidelines.
To promote transparency and replicability, students followed a clearly structured ChatGPT-supported writing workflow consisting of five sequential steps:

1. Task comprehension: Students began by identifying the purpose and intended audience of the email.
2. Prompt design: They then crafted targeted prompts, such as: – "Write a friendly email to my classmate asking for details about our upcoming project submission." – "Generate two versions of an email to a friend enquiring about the progress of a group project."
3. Model interaction: Students evaluated ChatGPT's responses using a provided rubric, focusing on aspects such as formality, tone, and structural coherence.
4. Revision and justification: Learners rewrote the AI-generated text in their own words, making at least two modifications and explaining their rationale, whether for tone adjustment, grammatical accuracy, or cultural appropriateness.
5. Reflection and feedback: Students summarized their takeaways from the interaction and described how ChatGPT's suggestions influenced and improved their drafts.

In addition, the teacher offered model prompts and scaffolded classroom exercises. Sample tasks included:

1. "Ask ChatGPT to rewrite your email opening in a more formal tone and compare it with your version."
2. "Use ChatGPT to identify possible errors in your subject line and propose alternatives."

When workstations were unavailable or malfunctioning, students accessed ChatGPT using their personal mobile devices. To safeguard privacy, no personal data was uploaded during the sessions; prompts were referenced using anonymous task IDs. This approach positioned each student as a human-in-the-loop evaluator of the model's generative output, enabling real-time assessment of its effectiveness for a targeted, applied writing task.
The lecturer delivered a two-hour micro-workshop on themes like prompt design, safety and ethical considerations, troubleshooting, and aligning AI outputs with textbook rubrics. Throughout the class, the teacher used model lessons and supported the students as needed.
The preliminary meetings aimed to establish baselines, standards, and readiness for the usage of AI tools. Students initially completed a diagnostic and baseline phase, which included a 40-minute pre-test on writing emails without AI and a brief requirements analysis survey that assessed their device access, confidence, and previous AI experience. Following that, students had a 30- to 40-minute orientation and ethics session in which they were told what AI can and cannot do, the importance of avoiding copy-paste submissions, and the necessity of revising outputs in their own words while documenting the changes and their reasoning. Privacy considerations were emphasized, such as using anonymous task codes and avoiding using names, emails, or personal IDs in prompts. Finally, a 20-25-minute micro-training session provided a thorough overview of the procedure, from prompt development to outline, sample generation, critique, and change.
Both groups attended two weekly classes (each lasting 100 minutes) that featured basic textbook tasks, with the experimental group receiving an additional 50 minutes of ChatGPT-assisted practice that rotated through various formal email writing components. Sessions typically began with a 30- to 40-minute model-analysis phase, followed by 40- to 50-minute guided practice on writing an appropriate subject line, salutation, information about the writer, appropriate contents, matter of concern, closing line, sign off, and signature. Following

that, the experimental group worked independently on lab PCs or mobile devices for 50 minutes while receiving teacher assistance via a prompted checklist. Meanwhile, the control group did parallel assignments on paper, with input from their instructors. Both groups utilized the same rubric for peer review and editing (30-40 minutes).

Sessions concluded with 15-20-minute exit tickets with one question and one improvement. Students' ChatGPT-based revisions were systematically tracked through reflection logs, in which they recorded their original drafts, the AI-generated suggestions, and their final revised versions. These logs provided tangible evidence of learner engagement and served as a valuable resource for post-intervention analysis.

Students revised outputs in their own words and documented changes made with ChatGPT for concept development, micro-explanations, alternative phrasing, and formal email writing. Sample prompts address the following topics: subject line, salutation, writer information, appropriate contents, matter of concern, closing line, sign off, and signature. It is explicitly stated that AI writing must be transformed rather than being submitted in its original form.

The post-test was administered under the same conditions as the pre-test: a 40-minute writing task with no AI support, graded using the same rubric. The findings demonstrated a statistically significant improvement in students' formal email writing skills compared to the pre-test. To gain qualitative insights, a stratified subset of the experimental group participated in 10–15-minute semi-structured interviews. Despite a few minor technical challenges, students reported positive experiences with the intervention, including improved writing techniques and confidence. All AI records were stored using anonymized IDs and were solely used for research purposes. Students in the control group dedicated the same amount of time to teacher-led drills, textbook exercises, and manual drafting/revision as they did to textbook modules, guided practice, and peer review.

3.5 Validity and reliability

3.5.1 Test validity

The test was initially developed and presented to a group of validators and specialists, totaling ten validators, to evaluate its suitability, the appropriateness of its domains to determine the intended objectives, the formulation of the domains in accordance with test design standards, the clarity of the domains' linguistic phrasing, and the clarity of the test instructions. The validators' comments, including deletions, adjustments, and additions, were integrated, resulting in a final test form with eight domains. The final test consisted of eight domains based on the validators' response, which included item deletions, revisions, and additions. Here are both the before and after versions of validators illustrated in Table 1.

Table 1: Experts' Observations

Stage	Prompt(s)	Domains Assessed
Test before modification	Write an email to your friend inviting him to participate in a football match.	1. Subject line 2. Salutation 3. Information about the writer 4. Matter of concern 5. Closing line 6. Signature
Test after modification	Write an email to your friend enquiring about a project. Or Write an email to your friend requesting him to provide notes of the last English class.	1. Subject line 2. Salutation 3. Information about the writer 4. Appropriate contents 5. Matter of concern 6. Closing line 7. Sign off 8. Signature

The test was also administered to a group of twenty students who were not part of the study. Pearson correlation coefficients were calculated between the scores on each question and the total test score. Table 2 below illustrates the results.

Table 2: Pearson Correlation Coefficients between Domain Scores and Total Test Score

Domains	Person correlation	P – value
1. Subject line	0.728**	0.000
2. Salutation	0.471*	0.036
3. Information about the writer	0.655**	0.002
4. Appropriate contents	0.639**	0.002
5. Matter of concern	0.814**	0.000
6. Closing Line	0.560*	0.010
7. Sign off	0.702**	0.001
8. Signature	0.530*	0.016

***. Correlation is significant at the 0.01 level (2-tailed)*

**. Correlation is significant at the 0.05 level (2-tailed).*

Table 2 reveals that the correlation coefficients between the questions and the total test score ranged from (0.471* to 0.814**), which were statistically significant at the (0.05) and (0.01) levels, demonstrating that the test is consistent. The Pearson correlation values indicate that all question scores are positively and significantly correlated with the total test score. The matter of concern ($r = 0.814$, $p < 0.001$), subject line ($r = 0.728$, $p < 0.001$), and sign off ($r = 0.702$, $p = 0.001$) had the highest correlations with overall performance, suggesting its importance. There are moderate correlations between information about the writer ($r = 0.655$, $p = 0.002$), appropriate contents ($r = 0.639$, $p = 0.002$), and the ending line ($r = 0.560$, $p = 0.010$). The weakest but still significant associations are found with salutation ($r = 0.471$, $p = 0.036$) and signature ($r = 0.530$, $p = 0.016$). All components contribute considerably to the overall score, but content-related factors (matter of concern, subject line, and sign off) have the biggest impact.

3.5.2 Test reliability

The test's reliability was calculated using Cronbach's Alpha, as shown in Table 3.

Table 3: Cronbach's Alpha Coefficient

Domain	Reliability Coefficient
Total Score	0.94

Cronbach's alpha reliability analysis revealed a strong internal consistency for the total test score ($\alpha = 0.94$), indicating that the test components accurately assess overall performance.

3.5.3 Difficulty coefficients

The difficulty coefficient for test questions was calculated according to Odeh [34] using the following formula: Difficulty Coefficient = Total Domain Scores / (Number of Students * Domain Score). Table 4 shows the difficulty coefficients for the test domains based on the exploratory sample results.

Table 4: Difficulty Coefficients for Test Questions

Domains	Difficulty Coefficient
1	0.65
2	0.75
3	0.65
4	0.65
5	0.65
6	0.70
7	0.65
8	0.55

The item analysis found that the exam domains' difficulty coefficients ranged between 0.55 and 0.75, indicating moderate overall complexity. Domain 2 was the most challenging (0.75), and Domain 8 was the easiest (0.55). The remaining items (Domains 1, 3, and 7) had moderate difficulty levels ranging from 0.65 to 0.70, indicating a balanced test design. Odeh [34] states that any item with a difficulty coefficient ranging from 0.20 to 0.80 is acceptable and should be retained in the test. The total average difficulty coefficient for the test was 0.66, indicating that the test is of medium complexity [34].

3.5.4 Discrimination coefficients

The discrimination coefficients for the objective test items were calculated according to the following formula [34]: Discrimination Coefficient = (Nu – NI) / N

- Nu = Number of students from the upper group who answered the question correctly.
- NI = Number of students from the lower group who answered the question correctly.
- N = Number of individuals in one of the groups.

To calculate the discrimination coefficient for each test question, the students were divided into two groups: an upper group with 50% of the highest test scores and a lower group with 50% of the lowest test scores (10 students in each group).
According to Odeh [34], measurement specialists have established reference values for judging test items as follows:

- Items with a negative discrimination coefficient are discarded.
- Items with a discrimination coefficient less than 0.20 are recommended for removal.
- Items with a discrimination coefficient of 0.20 or higher are accepted.

Table 5 shows the discrimination coefficients for the test questions.

Table 5: Discrimination Coefficients for Test Questions

Domain	Discrimination Coefficient
1	0.70
2	0.30
3	0.30
4	0.50
5	0.70
6	0.40
7	0.50
8	0.30

The discrimination analysis found that the test items differed in their capacity to distinguish between high- and low-performing students. Domains 1 and 5 showed high discrimination (0.70), showing a strong discrimination ability. Domains 4, 6, and 7 demonstrated moderate discrimination (0.40-0.50), but Domains 2, 3, and 8 had less discrimination (0.30). Overall, the test showed an appropriate level of item discrimination.

3.5.5 Normality of distribution

The Kolmogorov-Smirnov test was used to verify the normality of the distribution of the study sample's scores on the test in both the pre-test and post-test, as shown in the following Table 6.

Table 6: Kolmogorov-Smirnov Test for Normality of Distribution of Study Sample Scores in Pre-Test and Post-Test

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
pre	Control	0.112	30	0.200*	0.953	30	0.475
	Experiment	0.108	30	0.200*	0.982	30	0.663
post	Control	0.140	30	0.135	0.934	30	0.121
	Experiment	0.164	30	0.64	0.926	30	0.107

The Kolmogorov-Smirnov and Shapiro-Wilk normality tests revealed that both the control and experimental groups' pre-test and post-test scores were normally distributed ($p > 0.05$). These findings demonstrate that the normalcy assumption was met, allowing for the use of parametric tests in future investigations.

3.6 Data analysis

The researcher used the Statistical Package for the Social Sciences (SPSS) to answer the study questions, calculating:

- Pearson correlation coefficient for consistency validity.
- Cronbach's Alpha for reliability.
- Kolmogorov-Smirnov test for normality.
- T-test for independent samples.
- Eta squared for effect size.

4. Results

4.1 Research question 1: How effective is the use of an AI-powered tool for improving EFL students' business writing skills?

The t-test for independent samples was used to determine the significance of differences between the mean scores of the control and experimental groups in the pre-test, as shown in Table 7 and Figure 1.

Table 7: T-Test for Independent Samples to Determine Significance of Differences in Mean Scores of Controls and Experimental Groups in Pre-Test

Domain	Group	N	Mean	Std. Deviation	T	Df	Sig. (2-tailed)
1. Subject line	Control	30	0.42	0.4	0.360	58	0.720
	experiment	30	0.48	0.93			
2. Salutation	Control	30	0.42	0.37	0.171	58	0.865
	experiment	30	0.40	0.38			
3. Information about the author	Control	30	1.13	1.09	1.246	58	0.218
	experiment	30	0.80	1.06			
4. Contents	Control	30	2.83	1.05	0.496	58	0.622
	experiment	30	3.00	1.51			
5. Matter of concern	Control	30	0.22	0.34	0.191	58	0.849
	experiment	30	0.20	0.34			
6. Closing line	Control	30	0.72	0.68	0.094	58	0.925
	experiment	30	0.73	0.69			
7. Sign off	Control	30	0.60	0.33	0.589	58	0.558
	experiment	30	0.65	0.33			
8. Signature	Control	30	0.83	0.38	0.000	58	1.000
	experiment	30	0.83	0.38			
Total	Control	30	8.00	2.05	0.320	58	0.750
	experiment	30	7.80	2.75			

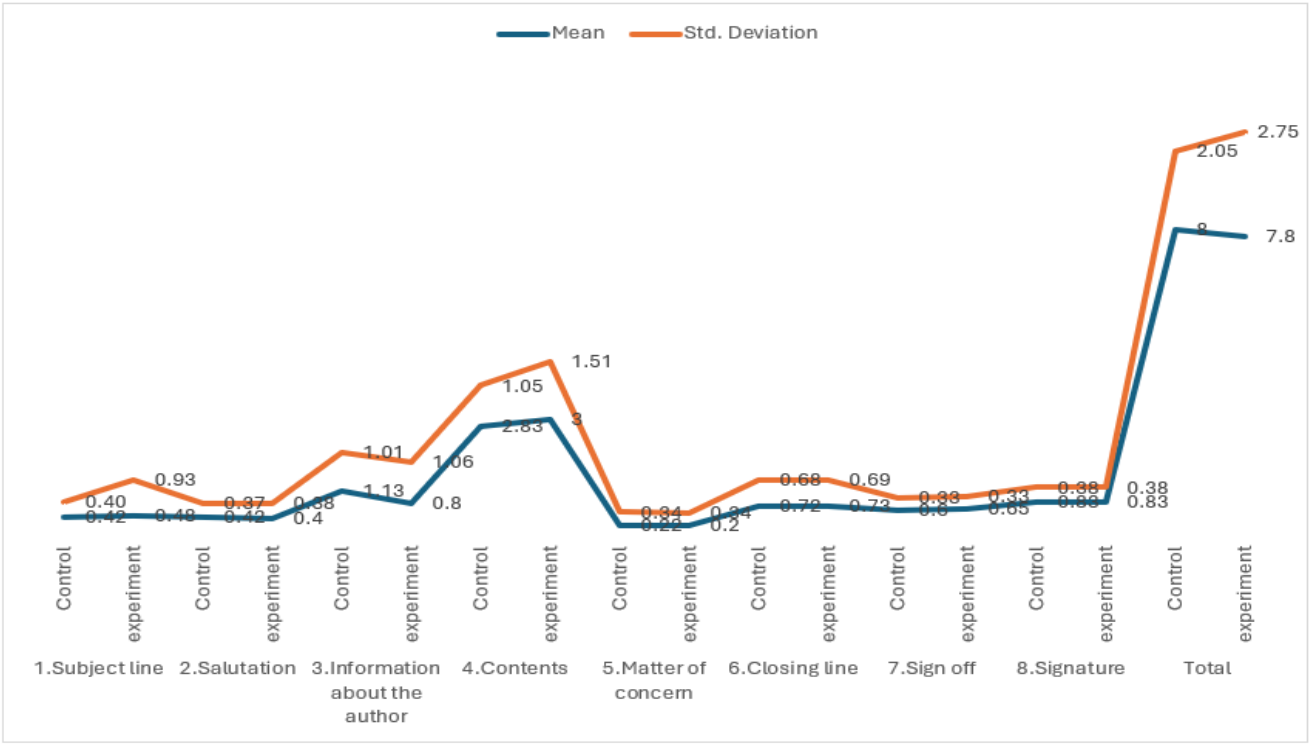


Fig. 1: Comparison of Pre-test Scores for Control and Experimental Groups

Looking at the overall total scores, the control group achieved a mean of 8.00 (SD = 2.05), while the experimental group recorded a mean of 7.80 (SD = 2.747). The t-test showed no significant difference between the two groups, t(58) = 0.320, p = 0.750.

The t-test for independent samples was used to determine the significance of differences between the mean scores of the control and experimental groups in the post-test. The analysis revealed statistically significant improvements in the experimental group across all eight domains of formal email writing. As detailed in Table 8 and Figure 2, the independent samples t-test demonstrated that the ChatGPT-mediated group significantly outperformed the control group, with the overall total score showing a massive effect size ($\eta^2 = 0.738$), indicating that the intervention accounted for nearly 74% of the variance in the post-test scores.

Table 8: T-Test for Independent Samples to Determine Significance of Differences in Mean Scores of the Control and Experimental Groups in Post-Test

Domain	Group	N	Mean	Std. Deviation	T	df	Sig. (2-tailed)	Eta Squared	Level
1. Subject line	Control	30	0.57	0.45	2.547	58	0.014	0.101	Medium
	experiment	30	0.83	0.36					
2. Salutation	Control	30	0.67	0.38	2.971	58	0.004	0.132	Large
	experiment	30	0.90	0.20					
3. Information about the author	Control	30	3.00	0.64	4.826	58	0.000	0.287	Large
	experiment	30	3.70	0.47					
4. Contents	Control	30	5.00	0.95	11.366	58	0.000	0.690	Large
	experiment	30	7.33	0.61					
5. Matter of concern	Control	30	1.03	0.63	4.274	58	0.000	0.240	Large
	experiment	30	1.73	0.64					
6. Closing line	Control	30	0.90	0.68	6.014	58	0.000	0.384	Large
	experiment	30	1.77	0.50					
7. Sign off	Control	30	0.58	0.4	2.359	58	0.022	0.088	Medium
	experiment	30	0.80	0.31					
8. Signature	Control	30	0.72	0.43	2.248	58	0.028	0.080	Small
	experiment	30	0.92	0.23					
Total	Control	30	13.23	1.43	12.785	58	0.000	0.738	Large
	Experiment	30	18.17	1.56					

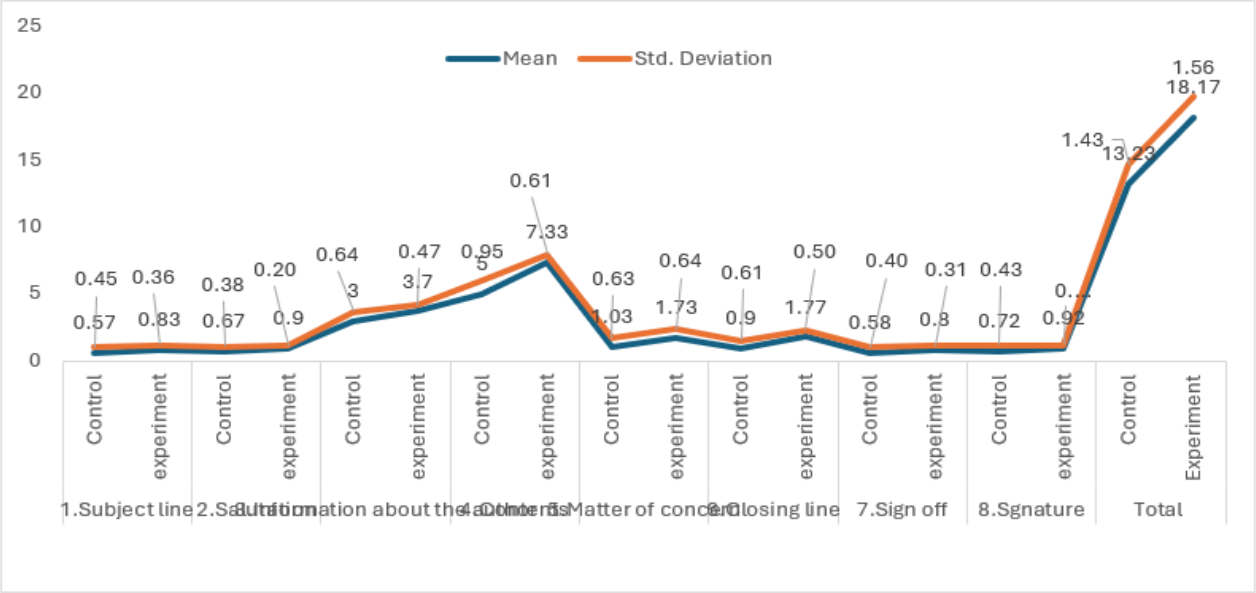


Fig. 2: Comparison of Post-test Scores for Control and Experimental Groups

The most dramatic improvements were observed in the substantive aspects of email composition. For instance, in 'Contents', the experimental group's mean score (M = 7.33, SD = 0.61) was vastly higher than the control group's (M = 5.00, SD = 0.95), a difference that was statistically significant ($t(58) = 11.37$, $p < 0.001$) with an exceptionally large effect size ($\eta^2 = 0.690$). Similarly, for the 'Closing Line', the experimental group (M = 1.77, SD = 0.50) far surpassed the control group (M = 0.90, SD = 0.61), with a highly significant result ($t(58) = 6.01$, $p < 0.001$) and a large effect size ($\eta^2 = 0.384$).

Even in domains with more modest gains, the results were clear and significant. For example, in 'Signature', the experimental group (M = 0.92, SD = 0.23) showed a small but statistically significant advantage over the control group (M = 0.72, SD = 0.43), $t(58) = 2.25$, $p = 0.028$, with a small effect size ($\eta^2 = 0.080$). This pattern confirms that the intervention had a pervasive positive impact, driving major improvements in core content and structure, while still yielding measurable gains in more mechanical components.

Overall, the findings confirm that the intervention was highly effective, resulting in improvements not only in structural elements such as salutations, subject lines, and signatures, but also in substantive areas such as content development and message clarity, both of which are critical for professional email communication.

These findings confirm that both groups began the intervention stage with comparable baseline performance in all domains of business formal email writing. The tiny numerical differences (for example, slightly higher control group means in Information about the author and slightly higher experimental group mean in Contents and Sign off) were not statistically significant. This similarity is critical because it ensures that any improvements observed in the post-test may be more reliably attributed to the instructional intervention rather than pre-existing ability gaps.

4.2 Research question 2: What are students' experiences regarding the integration of ChatGPT in the process of formal email writing?

The qualitative interview data were analyzed following Braun and Clarke's [30] six-phase thematic analysis framework. The findings highlight students' experiences with ChatGPT in formal email writing, alongside their broader perceptions of the tool's educational value. In the familiarization phase, the data were repeatedly read to identify patterns in how students described ChatGPT's role in guiding their writing. Initial coding captured meaningful features of the data, such as accuracy in writing subject lines, support in composing self-introductions, improved organization of content, and ease of use. For example, S1 noted that "ChatGPT helps in writing subject lines correctly," which was coded as accuracy in email components, while S5 stated that the tool "guides how to write about the author," which was coded as support in self-introduction. Similarly, S6 and S7 emphasized that ChatGPT "helps in writing relevant content" and "helps in writing better content," which were coded under content organization and relevance.

As coding progressed, broader themes were identified and organized into two overarching categories: experiential themes and opinion-based themes. The experiential themes centered on how students engaged with ChatGPT in writing specific components of formal emails. For example, S8 highlighted that ChatGPT "helps improve in writing good closings," and S9 added that the tool "makes it easy to write closing lines," both of which were grouped under professional and polite endings. S13 reflected on being "guided in understanding sign-off and signature placement," demonstrating ChatGPT's contribution to proper formatting and closure in emails. More generally, S14 and S15 described how ChatGPT "helped in learning more about formal email writing" and provided "rich vocabulary support, learning new words and expressions," illustrating the theme of learning and language development.

The opinion-based themes reflected students' perceptions of ChatGPT's broader utility. Many students viewed the tool as accessible and easy to use, with S7 describing it as "very easy" and S8 reinforcing that it is "very user-friendly." Efficiency and productivity also emerged as key themes, with S2 stating, "I think ChatGPT is a time-saving AI tool." Students additionally valued its problem-solving role, as S9 noted, "I believe ChatGPT can solve many formal email writing issues." However, some respondents emphasized the importance of using ChatGPT within a structured learning context; for example, S12 remarked, "I believe ChatGPT can be good if used regularly under teacher support," highlighting the theme of pedagogical integration. Finally, students expressed that ChatGPT fostered confidence and motivation in writing. As S15 put it, "ChatGPT is good to build confidence and face writing challenges," reflecting the attitudinal benefits of the tool.

The final thematic structure was reviewed to ensure clarity and distinction between themes. For instance, the theme of appropriate contents, drawn from S6 and S7's comments, was differentiated from the matter of concern, which, while not directly mentioned, was inferred to mean clearly stating the purpose of the email. Similarly, ease of use (S7, S8) was kept distinct from efficiency (S2) to emphasize usability versus time-saving benefits. Two experiential themes were therefore finalized: mastery of email components and learning and language

development. Five opinion-based themes were also confirmed: ease of use and accessibility, efficiency and productivity, problem-solving and writing accuracy, pedagogical integration, and confidence building and motivation.
Overall, the analysis demonstrates that students' experiences with ChatGPT were twofold. On the practical level, ChatGPT supported the mastery of email components such as subject lines, self-introductions, content organization, closings, and signatures. On the broader level, students reported that the tool was easy to use, efficient, supportive of problem-solving, beneficial when integrated with teacher guidance, and motivating in building confidence. The inclusion of direct student voices (e.g., S1, S5, S7, S12, S15) underscores the role of ChatGPT as both a technical aid and an affective support in formal email writing.

5. Discussion

5.1 Effectiveness of ChatGPT in enhancing EFL students' business formal email writing skills

The quantitative results provide robust, statistically significant evidence that ChatGPT-mediated instruction markedly enhanced the formal email writing skills of EFL students. The experimental group's superior performance, particularly in 'Contents' ($\eta^2 = 0.690$) and 'Closing Line' ($\eta^2 = 0.384$), demonstrates the intervention's powerful effect on the most critical aspects of professional communication: developing clear, coherent, and purposeful messages. These results suggest that ChatGPT-mediated instruction substantially enhanced the students' ability to write clear, concise, and professional emails.
Qualitative interview data support this, revealing themes of improved email components (e.g., subject lines, content), more effective closing remarks, vocabulary enrichment, and better sentence-level organization and coherence. However, several students expressed concerns about over-reliance on AI, emphasizing the need for teacher guidance to critically assess ChatGPT outputs, highlighting the essential role of pedagogical oversight in AI-assisted instruction.
Triangulation with prior literature further supports these results. Song and Song [32] found that ChatGPT's immediate, targeted feedback significantly enhanced Chinese EFL students' academic writing quality and motivation, particularly regarding content organization and coherence, consistent with the large effect sizes observed in Contents and Closing Line in this study. Similarly, Shang [31] reported that email-based writing tasks improved syntactic complexity, grammatical accuracy, and engagement among Taiwanese EFL learners, paralleling the improvements observed in salutations, closings, and sign-offs here. Marzuki et al. [26] found that AI writing tools enhanced Indonesian EFL students' ability to structure business correspondence, aligning with the current findings of improved content clarity and organization.
However, the literature also identifies limitations that contextualize these findings. Touvron et al. [28] and Al-Afnan et al. [27] caution that AI-generated emails may lack context-specific depth or client-specific information, reflecting students' concerns about over-reliance. The structured intervention in this study, which combined ChatGPT with teacher support and peer evaluation, likely mitigated these limitations, as evidenced by strong post-test improvements across all domains. Romadhon [29] similarly reported positive ESP student perceptions of ChatGPT for business formal email writing, emphasizing efficiency and genre awareness, which aligns with the current study's qualitative themes of ease of use and confidence development.
While the findings are encouraging, the study's reliance on a homogeneous sample of 60 Arabic-speaking undergraduates from Najran University inevitably constrains the broader applicability of the results. Cultural, linguistic, and institutional factors may shape how learners engage with AI tools and interpret their feedback. To assess the wider relevance of ChatGPT-supported instruction, future research should replicate similar interventions across more diverse EFL populations, encompassing varied linguistic backgrounds, proficiency levels, and academic disciplines. Such comparative studies could also investigate how AI-assisted writing aligns with the communicative norms and expectations of different professional and academic contexts. Overall, this study confirms ChatGPT's effectiveness in enhancing formal email writing skills, particularly in content development, clarity, and professional tone, while emphasizing the importance of pedagogical oversight to ensure AI is used as a supplementary tool rather than a replacement for critical thinking and human judgment.

5.2 Students' experiences and perceptions of using ChatGPT

The post-test results strongly confirm ChatGPT's effectiveness, showing the experimental group outperforming the control group across all eight categories. Students' positive perceptions in the interviews further reinforced these quantitative gains. Several positive themes like accessibility and usability, efficiency of time, building confidence, enhancement of vocabulary and structural development, learning motivation and engagement assistance, and independence emerged from the students' responses. At the same time, certain significant themes surfaced that need to be constructively addressed. These include students' concerns about becoming overly dependent on ChatGPT and broader concerns regarding ethics and its role in education.
Triangulation with existing research confirms these findings. Song and Song [32] reported that ChatGPT enhanced academic writing quality among Chinese EFL learners, particularly in content organization and coherence, reflecting the improvements observed in this study. Marzuki et al. [26] similarly found that AI tools improved Indonesian EFL learners' business correspondence, reinforcing students' reports of better subject lines and content structure (S1, S7). Romadhon [29] highlighted positive perceptions of ChatGPT for ESP students in business formal email writing, supporting qualitative themes of ease of use, efficiency, and genre awareness (S2, S7, S8). Literature also underscores the limitations of AI in writing tasks. Touvron et al. [28] and Al-Afnan et al. [27] cautioned about context-specific and emotional deficiencies in AI-generated emails, echoing student concerns about over-reliance. The intervention design of this study, combining ChatGPT with structured teacher guidance, ensured that students critically evaluated AI outputs, mitigating these risks. Moreover, both quantitative and qualitative results confirm that ChatGPT is a highly effective learning tool for enhancing EFL students' business formal email writing skills, particularly when combined with instructor support to foster critical thinking and ensure personalization. However, since the participants in this study belong to a specific cultural and linguistic group, future research could build on these findings by incorporating learners from diverse EFL backgrounds. Doing so would help uncover potential cross-cultural differences in how students perceive and utilize AI tools, offering deeper insight into the ways local educational practices and communicative norms influence engagement with AI-assisted writing. Bender et al. [33] highlight key ethical, social, and cultural concerns surrounding LLMs. Despite their ability to produce fluent, human-like text, LLMs lack genuine understanding or empathy, often replicate cultural biases from training data, raise environmental sustainability issues due to intensive computational demands, and risk user over-reliance by conflating linguistic fluency with accuracy or ethical validity.

5.3 Technical implications for LLM design and application

The quantitative and qualitative results of this study have clear implications for the applied science of LLMs. The model's strong performance in generating appropriate content, closing lines, and author information highlights the efficacy of its neural network in capturing and reproducing standard rhetorical patterns found in its training data. Conversely, its relative weakness in generating effective subject lines and sign-offs points to a specific technical limitation.

Subject lines require a high degree of creativity and summarization, often needing to be both informative and engaging, a nuanced task that may challenge the model's probabilistic generation approach. Similarly, culturally and contextually appropriate sign-offs (e.g., "Best regards" vs. "Sincerely yours") involve subtle pragmatic knowledge that the model may not consistently encode. These findings suggest that for LLMs to achieve true precision in specialized domains, future model fine-tuning or the development of context-sensitive prompting strategies may be necessary. Given the context-specific nature of this study, future research could explore LLM performance using culturally diverse datasets to reflect linguistic variation and pragmatic norms in global business communication. Such efforts would help ensure that emerging models better support the needs of a broader spectrum of EFL learners in a more equitable and context-aware manner.

The present study thus provides an empirical basis for directing both pedagogical and technical refinements in LLMs aimed at educational and professional communication applications.

6. Conclusion

This study explored AI-mediated precision in transforming EFL students' writing skills with special reference to formal email writing. The findings reveal that ChatGPT demonstrated both strengths and weaknesses: it performed well in content, salutations, author information, matters of concern, and closing lines, but showed limitations in subject lines, sign-off, and signature. Because of its low performance in three categories, ChatGPT should be reviewed for usage in a formal email writing lesson. Given the outcomes of ChatGPT tests in formal email writing, these models have significant potential for improving several aspects of corporate correspondence.

The findings show that they demonstrate linguistic skills, clarity of purpose, and relevance, all of which are essential for effective business communication. While ChatGPT can support many aspects of this process, it cannot replace human judgment, creativity, or empathy. Effective correspondence requires human touch, emotional intelligence, and a comprehensive understanding of the business context.

The study's findings highlight limitations, such as its exclusive focus on ChatGPT, which may not fully represent all AI technologies' capabilities, and its assessment of specific email scenarios and mechanical aspects, which ignore emotional tone, cultural nuances, and audience-specific adaptations, all of which are critical for authentic business communication. The sample study of 60 Arabic-speaking undergraduate EFL students at Najran University may not represent a wide range of requirements or professional settings. The four-month intervention period may not adequately represent the long-term impacts and sustainability of AI-assisted learning. Finally, while the quasi-experimental methodology was robust, it did not properly control external variables such as students' prior exposure to AI technologies or differing levels of desire, which may have influenced outcomes.

The study proposes a shift in EFL pedagogy toward technology-mediated instruction that incorporates AI and human interaction, such as ChatGPT, to minimize writing fear and increase motivation. The study also advocates incorporating ChatGPT into EFL writing curricula through the CALL framework, with AI serving as a supporting tool. Structured activities and teacher training should address issues with topic lines, sign-offs, and signatures. AI tool makers should improve emotional intelligence and audience adaptability for business communication by including various datasets and working with educators to adjust prompts for more nuanced results.

Moreover, this study demonstrates that the application of an advanced LLM like ChatGPT can significantly enhance the precision of a specific language task. By framing our research within the context of applied computational linguistics, we move beyond viewing AI as a mere black-box tool and begin to critically assess its architectural capabilities and constraints. We recommend that future work in this area actively collaborate between linguists, educators, and computer scientists to develop more refined, domain-specific models and to create optimized prompt libraries that can better leverage existing model architectures for educational purposes. This interdisciplinary approach is essential for harnessing the full potential of basic AI research in solving complex, applied problems in human communication.

The implications of this study extend beyond ChatGPT and the EFL classroom. The observed gains in writing clarity, coherence, and professional tone suggest that comparable AI-assisted pedagogical frameworks could be applied to other writing tools such as Gemini, WordTune, or Copilot across academic and professional communication settings. In multilingual or non-EFL contexts, these findings may support the integration of AI-mediated feedback systems to strengthen genre awareness, stylistic precision, and workplace communication competencies.

By positioning these outcomes within a broader educational and technological framework, the study highlights the transformative potential of AI-assisted instruction across disciplines. In domains such as business communication, technical writing, or professional correspondence training, the synergy between AI-generated feedback and human guidance may foster learner autonomy, linguistic accuracy, and communicative effectiveness. Future research should explore cross-linguistic and cross-disciplinary adaptations of this approach to assess its scalability and broader applicability.

Future research should analyze the usefulness of AI tools in EFL business writing, considering emotional and cultural elements, as well as longitudinal studies with varied populations to determine the long-term viability of AI-mediated advances. Educators should stress critical interaction with AI outputs, mandate clear training on content evaluation and revision, and use rubrics to promote customization and uniqueness. To promote responsible and inclusive use, institutions should develop ethical AI principles for education that include data protection, data anonymization, and fair access to technology like ChatGPT.

Future research should adopt more targeted and systematic designs to deepen understanding of AI-assisted writing. Cross-cultural comparative studies could investigate how learners from varied linguistic and cultural backgrounds interact with AI tools when composing business correspondence. Longitudinal approaches may reveal how sustained engagement with AI-supported writing influences learners' independent proficiency and critical editing skills over time. Experimental studies comparing ChatGPT with other AI systems, such as Gemini, Claude, or Copilot, could determine whether observed improvements are model-specific or broadly applicable across platforms.

From a technical perspective, further inquiry is needed into fine-tuning large language models for pragmatic competence and cultural sensitivity, especially within professional and business communication contexts. Enhancing models' ability to detect tone, politeness strategies, and context-dependent formality would significantly improve their effectiveness in intercultural exchanges. Collaborative efforts among applied linguists, AI developers, and educators could support the creation of specialized datasets and prompt libraries that incorporate socio-pragmatic cues and emotional nuance.

By following the recommendations, educators and researchers can capitalize on ChatGPT's potential to transform EFL business writing instruction while addressing its limitations, ultimately fostering a hybrid intelligence that combines AI efficiency with human judgment and creativity for effective professional communication.

Acknowledgment

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Growth Funding Program grant code (NU/GP/SEHRC/13/182-3).

References

[1] Pratama, Y. D. (2020). The role of writing in fostering critical thinking and creativity in EFL contexts. Indonesian Journal of Applied Linguistics, 10(2), 345–354. https://doi.org/10.17509/ijal.v10i2.28678
[2] Fitria, T. N., Hidayat, D. N., & Suparman, E. (2022). The impact of technology integration on EFL writing skills: A review of recent studies. Journal of English Language Teaching Innovation and Materials, 4(1), 56–67. https://doi.org/10.26418/jeltim.v4i1.45678
[3] Guo, K., & Wang, J. (2024). The impact of AI-driven tools on collaborative writing in EFL classrooms. Journal of Educational Technology Development and Exchange, 17(1), 23–34. https://doi.org/10.18785/jetde.1701.02
[4] Hoffman, R. (1996). Computer networks: Webs of communication for language teaching. In The power of CALL (pp. 55-78). Athelstan.
[5] Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), Multimedia language teaching (pp. 3–20). Logos International.
[6] Sabha, M. R. N. (2013). Using mobile-based email for English foreign language learners. Turkish Online Journal of Educational Technology, 12(2), 178–186.
[7] Jovic, M., & Mnasri, S. (2024). On the need to explicitize the unstated argument in cancer research: an ethnography on scientific argumentation. Humanities and Social Sciences Communications, 10(1), 1-9. https://doi.org/10.1057/s41599-023-01823-7
[8] Loukili, S., Fennan, A., & Elaachak, L. (2023). Applications of text generation in digital marketing: A review. In ACM International Conference Proceedings Series (pp. 123–130). ACM. https://doi.org/10.1145/3607720.3608451
[9] Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., & Stapp, J. (2022). Artificial intelligence in business communication: The changing landscape of research and teaching. Business and Professional Communication Quarterly, 85(1), 7–33. https://doi.org/10.1177/23294906221074311
[10] Nguyen, T. C. (2023). University teachers' perceptions of using ChatGPT in language teaching and assessment. In Proceedings of the AsiaCALL International Conference (Vol 4, pp. 116–128). https://doi.org/10.54855/paic.2349
[11] Chapelle, C., & Chapelle, C. A. (2001). Computer applications in second language acquisition. Cambridge university press.
[12] Qu, K., & Wu, X. (2024). ChatGPT as a CALL tool in language education: A study of hedonic motivation adoption models in English learning environments. Education and Information Technologies, 29(15), 19471-19503.
[13] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(3), 537–550. https://doi.org/10.1177/00336882231162868
[14] Elboshi, A. (2021). Web-enhanced peer feedback in ESL writing classrooms: A literature review. English Language Teaching, 14(4), 66–76. https://doi.org/10.5539/elt.v14n4p66
[15] Elliot, N., & Williamson, D. (2013). Assessing writing special issue: Assessing writing with automated scoring systems. Assessing Writing, 18(1), 1–6. https://doi.org/10.1016/j.asw.2012.11.002
[16] Nashruddin, N., Alam, F. A., & Tanasy, N. (2020). Perceptions of teacher and students on the use of e-mail as a medium in distance learning. Berumpun International Journal of Social, Politics and Humanities, 3(2), 182–194.
[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30(1), 5998-6008.
[18] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
[19] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901. https://arxiv.org/abs/2005.14165
[20] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730–27744. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf
[21] Gayed, J. M., Carlon, M. K. J., & Orii, Y. (2022). Exploring AI-mediated language learning: A case study of an AI-assisted writing application for EFL students. Journal of Educational Technology Systems, 51(2), 189–207. https://doi.org/10.1177/00472395221098765
[22] Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-mediated formal email writing. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (Article 474, pp. 1–13). ACM. https://doi.org/10.1145/3491102.3517731
[23] Ng, D. T. K., Al Balushi, A., & Leung, J. (2025). Impact of Gemini on EFL learners' writing skills and motivation: A mixed-methods study. Journal of Educational Computing Research, 62(3), 456–478. https://doi.org/10.1177/07356331241234567
[24] Ippolito, D., Yuan, A., Coenen, A., & Burnam, S. (2022). Wordcraft: An AI-assisted text editor for creative writing (Preprint). arXiv. https://doi.org/10.48550/arXiv.2210.12133
[25] Lam, Y. W., & Moorhouse, B. L. (2022). Using WordTune to support EFL students' academic writing: A case study. Technology, Pedagogy and Education, 31(4), 451–466. https://doi.org/10.1080/1475939X.2022.2068642
[26] Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. Cogent Education, 10(2), 1-17. https://doi.org/10.1080/2331186X.2023.2236469
[27] Al-Afnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. Journal of Artificial Intelligence and Technology, 3(1), 60–68. https://doi.org/10.37965/jait.2023.0184
[28] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv. https://doi.org/10.48550/arXiv.2302.13971
[29] Romadhon, R. (2024). AI in language learning: English for specific purposes students' perspectives on using ChatGPT for formal emails. Lingua, 20(2), 173–188. https://doi.org/10.34013/lingua.v20i2.1234
[30] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
[31] Shang, H. F. (2007). An exploratory study of e-mail application on FL writing performance. Computer Assisted Language Learning, 20(1), 79–96. https://doi.org/10.1080/09588220601118479

[32] Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in Psychology, 14, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843
[33] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21) (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922
[34] Odeh, R. (2005). Examining the difficulty coefficient in educational assessments. Journal of Educational Research, 98(4), 200–210. https://doi.org/10.1080/00220670509597512